## MIT Open Access Articles

## *Moving from postdictive to predictive kinetics in reaction engineering*

**Perspective**

## Moving from Postdictive to Predictive Kinetics in Reaction Engineering

William H. Green, Dept of Chem. Eng., Massachusetts Institute of Technology, 77 Massachusetts Ave. Rm. E17-504, Cambridge, MA 02139

whgreen@mit.edu

## 1 Introduction

Accurate quantitative chemical kinetic models are useful in many applications, ranging from design of chemical processes to building a consensus for international treaties (e.g. the Montreal Protocol that regulates production of ozone-destroying chemicals). The key components of most kinetic models are the list of reactions (with stoichiometric coefficients $\nu$), the function that describes the reaction rates *r(C,T)*, where $r_m$ represents the net flux of the *m*th reaction, and the corresponding instantaneous rate of change of the species $\omega = \nu\, r$ where $\omega_n$ represents the instantaneous rate of change of the concentration of the *n*th chemical species at position x,y,z and time t due to chemical reactions.[1,2] (N.B. Various symbols are used in the literature instead of $\omega$. Sometimes the matrix of stoichiometric coefficients $\nu$ is denoted ***S***.)  In the rate function, $C_m$ is the local instantaneous concentration of the *m*th chemical species, and T is the temperature. This 'chemical source term' $\omega$ is combined with other terms that represent transport of the species, energy, and momentum to build a model for the reacting system. The model can then be used to understand, optimize, and engineer the system.

If the feed is simple (only a few species, not many impurities) and the chemistry is very specific, the reaction kinetics can sometimes be accurately represented over the full range of interesting conditions using a simple rate law. For example, perhaps the *n*th species only participates in a single reaction, well-represented by an Arrhenius rate law, then

$$\omega_n = Ae^{-Ea/RT}\{C_iC_j - C_mC_n/K_{eq} \} \qquad \text{Eq. (1)}$$

In a very simple case like this, just a few laboratory experiments varying a handful of species concentrations and temperature might be sufficient to determine the small number of parameters that appear in $\omega$ (in this case A, Ea, and $K_{eq}(T)$). These are the types of simple systems that typically appear in undergraduate kinetics textbooks, and several simple systems of this type are indeed important in industrial practice. For these very simple 'textbook systems' the chemical engineering community knows exactly how to proceed.

However, very often the feed is complicated and the chemistry is unselective, so that there are a large number of reactions occurring simultaneously in the system, including secondary reactions of the primary products.[3] In these cases we are often unsure which molecules and reactive intermediates are present, and which reactions each of these species is involved in. It is not obvious how to model such complicated chemistry. There are two different main approaches: the conventional/historical postdictive approach (typically using small models) and the more recent predictive approach (often using very large detailed models), both described below.

### A. The postdictive approach to complicated chemistry

The conventional/historical postdictive approach, which dominated during the 20th century, can be summarized: "don't attempt to model all that chemistry detail, instead write a simple fitting model for $\omega$ that only involves a few species, and only has a few adjustable parameters". Examples of these simple models include the classic A $\rightarrow$ B $\rightarrow$ C models in textbooks and power-law rate expressions such as

$$\omega_n = k_{eff}[A]^\alpha[B]^\beta[C]^\gamma \qquad \text{Eq. (2)}$$

Levenspiel was a strong advocate for this simple-model postdictive approach, which he called the CRE approach [4], though of course Chemical Reaction Engineering includes many other approaches as well. Models like this have the important advantage that a small number of experimental data can be sufficient to determine the small number of parameters in the model. (Though of course just because the number of data are much larger than the number of parameters does not mean all the rate parameters can be determined from the available data, see for example [5,6].) A simple rate expression is also very convenient for solving reactor simulations and for mathematical [101,102,103] and numerical analysis (though the numerical aspect of the problem is much less important today, with modern computers, than it was in the 20th century). We call this model 'postdictive' since one normally must complete experiments to test if the proposed functional form actually fits, and to determine the numerical values of the parameters, before one has a model that can be used to predict anything.

The simple-model predictive approach has been used successfully in many cases and it is the most commonly used modeling approach in industry, but it has several important weaknesses. First, the simple models usually have little relation with the true reaction kinetics occurring in the system, and the model parameters (e.g. $k_{eff}$ $\alpha$, $\beta$, $\gamma$ in Eq. 2) do not have any clear relationship with the true elementary-step rate coefficients $k$(T). This means even if we knew or could estimate the true value of $k_i$(T) for some of the reactions occurring in the reactor, that knowledge would not help us build/improve the kinetic model – since elementary-step rate coefficients $k_i$(T) do not appear at all in the model!

The parameters in a rate expression like Eq. 2 should be viewed as empirical fitting parameters, not fundamental rate coefficients. Usually the parameter values are not transferrable to other systems; any significant change (e.g. adding an additive to the feed stream or the catalyst, significantly changing the

2

residence time or temperature) requires new experiments and a new fit. Because the simple model does not faithfully represent the true underlying chemistry, the model predictions tend to deviate from the system behavior as one moves away from the reaction conditions where the fitting data were obtained. Better extrapolative power would be expected if the rate expression conforms to the actual chemistry.[7] So while simple postdictive models can be excellent for interpolation around the training data, they are not reliable for extrapolation to new reaction conditions.

Often the parameter values in the expression for $\omega$ change as the process is scaled up, an indication that the parameters in the chemistry model include transport/mixing effects as well as the intrinsic chemistry. Indeed, this weakness of the conventional approach is a primary reason why pilot plant studies are required – when using simplistic fitting models, one usually cannot rely on a rate function $\omega$ determined at the laboratory scale. The fundamental reason is that the species A,B,C in Eq. (2) were usually chosen by the user because they were easy to measure in the reactor feed or effluent, not because A,B,C are the true key species controlling the chemistry. Commonly there are other unmeasured but much more reactive species X,Y,Z present inside the reactor that control the reaction rate, and the concentrations of X,Y,Z in different locations inside the reactor depend on many factors including the temperature distribution and mixing/flow/diffusion, not just on the measured [A],[B],[C] in the feed or effluent. The conventional/historical postdictive approach only requires a correlation between [A],[B],[C] and $\omega_n$ that holds with sufficient accuracy at the specific conditions encountered in this pilot plant. It is not seeking a general rate formula accurate at all possible reaction conditions.

The conventional/historical approach is fundamentally "postdictive": one needs to have experimental data on the system, at conditions close to the conditions of interest, before one can make any quantitative predictions, i.e. one needs to have measured the data before one can predict it. With this approach, pilot-plant experiments (to measure training data needed to determine the effective parameters in the model) are usually required as part of the design process. This is a big disadvantage of the conventional approach, since it implies a very significant commitment of money and time before one has a reliable model, and so can rationally decide whether or not to commercialize a new process.

Because the conventional postdictive approach has already been discussed thoroughly in textbooks and journals, for the remainder of this article we focus instead on the alternative 'predictive' modeling approach, which can provide a way to predict kinetics even before doing any experiments, and so perhaps avoid the need for expensive pilot plant experimentation.

B. The high-fidelity 'predictive' modeling approach

In principle, we could achieve higher fidelity using more detailed chemistry models. Instead of using the simple fitting forms used in conventional approaches, one could instead use elementary-step rate equations, Eqs. (3) and (4), which are thought to be accurate over a wide range of conditions.

3

$$\omega_j(\textbf{\textit{C}},T,...)= \Sigma\ \nu_{mj}\ r_m \qquad\qquad\qquad \text{Eq. (3)}$$

$$r_m = k_m \prod_{n=1}^{N}(\gamma_n C_n)^{-\nu_{mn}H(-\nu_{mn})} - \frac{k_m}{K_m} \prod_{n=1}^{N}(\gamma_n C_n)^{\nu_{mn}H(\nu_{mn})} \qquad\qquad \text{Eq. (4)}$$

In these equations, $r_m$ is the net rate of the *m*th elementary-step reaction, $k_m$ is its rate coefficient and $K_m$ is its equilibrium constant (in concentration units). *H* is the Heaviside function. $\nu_{mj}$ is the stoichiometry of the *j*th species in the *m*th reaction, with the convention that reactants are assigned negative values. The $\gamma$'s are activity coefficients (=1 for ideal mixtures).

An example contrasting the conventional/historical postdictive approach with this more detailed predictive approach for propane pyrolysis is given in Table 1. The larger model built on the physical rate law for the actual reactions occurring in the system is expected to have more fidelity to what is really happening in the system, and so better power of extrapolation, than an empirical fit to $\omega$.

As discussed in the Perspective by 2018 Wilhelm Award winner Linda Broadbelt [3], it is now often possible to write down all the important reactions in a detailed kinetic model using various automatic mechanism generation software packages [8-17]. These computer-generated reaction mechanisms are intended to include not just the molecules in the feed and the product streams, but also all the important reactive intermediates, as well as multiple side channels leading to byproducts (though sometimes some of the species are intentionally lumped [17]). The computer-generated reaction mechanisms can be quite large, often involving hundreds or thousands of reactions and a similarly large number of species. The corresponding reactor simulations typically include a differential-algebraic equation for each of the many state variables. These complicated simulations usually need to be solved using specialized numerical techniques because the system of equations is stiff. In the past solving such large simulations was very difficult, but with modern algorithms, software, and computers solving the simulation is often not the main challenge today.

Instead, the principal difficulty with the predictive approach – and the main reason why chemical engineers instead usually choose the postdictive approach despite its failings – is the need to determine or accurately estimate the large number of k's in the model (and also the thermochemical parameters *K* and **γ**). For any system with more than a few dozen reactions, it is expensive or even completely impractical to determine all the rate parameters from experiments. Globally-optimal least-squares fitting to kinetic data is possible only for systems with 10 or fewer adjustable parameters [6,107-108]. A few more parameters can be used if one does not need a guarantee of global optimality, but the fitting process often runs into problems [4]. Because kinetic models are so non-linear, and because the observables are usually only weakly sensitive to some of the parameters, least-squares fits to kinetic data are typically ill-determined, i.e. several different sets of rate parameters could fit the limited experimental data equally well. To some extent this problem can be alleviated by using orthogonal experiments to try to break the covariance of coupled parameters; good software for this sort of effort has been developed by the PrIMe project [5]. However, even for relatively small very heavily studied

4

reaction mechanisms such as combustion of $H_2$, the experimental data alone are insufficient to determine all the rate parameters [18]. Even if were possible to determine the parameters experimentally, the need for an extensive experimental campaign to measure rate parameters for each proposed new process would add significant cost and significantly slow the development of the new process.

However, there is a way out of this conundrum: we do not need to determine all the rate parameters from experiment. For many decades we have been able to estimate rate coefficients from analogy, correlations, or using the functional group approximation and linear free-energy relationships. Famous methods of this type have been described by Evans & Polanyi, Hammett, and Benson [19]. At the time those methods were derived, all the estimates were based in some way on experimentally measured rate coefficients. Typically they were only accurate for reactions quite similar to those already measured. But for many of the most important reaction types there are several rate measurements in the literature, so these techniques allow one to fill in many (but not all) of the rate parameters in a detailed kinetic model.

As discussed below, it is now possible to compute many rate coefficients and thermochemical parameters from first principles, using a combination of quantum chemistry, statistical mechanics, and rate theory, with accuracies comparable to experimental measurements. This provides a rapid inexpensive way to fill in the missing rate parameters. These calculations give us access to many more types of reactions than have ever been measured experimentally.

For example, in a recent high-fidelity model for $H_2$ combustion, it was found that including a large number of rate [20] and transport [21] parameters coming purely from quantum chemistry calculations, not inferred at all from experimental data, significantly improved the predictions of important properties such as flame speed [18]. Combining this new capability to compute model parameters from first principles with automatic reaction mechanism generation software [3, 8-17] makes it possible to predict the course of chemical reactions before doing any experiments, and to understand processes that are happening in an experimental system even if they are impractical to measure. This new predictive capability has many implications for the future of Reaction Engineering, for industrial practice, and for how chemical engineers should be educated, some of which are highlighted below.

Although it is now possible to predict chemical kinetics, this capability will only have real-world impact if it is practical for chemical engineers to make these predictions on a reasonable timescale, and the value of the predictions will greatly depend on their accuracy. These issues are discussed further below.

### C. Mixed predictive-postdictive models

The weaknesses of the conventional/historical purely-postdictive approach have been known for many decades, but only recently has it become practical to compute large numbers of high-accuracy rate coefficients from first principles (and then only for limited types of reactions). Over the past >30 years a

5

variety of modeling approaches have been proposed which combine features of both approaches, to try to achieve higher fidelity even when it is impossible to predict all the parameter values to desirable accuracy. One profitable direction has been to use what is known about the chemistry to construct functional forms for $\omega(C,T)$ which have the correct qualitative behavior and capture some of the correlations between the parameters; a classic example is the use of LHHW functional forms for $\omega$ instead of the less chemically-realistic power-law model shown in Eq. (2).

In other cases a realistic detailed-model structure has been employed, even though it is not practical to predict all the parameter values to sufficient accuracy before doing experiments. So then experiments are used to adjust a few of the parameters in the model (e.g. often the top N most-sensitive parameters are adjusted to fit the experimental data). Sometimes this approach reduces the number of experiments required to achieve acceptable accuracy, and in any case it is expected to increase the range of validity of the model (i.e. to improve its extrapolative power).

This century, many models have been published where some of the parameters are from the literature, some are estimates-by-analogy (e.g. using Benson's methods []), some are computed from first-principles, and some are obtained by fitting to experimental data. Sometimes these mixed models are very accurate, but even relatively inaccurate models can be very helpful in guiding future experimentation. Although these models are helpful in many ways, we note that it takes special effort to document the disparate origins and uncertainties of all these parameters, and it can be challenging to reliably estimate the uncertainties in predictions made using such a model.

We do not have the space here to comprehensively review these approaches, which have led to many very useful kinetic models, and are often the best available modeling option. We note only that hybrid approaches are most economically valuable if they allow one to avoid the need for expensive pilot plant experimentation, instead using only cheaper laboratory experiments. But achieving this requires demonstrating that the model predictions are accurate enough they can be relied on, which can be very challenging. Even if these models do not provide immediate economic benefits, they can be very helpful for advancing understanding of important systems with complicated chemical kinetics. For the most important reaction systems, where the societal value of an accurate kinetic model can be very high, it is not unusual for dozens of models to be developed, each building on learnings from earlier iterations.

## 2 Technical Issues

### A. Computations & Workflow: predictive detailed-model approach

The detailed-model approach typically starts with a small "seed" kinetic model, and gradually grows it by adding additional reactions and chemical species. Usually the corresponding rate and thermochemical

6

parameter values initially come from estimation methods (e.g. group-additivity for thermochemistry; linear-free-energy estimates for rate parameters [19]), but these parameters are not very accurate, so some of the values are refined using quantum chemistry. Only rarely have all the parameters in the chemistry model been computed from first-principles [22]. Often the people building the kinetic models are different from the people performing the quantum chemistry calculations, so the sequencing of the work has often been inefficient, driven primarily by practical considerations such as personnel availability and funding cycles. However, the increasing automation of steps in the workflow have partially ameliorated this problem: there are now several well-developed software packages for constructing reaction mechanisms [e.g. 8-16, 99], and there has also been continuous progress in automating the quantum chemistry calculations [22-28,94,95], making it increasingly practical for a single person to both build the kinetic model and refine its parameters. For example, in one recent study all the thermochemical parameters of all the molecules in a large kinetic model were automatically computed using high level quantum chemistry [23]. Several groups are developing the capability to do automatic calculations of the rate coefficients as well [24-28]. There have also been significant advances in methods for computing solvation effects.[29,30] There have also been continuing improvements in methods for estimating rate and thermochemistry parameters (including solvation and pressure-dependence of rate coefficients) [31-37] and in algorithms for solving kinetic simulations [38-46]. We suggest an efficient workflow combining all these software components to build predictive models in Figure 1.

Fig. 1 Proposed workflow for automated development of predictive models for reaction chemistry

### B. Limits on the Scope of Systems that can be Predicted Today

In order to predict important observables (e.g. conversion and selectivities to different products and by-products) with a kinetic model, we must be able to (1) construct the model, (2) predict values of the parameters in the model, (3) estimate the initial state of the system (including the feed composition), and (4) solve the model. In various cases each of these four requirements can be challenging, as discussed here:

1. Model construction is difficult if the reacting mixtures are very complex and the chemistry is not specific. In that case, the physical system will have a huge number of distinct species (e.g. molecules, surface-bound intermediates) each reacting in multiple ways. A model which is faithful to this complexity will be extremely large; in some cases more than a million reactions are expected to be important. (If one considers conformational changes to be reactions, as in protein folding or some polymer problems, the complexity is orders of magnitude worse.[47]) The process of constructing models this large can run into computer memory limitations [12,48,

7

96] and be slow enough that it needs parallelization [49]. This sort of problem is often encountered in systems where molecular weight growth (e.g. coke or soot formation, polymerization) is significant, because then the complexity of the molecules in the system, and the number of important reactions per molecule, increases as the reaction proceeds. Often it can take a very large number of reaction steps to get from the starting material to the macromolecular products of interest, and all these steps must be included in the model. Even if molecular weight growth is unimportant, unselective reactions of large molecules (e.g. degradation reactions driven by OH radicals or strong acids) also lead to a huge number of reactions & reactive intermediates in the model. The number of isomers grows combinatorially with the size of the molecules. So modeling methods which work very well for molecules in the gasoline range (6-8 carbons) may be impractical for molecules in the diesel range (14-24 carbons), because the heavier feed typically contains many more distinct molecules, and each heavy feed molecule can react to form a large number of reactive intermediates. The full-detail modeling approach, where the state of the system includes the concentration of every distinct molecule, is therefore unfeasible for heavier feeds such as bitumen, biomass, and waste plastic. To try to address this problem, various reduced-model representations of the large molecules have been proposed (Structure-Oriented Lumping [50], the Fragment method [51], method of moments [52], probability density functions [53]) but there remain many technical challenges.[98]

2. The existence of quick and reasonably reliable estimates for thermochemistry and reaction barriers greatly facilitates the construction of useful predictive models, e.g. by allowing one to rule out species or reactions that will not be energetically accessible at the reaction conditions. One can also use the estimated parameter values to construct an initial model; sensitivity and uncertainty analyses of that initial model can identify which parameter values need to be refined (e.g. by quantum chemistry calculations) to significantly tighten the error bars on the predictions.

Often good estimation procedures are available for well-studied chemistry, and for many of those well-studied systems it is already known which quantum chemistry methods are suitable / give accurate predictions of rate parameters. However, if the system to be modeled is very different from systems that have been studied previously, it can be challenging to develop reliable or even plausible estimates of the rate parameters, and the performance of different quantum chemistry methods on this new system may be unknown. At the time this is written, well-tested accurate quantum chemistry methods exist for most small closed shell organic molecules and for simple organic free radicals. But it is hard to accurately compute rate and thermo parameters for large molecules, for electronically-excited molecules, and for systems involving multiple transition metal atoms (due to strong electron correlation and many low-lying electronically excited states) or heavy elements (with significant relativistic and spin-orbit

8

effects). There is an abundance of prior work on hydrocarbon chemistry and significant prior work on certain types of organic chemistry, so in those cases estimation methods already exist for many of the rate and thermochemical parameters that appear in the models, and it is often known which quantum chemistry methods are particularly good or bad for computing these reactions. On the other hand, if one were to study a system containing iodine, sulfur, boron, bismuth and tellurium atoms, there would be no reliable estimation methods and one would have to compute almost every proposed species and reaction quantum mechanically. Additional work (and probably experimental validation) would also be required to determine which quantum chemistry methods are suitable for this less-studied system.

Although reliable thermochemistry estimation procedures exist for many organic molecules, often based on methods developed by Benson and co-workers [19], these methods have limited scope. For example, the predictions for the thermochemistry of molecules containing fused rings or multiple substitutions on an aromatic ring are still unreliable, despite significant efforts to improve them [54-56]. Also, the group values are unknown for many common functional groups involving heteroatoms, though in recent years many of these values have been filled in using quantum chemistry calculations [57-60]. Several researchers have proposed replacing Benson-type estimates with machine-learning models, this has shown some promise, but so far these models also have limited scope and significant uncertainties [35, 61-63].

3. Estimating the initial state of the system is often challenging. In heterogeneous catalysis, we often do not know the morphology of the catalytic surface during the reaction, much less which of the different surface sites are actually active for the reaction of interest. How to use the limited available information available on the catalyst to develop a quantitative estimate of the state of the catalytic surface is the subject of ongoing research. Note that even for freshly-prepared catalysts, the active sites are usually occupied by something, they are not completely vacant, and indeed it is commonly observed that different pre-treatment or 'activation' of a catalyst leads to different measured conversions.

   If we are performing reactions on complicated feeds such as biomass, petroleum fractions, or waste plastic, we generally do not know the molecular composition of the feed, though we usually have some partial information. However, there are well-developed "feed reconstruction" methods [64, 104-106] for constructing a reasonable estimate of the feed composition.

4. Convenient affordable methods exist to solve most simple kinetic models (e.g. <100 chemical species) in common idealized reactors, in packages such as CHEMKIN [65] and CANTERA [66]. Typically the idealized reactors are 0-d or 1-d, sometimes 2-d, but not three-dimensional (3-d). If the chemistry equations are not too stiff they can be solved using explicit methods, which have

9

been scaled to run efficiently on supercomputers even for 3-d systems [44]. But often kinetic model equations are very stiff because of the short lifetimes of some of the reactive intermediates. Also often the reactor model is formulated as a system of differential-algebraic equations (DAE), which often must be solved using stiff-solver techniques and which sometimes can be problematic (e.g. high-index equations, non-unique solutions).

The standard way to solve these stiff models is to first discretize the spatial degrees of freedom (most commonly using the finite volume formulation) and then use operator splitting to separate the transport and the chemistry terms in the discretized equations. Operator splitting allows one to solve the chemistry equations in different finite volumes in parallel, and to use specialized numerical methods for the transport terms. Today Strang splitting [67] is the most popular operator-splitting approach, but this can run into problems particularly for steady-state solutions [45,68] so in some cases other splitting methods can have advantages.[69] The operator-split stiff chemistry equations are usually solved using methods that require construction of the Jacobian of the chemistry terms. This often works well for systems with even several thousand chemical species. But solving the model equations using Jacobian-based methods can be very challenging if the system involves more than about 10,000 species (since then in each finite volume one must repeatedly construct and solve an ill-conditioned Jacobian involving more than 100 million entries, at significant computational cost and also incurring some numerical error due to the poor conditioning). If the model is formulated cleverly [38], the Jacobian can be quite sparse, and taking advantage of this sparsity can provide significant speed-up and also improve numerical accuracy. Several methods have been proposed to accelerate the solution of the chemistry equations. One line of work has been to just accelerate the numerical procedure, e.g. by using GPUs instead of CPUs [39,40] or by using better pre-conditioning and other numerical methods.[41,42] Other methods avoid the need for constructing a Jacobian at all, instead using specialized Jacobian-free stiff ODE solvers [41,43]. Yet other methods modify the model to reduce its stiffness enough so that Jacobian-free explicit ODE solvers can be employed.[70]

Another large body of work has been focused on simplifying the chemistry models ("model reduction") to make it easier to solve them. One method is to judiciously delete species and reactions from the large detailed model ('skeletalization').[71] Alternatively, one can rewrite the model in terms of a smaller set of state variables which may not represent the concentrations of individual species (e.g. one approach is to use eigenvectors of the Jacobian as the state variables).[72,100] In systems where the chemical composition or temperature varies dramatically it can be advantageous to use different chemical kinetic models to evaluate $\omega$ in different spatial locations or at different times ('adaptive chemistry').[73,74] Some of these model-reduction methods are so fast that they can be applied 'on-the-fly' whenever an evaluation of $r$ is required.[75-79]

10

Today it is often more difficult to build an accurate chemical kinetic model than it is to solve it, but in some cases (e.g. if the reactor must be modeled in 3-d, or if it is important to accurately capture the turbulence-chemistry interaction, or if the number of species is extremely large), solving the model can still be the bottleneck.

## C. Accuracy

### 1. Model vs. Data comparisons

Assessing the accuracy of a model's predictions is challenging whether one uses a predictive, postdictive, or hybrid approach.

In cases where a postdictive approach is employed, one usually needs to use most of the available pilot plant experimental data to train the model, so there are very few independent "test data" available for checking the accuracy of the model. Also, pilot plants are typically not as well controlled as laboratory instruments, so if there is a discrepancy between the model prediction and measured data point, it is often not clear whether to attribute that to imperfect knowledge of the reaction conditions, to an error in the model, or to the error bars on the experimental data. Even if it is decided that the discrepancy is primarily due to model error, it may not be clear whether that error arose because a poor functional form was used to fit the data, or because insufficient (and/or inaccurate) training data were used to determine the model parameters. And of course, the fact that one test data point is consistent or inconsistent with the model prediction does not provide much information about whether some other observable measured at some other reaction condition would be accurately predicted by the model.

With the predictive approach, the model predicts a very large number of observables, but typically only a few of these are actually measured in an experiment. If those limited measurements are consistent with the model predictions within the experimental error bars, one says the model has been "validated" at those specific reaction conditions, but that does not mean the other predictions of the model are sure to be correct, nor that the model predictions would be accurate at a different reaction condition. If the measurements are very different from the predictions, with deviations much larger than the expected experimental errors, one can conclusively say that the model predictions are incorrect for those specific observables at that specific reaction condition. In practice, being conclusive can be more difficult than it might appear at first glance, since none of the parameters in the model are known with perfect accuracy. One often has a handful of less-interesting (e.g. experiment-specific) parameters in the model which one wishes to optimize first before making the model-vs.-data comparison. In principle that should be done using global optimization techniques [6] since kinetic models are typically nonlinear and nonconvex.

11

However, even if one can clearly judge whether or not one prediction of a model is consistent with the available experimental data, this is not sufficient to draw general conclusions about the accuracy of the model's other predictions.

2. Uncertainty in Model Predictions due to Parameter Uncertainty

If one can estimate the uncertainties in the parameters in the kinetic model, one can, in principle, propagate those uncertainties through to the model's predictions. This makes it possible to assign uncertainties to each prediction. If this process assigns a very large uncertainty on some prediction, the user is warned not to have much confidence in that prediction. Uncertainty and related sensitivity analyses can be very helpful at identifying which parameters in the model should be improved.

It is challenging to accurately estimate the uncertainties of all the model parameters in a detailed model, a pre-requisite for performing uncertainty analysis. Some methods such as machine learning provide uncertainty estimates automatically, but these estimates are usually too optimistic.[61,63] Often, the uncertainties in different parameter values are highly correlated, so one needs to know not only the error bar on each parameter, but also each parameter's covariance with several others. This problem is endemic in large kinetic models, because several parameters are estimated based on the same underlying functional group values or linear free energy relationships. Recent papers have highlighted the importance of accounting for this type of correlation when computing uncertainties in the predictions.[80,81]

Less obviously, some types of experimentally-derived parameters are intrinsically correlated, because e.g. Arrhenius A and Ea values are both derived from the same measured k(T), [82] and individual-molecule thermochemistry is often derived from measurements of $\Delta H_{rxn}$ or $K_{eq}$ for a reaction involving several species. Any errors in the measured k(T), $\Delta H_{rxn}$, or $K_{eq}$ will affect several parameters, and those errors will be correlated. Correctly accounting for these correlations in the parameter uncertainties and using other independent measurements (or high-accuracy quantum calculations) to break the correlations can significantly tighten the error bars.[5,83]

Similarly, the errors in parameters derived from quantum chemistry calculations are also often correlated. For example, if one uses a quantum chemistry method that systematically underestimates the stability of C-C triple bonds, all of the thermochemistry values for molecules containing triple bonds computed using that method will be off in the same direction. This issue is well-known, and is the reason Bond-Additivity Corrections [84,85] and other adjustments are commonly employed when computing thermochemistry. However, if the quantum chemistry method miscomputes triple bonds, the computed barriers for reactions forming or consuming triple bonds are also likely to have systematic errors, and need some additional corrections.

12

At present in the literature there is a very small set of thermochemical and kinetic parameters with well-established uncertainties (sometimes even accounting for correlations), a larger set with guessed uncertainties, and a much larger set of published values or estimation procedures with no uncertainty estimates at all. Significant effort will be required to overcome this serious deficiency. In part what is required is a change in culture in the broad kinetic community, including funding agencies and journal reviewers, to encourage/require error bars and uncertainty estimates be attached to every published rate coefficient and thermochemical parameter.

3. Prediction Errors due to Model Truncation

An even more difficult problem is that it is possible that the kinetic model completely omitted an important reaction or species, in which case it would make incorrect predictions even if all the model parameters were known perfectly. Error due to omitting a reaction is called "Model Truncation Error"[86] to distinguish it from the more well-studied "Parameter Error" discussed above. At present we do not have good ways to estimate the magnitude of the Model Truncation Error, but we can detect it if we observe that adding additional species and reactions to a kinetic model significantly changes its predictions. The whole effort to automate reaction mechanism generation can be seen as an attempt to overcome the Model Truncation Error in hand-built models for complicated systems. Overcoming Model Truncation Error is often not hard for simple room-temperature laboratory systems starting from pure materials and run to low converstions, since there are a limited number of possible reactions with barriers low enough to be important at low T. However, controlling this error can be very difficult for systems with complicated feed mixtures run to high conversions with less selective chemistry.

4. Accuracy/Errors Depend on Reaction Conditions

Our goal is for the model to accurately emulate the true species production rate, i.e. we want the difference

$$||\omega_{model}(\mathbf{C},T) - \omega_{true}(\mathbf{C},T)||$$

to be small. Clearly the Parameter Error and Model Truncation Error both depend on the instantaneous local reaction conditions ($\mathbf{C}$,T). A chemical kinetic model can predict the conversion and all the major products to acceptable accuracy at certain reaction conditions but be completely wrong at other conditions. This suggests that each model should be published along with its 'range of validity', to warn users not to use the model where it is inaccurate. However, at present it is challenging to identify the true range of validity of a kinetic model; usually all one can say is that the model has been experimentally validated at a few specific conditions. For an early attempt to

13

develop a systematic algorithm for identifying the range of validity of computer-constructed models, considering only Model Truncation Error, see Song et al.[87]

# 3 Some Examples

### A. Overview. Why Small Molecules & High Temperature?

The detailed-model approach, with key parameters computed using quantum chemistry, has been able to predict the chemistry of several reaction systems even before experimental data were available.[88,89] In these cases, the conventional postdictive approach would be impossible, since there were no experimental data available to use to determine the model parameters.

To date, the detailed-model predictive chemistry approach has been most successful for high-temperature gas-phase systems comprised of relatively small molecules. Small molecules are easier to model because they each participate in a smaller number of reactions (fewer reactive sites, forming fewer reactive intermediates), and because CPU time required for high-accuracy quantum chemistry techniques scales very poorly with the number of atoms involved in a reaction. It is currently difficult to compute highly accurate rate coefficients for any reactions containing more than about 40 atoms. Existing quantum chemistry software is designed for computing reactions for isolated molecules, which is not too different from a molecule in the gas phase; correcting for condensed-phase effects increases the computational cost and reduces the accuracy.

High-temperature systems are often easier to model accurately than low-temperature systems, because the high-temperature systems are much less sensitive to small errors in computed or estimated energies. Consider a reaction where there is an error $\delta E$ in the computed or estimated enthalpy of formation of one of the reactants. Since the reaction barrier is computed as ($E_{TS} - E_{reactants}$), this error will cause the activation energy $E_a$ to be off by about $-\delta E$. Similarly, the computed $\Delta H_{rxn}$ depends on the computed ($E_{products} - E_{reactants}$), so this error will cause computed $\Delta G_{rxn}$ to be off by about $-\delta E$. Consequently the error $\delta E$ will cause errors in the computed k and $K_{eq}$ used in the model. If $\delta E$ is the only important error,

$$k_{model} = A \exp(-E_{a,model}/RT) \approx A \exp((-(E_{a,true} - \delta E))/RT) \approx k_{true} \exp(\delta E/RT)$$

$$K_{eq,model} = \exp(-\Delta G_{rxn,model}/RT) \approx \exp(-(\Delta G_{rxn,true} - \delta E)/RT) \approx K_{eq,true} \exp(\delta E/RT)$$

With many quantum chemistry methods the uncertainties in computed energies $|\delta E| > 4$ kJ/mole, so if T < 500 K the quantity in the exponential ($\delta E/RT$) is >1, and both k and $K_{eq}$ have uncertainty ranges of at least an order of magnitude. Such large errors in k's and $K_{eq}$'s can cause the chemical kinetic model to behave quite differently from the true system, e.g. it might mispredict which products dominate, and the model might predict some reactions to run backwards from their actual reaction direction. Starting from such a poor zero-order model makes it very challenging to identify what should be done to improve the model. But if T > 1000 K and $|\delta E| \sim 4$ kJ/mole, ($\delta E/RT$) is < ½, and so the full uncertainty

14

ranges of k and $K_{eq}$ are all less than a factor of 3. With such modest errors in the nominal parameter values, one can usually use sensitivity analysis to correctly identify a small number of important parameters, giving good direction to efforts to refine the model. As a consequence, it is often easier to give semi-quantitative predictions for high-temperature reaction systems (e.g. steam-cracking, pyrolysis, combustion, steam methane reforming) than it is for room-temperature systems (e.g. environmental, biological, and most synthetic organic chemistry).

Sometimes, of course, one can be lucky, e.g. it is not unusual for some observables to be predicted more accurately than one would expect, due to cancellation of errors. For example, your estimation method might give the wrong thermochemistry for a certain functional group, but if that particular functional group is conserved through the rate-controlling reaction steps, that error would not affect the important $K_{eq}$'s. Similarly, sometimes branching ratios between similar reactions are computed accurately even though the absolute rate coefficients are inaccurate, because the key competing reactions are mis-estimated by a similar factor. Often the errors in the conversion and the product selectivities are very different, since the conversion is sensitive to the absolute concentration of reactive species (e.g. radicals, active sites of a catalyst), but the selectivities are not. Sometimes this systematic error-cancellation in certain observables can be detected by uncertainty analyses [80,81]. But of course there are other cases where the errors in individual parameters combine in ways that make it quite difficult to accurately predict the observable of greatest interest.

B. Example #1: Methane Pyrolysis

Methane pyrolysis to manufacture ethyne and ethene (plus syngas) was developed in the 1950's and has been commercially practiced for decades. As the price differential between C2's and methane has increased in some locations, there has been increased interest in improving this process. Also, there is considerable current interest in methane pyrolysis as way to produce $H_2$ from natural gas without the $CO_2$ emissions associated with steam methane reforming. In both embodiments, the feasibility and economics of the process depends on the product slate: if the carbon-bearing products are valuable, or at least easy to handle, the economics can be favorable, but if carbonaceous solids are formed at an inconvenient location (e.g. plugging the reactor with coke) the process is much less appealing.

While 'methane pyrolysis' suggests that the important chemistry involves methane and other very small molecules, in fact both the conversion and the product distribution are sensitive to reactions of much heavier species present as minor components in the reacting mixture. Matheu et al. [90] showed that under the conditions of the experiments of Back et al.,[91] the key reactions involved C5 species, especially cyclopentadienyl radicals, formed as minor byproducts through a multi-step reaction sequence from the methane. The slow formation of C5 is the primary cause of the long observed experimental induction time of about 20 minutes – an extraordinarily long time at the very high temperature of the experiments. After the C5 is formed, it catalytically forms radicals, driving very rapid conversion of the methane. Matheu et al. used methods of predictive modeling quite similar to what

15

one would use today before experiments were performed, but in that case the experimental product data had been known for decades before the model was constructed.

In a more recent study, Gudiyella et al. [92] modeled a series of pilot plant experiments using a wide variety of feed compositions (mostly methane + X, but also some others) and several reaction conditions. The pilot plant reactor can be modeled fairly accurately as a series of simple CSTR and PFR reactors, but the system is extremely inhomogeneous: temperature inside the reactor varies by more than 1500 K from the hottest section to the quench zone.  The Gudiyella et al. study is more accurately described as 'predictive' than Matheu's studies because Gudiyella did not have access to experimental data from the pilot plant until after they had built the chemical kinetic model, comprised of 664 chemical species and 8,121 reactions (clearly a big detailed model). Although it contains ~30,000 parameters with predicted values, the Gudiyella et al. reactor model is not 100% predictive, since they used the experimentally measured rise in cooling water temperature in the pilot plant to calibrate one of the heat transfer parameters used in the model. But all the chemistry parameters are pure predictions.

The accuracy of the Gudiyella model can be judged by its predictions of the yield of $C_2H_2$, the main desired product, compared with the pilot plant measurements, see Fig. 2. Particularly noteworthy is how well the model predicts the changes in the yield of desired products as the feed composition is changed, in some cases quite dramatically. Those interested in this chemistry should also see the follow-on study of Chu et al. [93]. In Chu's study, more of the rate and thermo parameters were computed using high accuracy quantum chemistry methods, more reactions of high-molecular-weight species were included, and the chemical kinetic model was tested against much more detailed data from laboratory experiments, including measurements of minor byproducts up to $C_{12}H_8$. It is impressive that it is now possible to accurately predict the yield of C12 byproducts from a C1 feed – there are quite a large number of elementary-step reactions between C1 and C12, and the reaction network is complex including autocatalytic cycles. Apparently the computer program successfully identified at least the main reaction pathways connecting the reactants and products, and computed or estimated the crucial k's and $K_{eq}$'s to adequate accuracy.


C.   Example #2: Octane Improvers from Biomass

There is considerable interest in figuring out how best to use biomass as a transportation fuel, both to reduce greenhouse gas emissions from fossil fuel, and to monetize waste biomass to improve the rural economy. Since biomass is much more expensive to collect and transport than fossil fuels such as petroleum, it suffers from an economic disadvantage. But perhaps a fuel derived from biomass could have performance advantages that compensate for the poor economics? One variant on this idea is to use molecules derived from biomass as octane improvers in gasoline spark-ignited engines. Higher octane allows higher piston compression ratio, and so higher efficiency, making it possible for smaller,

16

lighter, cheaper engines to provide the same vehicle performance (acceleration). Ethanol and isobutanol are two commercial examples of octane-increasing gasoline components derived from biomass (plant starch). Many researchers are studying other octane-improving molecules derived from biomass, ideally from the non-food components of plants such as cellulose or lignin.

Zhang et al. [88] studied one class of biomass-derived molecules, substituted phenols derived from biomass pyrolysis. There are many different phenols which could be made this way; the key unknown question at the time of the study was which phenols were most effective at increasing octane. Zhang et al. constructed a model for the reactions occurring in the end gas of an engine cylinder at the conditions of the RON 100 test. Zhang et al. assumed a fuel containing 2% of any of six different phenols in butane, in a stoichiometric mixture with air. This system can be modeled as a simple homogeneous reactor, but with a strong time-variation in the pressure, local temperature, and local density due to the piston motion and the combustion occurring in other parts of the cylinder (as well as heat release from the partial oxidation occurring in the gases being modeled). The oxidation kinetics are even more nonlinear than the methane pyrolysis chemistry described in example #1, with a variety of coupled autocatalytic cycles forming radicals, each with high temperature sensitivity. In the engine case the main interest is to know the ignition delay (i.e. the induction period), since this correlates directly with the research octane number (RON).

The model built by Zhang et al. contained 1,465 species and 27,248 reactions. The model was built before the experimental data (measured in engine tests) were available; these are pure predictions. The goal is to predict ΔRON: how much adding 2% of each phenol to the base fuel changes the octane number. The correlation between the predicted changes in ignition delays and the experimental ΔRON numbers are shown in Fig. 3. The scales are chosen to be representative of the relationship between the two quantities.[88] The experimental RON measurements have uncertainties of about 0.5 of a RON unit. The model predictions are not quantitatively accurate, but appear to be sufficient to correctly identify the best additive and its approximate economic value as a octane-increasing gasoline additive.

## 4   Implications for the Future of Reaction Engineering
### A.   The Value of Predictive Mode

The nascent ability to do reaction engineering in "predictive mode", where a useful chemical kinetic model can be constructed, all its parameters computed, and the model solved, even before any experiments are performed, has the potential to dramatically change chemical engineering practice.

At present, most reaction engineering is done using the conventional postdictive simple-model paradigm. Someone considering developing a new chemical reaction process with that paradigm knows the models may not be reliable, and expects to spend a lot of money and time doing pilot-plant experiments before they will be able to evaluate whether or not the proposed process is feasible or

17

economically attractive.  (In contrast with reactions, many proposed separations can be evaluated satisfactorily using computer models calibrated by laboratory experiments, without the need for pilot plant experimentation.) This need for a significant investment of money and time, with no guarantee of success, is a significant barrier to innovation. Because it is so expensive and risky to try to develop a new chemical reaction process, many reaction processes engineered in the 1950s and 1960s are still commercially practiced, even though they almost certainly are not optimal today. Working in "predictive mode" using the detailed-model paradigm may make it cost-effective to change or optimize some of these processes. Even if it is not practical to predict accurate values for all the parameters in the model, the models can be very helpful for guiding the thinking of the researchers involved. In some cases the model allows one to infer upper limits on the performance of a process even if one does not know all the parameters accurately [97]. Knowing this upper bound can allow one to assess whether there is room for significant improvement in an existing process.

There are also many important processes which cannot be addressed effectively using the postdictive paradigm because it is impossible or too expensive to perform the experiments needed to determine the effective parameters. For example, many processes in the environment and in geochemistry occur on such long timescales that they are experimentally inaccessible – but it still may be possible to model them using "predictive mode".

In other cases, experiments are possible, but so expensive that model predictions might be preferred. For example, it would be very interesting to know which of all the many possible biofuel and solar fuel compositions would perform best in a high-efficiency jet engine, as a possible method of sustainable low-carbon civil aviation. Unfortunately, it is expensive to test even a single new fuel composition, in part because one needs to synthesize/manufacture a large volume of the new proposed fuel in order to perform the test, and also because of the high capital cost and significant operational/maintenance costs of running experiments with an instrumented advanced jet engine. The cost of such an experimental campaign to determine the best future jet fuel out of all the feasible compositions would be in the billions of dollars – not impossible, indeed probably worthwhile given the value to society of finding the right solution - but certainly a lot more than the budget available to most R&D managers.

Of course, while predictive mode is expected to gradually become more important, there are also many situations where the conventional postdictive approach (based heavily on fitting experiments) will continue to be preferred. For some systems it is not practical to build a sufficiently accurate predictive model, and in other cases the experimental data needed to build an empirical model are easy and inexpensive to obtain. Examples of the former include low-temperature reactions involving transition-metal-containing heterogeneous catalysts (since currently there is no affordable quantum chemistry method to predict the reaction barriers to sufficient accuracy) or systems with poorly characterized catalysts or feeds. Examples of the latter include systems amenable to high-throughput experimentation, particularly microfluidic systems where there is no difficulty in ensuring good mixing and temperature uniformity, so the inferred empirical rate expression $\omega(\mathbf{C},T)$ is free of artifacts due to

18

inhomogeneity. In a few systems there is a vast amount of experimental data freely available (e.g. from sensors in chemical plants), and it may be more convenient and accurate to use rather complicated empirical models derived using Machine Learning methods instead of using any of the kinetic modeling approaches discussed in this paper.

So for the next few decades, chemical engineers will need to be able to work using several rather different modeling paradigms.

### B. Change in Focus of Reaction Engineering

Historically and still today, the focus of Reaction Engineering has been mostly on the reactor design and related transport/mixing issues, not on the reaction chemistry. This is reasonable if one expects to model the reactions using a simplistic A $\rightarrow$ B $\rightarrow$ C model, with all the rate parameters derived from fitting to experiments. However, this classic approach typically does not take advantage of the chemical knowledge already available, nor of our growing ability to accurately and inexpensively compute key parameters from first principles. Also, because of the efforts of many chemical engineers over the last 100 years, the conventional reaction engineering step is now often quite straightforward, not the intellectual and computational challenge it was in the 20th century.

This all suggests that it is time for the focus of Reaction Engineering to shift, to try to bring in more chemistry knowledge, including methods for rapidly computing rate parameters from first principles. This will allow our field to address a broad range of important questions that were sometimes considered out-of-scope in academic Reaction Engineering, e.g.

1) What would happen if the feed composition was changed to include a new molecule (e.g. an additive)?
2) What would happen if the feed were changed to different molecules?
3) Can we predict what would happen if the catalyst were changed?
4) Is there a fundamental reason why my process always makes undesired impurity X? If not, what should I do to prevent production of X?
5) How could one change the process conditions, the catalyst, and the feed to obtain optimal results (high yield, low impurities, low energy input, low emissions)? Is the current process stuck in a local optimum, but not the globally optimal process?

These are often the main questions in industry, in part because today's methods of Reaction Engineering have already resolved many of the questions about reactor design and local optimization of a process. These new questions should be the focus of a new century of Reaction Engineering research.

### C. How to make this big shift in the Focus of Reaction Engineering?

Current undergraduate Chemical Engineers are generally well-trained in the conventional simple-model postdictive paradigm, and comfortable taking advantage of the well-developed methods for reactor

design. So they are well-prepared to tackle the problems that formed the focus of classical Reaction Engineering using simple models for the chemistry.

However, most undergraduate chemical engineers have gaps in their education that make it difficult for them to use the higher-fidelity predictive (or hybrid) approaches to modeling chemistry. To start with, most chemical engineers are not that comfortable with chemistry details. Perhaps they only took two or three semesters of chemistry as freshmen and sophomore undergraduates, and probably many chemical engineers haven't opened a chemistry textbook since turning age 21. Similarly, most of today's practicing chemical engineers have probably never computed a rate coefficient from first principles, and don't have familiarity with modern software for kinetic modeling, constructing reaction mechanisms, or estimating rate parameters. Even fewer have experience using machine learning techniques for estimating molecular and reaction parameters from large data sets.

To effectively take advantage of the new approach, not only do the predictive-kinetics techniques need to be well-developed and demonstrated, with convenient software, but also a significant effort will be required in education and training, including changes to the curriculum of Reaction Engineering. Today almost all practicing reaction engineers are doing their design and simulation work on the computer, and many have access to excellent software for solving large chemical kinetic models in all the conventional reactor types. Most also have access to excellent software for optimization and for fitting. In most cases the main challenge is therefore how to obtain/construct the kinetic model: once one has a good chemistry model, the design problem can be straightforwardly solved on the computer.

However, our current curriculum almost completely ignores this issue – usually the kinetic model and rate parameters are given in the problem statement – and instead we teach the students how to solve reaction engineering problems by hand, often by making a lot of approximations. There is certainly a lot of pedagogical value in doing a calculation by hand, but it is a shame to spend a lot of time teaching calculation methods that students will never actually do in practice, while ignoring the key problem they will actually face. Of course, making a major change in curriculum is always difficult, and great care is needed since the current chemical engineering curriculum has proven its value over many decades. But unless the chemical engineering curriculum changes, it will be hard to change how Reaction Engineering is done in practice.

Some may say that it is not the chemical engineer's job to develop the chemistry model, instead have a chemist do that. This is an option, but most of today's chemists are even less prepared to build and use large computer models for chemical kinetics than the engineers. Besides, we Reaction Engineers are the ones who 'own' the problem of designing a reaction systems; it behooves us to learn how to solve the Reaction Engineering problem in the best way possible, even if it involves the use of new tools and techniques, and even if it requires more attention to chemistry.

## 5 Summary

20

Recent advances in capabilities for constructing accurate reaction mechanisms (described in the AIChE Perspective by the 2018 Wilhelm Award winner [3] and elsewhere [8-16, 99]), combined with advances in quantum chemistry, rate theory, numerical methods, machine learning, and in computer power, have now made it possible to address many Reaction Engineering problems using much more accurate models.  These detailed kinetic models have much higher fidelity to the actual chemistry occurring in the system than was practical in the past. Iterative computer construction of the reaction mechanisms provides a systematic method for reducing the Model Truncation Error due to omission in the model of some of the reactions and intermediates that exist in the real system. Good initial estimates of the parameters, followed by iterative refinement of the most sensitive parameters, allows a systematic way to reduce the effects of Parameter Error on the model predictions.  For some systems, all the key model parameters can be computed/estimated accurately before doing any experiments, allowing a complete change in paradigm to 'predictive mode', with many advantages especially for design of new systems. For systems where this approach is feasible, it is highly recommended. At present this works very well for high-temperature gas-phase systems, as shown in the Examples, but there is reason to believe it will soon be practical to use this approach at somewhat lower temperatures and in liquid phase systems.

However, for many important systems it is not yet possible/practical to accurately compute/estimate all the important parameters. In some cases this is due to the large number of species in the system.  To some extent this complexity issue can be overcome by better use of supercomputers, but for systems containing large molecules and unselective reactions different modeling approaches are needed.[50,51,98] A more fundamental and important challenge is the lack of high-accuracy methods for computing the barriers of reactions on heterogeneous catalysts. There are also similar problems with computing ionic reactions in condensed phase to high accuracy. Hopefully ongoing improvements in quantum chemistry techniques and in methods for including condensed-phase effects on reaction parameters, along with the continued growth in computing power, will soon make it practical to predict these reactions to the same high accuracy as is now possible for reactions of neutral organic molecules and radicals.

However, even after it becomes practical to accurately compute the barrier for any specific reaction, there will still be significant uncertainties due to incomplete knowledge of the boundary conditions: feed composition, initial morphology and condition of the catalyst. In some cases, the catalyst is known to change morphology during the reaction, making the simulations even more challenging. Because of these issues, it will remain difficult to make accurate predictions for some systems, including many catalytic systems, long after it becomes routine to use supercomputers to enumerate all the reactions and accurately compute all the rate parameters on well-ordered surfaces.

Here we have highlighted some of the many technical as well as pedagogical/training challenges that must be overcome as Reaction Engineering transitions away from reliance on highly empirical postdictive reaction rate models (based on expensive experimentation often in pilot plants), to instead use higher-fidelity detailed chemistry models using more fundamental rate parameters. We hope this

21

presentation of both the current capabilities of Predictive Kinetics and its remaining challenges will inspire future efforts to achieve its promise.

References

[1] Kee RJ, Coltrin ME, Glarborg P. Chemically Reacting Flow: Theory and Practice. Hoboken, NJ: John Wiley & Sons, 2003.

[2] Fox RO. Computational Models for Turbulent Reacting Flows. Cambridge University Press, 2003.

[3] Vernuccio S, Broadbelt, LJ. Discerning complex reaction networks using automated generators. AIChE J. 2019; 65:e16663 https://doi.org/10.1002/aic.16663

[4] Levenspiel O. Chemical Reaction Engineering. Ind Eng Chem Res 1999; 38:4140-4143. https://doi.org/10.1021/ie990488g

[5] Russi T, Packard A, Frenklach M. Uncertainty quantification: Making predictions of complex reaction systems reliable. Chem Phys Lett 2010; 499:1-8. https://www.sciencedirect.com/science/article/abs/pii/S0009261410012212

[6] Singer AB, Taylor JW, Barton PI, Green WH. Global Dynamic Optimization for Parameter Estimation in Chemical Kinetics. Journal of Physical Chemistry A 2006; **110**:971-976. https://doi.org/10.1021/jp0548873

[7] Yablonsky G. Comments on a Commentary by Prof. Levenspiel. Ind Eng Chem Res 2000; 39:3120. https://doi.org/10.1021/ie001098t

22

[8] Warth V, Battin-Leclerc F, Fournet R, Glaude PA, Come GM, Scacchi, G. Computer based generation of reaction mechanisms for gas-phase oxidation. *Computers and Chemistry* 2000; 24:541-560.

[9] Gao CW, Allen JW, Green WH, West RH. Reaction Mechanism Generator: Automatic Construction of Chemical Kinetic Mechanisms. Comp. Phys. Comm. 2016; 203:212-225. doi: 10.1016/j.cpc.2016.02.013

[10] Battin-Leclerc F, Simmie JM, Blurock E. Cleaner Combustion – Developing Detailed Chemical Kinetic Models, Green Energy and Technology. Springer **2013**. ISBN: 978-1-4471-5306-1, doi: 0.1007/978-1-4471-5307-8.

[11] Broadbelt LJ, Stark SM, Klein MT. Computer generated reaction modelling: Decomposition and encoding algorithms for determining species uniqueness. *Computers and Chemical Engineering* 1996; 20:113-129.

[12] Susnow RG, Dean AM, Green WH, Peczak P. Rate-Based Construction of Kinetic Models for Complex Systems. Journal of Physical Chemistry A 1997; 101:3731-3940.

[13] Van de Vijver R, Vandewiele NM, Bhoorasingh PL, Slakman BL, Seyedzadeh Khanshan F, Carstensen H-H, Reyniers M-F, Marin GB, West RH, Van Geem KM, Automatic mechanism and kinetic model generation for gas-and solution-phase processes: a perspective on best practices, recent advances, and future challenges. Int J Chem Kinet 2015; 47:199-231. https://doi.org/10.1002/kin.20902

[14] Vandewiele NM, Van Geem KM, Reyniers M-F, Marin GB. Genesys: Kinetic Model Construction using Chem-Informatics. Chem Eng J 2012; 207:526-538.

[15] Gupta U, Le T, Hu W-S, Bhan A, Daoutidis P. Automated network generation and analysis of biochemical reaction pathways using RING. Metabolic Engineering. 2018; 49:84-93.

[16] Goldsmith CF, West RH. Automatic Generation of Microkinetic Mechanisms for Heterogeneous Catalysis. Journal of Physical Chemistry C. 2017; 121:9970-9981. https://doi.org/10.1021/acs.jpcc.7b02133

[17] Ranzi E, Dente M, Goldaniga A, Bozzano G, Faravelli T. Lumping procedures in detailed kinetic modeling of gasification, pyrolysis, partial oxidation and combustion of hydrocarbon mixtures. Progress in Energy and Combustion Science. 2001; 27:99-139. http://doi.org/10.1016/S0360-1285(00)00013-7

[18] Konnov AA. Yet another kinetic mechanism for hydrogen combustion. Combustion and Flame. 2019; 203:14–22. https://doi.org/10.1016/j.combustflame.2019.01.032

[19] Benson SW. Thermochemical Kinetics (2nd ed.). New York: John Wiley & Sons, 1976.

23

[20] Burke MP, Klippenstein SJ, Ephemeral collision complexes mediate chemically termolecular transformations that affect system chemistry. Nature Chemistry 2017; 9:1078.

[21] Jasper AW, Miller JA. Lennard–Jones parameters for combustion and chemical kinetics modeling from full-dimensional intermolecular potentials. Combustion and Flame 2014; 161:101-110. https://doi.org/10.1016/j.combustflame.2013.08.004

[22] Sabbe MK, Van Geem KM, Reyniers M-F, Marin GB. First principle-based simulation of ethane steam cracking. AIChE Journal 2011; 57:482-496. **https://doi.org/10.1002/aic.12269**

[23] Keceli M , Elliott S, Li Y-P, Johnson MS, Cavallotti C, Georgievskii Y, Green WH, Pelucchi M, Wozniak JM, Jasper AW, Klippenstein SJ. Automated Computational Thermochemistry for Butane Oxidation: A Prelude to Predictive Automated Combustion Kinetics. Proceedings of the Combustion Institute 2019; **37**: 363-371. https://doi.org/10.1016/j.proci.2018.07.113

[24] Van de Vijver R, Zádor J. KinBot: Automated stationary point search on potential energy surfaces, Computer Physics Communication, 2019; 106947. https://doi.org/10.1016/j.cpc.2019.106947

[25] Cavallotti C, Pelucchi M, Georgievskii Y, Klippenstein SJ. EStokTP: Electronic Structure to Temperature- and Pressure-Dependent Rate Constants-A Code for Automatically Predicting the Thermal Kinetics of Reactions J Chem Theory Comput, 2019; 15(2):1122-1145. **doi: 10.1021/acs.jctc.8b00701**

[26] Jacobson LD, Bochevarov AD, Watson MA, Hughes TF, Rinaldo D, Ehrlich S, Steinbrecher TB, Vaitheeswaran S, Philipp DM, Halls MD, Friesner RA. Automated Transition State Search and Its Application to Diverse Types of Organic Reactions. J Chem Theory Comput 2017; 13: 5780–5797.

https://doi.org/10.1021/acs.jctc.7b00764

[27] Dewyer AL, Arguelles AJ, Zimmerman PM. Methods for exploring reaction space in molecular systems. WIREs Computational Molecular Science 2018; 8: e1354 **https://doi.org/10.1002/wcms.1354**

[28] Grambow CM, Jamal A, Li YP, Green WH, Zador J, Suleimanov YV, Unexpected Unimolecular Reaction Pathways of a γ-Ketohydroperoxide from Combined Application of Automated Reaction Discovery Methods. Journal of the American Chemical Society 2018; 140:1035-1048. http://dx.doi.org/10.1021/jacs.7b11009

[29] Klamt A, Diedenhofen M. Calculation of Solvation Free Energies with DCOSMO-RS. J. Phys. Chem. A 2015; 119(21): 5439–5445. https://doi.org/10.1021/jp511158y

[30] Mellmer MA, Sanpitakseree C, Demir B, Bai P, Ma K, Neurock M, Dumesic JA. Solvent-enabled control of reactivity for liquid-phase reactions of biomass-derived compounds. Nature Catalysis. 2018; 1:199-207.

[31] Sumathi R, Green WH. *A priori* Rate Constants for Kinetic Modeling. Theoretical Chemistry Accounts 2002; 108:187-213.

[32] Allen JW, Goldsmith CF, Green WH. Automatic Estimation of Pressure-Dependent Rate Coefficients. Physical Chemistry Chemical Physics 2012; 14(3):1131-1155.

[33] Chung Y, Gillis RJ, Green WH. Temperature-dependent solvation free energy estimation from minimal data. AIChE Journal 2020; 66:e16976. **https://doi.org/10.1002/aic.16976**

[34] Lay, TH, Bozzelli JW, Dean AM, Ritter ER. HYDROGEN ATOM BOND INCREMENTS (HBI) FOR CALCULATION OF THERMODYNAMIC PROPERTIES OF HYDROCARBON RADICAL SPECIES. Journal of Physical Chemistry 1995; 99(39):14514-14527. DOI: 10.1021/j100039a045

[35] Grambow CM, Li YP, Green WH. Accurate Thermochemistry with Small Datasets: A Bond Additivity Correction and Transfer Learning Approach. Journal of Physical Chemistry A 2019; 123:5826-5835. https://doi.org/10.1021/acs.jpca.9b04195

[36] Grambow CM, Pattanaik L, Green WH. Deep Learning of Activation Energies. Journal of Physical Chemistry Letters 2020; 11:2992-2997. https://dx.doi.org/10.1021/acs.jpclett.0c00500

[37] Buerger P, Akroyd J, Mosbach S, Kraft M. A systematic method to estimate and validate enthalpies of formation using error-cancelling balanced reactions. Combust Flame 2017; 187:105-121. DOI: 10.1016/j.combustflame.2017.08.013

[38] Schwer DA, Tolsma JA, Green WH, Barton PI. On Upgrading the Numerics in Combustion Chemistry Codes. Combustion and Flame 2002; 128:270-291.

[39] Sewerin F, Rigopoulos S. A methodology for integration of stiff chemical kinetics on GPU. Combust Flame 2015; 162:1375-1394. https://doi.org/10.1016/j.combustflame.2014.11.003

[40] Shi Y, Green WH, Wong HW, Oluwole OO. Redesigning Combustion Modeling Algorithms for the Graphics Processing Unit (GPU): Chemical Kinetic Rate Evaluation and Ordinary Differential Equation Integration. Combustion & Flame 2011; 158(5):836-847.

[41] Long AE, Speth RL, Green WH. Ember: an open-source, transient solver for 1-d reacting flow using large kinetic models, applied to strained extinction. Combustion & Flame 2018; 195:105-116. https://doi.org/10.1016/j.combustflame.2018.05.001

[42] McNenly MJ, Whitesides R, Flowers D. Faster solvers for large kinetic mechanisms using adaptive preconditioners. Proc Combust Inst 2015; 35(1):581-587. DOI: 10.1016/j.proci.2014.05.113

[43] Morii Y, Terashima H, Koshi M, Shimizu T, Shima E. ERENA: A fast and robust Jacobian-free integration method for ordinary differential equations of chemical kinetics. J Comput Phys 2016; 322:547-558. https://doi.org/10.1016/j.jcp.2016.06.022

[44] Yoo CS, Luo Z, Lu T, Kim H, Chen JH. A DNS study of ignition characteristics of a lean iso-octane/air mixture under HCCI and SACI Conditions. Proc Combust Inst 2013; 34(2):2985-2993. https://doi.org/10.1016/j.proci.2012.05.019

[45] Tosatto L, Bennett BAV, Smooke MD. Parallelization strategies for an implicit Newton-based reactive flow solver. Combustion Theory and Modelling 2011; 15:455–486.

[46] Kundu P, Ameen MM, Som S. Importance of turbulence-chemistry interactions at low temperature engine conditions. Combust Flame 2017; 183:283-298. https://doi.org/10.1016/j.combustflame.2017.05.025

[47] Lei H, Duan Y. Kinetics and Thermodynamics of Protein Folding, Thermodynamics - Kinetics of Dynamic Systems, Juan Carlos Moreno Piraján, IntechOpen, 2011. DOI: 10.5772/23324. https://www.intechopen.com/books/thermodynamics-kinetics-of-dynamic-systems/kinetics-and-thermodynamics-of-protein-folding

[48] Han K, Green WH, West RH. On-the-Fly Pruning for Rate-Based Reaction Mechanism Generation. Computers and Chemical Engineering 2017; 100:1-8. http://dx.doi.org/10.1016/j.compchemeng.2017.01.003

[49] Jocher A, Vandewiele NM, Han K, Liu M, Gao CW, Gillis RJ, Green WH. Scalability strategies for automated reaction mechanism generation. Computers & Chemical Engineering 2019; 131:106578. https://doi.org/10.1016/j.compchemeng.2019.106578

[50] Quann RJ, Jaffe SB. Structure-Oriented Lumping: describing the chemistry of complex hydrocarbon mixtures. Ind Eng Chem Res 1992; 31:2483–2497. https://doi.org/10.1021/ie00011a013

[51] Han K, Green WH. A Fragment-based Mechanistic Kinetic Modeling Framework for Complex Systems. *Industrial & Engineering Chemistry Research* 2018; **57**:14022–14030. https://pubs.acs.org/doi/abs/10.1021/acs.iecr.8b02870

[52] Mastan E, Zhu S. Method of Moments: a versatile tool for modeling polymerization kinetics. European Polymer Journal 2015; 68:139-160. https://doi.org/10.1016/j.eurpolymj.2015.04.018

[53] Sokolovski D, Rusconi S, Akhmatskaya E, Asua JM. Non-Markovian models of the growth of a polymer chain. Proc Royal Soc A 2015; 471:20140899. https://doi.org/10.1098/rspa.2014.0899

[54] Ince A, Carstensen HH, Reyniers MF, Marin GB. First-principles based group addivity values for thermochemical properties of substituted aromatic compounds. AIChE J 2015; 61:3858-3870. **https://doi.org/10.1002/aic.15008**

[55] Han K, Jamal A, Grambow CM, Buras ZJ, Green WH. An Extended Group Additivity Method for Polycyclic Thermochemistry Estimation. International Journal of Chemical Kinetics 2018; 50:294-303. **https://doi.org/10.1002/kin.21158**

[56] Yu J, Sumathi R, Green WH. Accurate and Efficient Method for Predicting Thermochemistry of Polycyclic Aromatic Hydrocarbons. Journal of the American Chemical Society 2004; 126(39):12685-12700.

[57] Vandeputte AG, Sabbe M, Reyniers MF, Marin GB. Modeling the gas-phase thermochemistry of organosulfur compounds. Chemistry Eur J 2011; 17:7656-7673. **https://doi.org/10.1002/chem.201002422**

[58] Gillis RJ, Green WH. Thermochemistry Prediction and Automatic Mechanism Generation for Oxygenated Sulfur Systems: A Case Study of Dimethyl Sulfide Oxidation. ChemSystemsChem 2020; 2: e1900051. http://dx.doi.org/10.1002/syst.201900051


[59] Da Silva G, Moore EE, Bozzelli JW. Quantum chemical study of the structure and reactivity of the five-membered nitrogen-containing heterocycles and their anions and radicals. J Phys Chem A 2006; 110:13979-13988.

[60] Wong HW, Alva Nieto JC, Swihart MT, Broadbelt LJ. Thermochemistry of Silicon-Hydrogen compounds generalized from quantum chemical calculations. J Phys Chem A 2004; 108:874-897.

[61] Yalamanchi KK, van Oudenhoven VCO, Tutino F, Monge-Palacios M, Alshehri A, Gao X, Sarathy M. *J. Phys. Chem. A* 2019; 123:8305–8313. Machine Learning to Predict Standard Enthalpy of Formation of Hydrocarbons. https://doi.org/10.1021/acs.jpca.9b04771


[62] Li YP, Han K, Grambow CA, Green WH. Self-evolving machine: A Continuously Improving Model for Molecular Thermochemistry. *J. Phys. Chem. A* 2019; **123**:2142-2152. https://pubs.acs.org/doi/abs/10.1021/acs.jpca.8b10789

[63] Scalia G, Grambow CM, Pernici B, Li YP, Green WH. Evaluating Scalable Uncertainty Estimation Methods for DNN-Based Molecular Property Prediction. Journal of Chemical Information and Modeling 2020; 60:2697–2717. https://doi.org/10.1021/acs.jcim.9b00975

[64] Deniz CU, Yasar M, Klein MT. Stochastic reconstruction of complex heavy oil molecules using an artificial neural network. Energy Fuels 2017; 31(11):11932–11938. https://doi.org/10.1021/acs.energyfuels.7b02311

[65] Kee, RJ, Rupley, FM, Miller, JA. Chemkin-II: A Fortran chemical kinetics package for the analysis of gas-phase chemical kinetics. 1989. https://www.osti.gov/biblio/5681118-chemkin-ii-fortran-chemical-kinetics-package-analysis-gas-phase-chemical-kinetics. CHEMKIN is distributed commercially by ANSYS.

[66] David G. Goodwin, Raymond L. Speth, Harry K. Moffat, and Bryan W. Weber. Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. https://www.cantera.org, 2018. Version 2.4.0. doi:10.5281/zenodo.1174508

[67] Strang G. On the construction and comparison of difference schemes. SIAM Journal on Numerical Analysis 1968; 5:506-517.

[68] Schwer DA, Lu P, Green WH, Semiao V. A Consistent-Splitting Approach to Computing Stiff Steady-State Reacting Flows with Adaptive Chemistry. *Combustion Theory and Modelling 2003;* **7**:383-399.


[69] Speth RL, MacNamara S, Green WH, Strang G. Balanced Splitting and Rebalanced Splitting. SIAM Journal on Numerical Analysis 2013; 51:3084–3105.

[70] Lu T, Law CK, Soo CS, Chen JH. Dynamic stiffness removal for direct numerical simulations. Combust Flame 2009; 156: 1542-1551. https://doi.org/10.1016/j.combustflame.2009.02.013

[71] Mitsos A, Oxberry GM, Barton PI, Green WH. Optimal Automatic Reaction and Species Elimination in Kinetic Mechanisms. Combustion & Flame 2008; 155:118-132.


[72] Goussis DA, Maas U. Model Reduction for Combustion Chemistry. In: Echekki T., Mastorakos E. (eds) Turbulent Combustion Modeling. Fluid Mechanics and Its Applications 2011; vol. 95. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0412-1_9

[73] Schwer DA, Lu P, Green WH. An Adaptive Chemistry Approach to Modeling Complex Kinetics in Reacting Flows. Combustion & Flame 2003; 133:451-465.


[74] Lovas T, Mauss F, Hasse C, Peters N. Modeling of HCCI combustion using adaptive chemical kinetics. SAE 2002-01-0426. https://doi.org/10.4271/2002-01-0426

[75] Lu T, Law CK. On the applicability of directed relation graphs to the reduction of chemical mechanisms. Combust Flame 2006; 146:472-483. https://doi.org/10.1016/j.combustflame.2006.04.017

[76] Pepiot P, Pitsch H. Systematic Reduction of Large Chemical Mechanisms. Presented at the 4th Joint Meeting of the US Sections of the Combustion Institute, Philadelphia, 2005. https://web.stanford.edu/group/pitsch/publication/PepiotJMUSSCI2005.pdf

[77] Zheng XL, Lu T, Law CK. Experimental counterflow ignition temperatures and reaction mechanisms of 1,3-butadiene. Proc Combust Inst 2007; 31:367-375. doi:10.1016/j.proci.2006.07.182

[78] Liang L, Stevens JD, Farrell JT. A dynamic adaptive chemistry scheme for reactive flow calculations. Proc Combust Inst 2009; 32(1):527-534. DOI:10.1016/j.proci.2008.05.073

[79] Gou X, Chen A, Sun W, Ju Y. A dynamic adaptive chemistry scheme with error control for combustion modeling with a large detailed mechanismCombust Flame 2013, 160:225-231. https://doi.org/10.1016/j.combustflame.2012.10.015

[80] Fridlyand A, Johnson MS, Goldsborough SS, West RH, McNenly MJ, Mehl M, Pitz WJ. The role of correlations in uncertainty quantification of transportation relevant fuel models. Combust Flame 2017; 180:239-249. https://doi.org/10.1016/j.combustflame.2016.10.014

[81] Gao CW, Liu M, Green WH. Uncertainty Analysis of Correlated Parameters in Automated Reaction Mechanism Generation. International Journal of Chemical Kinetics 2020; 52:266-282. http://dx.doi.org/10.1002/kin.21348

[82] Nagy T, Turanyi T. Uncertainty of Arrhenius Parameters. Int J Chem Kinet 2011; 43:359-378. DOI: 10.1002/kin.20551

[83] Ruscic B, Pinzon RE, Morton ML, von Laszevski G, Bittner SJ, Nijsure SG, Amin KA, Minkoff M, Wagner AF. Introduction to Active Thermochemical Tables. *J. Phys. Chem. A* 2004; 108:9979–9997 https://doi.org/10.1021/jp047912y

[84] Anatharaman B, Melius CF. Bond-Additivity Corrections for G3B3 and G3MP2B3 Quantum Chemistry Methods. *J. Phys. Chem. A* 2005; 109:1734–1747. https://doi.org/10.1021/jp045883l

[85] Petersson GA, Malick DK, Wilson WG. Calibration and comparison of the Gaussian-2, complete basis set, and density functional methods for computational thermochemistry. J. Chem. Phys. 1998; 109:10570. https://doi.org/10.1063/1.477794

[86] Green WH. Automatic Generation of Reaction Mechanisms. in "Mathematical Modeling of Complex Reaction Systems", ed. by Tiziano Faravelli, Flavio Manenti, and Eliseo Ranzi, *Computer-Aided Chemical Engineering* 2019; **45**: 259-294. https://www.sciencedirect.com/bookseries/computer-aided-chemical-engineering/vol/45/suppl/C

29

[87] Song J, Stephanopoulos G, Green WH. Valid Parameter Range Analyses for Chemical Reaction Kinetic Models. *Chemical Engineering Science* 2002; **57**: 4475

[88] Zhang P, Yee NW, Filip SV, Hetrick CE, Yang B, Green WH. Modeling Study of the anti-knock tendency of substituted phenols as additives: An application of the Reaction Mechanism Generator (RMG). Physical Chemistry Chemical Physics 2018; 20:10637-10649.
https://dx.doi.org/10.1039/C7CP07058F

[89] Hansen N, Merchant SS, Harper MR, Green WH. The Predictive Capability of an Automatically Generated Combustion Chemistry Mechanism: Chemical Structures of Premixed *iso*-Butanol Flames. Combustion and Flame 2013; 160: 2343-2351.

[90] Matheu DM, Dean AM, Grenda JM, Green WH. Mechanism Generation with Integrated Pressure-Dependence: A New Model for Methane Pyrolysis. Journal of Physical Chemistry A 2003; 107:8552-8565.

[91] Back MH, Back RA. Thermal decomposition and reaction of methane. L.F. Albright, B.L. Crynes, W.H. Corcoran (Eds.), Pyrolysis: Theory and Industrial Practice, Academic Press, New York (1983), pp. 1-24

[92] Gudiyella S, Buras ZJ, Chu TC, Lengyel I, Pannala S, Green WH. A modeling study of high-temperature pyrolysis of natural gas. Industrial & Engineering Chemistry Research 2018; 57(22):7404–7420. https://pubs.acs.org/doi/abs/10.1021/acs.iecr.8b00758

[93] Chu, TC, Buras ZJ, Oßwald P, Liu M, Goldman MJ, Green WH. Modeling of aromatics formation in fuel-rich methane oxy-combustion with an automatically generated pressure-dependent mechanism. Phys. Chem. Chem. Phys. 2019; 21:813-832. https://doi.org/10.1039/C8CP06097E

[94] Broadbelt LJ, Stark SM, Klein MT. Computer-generated reaction networks: on-the-fly calculation of species properties using computational quantum chemistry. Chem Eng Sci 1994; 49:4991-5010.
https://doi.org/10.1016/0009-2509(94)00326-2

[95] Magoon GR, Green WH. Design and implementation of a next-generation software system for on-the-fly quantum and force field calculations in automated reaction mechanism generation. Computers & Chemical Engineering 2013; 52:35-45.

[96] Klinke DJ, Broadbelt LJ. Mechanism reduction during computer generation of compact reaction models. AIChE J 1997; 43:1828-1837. **https://doi.org/10.1002/aic.690430718**

30

[97] Su YS, Ying J, Green WH. Upper Bound on the Yield for Oxidative Coupling of Methane. J Catal 2003; 218:321-333.

[98] Klein MT, Hou G, Bertolacini R, Broadbelt LJ, Kumar A. Molecular Modeling in Heavy Hydrocarbon Conversions. Taylor & Francis (Boca Raton) 2006. ISBN 0-8247-5851-X

[99] Rangarajan S, Bhan A, Daoutidis P. Language-Oriented Rule-Based Reaction Network Generation and Analysis: Description of RING. Comp. Chem. Eng. 2012; **45**:114.

[100] Vora N, Daoutidis P. Nonlinear Model Reduction of Chemical Reaction Systems. *AIChE J*. 2001; 47:2320. **https://doi.org/10.1002/aic.690471016**

[101] Aris R. Prolegomena to the rational analysis of systems of chemical reactions. Arch. Ration. Mech. Anal. 1965; 19(2):81–99.

[102] Gorban AN, Yablonsky GS. Three Waves of Chemical Dynamics. Math. Model. Nat. Phenom. 2015; 10(5):1–5. DOI: 10.1051/mmnp/201510501

[103] Marin GB, Yablonsky GS, Constales D. Kinetics of chemical reactions 2$^{nd}$ edition. Wiley-VCH 2019.

[104] Ranzi E, Pierucci S, Dente M, van Goetham M, van Meeuwen D, Wagner E. Correct Molecular Reconstruction of Cracking Feeds: a Need for the Accurate Prediction of Ethylene Yields. Chem. Eng. Trans. 2015; 43:871-876. DOI: 10.3303/CET1543146

[105] Conjeevaram Krishnakumar NK. A Bayesian approach to feed reconstruction. M.S. thesis, MIT 2013. https://hdl.handle.net/1721.1/82414

[106] Pernalete CG, van Baten J, Urbina JC, Arevalo JF. A molecular reconstruction feed characterization and CAPE OPEN implementation strategy to develop a tool for modeling HDT reactors for light petroleum cuts. Computer Aided Chem. Eng. 2015; 37:359-364.

[107] Wilhelm ME, Le AV, Stuber MD. Global Optimization of Stiff Dynamical Systems. AIChE J. 2019; 65(12):e16836. Doi.org/10.1002/aic.16836

[108] Papamichail I, Adjiman CS. Global Optimization of Dynamic Systems. Comp. Chem. Eng. 2004; 28(3):403-415.

32

Table 1 "conventional postdictive" vs. "detailed predictive" rate expressions for propane cracking
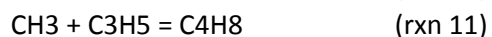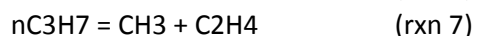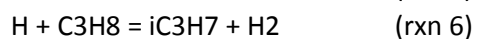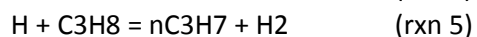
***Conventional (postdictive)***

$r_{C3H6} = A_1 \exp(-E_1/RT) [C_3H_6]^{\alpha_1} [C_2H_4]^{\beta_1} [H_2]^{\gamma_1}$

$r_{C2H4} = A_2 \exp(-E_2/RT) [C_3H_6]^{\alpha_2} [C_2H_4]^{\beta_2} [H_2]^{\gamma_2}$

$r_{H2} = A_3 \exp(-E_2/RT) [C_3H_6]^{\alpha_2} [C_2H_4]^{\beta_2} [H_2]^{\gamma_2}$

3 state variables and equations. 15 parameters to fit.

***Detailed (predictive)***

| | |
|---|---|
| C3H8 = C2H5 + CH3 | (rxn 1) |
| C2H5 = H + C2H4 | (rxn 2) |
| CH3 + C3H8 = nC3H7 + CH4 | (rxn 3) |
| CH3 + C3H8 = iC3H7 + CH4 | (rxn 4) |
| H + C3H8 = nC3H7 + H2 | (rxn 5) |
| H + C3H8 = iC3H7 + H2 | (rxn 6) |
| nC3H7 = CH3 + C2H4 | (rxn 7) |
| iC3H7 = C3H6 + H | (rxn 8) |
| C3H6 = H + C3H5 | (rxn 9) |
| CH3 + CH3 = C2H6 | (rxn 10) |
| CH3 + C3H5 = C4H8 | (rxn 11) |
| CH3 + H = CH4 | (rxn 12) |
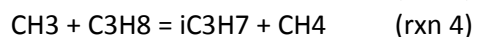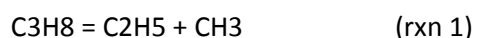
13 state variables and equations. 12 k(T,P)'s, 13 NASA polynomials: 206 parameters. Too many to determine by fitting to experimental data.

33

Fig. 1 Proposed workflow for automated development of predictive models for reaction chemistry, where the quantum calculations reduce Parameter Errors and the reaction mechanism expansions by RMG reduce the Mechanism Truncation Error. At the time this is written, the steps in the solid arrow are fully automated, but the steps with dashed arrows are only semi-automated. Several groups are working to make the loop fully automatic, so it would be practical to go through the full cycle hundreds or thousands of time, to automatically build accurate predictive chemical kinetic models.
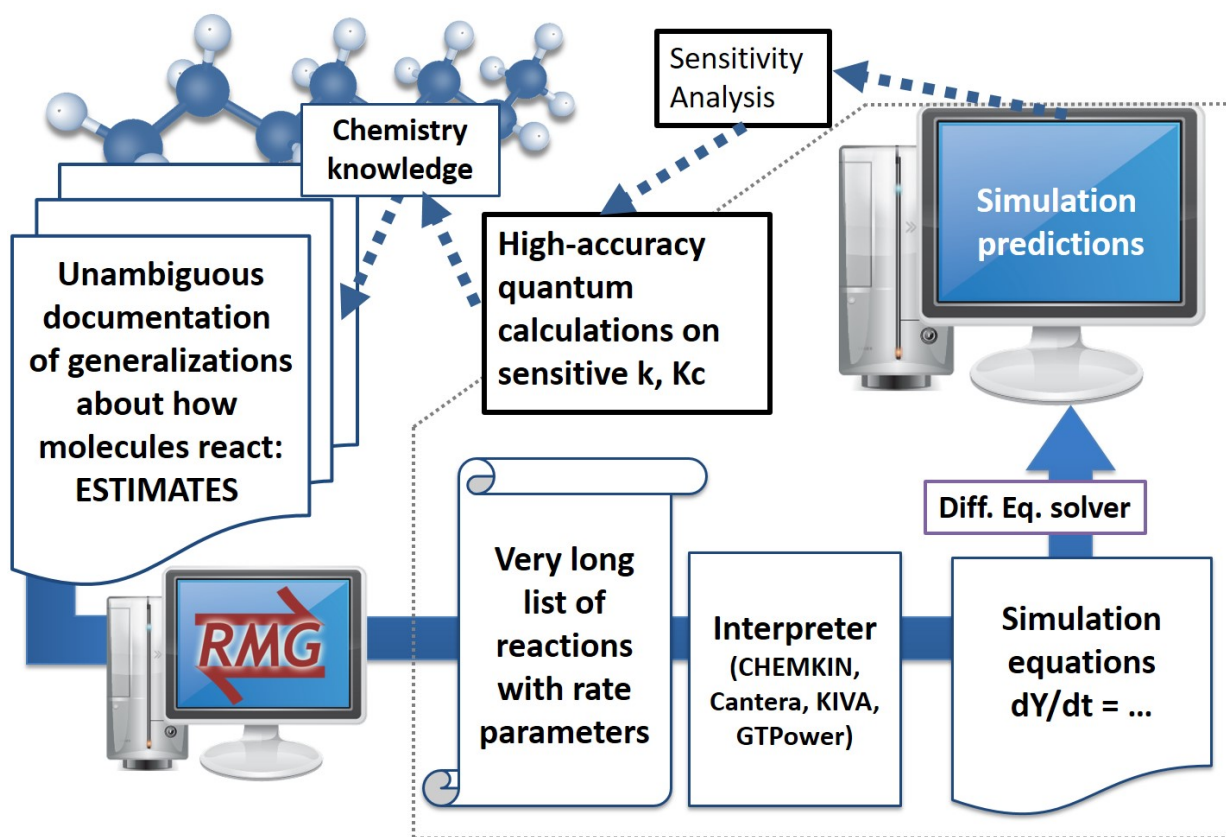
Fig. 2 Predicted vs. Measured Yields of the $C_2H_2$ product in the high-temperature pyrolysis of natural gas and several other feeds at various pilot plant reaction conditions, from a model constructed using RMG in summer 2017. The dashed lines indicates a plus or minus 50% discrepancy between the prediction and the measurement. In all cases the predicted $C_2H_2$ yields are within a factor of 2 of what was measured in the pilot plant, and in most cases the predictions are within 30% of the measured yields. For updated models with many more predictions and comparisons with both pilot plant and laboratory experiments, and full discussion, see [92,93].
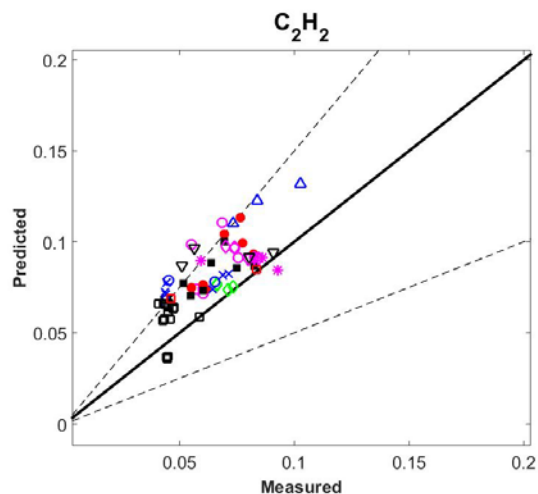


35

Fig. 3 Predicted changes in ignition delays in the end gas in RON100 engine test vs. experimentally measured changes in RON due to addition of 2% of each phenol. The predictive chemical kinetic model correctly identifies which of the six phenols is the best for delaying unwanted ignition (knock) and so increasing the octane number of gasoline.

37