Process Informatics Tools for Predictive Modeling: Hydrogen Combustion

XIAOQING YOU, ANDREW PACKARD, MICHAEL FRENKLACH

Department of Mechanical Engineering, University of California, Berkeley, CA 94720

Received 19 July 2011; revised 11 October 2011; accepted 19 October 2011

DOI 10.1002/kin.20627
Published online 4 December 2011 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Predictive modeling of chemical reaction systems is facilitated through development of an automated data-centric infrastructure, Process Informatics Model (PrIMe). Modeling and analysis tools were built to bridge the PrIMe Data infrastructure with DataCollaboration, a framework designed to make inferences from experimental observations in the context of an underlying model. The developed tools were integrated with the PrIMe Workflow Application, allowing users to create and run drag-and-drop applications on the fly. Organizing the data, linking the data to scientific methods, and automating the analysis offer new venues of scientific inquiry, such as evaluating the consistency of heterogeneous data records, making uncertainty-quantified predictions, quantifying the contribution of newly obtained or even hypothetical data to the question of interest, or testing similar "what–if" scenarios. The developed system is demonstrated for the chemical kinetics of hydrogen combustion. © 2011 Wiley Periodicals, Inc. Int J Chem Kinet 44: 101–116, 2012

INTRODUCTION

Kinetic modeling of complex chemical reaction systems has been accepted as a common approach in the fields of combustion, atmospheric chemistry, material synthesis, astrophysics, and system biology. To reliably develop predictive reaction models for such complex systems requires integration of large amounts of theoretical, computational, and experimental data collected by numerous researchers and often from different disciplines. The integration entails assessment of the consistency of the data, validation of models, and quantification of uncertainties for model predictions [1]. The problem complexity and the volume and heterogeneity of the data call for a systems approach [2,3].

Process Informatics Model (PrIMe) is one such data-centric system framework to predictive chemical reaction models and modeling [3,4]. PrIMe aims at development of a cyber-based infrastructure for organization of scientific data and applications that facilitate collaborative research in collecting, assembling, modeling, and transforming data into knowledge. PrIMe is composed of two principal constituents: a data depository and a collection of tools. Prior efforts during the past several years were devoted to building PrIMe Data Models, digitizing reaction models and experimental data (PrIMe Warehouse), building the

Correspondence to: Michael Frenklach; e-mail: myf@me.berkeley.edu.

Contract grant sponsor: National Science Foundation Chemistry Division, Cyber-Enabled Chemistry.
Contract grant number: CHE-0535542.
Contract grant sponsor: Air Force Office of Scientific Research.
Contract grant numbers: FA9550-08-1-0003 and FA9550-10-1-0450.
Contract grant sponsor: Office of Energy Research, Office of Basic Energy Sciences, Chemical Sciences, Geosciences and Biosciences Division of the U.S. Department of Energy.
Contract grant number: DE-AC03-76F00098.
Supporting Information is available in the online issue at wileyonlinelibrary.com.

integrating infrastructure (PrIMe Workflow), and developing the scientific methodology of predictive modeling (DataCollaboration). In the present study, we focus on the development of tools, interfaces, and additional data models enabling and automating the DataCollaboration methodology within the PrIMe infrastructure. The underlying scientific aim of the development is to foster a new approach to predictive modeling. For a specific platform to illustrate the ideas and methodology, we selected a scientifically challenging problem: reliable modeling of hydrogen combustion.

DataCollaboration is a set-based data analysis method that puts models, theory, and experimental data on the same footing [5], applicable to any data-and-model system. DataCollaboration can establish consistency or inconsistency of a data-and-model system [6], discriminate among alternative hypotheses [7], quantify uncertainties of model predictions [1,5,8], analyze sensitivities to uncertainty levels [9], and optimize a model considering all uncertainties [10]. Numerical efficiency is attained through the use of surrogate models in numerical algorithms of DataCollaboration. The surrogate models are developed by Solution Mapping—approximation of model responses via computer experiments and regression [2,11,12]. Building surrogate models and organizing data are not simple tasks and, as experience showed [3], become error prone with the increase in the (reaction) system size. It is beneficial, therefore, to automate the process: this will enforce proper linking, hide unnecessary details, and focus the user on the scientific issues at hand. Having a simpler-to-operate environment will bring modeling tools into the hands of experimenters.

Thus, one of the principal objectives of the present study is to develop a user-friendly PrIMe infrastructure for the automation of DataCollaboration. To attain this objective, we designed and implemented new PrIMe Data Models, computational components, and component interfaces. To attain true automation and on-the-fly operation, we had to limit the scope at this stage to shock-tube and flow-reactor experiments, as computer simulations for these experiments run on a time scale of a second. By comparison, simulations of laminar premixed flames, the next level of experimental complexity, are on the order of 10–100 min. While the developed infrastructure can be readily extended to include such data in the analysis, the automation will require the use of massively parallel computers, a vibrant subject and a challenge for future development.

The second principal objective of the present study is to demonstrate the developed tools and infrastructure with a real-life problem, and thereby to show the benefits of a more rigorous approach, such as DataCollaboration, to data analysis and model building. We choose a hydrogen combustion model as the working platform. First, its reaction set forms the foundation of all other fuel combustion models. Second, hydrogen combustion in itself is of practical interest in light of prospective utilization of hydrogen as a clean fuel. And finally, the relatively small size of the hydrogen reaction set makes it often a test case of choice for complex simulations, such as turbulent combustion and the automation prospect of the present study.

Even though the hydrogen model is relatively small, the combustion characteristics of hydrogen are quite complex. Over the years, many kinetic models for hydrogen combustion have been suggested, as reviewed by Ströhle and Myhrvold [13]. These models need to be reassessed every time new experimental, thermodynamic, or kinetic data become available. Our objective here is not just to automate a "single computation," but to automate the entire process from new data availability to an automated reporting of the improved prediction. A highlight of the present work is the demonstration of a workflow that accomplishes such a task.

The article is organized as follows. We begin by briefly outlining the methodology of DataCollaboration, and then the organization of the data and the basic architecture of the computational infrastructure, including both existing and newly developed pieces. Next, we describe the hydrogen combustion model, introduce automation tools for uncertainty-quantified (UQ) model building, and apply the developed tools to the hydrogen combustion model. Finally, we summarize our work.

DATA-COLLABORATION METHOD

DataCollaboration is a framework designed to make inferences from experimental observations in the context of an underlying model [5,6,8,9,14]. An observable, Y, is experimentally measured (here a training-set target) and separately predicted by a model with a set of model parameters $\mathbf{x}$. $M(\mathbf{x})$ is a model prediction of observable Y, and d is its measured value. We assume that the discrepancy between the measurement and its model prediction has an uncertainty bounded by l from below and by u from above, $l \leq M(\mathbf{x}) - d \leq u$. The triple of measurement d, uncertainty l and u, and model $M(\mathbf{x})$ for a given set of conditions is referred to as a *dataset unit*. A collection of dataset units, whose elements are indexed by e ranging from 1 to m, is referred to as a *dataset*.

In the present case of a chemical reaction system, it is presumed the existence of an overall, "full" (in the sense of being "tentatively entertained" [1]) chemical-kinetics model designed to predict all experimental

observations of the dataset. A dataset-unit model, M_e, is designed to simulate a given observable, Y_e, at the specific physical conditions of the experiment. These models are developed through computer experiments using shock-tube and flow-reactor codes running the full chemical kinetics model. The resulting expressions, referred to as *surrogate* models, may and usually do exhibit dependence on only a subset of the model parameters, and such a subset is referred to as *active variables*, $\mathbf{x}_e$. Different dataset models have different active variable sets, some in common and some unique to each. Their union is denoted as $\mathbf{x} = \bigcup_{e=1}^{m} \mathbf{x}_e$. The value x_i of each variable is bounded by prior knowledge, expert-assessed uncertainties of the form $x_{i,\min} \le x_i \le x_{i,\max}$. Collectively, all dataset parameter values, $\mathbf{x}$, span the prior-knowledge "hypercube" $H = \{\mathbf{x} \in \Re^n : x_{i,\min} \le x_i \le x_{i,\max}, i = 1, 2, \ldots, n\}$. The subset of the hypercube satisfying $l_e \le M_e(\mathbf{x}_e) - d_e \le u_e$ for all dataset units is referred to as the *feasible set*, F. The dataset is said to be *consistent* if F is nonempty.

PrIMe DATA MODELS

To archive and organize reaction models and data pertinent to the field of combustion, the Extensible Markup Language (XML) standards and schemas, the PrIMe Data Models, were developed in previous studies [3,15–19] and documented at the PrIMe portal [4]. The previously developed Data Models include *bibliography* for referencing data sources, *chemicalElement* for chemical elements, *chemicalSpecies* for chemical species, *thermodynamicPolynomials* for species thermodynamics, *reaction* for chemical reactions, *reactionRate* for reaction rate expressions and parameters, *chemicalModel* for chemical reaction models, as well as *experiment* for experimental observations, *dataAttribute* for specific features of experimental data, and *instrumentalModel* for capturing conversion of raw experimental observations into modelable targets. The present study necessitated further development of *dataAttribute* and design of four new components, described below: *optimizationVariable*, *optimizationVariableBounds*, *surrogateModel*, and *dataset*. All instances of PrIMe Data Models are represented with XML files stored in the PrIMe Data Warehouse, and each file is assigned a unique identifier, *primeID*. Data models are interconnected to one another through linking by respective primeIDs. A diagram of the PrIMe Data Models of relevance to the present work is shown in Fig. 1.

The *dataAttribute* model is designed to extract a specific feature, the observable target Y, from experimental measurements. It archives the experimental conditions (often required for numerical simulation during the later DataCollaboration calculations) of the observable as well as its measured value and estimated uncertainty. We designed several formats to capture various experimental targets, including species concentration at a given reaction time, peak value of a species profile, peak location, rate of change in species concentration, ignition delay, and laminar flame speed. Whereas the details vary for different types of targets, each *dataAttribute* XML file contains three main parts: *propertyLink*, which refers to an experimentally observed property; *feature*, which records the feature type; and *dataAttributeValue*, which stores the observable value, d, and associated uncertainty bounds,

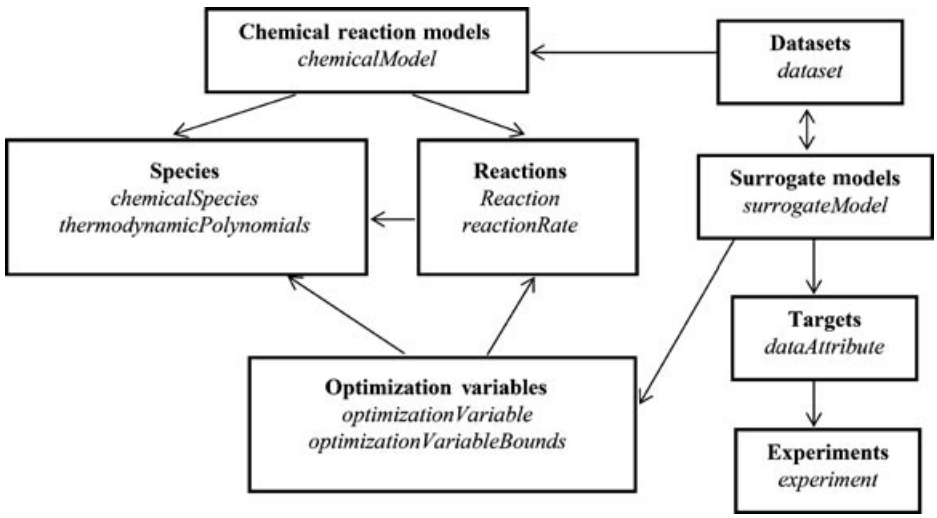

Figure 1 Diagram of the PrIMe Data Models of relevance to the present work. The arrows indicate linking through *primeID*s.

d_l and d_u. An example is presented in Fig. S1 in the Supporting Information.

A chemical reaction model is composed of reaction steps, each defined by its rate-law expression and changes in thermodynamic properties; altogether, these comprise the model parameters. Those among them that are active are captured by the *optimizationVariable* and *optimizationVariableBounds* data models. These are data models designed to archive information on pertinent active variables and their respective bounds—the data to be used for building surrogate models followed and to perform DataCollaboration analysis. These data models specify the links to corresponding parameters, those of the reaction-rate or species-thermodynamics properties. Examples are shown in Fig. S2 in the Supporting Information.

The *surrogateModel* is designed to store coefficients of the surrogate model, $M_e(\mathbf{x}_e)$, a quadratic function in our case, and links to the respective optimization variables, *dataAttribute*, and *dataset* records. An example is shown in Fig. S3 in the Supporting Information.

The *dataset* model contains a *chemicalModel* link and a set of *surrogateModel* links. As the *surrogateModel* is linked with *dataAttribute*, *optimizationVariable*, and *optimizationVariableBounds,* the *dataAttribute* and *surrogateModel* pairs form dataset units for DataCollaboration. The *chemicalModel* link of the *dataset* allows the information associated with the reaction model to be extracted from the associated interlinked XML files. Examples of the *chemicalModel* and *dataset* XML files of the trial model of hydrogen combustion are shown in Figs. S4 and S5 in the Supporting Information.

SOFTWARE INFRASTRUCTURE

Workflow

The central feature of the PrIMe infrastructure is the PrIMe Workflow Application (PWA) [20], a cyber-based computational architecture designed for linking all PrIMe components together, and, most importantly, automation of the data flow. An overall diagram of the software infrastructure is shown in Fig. 2. The infrastructure enables interactions with different users, having diverse objectives and varying roles. In this presentation, we differentiate among a scientist/engineer ("user") solving a data-analysis problem and a numerical analyst ("developer") contributing his/her computational tools intended for solving such problems.

The practical implementation of the PWA is Web based and is built on the Microsoft.NET platform, enabling the user to access and analyze the PrIMe data remotely. It is designed as a user-friendly drag-and-drop application that does not require any programming experience on the part of the user. The user can start with a prebuilt workflow project or create a custom one, by selecting from a set of available objects, each representing a component. An example of a PWA project is shown in Fig. S6 in the Supporting Information.

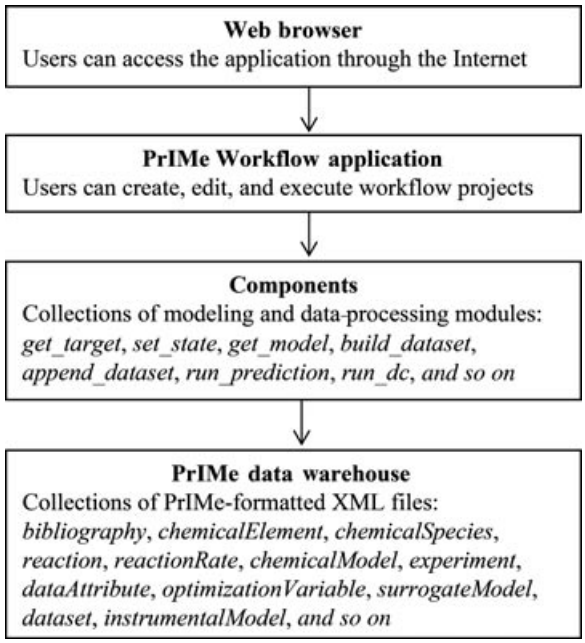

Figure 2 An overall diagram of the PrIMe software infrastructure.

Workflow Components

Every conceivable operation on the data, like retrieval of a particular record or a scientific calculation, constitutes a PWA component. The PWA components can be submitted by the developer via a menu-driven interface. This feature supports sharing of data-analysis and scientific-simulation codes. For instance, the developer may "link" code implementing the High-Dimensional Model Representation (HDMR) of Rabitz and coworkers [21]. In this way, the community will be able to apply different methods of analysis to the same data and meaningfully compare the results.

The current PWA architecture supports three types of components, those written in MATLAB [22], programmed using C# [23], and those installed on a remote (Windows or LINUX) platform. The former two types of components are executed on the fly, on the client machines. The C# components do not require any installation by the user. To run the MATLAB

components, the user needs to install a freely distributed MATLAB Compiler Runtime, which is automated at the PrIMe portal [4]. The data interfaces between components are implemented using a binary XML technology, HDF5. The scientific components developed as part of the present work were built using MATLAB's object-oriented programming language [22]. An example of a PrIMe component is shown in Fig. S7 in the Supporting Information. Several video tutorials available at the PrIMe portal [4] demonstrate the PWA operation.

KINETIC MODEL OF HYDROGEN COMBUSTION

The goal of our work is the automation of the model building, placing the emphases on rigorous uncertainty quantification and hence model predictiveness. Our current focus is chemical kinetic modeling of fuel combustion: It presents just the right combination of knowledge-based models and high-quality experimental data, and yet despite the advanced state of these two subfields, on their own, has not brought the field to the sought-after level of reliability in application of models to practical use [3]. To begin our quest of automation, we selected one of the smallest and better-studied systems, hydrogen combustion. The reaction set of this system forms the foundation of essentially all combustion models and, on its own, is of interest to the community focusing on hydrogen economy [25,26].

It is pertinent to emphasize that our objective is not to develop another or a "better mechanism" for hydrogen combustion, but to establish an automated infrastructure enabling building models of chemical kinetics (currently of hydrogen combustion) with rigorously assessed predictability. In this vein, the quest for automation is not just to make the process user-friendly, but dictated by the necessity to handle large amounts of disparate data, a process prone to errors if carried out manually [3]. We will return to this point later on, and now continue with the details of the selected reaction set.

A large number of detailed reaction models for hydrogen combustion can be found in the literature, including several models that were validated against experimental data [27–30]. We started with this collection of data, augmented with our own evaluation of the new thermodynamic and rate data, as well as their respective uncertainties. Recall that the starting point for the DataCollaboration analysis is creation of a dataset, which must include uncertainties of experimental targets and model parameters. In other words, in our approach, the focus is not on selecting the "right" parameter values, but on assigning their realistic bounds.

The trial model we compiled consists of 10 species and 21 elementary reactions. The species thermochemical data are listed in Table SI in the Supporting Information, with the most recent updates incorporated from Burcat and Ruscic [31], especially for OH and HO_2 radicals. The reaction rate coefficients and the respective uncertainty bounds are listed in Table I. Among the 21 reactions, 14 rate parameters and the estimated bounds of the preexponential factors were taken from GRI-Mech 3.0 [32], except the upper bound for $H + HO_2 \rightarrow OH + OH$ was lowered from the original value of 3.23 to a value of 1.2, reflecting the newer recommendations [29,30]. The rate parameters of the remaining seven reactions were updated as follows: For reaction $O + H + M \rightarrow OH + M$, we took the rate expression given by Tsang and Hampson [33]. The rate expression of reaction $H + O_2(+M) \rightarrow HO_2(+M)$ was taken from the trial H_2/CO combustion model of Davis et al. [28], who derived the rate expression on the basis of Troe's study [34]. For reaction $OH + H_2O_2 \rightarrow H_2O + HO_2$, we took the rate expression given by Hong et al. [35]. For reaction $OH + HO_2 \rightarrow H_2O + O_2$, we took the rate expression proposed by Baulch et al. [36], which is consistent with a recent experiment study of Hong et al. [37]. The rate coefficient of reaction $H + HO_2 \rightarrow H_2 + O_2$ was derived from its reverse reported by Michael et al. [38], whose results agree well with a recent study [39]. For reaction $OH + OH(+M) \rightarrow H_2O_2(+M)$, we derived the rate expression from its reverse given in the H_2/O_2 model of Hong et al. [29] and estimated the uncertainty bounds on the basis of both early and recent studies [29,32,35]. For reaction $O + OH + M \rightarrow HO_2 + M$, we took the value from Saxena and Williams [40] and uncertainty bounds from Burke et al. [41], who showed that this reaction impacted significantly the mass burning rate of lean high-pressure hydrogen flames. In the present study, variables are defined as the normalized forms of rate preexponential factors, as multipliers to the overall rate coefficients (see the description and equations in Appendix A of [5] and [12]).

The other important ingredient of the dataset to be used with the DataCollaboration method is a set of experimental observations (targets) and their respective uncertainties. For this purpose, we use a set of shock-tube ignition delays and flow-reactor H_2 consumption rates assembled recently by Davis et al. [28]. We exclude the flame data at this time. The PrIMe workflow infrastructure is fully extensible and designed to handle all kind of targets including flames. Yet, efficient automation of the surrogate model building with such codes must await future developments, which are not the focus of the present paper.

Table I H_2/O_2 Reaction Model and Parameters of the Rate Coefficients, $AT^n e^{-E/RT}$, in units of cm^3, mol, s, cal, K

	Reactions	A	n	E	lb^a	ub^a	Reference	opt^a
1	$H + O_2 = O + OH$	2.65×10^{16}	−0.6707	17041	0.87	1.15	[31]	1.12
2	$O + H_2 = H + OH$	3.87×10^{4}	2.7	6260	0.63	1.6	[31]	1.00
3	$OH + H_2 = H + H_2O$	2.16×10^{8}	1.51	3430	0.77	1.3	[31]	0.78
4	$OH + OH = O + H_2O$	3.57×10^{4}	2.4	−2110	0.63	1.6	[31]	1.00
5[b]	$H + H + M = H_2 + M$	1.00×10^{18}	−1.0	0	0.33	3	[31]	1.00
	$H + H + H_2 = H_2 + H_2$	9.00×10^{16}	−0.6	0				
	$H + H + H_2O = H_2 + H_2O$	6.00×10^{19}	−1.25	0				
6[c]	$O + O + M = O_2 + M$	1.20×10^{17}	−1.0	0	0.5	2	[31]	1.00
7[d]	$O + H + M = OH + M$	4.71×10^{18}	−1.0	0	0.5	2	[32]	1.00
8[e]	$H + OH + M = H_2O + M$	2.20×10^{22}	−2.0	0	0.5	2	[31]	1.00
9[f]	$H + O_2 + M = HO_2 + M$	5.75×10^{19}	−1.4	0	0.83	1.2	[27,33]	1.00
	$H + O_2 = HO_2$	4.65×10^{12}	0.44	0				
10	$H + HO_2 = O + H_2O$	3.97×10^{12}	0.0	671	0.33	3	[31]	1.00
11	$H + HO_2 = H_2 + O_2$	2.99×10^{6}	2.12	−1172	0.8	1.25	[34]	1.00
12	$H + HO_2 = OH + OH$	8.40×10^{13}	0.0	635	0.81	1.2	[31]	0.91
13	$O + HO_2 = OH + O_2$	2.00×10^{13}	0.0	0	0.5	2	[31]	1.00
14	$OH + HO_2 = H_2O + O_2$	2.89×10^{13}	0.0	−497	0.5	2	[35]	1.00
15	$HO_2 + HO_2 = H_2O_2 + O_2$	1.30×10^{11}	0.0	−1630	0.5	2	[31]	1.00
	$HO_2 + HO_2 = H_2O_2 + O_2$	4.20×10^{14}	0.0	12000				
16[g]	$OH + OH + M = H_2O_2 + M$	1.46×10^{11}	0.868	−8548	0.5	5.0	[28,36]	0.58
	$OH + OH = H_2O_2$	8.71×10^{9}	0.869	−2191				
17	$H + H_2O_2 = H_2O + OH$	1.00×10^{13}	0.0	3600	0.5	2	[31]	1.00
18	$H + H_2O_2 = HO_2 + H_2$	1.21×10^{7}	2.0	5200	0.5	2	[31]	0.68
19	$O + H_2O_2 = HO_2 + OH$	9.63×10^{6}	2.0	4000	0.33	3	[31]	1.00
20	$OH + H_2O_2 = H_2O + HO_2$	1.74×10^{12}	0.0	318	0.83	1.2	[36]	1.00
	$OH + H_2O_2 = H_2O + HO_2$	7.59×10^{13}	0.0	7272				
21[d]	$O + OH + M = HO_2 + M$	8.00×10^{15}	0.0	0	0.125	12.5	[37,38]	1.00

[a] $lb = A_{min}/A$ and $ub = A_{max}/A$ are relative lower and upper bounds of A, respectively; $opt = A_{opt}/A$ is the optimized relative value of A obtained in this study.

[b] Efficiency factors: Ar = 0.63.

[c] Efficiency factors: H_2 = 2.4; H_2O = 15.4; Ar = 0.83.

[d] Efficiency factors: H_2 = 2; H_2O = 12; Ar = 0.7.

[e] Efficiency factors: H_2 = 0.73; H_2O = 3.65; Ar = 0.38.

[f] Efficiency factors: H_2O = 12; Ar = 0.53; Troe parameters: $a = 0.5$, $T^{***} = 10^{-30}$, $T^{*} = 10^{30}$, $T^{**} = 10^{100}$.

[g] Efficiency factors: H_2 = 2; H_2O = 6; Ar = 0.67; Troe parameters: $a = 1.0$, $T^{***} = 10^{-30}$, $T^{*} = 10^{30}$, $T^{**} = 10^{30}$.

The selected experimental targets, their observed values, d, and estimated uncertainty bounds are listed in Table II. The upper and lower bounds, d_u and d_l, respectively, demarcate 2σ intervals. They are calculated in the following manner. For the flow-reactor targets, *flw*, the upper/lower bounds are $d \pm (2\times\sigma)$, where the value of σ is taken from Davis et al. [28]. For the ignition-delay targets, the same approach often results in negative lower bounds. Hence, we calculated the 2σ bounds by transforming σ, reported by Davis et al. [28], to the logarithmic axis of the ignition delay,

$$d_u = 10^{\log_{10} d + 2[\log_{10}(d+\sigma) - \log_{10} d]} = d\,(1 + \sigma/d)^2$$

$$d_l = 10^{\log_{10} d - 2[\log_{10} d - \log_{10}(d-\sigma)]} = d\,(1 - \sigma/d)^2$$

The last two columns of Table II report results of the numerical simulations with the detailed kinetics model, those obtained with the initial (Trial, x_0) and optimized (Opt, x_{opt}) set of rate coefficients; the corresponding parameter values for the two cases are reported in Table SII in the Supporting Information.

AUTOMATION OF UQ-MODEL BUILDING: HYDROGEN OXIDATION SYSTEM

System and Data

In this section, we apply the developed PrIMe tools to the hydrogen oxidation system described in the preceding section. The required data—the XML files of

Table II Experimental Targets of the Dataset and Model Predictions

	Composition (% mol)			Ignition delay time,[a] *ig*							
Target	H_2	O_2	Diluent	T_5 (K)	P_5 (atm)	d (μs)	d_l (μs)	d_u (μs)	Reference	Trial (μs)	Opt (μs)
1a	6.67	3.33	Ar	1051	1.729	231	53	533.	[41]	319	293
1b	6.67	3.33	Ar	1312	2.008	50	12.5	112.5	[41]	61	58
2a	20	10	Ar	1033	0.518	238	40	600	[42]	344	319
2b	20	10	Ar	1510	0.493	29	2.8	82.8	[42]	37	36
3	0.5	0.25	Ar	1754	33	10	4.9	16.9	[43]	13	12
4a	2	1	Ar	1189	33	293	127	527	[43]	223	211
4b	2	1	Ar	1300	33	11	4.5	20.5	[43]	23	20.5
5	0.1	0.05	Ar	1524	64	54	15.6	115.6	[43]	66	62
				Flow reactor, *flw*							
				T_0 (K)	P_0 (atm)	d (ppm/ms)	d_l (ppm/ms)	d_u (ppm/ms)		Trial (ppm/ms)	Opt (ppm/ms)
1[b]	1.18	0.61	N_2	914	15.7	8.6	7.6	9.6	[44]	13.5	9.6
2[b]	1.01	0.52	N_2	935	6	13	12	14	[44]	14.9	12.0
3[c]	0.95	0.49	N_2	934	3.02	19	17	21	[44]	31.5	20.2
4[b]	0.5	0.5	N_2	880	0.3	261	221	301	[44]	341.1	300.9

[a] Ignition delays of 1a, 1b, 2a, and 2b are defined by a rapid rise in pressure. The rest are defined by the maximum of d[OH]/dt.

[b] Target value is $\Delta x_{H_2}/\Delta t|_{C_{H_2}=0.4}^{C_{H_2}=0.6}$, where x is the mole fraction and C is the fractional conversion.

[c] Target value is $\Delta x_{H_2}/\Delta t|_{C_{H_2}=0.85}^{C_{H_2}=0.95}$, where x is the mole fraction and C is the fractional conversion.

the hydrogen combustion model {Tables I and SI (in the Supporting Information)} and 12 experimental targets (Table II) had been archived in the PrIMe Data Warehouse. In the following, we will create a dataset from these data records, build surrogate models, and carry out DataCollaboration analyses: test dataset consistency, make predictions with the data-and-model system, and perform model optimization.

Creation of a Dataset

To automate uncertainty-quantified model building using DataCollaboration, the first task to accomplish is to assemble a dataset from the data archived and annotated in the PrIMe Warehouse. This assembly is accomplished with a workflow that consists of three components: *get_target*, *get_model*, and *build_dataset*, as shown in Fig. 3. Execution of this workflow opens the Search-component forms, one (Component 1 in Fig. 3) customized to search the PrIMe Data Warehouse for a *dataAttribute* XML file and another (Component 2 in Fig. 3) for a *chemicalModel* archived XML file. After the user selects the appropriate records, these search components create target and model HDF5 files, respectively, and store them on the user's machine.

The *build_dataset*, the main component of this workflow (component 3 in Fig. 3), is executed next. It first extracts the information stored in the target and model HDF5 files and creates input objects for solution of differential equations, representing the reaction model. Next, a screening sensitivity analysis is performed for the selected target with respect to model parameters, in our case, with respect to the preexponential factors of the reaction rate coefficients. The objective of this step is to rapidly identify prospective active variables. This is efficiently accomplished by using saturated factorial designs [42]. In our implementation, we use Hadamard matrices, whose elements are either $+1$ or -1 and whose columns are mutually orthogonal.

Figure 4 displays the parameter sensitivity coefficients and impact factors. The specific snapshot shown in this figure was obtained by selecting, upon execution of Component 1 in Fig. 3, target file a00000092.xml documenting hydrogen ignition delay measured in a shock tube [43] and then, upon execution of Component 2, the hydrogen model file m00000004.xml. The right-hand panel of Fig. 4 has two bar charts. The one on the left shows ranked sensitivity coefficients, displayed also as a list on the left panel of the figure. The second bar chart shows corresponding impact factors, computed as a product of the absolute value of sensitivity coefficient and the corresponding span; the latter being the square root of the ratio of the upper and lower bounds of the reaction rate-coefficient preexponential factors, listed in the left panel of the figure (further details of this definition can be found in [5] and [12]).

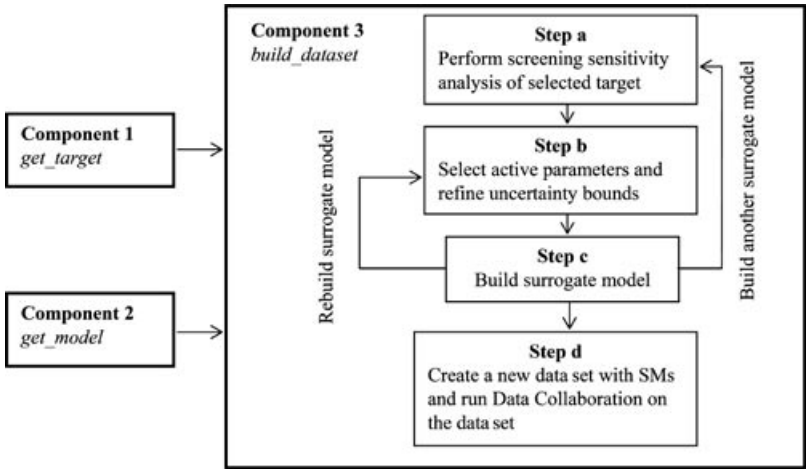

Figure 3 Component diagram of the PrIMe workflow to create a dataset from scratch and run DataCollaboration. SM denotes a surrogate model.

The left panel of Fig. 4 has a column titled "Selected," with several top rows checked and the corresponding bars colored red on the right-panel charts. These are top candidates for active variables to be used at the next stage, building surrogate models. The user may deselect some or select additional ones, triggering simultaneous color changes of the corresponding bars on the right-panel charts. The active variables selected for the hydrogen oxidation system are listed in Table III. The user may also change the parameter bound values. In this way, the user may combine the computed sensitivities with additional considerations or perform additional tests. After the user is content with the set of active variables and their bounds, the user can proceed to the next stage of Component 3 by clicking the "build surrogate model" button at the bottom of the right-hand panel.

Table III Active Variables with Sensitivity Rankings for Each Target

		ig								*flw*			
	Reactions	1a	1b	2a	2b	3	4a	4b	5	1	2	3	4
1	$H + O_2 = O + OH$	1	1	1	1	1	1	1	1	1	1	2	7
2	$O + H_2 = H + OH$	2	2	2	2	2		9	2	12			
3	$OH + H_2 = H + H_2O$	5	3	4	3	3		7	3	6	7	5	17
4	$OH + OH = O + H_2O$											11	11
5	$H + H + M = H_2 + M$	7		10	8		4	3		11	6	15	5
6	$O + O + M = O_2 + M$		6	8	6			10				16	
7	$O + H + M = OH + M$	9		9	5			5		9	13		2
8	$H + OH + M = H_2O + M$	8	5	6	10				10		4	9	
9	$H + O_2(+M) = HO_2(+M)$	3	10	7		5	2	2	4	3	2	1	9
10	$H + HO_2 = O + H_2O$					6	5	4	7	15	17	17	4
11	$H + HO_2 = H_2 + O_2$	4	4	3	4	4	9		5	4	3	7	1
12	$H + HO_2 = OH + OH$	10				8	3	8	6	8	8	3	13
13	$O + HO_2 = OH + O_2$				7		7	6		5	10	13	3
14	$OH + HO_2 = H_2O + O_2$		8							2	5	4	
15	$HO_2 + HO_2 = H_2O_2 + O_2$				9	9	10			14	16	12	10
16	$OH + OH(+M) = H_2O_2(+M)$		9			7	8			13	12	10	14
17	$H + H_2O_2 = H_2O + OH$									16	11	14	15
18	$H + H_2O_2 = HO_2 + H_2$					10	6		9	7	9	8	
19	$O + H_2O_2 = HO_2 + OH$	6	7	5					8	17			6
20	$OH + H_2O_2 = H_2O + HO_2$									10	15	6	16
21	$O + OH + M = HO_2 + M$	11	14	11	17	18	20	17	13		14		12

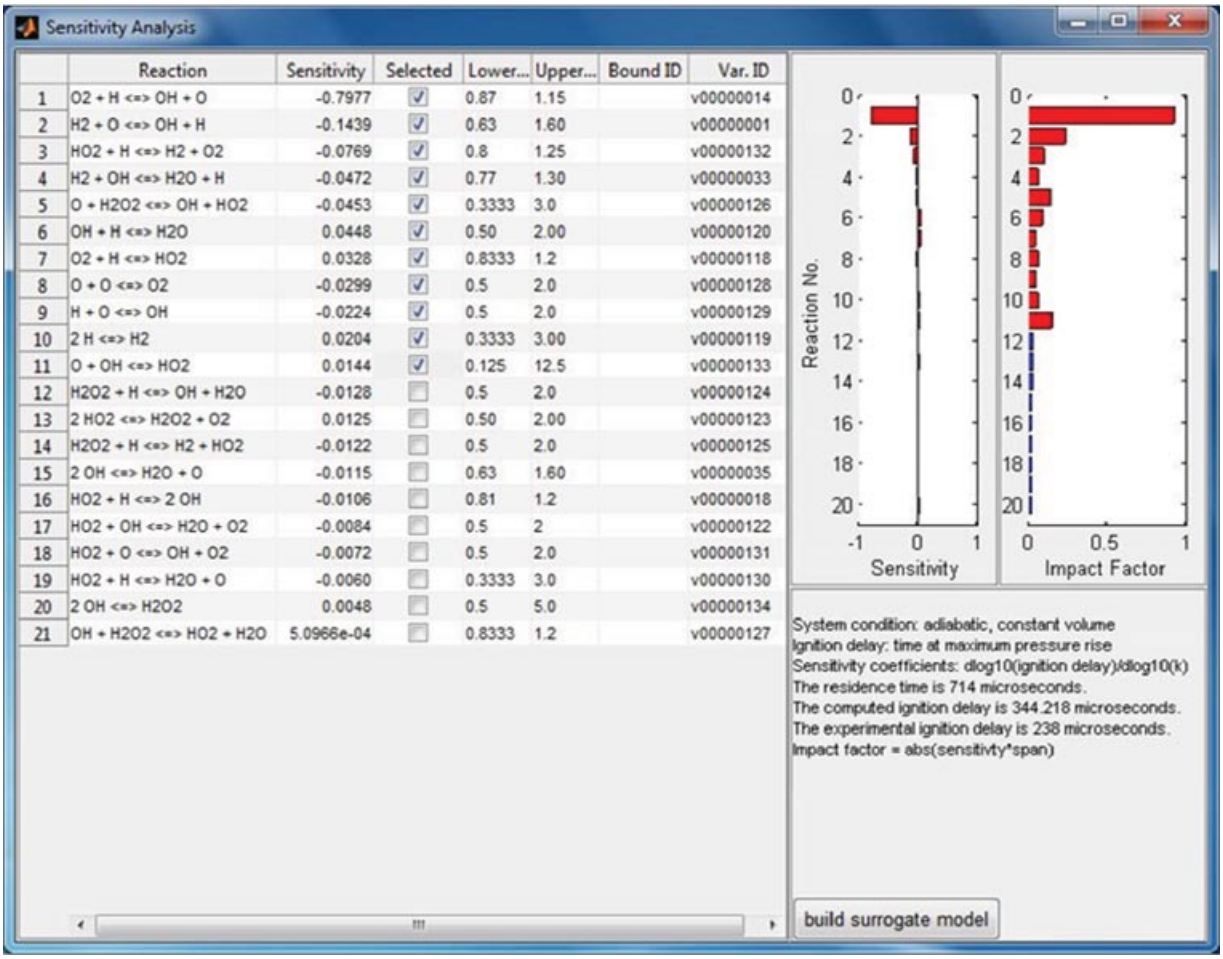

Figure 4 A snapshot showing the results of the screening sensitivity analysis.

The surrogate models are built on the fly, similar to the sensitivity analysis, by performing a series of computational runs according to a design matrix. At this stage, we perform space-filling sampling employing Latin hypercube designs [44]. By default, the number of design points (runs) is set to be twice the number of coefficients in the quadratic surrogate model to be built, but the user may change the number of runs in a pop-up window upon the "build surrogate model" button clicked. The values of the responses, computed with the "full" chemical model, are fitted into a quadratic in active variables (if quadratic is not sufficiently accurate to fit the response, DataCollaboration has a provision for piecewise quadratic approach [7], which opens the door to a wider range of surrogate models [21,44,45] that we plan to build into the future releases). The results are displayed for the user perusal, as demonstrated in Fig. 5, where the top-left panel lists the active variables, the bottom-left panel displays the computed coefficients of the surrogate model, and the top-right panel plots relative differences between the full kinetic and surrogate models. At the same time, a *surrogateModel* XML file, storing the surrogate model for further use, is generated automatically. If *optimizationVariable* or *optimizationVariableBounds* XML file of an active variable does not exist in the PrIMe Data Warehouse, it is also generated at this time. All the created XML files are saved on the user machine.

Examining the results, such as those depicted in Fig. 5, the user has the option of repeating the surrogate-building step by closing this window and making changes in the selection of active variables, their bounds, and in the number of runs to perform. The user may also choose to build another surrogate model for a different experimental target, or when all new desired surrogate models are built, to move to the next step of creating a dataset and run DataCollaboration by clicking the corresponding button in the bottom-right panel of Fig. 5. If the user chooses to build another

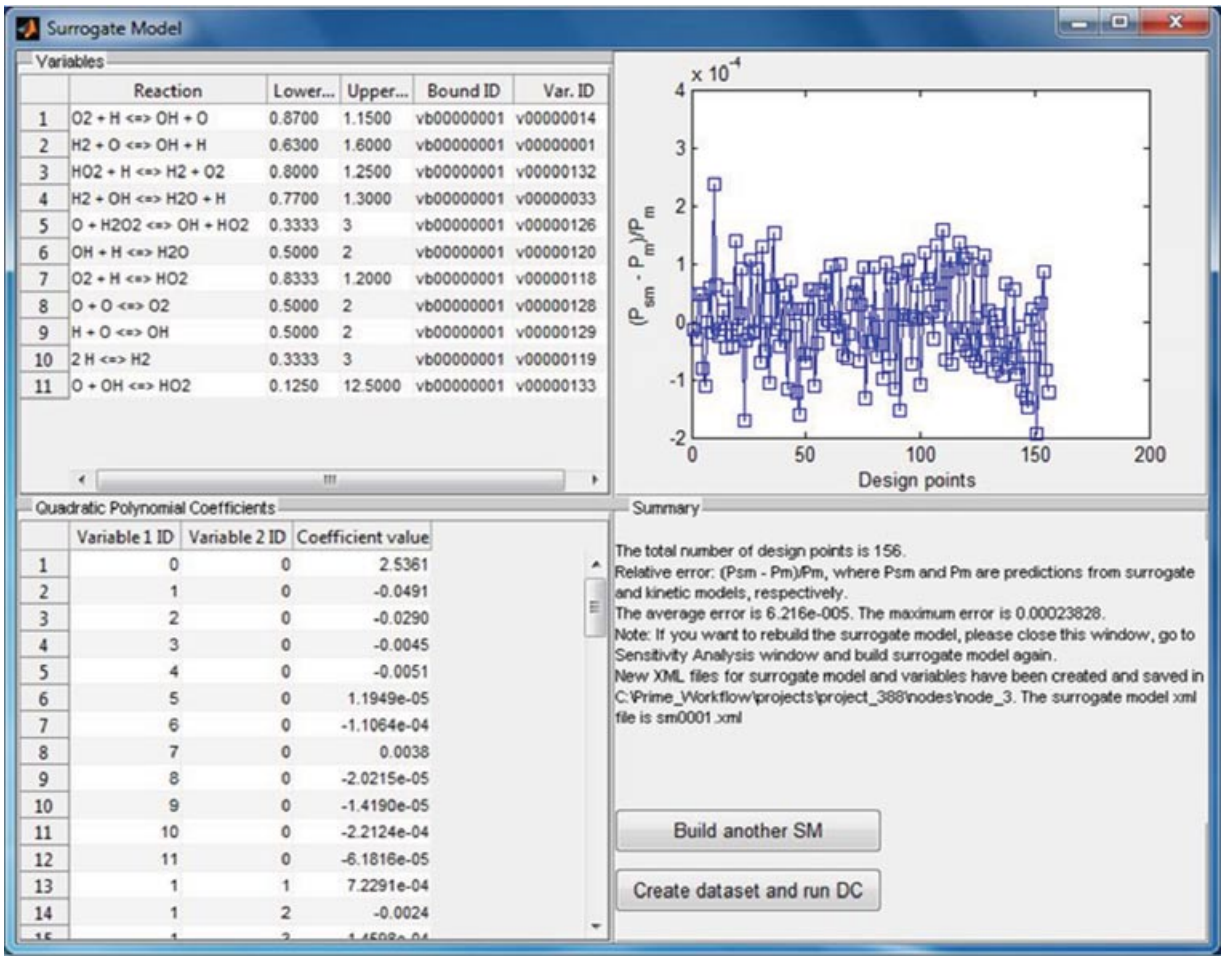

Figure 5 A snapshot showing results of building a surrogate model for target "a00000092.xml."

surrogate model, the search window appears for the user to select another *dataAttribute* target, and the previous steps of sensitivity analysis and surrogate model building are repeated. If the user chooses to run DataCollaboration, the dataset XML and HDF5 files are created with all *surrogateModel* XML files stored in the local directory, and the DataCollaboration window (see Fig. S8 in the Supporting Information) is initiated for the analysis of the dataset.

In the present work, we built surrogate models for all 12 targets by repeatedly clicking the "Build another SM" button in Fig. 5. After that, we clicked the "Create dataset and run DC" button in Fig. 5. A dataset XML file was generated with links to all the 12 *surrogateModel* XML files, and the DataCollaboration application was initiated.

Thus far, we have addressed how to create a dataset from scratch. After building a satisfactory dataset, the generated PrIMe-formatted XML files may be submitted to the PrIMe data warehouse and thus accessible by and shared with other users. In another scenario, one might want to append additional surrogate models to an archived dataset. This can be accomplished by creating a workflow project consisting of three components: *get_target*, *get_dataset*, and *append_dataset*, as shown in Fig. 6. Instead of using *get_model*, as was in the workflow of Fig. 3, we now use *get_dataset* component to search for an archived *dataset* XML file through the Search-component form. After the user selects the appropriate record, the search component creates a dataset HDF5 file and stores the dataset information on the user's machine. The *append_dataset* component in Fig. 6 reads the target and dataset HDF5 files. The chemical model is accessible through the model link node of the dataset file. The steps of *append_dataset* are similar to those of *build_dataset*, differing only in step d, where the *surrogateModel* XML files are appended to the existing dataset.

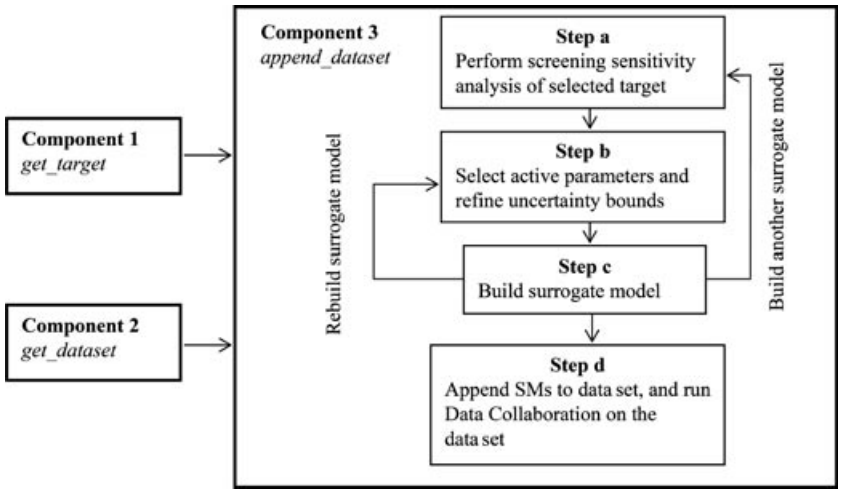

Figure 6 Component diagram of the PrIMe workflow to append surrogate models to an existing dataset and run DataCollaboration.

Testing Consistency of a Dataset

Once a dataset is created, one of the first tasks would be to determine if the data-and-model system is mutually consistent [6]. In the present case of the hydrogen oxidation dataset, the consistency analysis tests whether there exists at least one set of parameter values within their respective ranges of uncertainties (as listed in Table I) such that the kinetic model is able to predict the selected observable targets each within their respective ranges of uncertainties (as listed in Table II).

The PrIMe workflow to accomplish this task consists of two components, *get_dataset* and *run_consistency*, as shown in Fig. 7. The *get_dataset* component allows the user to select a desired dataset from those archived in the PrIMe Data Warehouse, hydrogen–oxygen in our case. The *run_consistency* component initiates the Consistency application window (see Fig. S9 in the Supporting Information) for testing the consistency of the uploaded dataset.

After testing the consistency of the hydrogen oxidation dataset, by double clicking on the result line, we may open the Consistency results window shown in Fig. 8. The displayed (left panel) consistency measure [6] falls into the range [0.32, 0.33]. These values indicate that the dataset is consistent and will remain consistent even if all parameter and target uncertainties are decreased by 32% and that the dataset will become inconsistent if the uncertainties are decreased by 33%.

The right panel of Fig. 8 displays computed impact factors, computed with the DataCollaboration methodology documented in [9]. Examination of this information reveals, for instance, that among the dataset 12 targets, the upper bound of the ignition-delay target 4b has the largest impact on the range of the consistency measure, and the upper bounds of preexponential factors of the $H + O_2 \rightarrow O + OH$ and $O + H_2 \rightarrow H + OH$ reaction rate coefficients are the two most influential parameters. It is also instructive to note that the top impact factor of the target experimental uncertainties (2.898, shown in the top upper corner of Fig. 8) is more than 50 times larger than the top impact factor of parameter uncertainties (0.05329, shown in the middle right of Fig. 8).

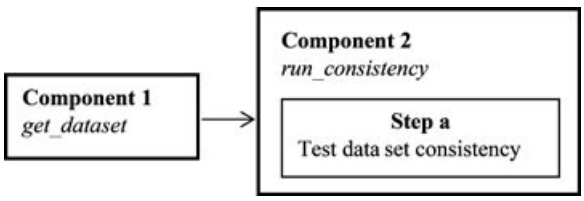

Figure 7 Component diagram of the PrIMe workflow to test consistency of an existing dataset.

Prediction on the Feasible Set

Having confirmed the dataset consistency, we can use the dataset to make predictions for yet unmeasured properties. Continuing our example, we set the objective of predicting ignition delay for a 10% H_2—10% O_2—10% H_2O—70% Ar mixture at the initial temperature of 1500 K and pressure of 1 atm. The PrIMe workflow for this task consists of three components: *set_state*, *get_dataset*, and *run_prediction*, as shown in Fig. 9. The *set_state* component opens a window that allows the user to specify the initial state of the mixture by entering the mixture composition, temperature, and pressure with menu-driven operations (a snapshot

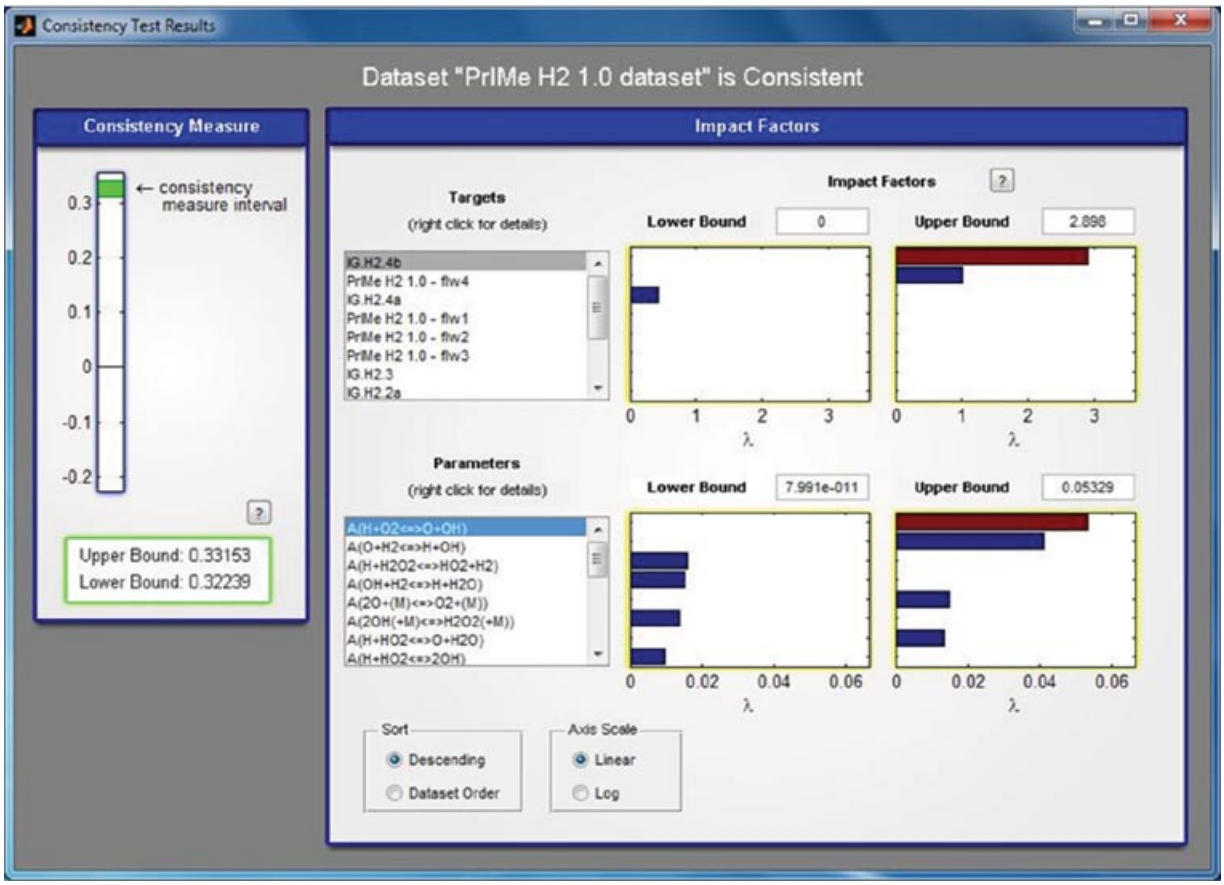

Figure 8 A snapshot showing the Consistency analysis results for the hydrogen oxidation dataset.

is shown in Fig. S10 in the Supporting Information). Upon the completion of the data entry and closure of the component window, the *set_state* proceeds by creating an interface, HDF5 file that stores the state information in an object-based organization. The *get_dataset* component, executed next, retrieves the desired dataset of hydrogen oxidation. The next step of the workflow is the execution of the *run_prediction* component. It retrieves the initial state and dataset records by loading the HDF5 files created by the respective components and stored in a local, user directory. The chemical model is retrieved through the model link stored in the dataset file. When all the data are loaded, the *run_prediction* component builds a surrogate model for the ignition delay under conditions described, appends it to the dataset, and opens the Prediction window (see Fig. S11 in the Supporting Information), which in turn launches the application for determining model prediction bounds.

The results produced by the execution of this workflow are shown in the Prediction results window (Fig. 10). The left panel of this figure displays the *interval* predicted for the property in question, ignition delay time for the given mixture at the specified initial conditions [21, 29] μs. The darker-green areas of the graphically displayed interval designate the uncertainty of the DataCollaboration method itself. These uncertainties can be reduced, bringing the inner and outer intervals closer to each other, by performing additional DataCollaboration calculations enabled by user-selected options in the application window.

The key point to emphasize here is that the prediction is not a single number, but an interval; its width originates from uncertainties of 21 model parameters (Table I), 12 experimental targets (Table II), and the intrinsic correlations among them.

Furthermore, the right panel of Fig. 10 displays computed impact factors [9], analogously to the Consistency results window. Examination of this information reveals, for instance, that among the dataset 12 targets, the upper bound of the flow reactor target 4 has the largest impact on the prediction range, and the upper bounds of the preexponential factors of the H $+ O_2 \rightarrow$ O $+$ OH and O $+ H_2 \rightarrow$ H $+$ OH reaction rate coefficients are the two most influential parameters. Also noteworthy in this case is that the top impact

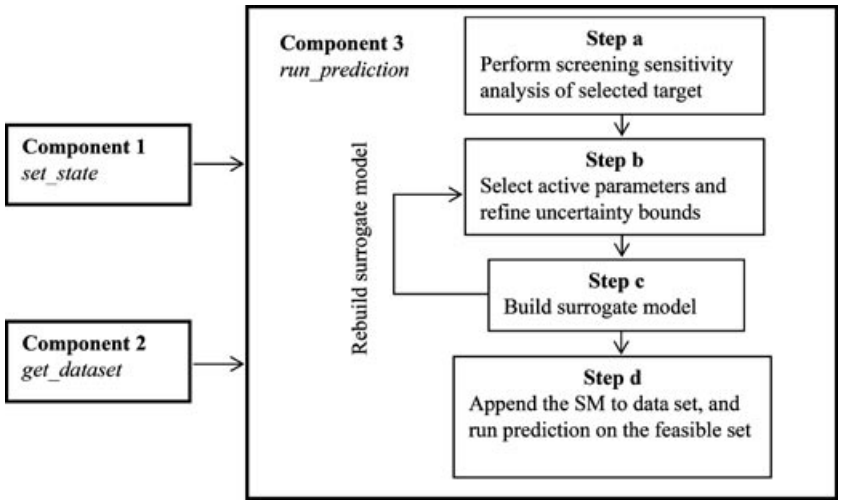

Figure 9 Component diagram of the PrIMe workflow to determine model prediction bounds for user-specified initial conditions and dataset.

factor of the target experimental uncertainties (0.2152) is more than 10 times larger than the top impact factor of the parameter uncertainties (0.01828).

Parameter Optimization on the Feasible Set

One may be interested in obtaining a reaction model with parameters fitted to a selected (by the user) a set of training data, i.e., a set or a subset of the dataset targets. The PrIMe workflow to accomplish this task is shown in Fig. 11. The *get_dataset* component retrieves the hydrogen–oxygen dataset, as described in the preceding section, and executes the *run_optimization* component that initiates the Optimization application (Fig. S12 in the Supporting Information).

The current application implements four optimization methods that the user can select: weighted least squares of the residuals minimized over the prior-knowledge hypercube H(LS-H), weighted least squares of the residuals minimized over the feasible set F(LS-F), minimization of the sum of absolute parameter deviations over the feasible set F(1N-F), and multiobjective optimization over the feasible set (MO-F) [10]. The user has also an option of assigning weights for individual optimization targets.

We demonstrate the workflow shown in Fig. 11 with the 1N-F method. Our objective is stated as follows. We consider the initial, trial model of hydrogen combustion (Table I). Its parameter vector ($\mathbf{x}_0$ in Table SII in the Supporting Information) is outside the feasible set, because for five (*ig*4b, *flw*1, *flw*2, *flw*3, and *flw*4) of the 12 targets the model predictions are outside their respective ranges of uncertainty (Table II). We now ask for a new parameter vector $\mathbf{x}_{opt}$, closest to $\mathbf{x}_0$ but contained in the feasible set, so that the model predictions with $\mathbf{x}_{opt}$ for each of the 12 experimental targets are within their respective uncertainty bounds.

The results produced by the execution of this workflow are shown in the Optimization results window (Fig. 12). The top panel of this figure displays initial and optimized parameter values, and the bottom panel shows the experimental value, its deviation from the model prediction, uncertainty range, and statistical weight of each target. Further information can be viewed by pressing the "Parameter Table" and "Data Table" buttons. The preexponential factors and model predictions obtained after the 1N-F optimization are documented in Tables I and II, respectively.

Examination of Fig. 12 reveals that the optimized parameter values of $OH + H_2 \rightarrow H + H_2O$ and $H + O_2 \rightarrow O + OH$ reactions have the largest deviations from their respective initial values. To further assess the performance of the trial and the optimized models, we compared model predictions with experimental data (reported in Figs. S13 and S14 in the Supporting Information). The results indicate that with only five rate parameters (the top five in Fig. 12) changed within their respective ranges of uncertainties; the optimized model predicts the experimental targets closer than the one with the initial, trial set of parameter values, both reported in the last two columns of Table II. For example, for target 4b, the trial model gives ignition delay time 23 μs, which is outside the experimental uncertainty interval, whereas the optimized model produces a value matching the upper bound of

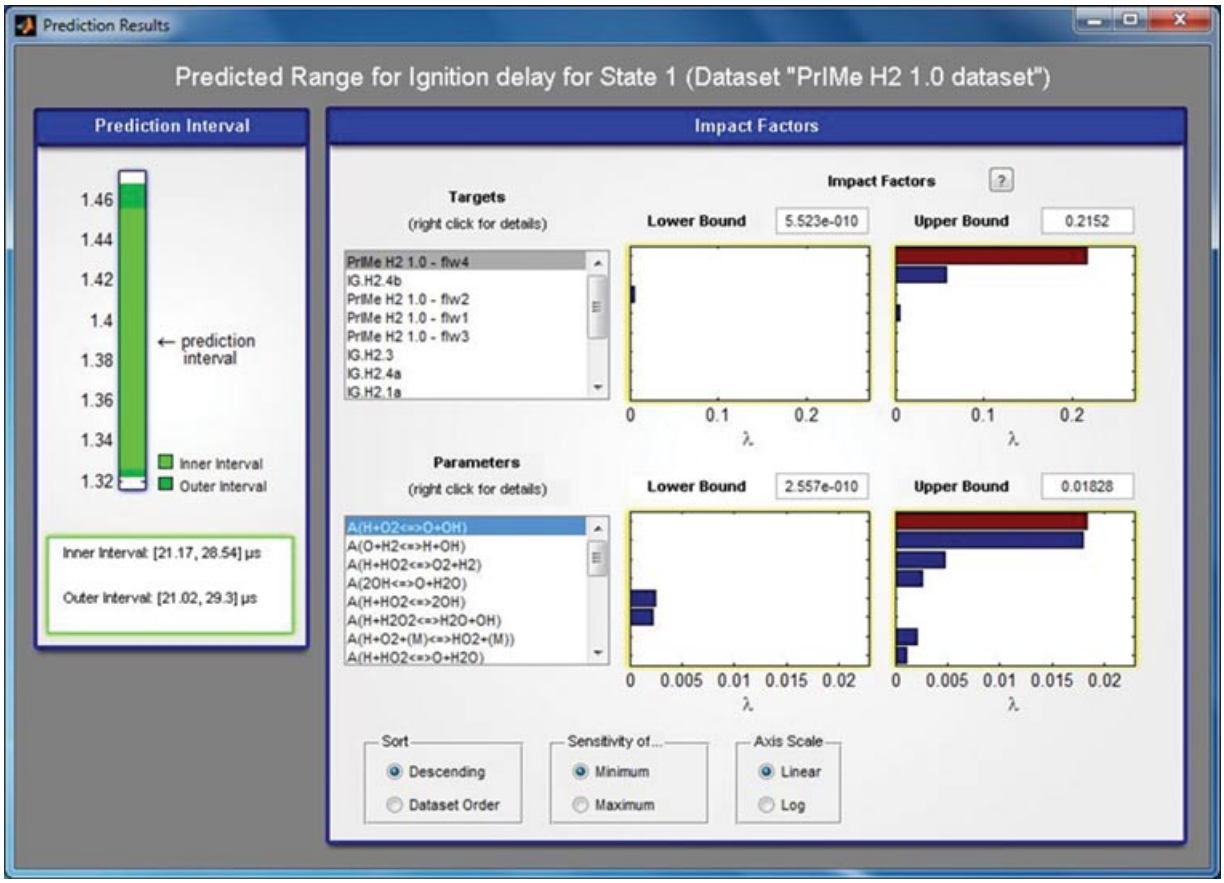

Figure 10 A snapshot showing the model prediction results for the hydrogen oxidation dataset.

the experiment. For all flow reactor targets, predictions with the trial model fall outside the uncertainty intervals of the experimental targets, whereas the optimized model brings the predictions to within these intervals.

SUMMARY

In this study, we developed process informatics tools for the automation of predictive modeling and demonstrated them with the case of hydrogen combustion. Among the new developments are (a) XML-based data models—*optimizationVariable*, *optimizationVariableBounds*, *surrogateModel*, and *dataset*—designed to properly archive and curate the pertinent experimental, theoretical, and uncertainty data within the PrIMe infrastructure; (b) a concrete dataset for the hydrogen combustion system, built employing the PrIMe data models; (c) interfaces linking simulation and uncertainty-quantification tools to the archived data; and (d) graphical user interfaces facilitating user-friendly operation of predictive modeling tools.

The new developments enable the user to pursue the following scientific quests:

1. create of a new *dataset*—a data-centric infrastructure uniting measurements, models, and uncertainties;
2. augment an existing dataset with a new experimental target;
3. test consistency of a dataset;
4. make the uncertainty-quantified prediction for a given property (currently ignition delay) at user-specified experimental conditions, where the prediction is an interval of values conditioned to satisfying all the uncertainty constraints, experimental and theoretical; and
5. build an optimized model, with user selection of an optimization criterion, targets to include in the training set, and statistical weights assigned to these optimization targets.

The developed PrIMe infrastructure and based on it automation allows, for instance, the user to create a new chemical reaction model, optimized to his/her

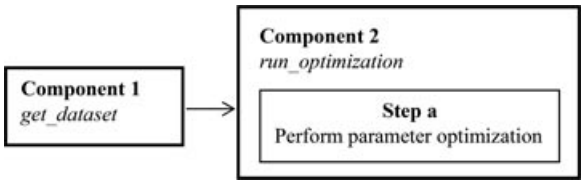

Figure 11 Component diagram of the PrIMe workflow to run parameter optimization on an existing dataset.

preferences, with "a push of a button" and get the result within a minute or two—the task that took otherwise months to accomplish in the past.

Organizing the data, linking the data to scientific methods, and automating the analysis not only speeds up the analysis but, more importantly, offers new venues of scientific endeavor, such as evaluating the consistency of data records, making uncertainty-quantified predictions, quantifying the contribution or impact of the newly obtained or hypothetical data to the question of interest, or testing various other "what–if" scenarios.

The methodology and infrastructure demonstrated here is not limited to hydrogen combustion or chemical kinetics, but may have potential applications in other fields as well. We envision that the infrastructure will be continuously evolving based on user community experiences.

The authors gratefully acknowledge contributions of Trent Russi and William Speight in the development of DataCollaboration graphical interface and Michael Gutkin in the development of the PrIMe workflow.

SUPPORTING INFORMATION

The species thermodynamics data, examples of PrIMe-formatted XML files, snapshots of the graphical user interfaces, and comparison of the trial and optimized hydrogen oxidation models are available in the online issue at wileyonlinelibrary.com. Also, a series of written and video tutorials are available at the PrIMe

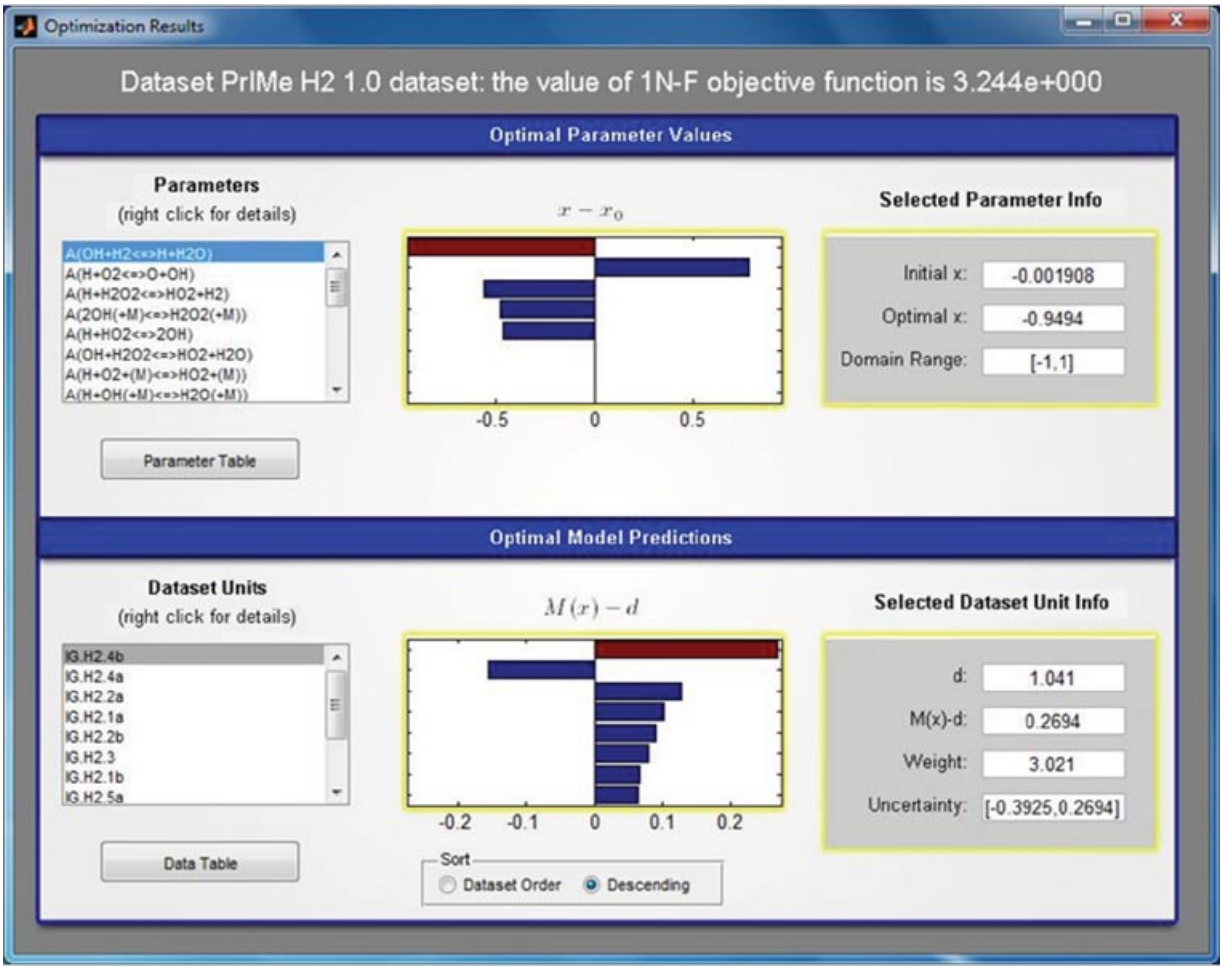

Figure 12 A snapshot showing the parameter optimization results for the hydrogen oxidation dataset.

portal [4], including a video demonstrating the prediction scenario described in the present article.

BIBLIOGRAPHY

1. Russi, T.; Packard, A.; Frenklach, M. Chem Phys Lett 2010, 499, 1–8.
2. Frenklach, M.; Wang, H.; Rabinowitz, M. J. Prog Energy Combust Sci 1992, 18, 47–73.
3. Frenklach, M. Proc Combust Inst 2007, 31, 125–140.
4. Process Informatics Model; available at: http://primekinetics.org, accessed September 30, 2011.
5. Frenklach, M.; Packard, A.; Seiler, P.; Feeley, R. Int J Chem Kinet 2004, 36, 57–66.
6. Feeley, R.; Seiler, P.; Packard, A.; Frenklach, M. J Phys Chem A 2004, 108, 9573–9583.
7. Feeley, R.; Frenklach, M.; Onsum, M.; Russi, T.; Arkin, A.; Packard, A. J Phys Chem A 2006, 110, 6803–6813.
8. Frenklach, M.; Packard, A.; Seiler, P. In Proceedings of American Control Conference, Anchorage, AK, 2002; pp. 4135–4140.
9. Russi, T.; Packard, A.; Feeley, R.; Frenklach, M. J Phys Chem A 2008, 112, 2579–2588.
10. You, X. Q.; Russi, T.; Packard, A.; Frenklach, M. Proc Combust Inst 2011, 33, 509–516.
11. Frenklach, M. Combust Flame 1984, 58, 69–72.
12. Frenklach, M.; Packard, A.; Feeley, R. In Comprehensive Chemical Kinetics; Carr, R. W., Ed.; Elsevier, Amsterdam, 2007; pp. 243–291.
13. Ströhle, J.; Myhrvold, T. Int J Hydrogen Energy 2007, 32, 125–135.
14. Seiler, P.; Frenklach, M.; Packard, A.; Feeley, R. Optim Eng 2006, 7, 459–478.
15. Djurisic, Z. M.; Amusin, D.; Bereknyei, T.; Allison, T. C.; Frenklach, M. Paper 05F-43. In 2005 Fall Meeting of the Western States Section of the Combustion Institute, Stanford, CA, 2005.
16. Djurisic, Z. M.; Frenklach, M.; Golden, D. M.; Gupta, A.; Davidson, D. F.; Wang, H. Poster paper 2C29. In 31st International Symposium on Combustion, Heidelberg, Germany, 2006.
17. Djurisic, Z. M.; Langston, M.; Frenklach, M. Paper No. C11. In 5th U.S. National Meeting of the Combustion Institute, San Diego, CA, 2007.
18. Djurisic, Z. M.; Frenklach, M. Poster Paper W5P051. In 32nd International Symposium on Combustion, Montreal, Canada, 2008.
19. Yeates, D.; You, X. Q.; Frenklach, M. In 6th U.S. National Meeting of the Combustion Institute, Ann Arbor, MI, 2009.
20. Frenklach, M.; Gutkin, M. AFOSR Technical Report, University of California, Berkeley, CA, July 14, 2009.
21. Li, G.; Rosenthal, C.; Rabitz, H. J Phys Chem A 2001, 105, 7765–7777.
22. MATLAB; version R2010a, The Mathworks, Inc.: Natick, MA, 2010.
23. Sharp, J. Visual C# 2010; Microsoft: Redmont, WA, 2010.
24. The HDF Group; Available at: http://www.hdfgroup.org/HDF5, accessed September 30, 2011.
25. Chaos, M.; Dryer, F. L. Combust Sci Technol 2008, 180, 1053–1096.
26. Sinaki, M. Y.; Matida, E. A.; Hamdullahpur, F. Int J Hydrogen Energy 2011, 36, 2936–2944.
27. Conaire, M.; Curran, H. J.; Simmie, J. M.; Pitz, W. J.; Westbrook, C. K. Int J Chem Kinet 2004, 36, 603–622.
28. Davis, S. G.; Joshi, A. V.; Wang, H.; Egolfopoulos, F. Proc Combust Inst 2005, 30, 1283–1292.
29. Hong, Z.; Davidson, D. F.; Hanson, R. K. Combust Flame 2011, 158, 633–644.
30. Li, J.; Zhao, Z.; Kazakov, A.; Dryer, F. L. Int J Chem Kinet 2004, 36, 566–575.
31. Burcat, A.; Ruscic, B. ftp://ftp.technion.ac.il/pub/supported/aetdd/thermodynamics/, accessed September 30, 2010.
32. Smith, G. P.; Golden, D. M.; Frenklach, M.; Moriarty, N. W.; Eiteneer, B.; Goldenberg, M.; Bowman, C. T.; Hanson, R.; Song, S.; Gardiner, W. C., Jr.; Lissianski, V.; Qin, Z. GRI-Mech 3.0; Available at: http://www.me.berkeley.edu/gri_mech/, accessed September 30, 2010.
33. Tsang, W.; Hampson, R. F. J. Phys Chem Ref Data 1986, 15, 1087–1279.
34. Troe, J. Proc Combust Inst 2000, 28, 1463–1469.
35. Hong, Z.; Cook, R. D.; Davidson, D. F.; Hanson, R. K. J Phys Chem A 2010, 114, 5718–5727.
36. Baulch, D. L.; Cobos, C. J.; Cox, R. A.; Esser, C.; Frank, P.; Just, T.; Kerr, J. A.; Pilling, M. J.; Troe, J.; Walker, R. W.; Warnatz, J. J Phys Chem Ref Data 1992, 21, 411–734.
37. Hong, Z.; Vasu, S. S.; Davidson, D. F.; Hanson, R. K. J Phys Chem A 2010, 114, 5520–5525.
38. Michael, J. V.; Sutherland, J. W.; Harding, L. B.; Wagner, A. F. Proc Combust Inst 2000, 28, 1471–1478.
39. Harding, L. B.; Klippenstein, S. J.; Jasper, A. W. Phys Chem Chem Phys 2007, 9, 4053–4054.
40. Saxena, P.; Williams, F. A. Combust Flame 2006, 145, 316–323.
41. Burke, M. P.; Chaos, M.; Dryer, F. L.; Ju, Y. Combust Flame 2010, 157, 618–631.
42. Box, G. E. P.; Hunter, W. G.; Hunter, J. S. Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building; Wiley: New York, 1978.
43. Cohen, A.; Larsen, J. Report AD0667362, US Army Ballistics Research Laboratory, 1967; p. 1386.
44. Santner, T. J.; Williams, B. J.; Notz, W. The Design and Analysis of Computer Experiments; Springer: New York, 2003.
45. Forrester, A.; Sóbester, A.; Keane, A. Engineering Design via Surrogate Modelling; Wiley: Chichester, 2008.