OPTIMIZATION AND CONSISTENCY OF A REACTION DATASET

M. Frenklach, R. Feeley, A. Packard

Department of Mechanical Engineering, University of California at Berkeley, Berkeley, CA 94720-1740

and P. Seiler

Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana, IL

Understanding a variety of natural phenomena and industrial processes relies on the knowledge of chemical reaction mechanisms and kinetics. Endeavor in such cases begins with identification of underlying reaction pathways and fundamental mechanisms. When sufficient data accumulate, the interest often shifts to practical applications, motivating the development of mechanistic models.

The "textbook" approach to the development of mechanistic reaction models consists in conjecturing the reaction mechanism, expressing it in a suitable mathematical form, and comparing the predictions of the constructed model to available experimental observations. Typically, such tests result in a mixed outcome, some showing a reasonably close agreement and some not. The apparent inconsistency between the model and experiment obtained in the latter case is argued then to imply either that the model is inadequate or that the experiment (or rather its interpretation) is incorrect.

In some areas, such as heterogeneous catalysis and biochemical systems, the fundamental reaction mechanisms are largely unknown and establishing them form the challenge of the current research. Yet, in other fields, such as atmospheric chemistry and combustion of small hydrocarbons like methane, there are broad consensus over the reaction pathways underlying the mechanisms and the inadequacy of the kinetic models essentially rests in their parameter values. In the following we assume the latter situation.

If the kinetic parameters of such a "known" mechanism were known exactly, then a direct comparison of model prediction with a given experiment, within its uncertainties, would decisively indicate whether that experiment is consistent or inconsistent with the model. In reality, however, the model parameters themselves have uncertainties, and they have to be included in the analysis.

In principle, the parameter identification of chemical kinetic models can be posed as classical statistical inference: given a mathematical model and a set of experimental observations for the model responses, determine the best-fit parameter values, usually those that produce the smallest deviations of the model predictions from the measurements. The validity of the model and the identification of outliers is then determined using analysis of variance. The difficulty of applying standard statistical methods lies in the fact that chemical kinetics models are stated in the form of differential equations that do not possess a closed-form solution. Further complications arise from the highly "ill-structured" character of the objective function, with long and narrow valleys, resulting in ill-conditioned optimization and lack of unique solution.

The optimization problem for general, non-linear dynamic models has been addressed with a series of numerical methods: "direct" gradient search, gradient search based on sensitivities, solution mapping, genetic algorithms, and Monte Carlo techniques. In some cases, it was coupled with statistical inference and estimation of confidence regions. Recent developments also include formulation of the problem in the form of error propagation: given a set of uncertainty ranges of model parameters, estimate the intervals of variations for model predictions.

All of the above methods essentially view the problem as a two-step process: estimation of model parameters from fitting a selected set of experimental data followed by exercise of the obtained model, either as validation against an additional set of experiments or making predictions outside the experimentally accusable conditions.

Recently, we have pursued a different approach, which we call *data collaboration*.[1] In this approach we focus not on parametrization of the parameter uncertainty region, which the above methods engage in and rely upon, but rather on transferring the uncertainties of the "raw" (experimental) data into the model *directly*.[2] Doing so allows harvesting substantially more of the information content of the data, and to obtain more realistic bounds of model predictions. Our approach is anchored in the concept of a *dataset* that unites all the pertinent experimental data and the mechanistic knowledge for a given system, and the numerical analysis is based on combination of Solution Mapping and new developments of the Robust Control Theory. This numerical methodology avoids unnecessary over-constraining of model parameters, which plagues many other techniques due to inherent correlations among parameters, while exploring more closely the true *feasible set* of the parameter space in a computationally efficient manner.

The present work expands further on these ideas. New mathematical developments allow us to address the mutual consistency of experimental observations, within the framework of a dataset.

Acknowledgement. The work was supported by National Science Foundation, Information Technology Research Program, Grant No. CTS-0113985.

References

(1) Frenklach, M.; Packard, A.; Seiler; P.; Feeley, R. *Int. J. Chem. Kinet.* **2004**, in press.

(2) Frenklach, M.; Packard, A.; Seiler; P. In *Proceedings of the American Control Conference*, IEEE Catalog Number 02CH37301C, 2002, pp. 4135-4140.