

114c A Statistical Methodology for Building Catalytic Reaction Models with High Throughput Experimentation

Shuo-Huan Hsu, Gary E. Blau, James M. Caruthers, W. Nicholas Delgass, Fabio H. Ribero, and Venkat Venkatasubramanian

High Throughput Experimentation (HTE) generates an enormous amount of data, where the challenge is to extract information from this data. Mathematical models of the reactions occurring on the surface of the catalyst represent the knowledge contained in the data, including both the reaction mechanisms and the rate constants. Such models are inherently nonlinear (e.g. Arrhenius temperature dependence) and all kinetic data is contaminated with experimental error. Existing commercial packages for statistical data analysis (e.g. SAS, S-Plus, etc.) employ linear error analysis methods, which can lead to highly misleading conclusions. Even packages which permit nonlinear parameter estimation (e.g. Athena, AcslExtreme Optimize, etc) frequently converge to the wrong answer, fail to discriminate rival models or give an inaccurate representation of the quality of the parameter estimates. In some cases this is a consequence of the statistical assumptions surrounding the errors in the data. In others, the data from HTE experimentation is poorly designed.

The speed and storage capacity of modern computers now enable Monte Carlo based statistical methods that allow proper nonlinear analysis free from the assumptions of the classical approach. To this end we have developed a modeling system that uses the maximum likelihood method to estimate the parameters in the kinetic model simultaneously with the parameters in the statistical error model. Statistical modeling is used to accommodate data heteroscedasticity. Bayesian methods are used to generate true posterior probability distributions for the parameters along with their maximum likelihood estimates using Markov Chain Monte Carlo (MCMC) methods [1]. As part of our model building tool kit we have included procedures to suggest the experiments necessary to discriminate rival models and, once the best model is found, to improve the quality of the parameter estimates. For the former, experiments are located where rival models exhibit their greatest differences. For the latter, experiments are located where the uncertainties in the posterior parameter distributions for the model selected are minimized either by reshaping the confidence region (E-optimal design) and/or by minimizing the volume (D-optimal design). The framework is demonstrated for a number of typical catalyst model building problems, where power of these nonlinear statistical methods in combination with HTE data will be shown.

[1] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*: Chapman & Hall/CRC, 1996.