

Bayesian inference of chemical kinetic models from proposed reactions

Nikhil Galagali^a, Youssef M. Marzouk^{b,*}

^a*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

^b*Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Abstract

Bayesian inference provides a natural framework for combining experimental data with prior knowledge to develop chemical kinetic models and quantify the associated uncertainties, not only in parameter values but also in model structure. Most existing applications of Bayesian model selection methods to chemical kinetics have been limited to comparisons among a small set of models, however. The significant computational cost of evaluating posterior model probabilities renders traditional Bayesian methods infeasible when the model space becomes large. We present a new framework for tractable Bayesian model inference and uncertainty quantification using a large number of systematically generated model hypotheses. The approach involves imposing point-mass mixture priors over rate constants and exploring the resulting posterior distribution using an adaptive Markov chain Monte Carlo method. The posterior samples are used to identify plausible models, to quantify rate constant uncertainties, and to extract key diagnostic information about model structure—such as the reactions and operating pathways most strongly supported by the data. We provide numerical demonstrations of the proposed framework by inferring kinetic models for catalytic steam and dry reforming of methane using available experimental data.

Keywords: Bayesian inference, chemical kinetics, model selection, Markov chain Monte Carlo, adaptive MCMC, online expectation maximization

1. Introduction

A detailed elementary chemical kinetic model is a critical component of simulation tools in a wide range of applications, including combustion, catalysis, electrochemistry, and biochemistry. The development of kinetic models for surface reactions, charge transfer reactions, and biological reactions is particularly challenging, because there is typically little understanding about the operating reaction pathways and often no knowledge about rate parameter values in these settings. For instance, there exist a number of competing hypotheses about H₂ and CO oxidation mechanisms for a solid-oxide fuel cell (Hanna et al., 2014). The standard approach to building models in such a case is to postulate reaction models and to compare them based on their ability to reproduce indirect system-level experimental

*Corresponding author

Email addresses: nikhilg1@mit.edu (Nikhil Galagali), ymarz@mit.edu (Youssef M. Marzouk)

data. Data-driven model comparison thus involves defining a metric of fit, e.g., the sum of squared residuals, and selecting a model that minimizes this metric, e.g., least-squares or regularized least-squares fitting (Hanna et al., 2014; Vogler et al., 2009; Yurkiv et al., 2011). The principal limitation of the least-squares fitting approach is that it only identifies a single “best” model and yields point estimates of the underlying parameter values, without providing a meaningful description of the uncertainty among competing models and in their parameters. In addition, the estimated parameter values are usually only local optima of the fitting metric, although methods that base model comparison on global optima do also exist (Feeley et al., 2006).

Bayesian model inference overcomes the limitations of a least-squares fitting approach by providing a rigorous method for fusing available experimental data with prior knowledge, to yield a fuller description of model and parameter uncertainties (Bernardo and Smith, 2000; Gelman et al., 2004; Sivia, 2006). The application of Bayesian model inference to chemical kinetics, however, presents a significant computational challenge. Model discrimination in Bayesian analysis is based on computing model probabilities conditioned on available data, i.e., *posterior* model probabilities. Rigorous computation of posterior model probabilities requires evaluation of a high-dimensional integral for each model. A number of methods exist in the literature for this purpose (Chib and Jeliazkov, 2001; Friel and Pettitt, 2008; Kass and Raftery, 1995; Newton and Raftery, 1994), but they are computationally taxing. Applications of these methods to chemical kinetics have been pursued in combustion modeling and systems biology when the number of competing models is small (Braman et al., 2013; Vyshemirsky and Girolami, 2008; Vyshemirsky et al., 2006; Wilkinson, 2011; Xu et al., 2010; Zechner et al., 2014). When the number of competing models becomes large, however, the above methods become computationally infeasible. Yet a more systematic approach to model inference naturally requires appraising a combinatorially large number of models: instead of a few model hypotheses, one might start with a list of proposed reactions, for example, and form a collection of plausible models by considering all valid combinations of the proposed reactions.

In this paper, we present a systematic Bayesian inference framework for the development and uncertainty quantification of chemical kinetic models, one that systematically generates and compares a comprehensive set of model hypotheses. The framework makes use of the following observation. Reacting flow or process models with chemical kinetic subcomponents almost always yield outputs that are nonlinear functions of the reaction rate parameters. Nevertheless, if the rate of each elementary reaction is described by the law of mass action, e.g.,

$$\text{Rate} = -k[X][Y], \quad (1)$$

where k is the reaction rate constant and $[X]$ and $[Y]$ are reactant concentrations in appropriate units, it is possible to *eliminate* a reaction from a kinetic model simply by setting the corresponding rate constant to zero. Thus, given a set of proposed reactions, one can switch between kinetic models by setting the appropriate rate constants either to zero or nonzero values. Then, using a sampling scheme to explore the model space enables the estimation of posterior model probabilities, in proportion to the frequency of model visits.

Computations in our new framework are based on recent work in the statistics literature by Ji and Schmidler (2013), which proposes an efficient numerical approach to the nonlinear Bayesian variable selection problem. Effectively, our framework proposes to treat the kinetic

model inference problem as a particular form of variable selection. We impose point-mass mixture priors on the rate constants of all proposed elementary reactions and use an adaptive Markov chain Monte Carlo (MCMC) method to efficiently explore the model space, along with the rate parameter values (Ji and Schmidler, 2013; Mitchell and Beauchamp, 1988). MCMC is a class of sampling methods that enables exploration of complex multi-dimensional probability distributions using an easily-sampled proposal distribution coupled with an accept-reject step (Robert and Casella, 2004). The efficiency of posterior exploration, and in turn the quality of model probability and parameter value estimates, is directly linked to how well the proposal distribution approximates the posterior distribution. Following Andrieu and Moulines (2006), we use an adaptive independence MCMC algorithm based on online expectation maximization (EM) for posterior exploration. Both online EM and the adaptive MCMC method of Ji and Schmidler (2013) use a stochastic approximation method to continually update proposal parameters using posterior samples. While the algorithm in (Ji and Schmidler, 2013) relies on Monte Carlo estimates of gradients, the online EM approach adopts the two-step procedure of first updating estimates of posterior summaries (E-step) and then adapting parameters via simple analytical expressions (M-step). A related sampling approach, reversible-jump MCMC (Green, 1995), has recently been used to infer kinetic models in systems biology (Oates et al., 2012). But designing effective proposal distributions for between-model moves in a reversible-jump algorithm can be quite difficult in practice (Green and Hastie, 2009). In contrast, coupling our nonlinear variable selection framework with an adaptive MCMC algorithm provides an easy-to-use and efficient inference methodology.

Note that the modeling of species interactions in our work contrasts with an approach commonly used in systems biology (Friedman and Koller, 2003; Sachs et al., 2002, 2005), where the relationships between species are described probabilistically using Bayesian networks. In the latter approach, interactions between species (i.e., edges in the Bayesian network) are modeled using Gaussian or multinomial probability distributions; this assumption makes the calculation of individual posterior model probabilities analytically tractable. By contrast, the interactions between species in our work are modeled using systems of differential equations defined by the law of mass action. These equations can be further embedded into differential equation models that describe convective and diffusive transport, surface interactions, and other physical phenomena that affect experimental observations.

This paper is divided into six sections. In Section 2, we discuss different approaches for model inference and highlight advantages of the Bayesian methodology. Section 3 details our new Bayesian formulation for the inference of chemical kinetic models. Section 4 describes the adaptive MCMC methodology used for efficient exploration of the posterior distribution in this model inference approach. In Section 5, we apply our approach to examples involving the catalytic reforming of methane, with both real and synthetic data sets. The final section provides a brief summary and conclusions.

2. Model inference

Model inference can be defined informally as the assessment of models to ascertain the degree to which each is supported by available data. A prerequisite for model inference is the availability of (i) plausible models and (ii) relevant data to discriminate among the

models. It is important at this stage to distinguish *model inference* from the common practice of *model reduction* in chemical kinetics (Bhattacharjee et al., 2003; Oluwole et al., 2006). Model reduction refers to a systematic reduction in the size of a large kinetic model so as to reproduce model outputs within a specified tolerance. Such a procedure, however, assumes that an accurate model (i.e., the full kinetic model) is already known and fixed. And, crucially, it does not take experimental data into account during reduction.

2.1. What is a model?

A model of a physical process describes a specific collection of input-output relationships. In particular, a model describes how some pre-specified *quantities of interest* are related to input variables. As a result, a model may preclude the description of quantities for which it has not been specifically built. Figure 1 shows a typical chemically reacting flow model. This model—consisting of governing equations expressing conservation laws, chemical reactions, and thermo-kinetic parameters—may relate inputs such as concentration C_{in} , temperature T_{in} , pressure P , and applied voltage ΔV to observables such as concentration C_{out} , ignition delay τ_{ign} , and current I .

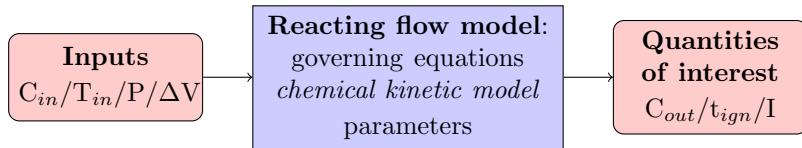


Figure 1: Chemically reacting flow model

2.2. Classical approaches to model inference

Performing model inference purely on the basis of agreement between model predictions and experimental observations runs the risk of fitting to noise in the data (Bishop, 2007; Hastie et al., 2009), i.e., over-fitting. Because a more complex model can more easily adapt to any available data, a suitable model inference criterion is one that rewards goodness of fit but also penalizes model complexity. This guiding principle for the assessment of models is encapsulated by *Occam's razor* (Mackay, 2003).

A wide variety of methods have been suggested for model inference. Usually, the data available in chemical kinetics settings are limited. Consequently, methods such as k -fold cross validation, which work best when data are plentiful, are not suitable (Arlot and Celisse, 2010; Hastie et al., 2009); in a data-poor context, the cross validation metric is noisy and its results can be highly variable. A method that explicitly incorporates an Occam's razor is more useful for the data-deficient case that one typically encounters in chemical kinetics.

Classical (i.e, non-Bayesian) model inference methodologies such as the Akaike information criterion (AIC), Mallow's C_p , and penalized least-squares (Akaike, 1974; Burnham and Anderson, 2002; Mallows, 1973) impose an Occam's razor by selecting models via the following type of optimization problem:

$$M^* = \arg \min_M \| \mathcal{D} - G_M(k_M) \|_2^2 + \alpha h(M), \quad (2)$$

where M^* is the optimal model, $\|Y - G_M(k_M)\|_2$ is the data-model discrepancy with \mathcal{D} representing the available data and $G_M(k_M)$ the model predictions, $h(M)$ is a measure of model complexity, and α is the penalty on model complexity. These optimization approaches tend to be *ad hoc*, as the results may depend strongly on the penalty parameter α and it is unclear how to choose it systematically (Berger and Pericchi, 2001). Moreover, these methods yield point estimates (i.e., they select a single model) and thus do not characterize uncertainty in the model M .

The Bayesian approach to model inference, in contrast, contains an automatic Occam's razor and yields a full description of posterior model and parameter uncertainties (Mackay, 2003). The computed posterior model probabilities can directly be interpreted as the degree to which different models are supported by available data (Berger and Pericchi, 2001). The effective use of these model probabilities then depends on the application context or goals at hand; subsequent decisions or modeling predictions could be based on the single model with the highest probability, on a pool of high probability models, or on all the models weighed according to their posterior probabilities (Hoeting et al., 1999).

2.3. Bayesian approach to model inference

Bayesian statistics provides a rigorous inferential framework for assimilating noisy and indirect data, a natural mechanism for incorporating prior knowledge from different sources, and a full description of uncertainties in model structure and parameter values (Bernardo and Smith, 2000; Gelman et al., 2004; Sivia, 2006). Inference is based on Bayes' rule of probability:

$$p(k|\mathcal{D}) = \frac{p(\mathcal{D}|k)p(k)}{p(\mathcal{D})}, \quad (3)$$

where k is the parameter being inferred, $p(k|\mathcal{D})$ is the posterior probability density of k conditioned on available data \mathcal{D} , $p(\mathcal{D}|k)$ describes the probability of observing \mathcal{D} given the parameter value k , and $p(k)$ is the prior probability density of parameter k . The denominator $p(\mathcal{D})$, commonly referred to as the *evidence* or marginal likelihood, is the probability of the observed data. Sampling the posterior distribution yields a description of uncertainty in the parameters and allows the estimation of posterior summaries such as the mean and standard deviation. Posterior exploration by sampling is usually non-trivial and typically carried out via a carefully designed MCMC procedure (Andrieu et al., 2003; Gilks et al., 1996; Robert and Casella, 2004; Tierney, 1994). Examples of Bayesian parameter inference applied to chemical kinetics are given in (Prager et al., 2013; Vahteristo et al., 2008, 2013).

The comparison of *models* in the Bayesian paradigm relies on the computation of posterior model probabilities $p(M|\mathcal{D})$. Applying Bayes' rule to a collection of models $\{M_j\}$, the posterior probability of model M_j is given by

$$\begin{aligned} p(M_j|\mathcal{D}) &\propto p(M_j)p(\mathcal{D}|M_j) \\ &\propto p(M_j) \int p(\mathcal{D}|\bar{k}_j)p(\bar{k}_j|M_j)d\bar{k}_j, \end{aligned} \quad (4)$$

where $p(M_j)$ is the prior model probability, $p(\mathcal{D}|M_j)$ is the evidence of model M_j , and \bar{k}_j is the vector of parameters in model M_j . The evidence $p(\mathcal{D}|M_j)$ incorporates an automatic Occam's razor that penalizes unnecessary model complexity; see Mackay (2003) for details.

As discussed in Section 2.2, Bayesian model inference offers many advantages over other approaches. Yet a central challenge of applying the Bayesian methodology is its computational cost. Comparisons based on an individual calculation of the evidence for each plausible model—either by model-specific MCMC runs or by using numerical quadrature—quickly become infeasible if the number of models becomes large. Chemical kinetics is particularly prone to this difficulty since the number of possible models can grow exponentially with N , the number of proposed reactions. Alternatives such as Laplace approximations and the Schwarz criterion have been suggested (Burnham and Anderson, 2002; Mackay, 2003; Schwarz, 1978), but they involve approximations of the posterior distribution. In general, very little can be said about the quality of inference based on these approximate methods. On the other hand, across-model sampling offers a solution in cases where the number of models is large (Carlin and Chib, 1995; Dellaportas et al., 2002; Godsill, 2001; Green, 1995). These methods work by making the MCMC sampler jump between models to explore the joint space of models and parameters. Model probabilities are estimated from the number of times the sampler visits each model. The prohibitively high cost of model comparisons based on the computation of evidence for each model is avoided, as the sampler visits each model in proportion to its posterior probability. The drawback of across-model sampling schemes, however, is that they require extensive tuning or pre-existing knowledge of the posterior distribution in order to achieve efficient exploration (Oates et al., 2012). The method we use in this paper for posterior exploration is an *adaptive* across-model sampling scheme, wherein the proposal distribution in the MCMC algorithm continuously adapts based on previous posterior samples to avoid manual tuning (Ji and Schmidler, 2013).

3. Bayesian inference of chemical kinetic models

We now present the details of our new framework for systematic and rigorous inference of chemical kinetic models. As we saw in Section 1, the law of mass action gives the rate of a chemical reaction (say $X + Y \rightarrow Z$) as the product of a reaction-specific rate constant k with reactant concentrations $[X]$ and $[Y]$. The rate constant k is expressed in Arrhenius form as

$$k = AT^n \exp\left(-\frac{E_a}{RT}\right), \quad (5)$$

where A is the pre-exponential factor, E_a is the activation energy, n is the temperature exponent, R is the universal gas constant, and T is temperature. In this paper, we treat k as the combined unknown parameter; it is also possible to infer A , E_a , and n separately (given observations over a range of temperatures) but we leave such a treatment for subsequent work.

In any chemically reacting process, the rates of the individual elementary reactions in the kinetic model together determine the values of the observables. And reacting flow models are seldom linear; that is, the model outputs depend nonlinearly on the elementary reaction rates and on the rate constants. Nevertheless, the mass-action rate expression (1) enables a reaction to be eliminated from the kinetic model simply by setting the corresponding rate constant to zero. As a result, a sampling scheme over a parameter space of *fixed* dimension can be employed to compute posterior model probabilities. For example, consider a setting where we have $N = 5$ postulated elementary reactions with rate constants $k_1, k_2, k_3, k_4,$

and k_5 . Assume, for simplicity, that the reacting flow model outputs can be produced from the inputs via any of the five reactions. Thus we wish to compare $2^5 - 1$ kinetic models that are *a priori* plausible. The key idea is to recognize that switching from a model M_i (for instance, comprising reactions 1, 2, and 5) to a model M_j (for instance, comprising reactions 3 and 4) requires that the parameter vector $\bar{k} \equiv (k_1, k_2, k_3, k_4, k_5)$ change from $(a, b, 0, 0, c)$ to $(0, 0, d, e, 0)$, where a, b, c, d , and e are nonzero rate constants for each reaction.

The first step in developing a sampling scheme that turns reactions off and on is to impose point-mass mixture priors on the rate constants k_i (Ji and Schmidler, 2013; Mitchell and Beauchamp, 1988). For simplicity, in the subsequent numerical demonstrations we will take the priors to be independent in each dimension (i.e., for each reaction), such that $p(\bar{k}) = \prod_{i=1}^N p_i(k_i)$. We note, however, that priors can certainly be designed to reflect any additional information, i.e., knowledge that necessitates the joint inclusion and exclusion of reactions. In any case, a point-mass mixture prior is given by

$$p_i(k_i) = w_{0,i}\delta(k_i) + w_{1,i}\mathcal{C}_i(k_i), \quad (6)$$

where $w_{0,i}$ and $w_{1,i} = 1 - w_{0,i}$ are weights of the two prior components. $\delta(k_i)$ is a probability atom (a point with unity probability mass) at zero and $\mathcal{C}_i(k_i)$ is the continuous component of the prior distribution. The continuous component of the prior probability distribution describes any prior information about the values that the rate constant can take and is often elicited from experts. If no such information exists, $\mathcal{C}_i(k_i)$ may be a uniform or log-uniform distribution over all positive real numbers (an ‘uninformative’ prior). In any case, Bayesian inference and indeed our framework allow the model developer significant flexibility in setting the prior distribution based on his or her subjective belief or any pre-existing information. The weights $w_{0,i}$ and $w_{1,i}$ are prior beliefs about reaction i being included or excluded, respectively, from the inferred model. The model developer may use these weights to impose prior information about the importance of this reaction in modeling the reacting flow model output.

It is instructive to discuss two specific cases. First, if the model developer has no prior preference for the inclusion or exclusion of a reaction, then an appropriate choice for the weights is an indifference prior setting of $w_{0,i} = w_{1,i} = 0.5$. In contrast, if the model developer believes that reaction i should definitely be part of the inferred model, then he/she can set $w_{0,i}$ to zero and $w_{1,i}$ to one. Note that if all the reactions are assigned a prior inclusion probability of $w_1 = 1.0$, then the model inference framework reduces to the familiar Bayesian parameter inference problem.

Letting \mathcal{D} denote the available data, an application of Bayes’ rule to the parameter vector \bar{k} yields

$$p(\bar{k}|\mathcal{D}) \propto p(\mathcal{D}|\bar{k})p(\bar{k}). \quad (7)$$

Here $p(\mathcal{D}|\bar{k})$ is viewed as a function of \bar{k} : it is the likelihood function, which reflects the discrepancy between the data \mathcal{D} and the model prediction at the specified \bar{k} . The precise form of the likelihood function depends on the noise model used to describe the data. For instance, in the examples of Section 5, we use an additive Gaussian model, yielding

$$\mathcal{D} = G(\bar{k}) + \epsilon. \quad (8)$$

Here $G(\bar{k})$ is the prediction of the *forward* model (the chemically reacting flow model) at the specified parameter value \bar{k} , and ϵ reflects a combination of observational noise and

model errors. We assume that every component of ϵ_j of ϵ is independent with mean zero and variance σ^2 , $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$. Because the prior distribution on each k_i is a point-mass mixture (6), the resulting posterior distribution of \bar{k} is also a mixture distribution over the product space of all reactions, where each component of the mixture contains a different set of non-zero rate constants and thus represents a different model. Sampling the posterior distribution of \bar{k} implies visiting posterior mixture components in proportion to their probabilities. Therefore, a scheme that samples $p(\bar{k}|\mathcal{D})$ will not only provide a full description of uncertainties in rate constant values, but will also yield estimates of the posterior model probabilities proportional to the number of times each posterior mixture component is visited.

4. Posterior exploration by Markov chain Monte Carlo

4.1. Sampling using an independence Metropolis-Hastings algorithm

The multi-dimensional posterior distribution of the parameter vector \bar{k} obtained in Section 3 cannot be sampled directly; because of the nonlinear forward model $G(\bar{k})$, the likelihood does not have a standard form and certainly is not conjugate to the prior distribution. However, simulating posterior samples is possible using the independence Metropolis-Hastings (MH) algorithm (Robert and Casella, 2004; Tierney, 1994), which is a particular type of MCMC sampling scheme. MCMC methods work by iteratively generating samples from a Markov chain whose limiting distribution is the posterior distribution. This process involves simulating a sample from a *proposal distribution*—a distribution that can be directly sampled—and passing the proposed sample through an accept-reject step. In an independence Metropolis-Hastings algorithm, the proposal distribution at each step is *independent of the current location of the chain*. Algorithm 1 describes the independence sampler using pseudocode. We note that another commonly-used class of MCMC algorithms, random-walk Metropolis-Hastings, is not suitable for our problem because its chains will tend to remain “stuck” in the point mass component of a parameter posterior unless the support of the continuous component is very close to zero.

Algorithm 1 The independence Metropolis-Hastings algorithm

```

1: Given: Data  $\mathcal{D}$ , prior density  $p(\bar{k})$ , likelihood function  $p(\mathcal{D}|\bar{k})$ , proposal  $q(\bar{k})$ , number
   of steps  $T$ 
2: Initialize  $\bar{k}^0$ 
3: for  $t = 0$  to  $T - 1$  do
4:   Sample  $u \sim \mathcal{U}_{[0,1]}$ 
5:   Sample  $\bar{k}^* \sim q(\bar{k}^*)$ 
6:   if  $u < \mathcal{A}(\bar{k}^t, \bar{k}^*) = \min \left\{ 1, \frac{p(\bar{k}^*|\mathcal{D})q(\bar{k}^t)}{p(\bar{k}^t|\mathcal{D})q(\bar{k}^*)} \right\}$  then
7:      $\bar{k}^{t+1} = \bar{k}^*$ 
8:   else
9:      $\bar{k}^{t+1} = \bar{k}^t$ 
10:  end if
11: end for
```

The Metropolis-Hastings algorithm’s efficiency in exploring the posterior distribution rests on the design of an effective proposal distribution. “Efficiency” in this context refers to how effectively the Markov chain explores the posterior—i.e., how nearly independent its states are—which translates directly into the Monte Carlo error of a sample-based posterior estimate. A good proposal distribution will require fewer posterior density evaluations to achieve a given error. Recall that computation of the posterior density $p(\bar{k}|\mathcal{D})$ for a proposed parameter value involves evaluating the likelihood $p(\mathcal{D}|\bar{k})$, which in turn requires solving the forward model. Restricting the number of forward model solves is especially important in the present application context, because detailed models of chemically reacting flow are computationally expensive.

Since the marginal posterior distribution of each parameter k_i is a mixture of a point mass and continuous components, the proposal distribution for each k_i is taken to be an independent point-mass mixture distribution of the form:

$$q(k_i; \psi_i) = b_{i,0}\delta(k_i) + \sum_{m=1}^M b_{i,m}q_m(k_i; \theta_{i,m}). \quad (9)$$

In the above equation, $\delta(k_i)$ is a point mass at zero, $q_m(k_i; \theta_{i,m})$ are continuous components of the proposal distribution, and $\bar{\psi} \equiv (b_{i=1:N, m=0:M}, \theta_{i=1:N, m=1:M})$ comprises all the parameters describing the proposal distribution. Recall that N is the number of proposed reactions, and thus the dimension of the posterior distribution. The number of continuous components M in each dimension is a choice left to the user. Increasing M can potentially improve the approximation of the posterior by the proposal, especially if the continuous part of the posterior distribution is itself multimodal. This is desirable, because a good proposal distribution for independence Metropolis-Hastings is generally one that approximates the posterior as closely as possible. But higher values of M increase the number of parameters needed to describe the proposal distribution, which can affect the cost and convergence rate of the proposal adaptation scheme discussed in Section 4.2. Choosing an independent proposal distribution for each parameter k_i means that the joint proposal distribution is given by

$$q(\bar{k}; \bar{\psi}) = \prod_{i=1}^N q(k_i; \psi_i). \quad (10)$$

4.2. Adaptive MCMC by online expectation maximization

As noted above, efficient sampling suggests that we choose the proposal parameters $\bar{\psi}$ so that (10) closely approximates the posterior. Of course, the true posterior distribution is not characterized a priori; its exploration is in fact the goal of MCMC. A useful strategy for improving sampling efficiency is, then, to continuously adapt the proposal parameters based on past samples from the MCMC chain. Algorithms of this kind are known as *adaptive MCMC* and require additional theoretical analysis to guarantee convergence to the target distribution (Robert and Casella, 2004). A commonly used adaptation criterion is to tune the proposal parameters to minimize the Kullback-Leibler (KL) divergence from the posterior distribution to the proposal distribution (Andrieu and Moulines, 2006; Ji and Schmidler, 2013). We adopt this strategy here and detail the adaptive independence Metropolis-Hastings algorithm as follows.

Formally, the optimal proposal parameters are given by

$$\bar{\psi}^* = \arg \min_{\bar{\psi}} \mathcal{D}_{KL}(p(\bar{k}|\mathcal{D})\|q(\bar{k};\bar{\psi})) = \arg \min_{\bar{\psi}} \int p(\bar{k}|\mathcal{D}) \log \left(\frac{p(\bar{k}|\mathcal{D})}{q(\bar{k};\bar{\psi})} \right) d\bar{k}. \quad (11)$$

Since this objective function involves integration over the posterior distribution $p(\bar{k}|\mathcal{D})$, finding a solution before exploring the posterior is difficult. An effective strategy is to use a stochastic approximation method (Kushner and Yin, 1997; Robbins and Monro, 1951) that couples posterior exploration with the solution of the minimization problem. A generic stochastic approximation method for problem (11) involves iteratively (i) simulating a batch of samples from the posterior distribution to estimate the KL divergence above, then (ii) using those results to update the proposal parameters. Under conditions explained by Andrieu et al. (2005), the proposal parameters converge to the optimal solution of (11) asymptotically. Within this general procedure, one could consider two possible instantiations. The first is stochastic gradient descent: simulate a finite number of samples from the posterior distribution and use them to compute a noisy estimate of the gradient of the objective in (11) with respect to $\bar{\psi}$; then take a step in the negative-gradient direction to update the parameters in each iteration. This approach is detailed in the paper by Ji and Schmidler (2013). The second approach involves solving (11) using a method called online expectation maximization (EM) (Andrieu and Moulines, 2006). Online EM alternately uses posterior samples to update estimates of the expectation of the logarithm of *complete-data likelihood* (E-step) and then directly adapts the proposal parameters using analytical expressions (M-step). We found the online EM approach to be more robust in practice, and have thus adopted it for this work. (See Appendix A.3 for more details on the complete-data likelihood.)

Here, we describe the expressions used to update the proposal parameters using the online EM algorithm. A detailed derivation of the online EM algorithm applied to point-mass mixture priors can be found in Appendix A. We consider the case where the reaction rate parameter vector is N -dimensional, i.e., $\bar{k} = (k_1, k_2, \dots, k_N)$, where T samples are simulated from the posterior distribution $p(\bar{k}|\mathcal{D})$ between each proposal parameter update, and where the M continuous components of the proposal distribution are Gaussian, resulting in a proposal of the form:

$$q(k_i; \psi_i) = b_{i,0}\delta(k_i) + \sum_{m=1}^M b_{i,m}\mathcal{N}_m(k_i; \theta_{i,m}). \quad (12)$$

A non-adaptive component \tilde{q} must also be added to the proposal distribution to satisfy conditions for the convergence of the adaptive MCMC algorithm to the posterior distribution (see Appendix A.3 for details). Thus, the overall proposal in each dimension is given by

$$q_s(k_i) = \lambda_i \tilde{q}(k_i, \tilde{\psi}_i) + (1 - \lambda_i)q(k_i; \psi_i), \quad (13)$$

where $0 < \lambda_i < 1$ and $\tilde{\psi}_i$ is a fixed set of proposal parameter values. At each step n of the online EM algorithm:

1. Simulate T samples $\bar{k}^1, \bar{k}^2, \dots, \bar{k}^T$ from the posterior $p(\bar{k}|\mathcal{D})$ using an independence Metropolis-Hastings algorithm with the current proposal parameter values.

2. Compute (for all parameters $i = 1 \dots N$)

For $m = 0$ to M :

$$O_{i,m} = \frac{1}{T} \sum_{t=1}^T \gamma(z_{i,m}^t),$$

For $m = 1$ to M :

$$P_{i,m} = \frac{1}{T} \sum_{\substack{t=1 \\ k_i^t \neq 0}}^T \gamma(z_{i,m}^t), \quad Q_{i,m} = \frac{1}{T} \sum_{\substack{t=1 \\ k_i^t \neq 0}}^T \gamma(z_{i,m}^t) k_i^t, \quad R_{i,m} = \frac{1}{T} \sum_{\substack{t=1 \\ k_i^t \neq 0}}^T \gamma(z_{i,m}^t) (k_i^t)^2,$$

where

$$\gamma(z_{i,m}^t) = \begin{cases} 1 & \text{if } k_i^t = 0 \text{ and } m = 0 \\ 0 & \text{if } k_i^t = 0 \text{ and } m \neq 0 \\ 0 & \text{if } k_i^t \neq 0 \text{ and } m = 0 \\ \frac{b_{i,m} \mathcal{N}(k_i^t; \mu_{i,m}, \sigma_{i,m}^2)}{\sum_{m'=1}^M b_{i,m'} \mathcal{N}(k_i^t; \mu_{i,m'}, \sigma_{i,m'}^2)} & \text{if } k_i^t \neq 0 \text{ and } m \neq 0. \end{cases} \quad (14)$$

3. Set $\eta_n = 1/n$ and update the running posterior summaries as

$$\begin{aligned} S_n^{O_{i,m}} &= S_{n-1}^{O_{i,m}} + \eta_n (O_m - S_{n-1}^{O_{i,m}}) \\ S_n^{P_{i,m}} &= S_{n-1}^{P_{i,m}} + \eta_n (P_{i,m} - S_{n-1}^{P_{i,m}}) \\ S_n^{Q_{i,m}} &= S_{n-1}^{Q_{i,m}} + \eta_n (Q_{i,m} - S_{n-1}^{Q_{i,m}}) \\ S_n^{R_{i,m}} &= S_{n-1}^{R_{i,m}} + \eta_n (R_{i,m} - S_{n-1}^{R_{i,m}}). \end{aligned} \quad (15)$$

4. Solve for new proposal parameters:

$$\begin{aligned} b_{i,m} &= \frac{S_n^{O_{i,m}}}{\sum_{m'=0}^M S_n^{O_{i,m'}}} \\ \mu_{i,m} &= \frac{S_n^{Q_{i,m}}}{S_n^{P_{i,m}}} \\ \sigma_{i,m}^2 &= \frac{\mu_{i,m}^2 S_n^{P_{i,m}} - 2\mu_{i,m} S_n^{Q_{i,m}} + S_n^{R_{i,m}}}{S_n^{P_{i,m}}}. \end{aligned} \quad (16)$$

4.3. Random-scan independence Metropolis-Hastings algorithm

A straightforward application of the adaptive independence MH algorithm described so far has one important inefficiency. In the parameter sampling step of Algorithm 1, all the parameters are proposed jointly and then passed through an accept-reject step. This approach can lead to a very high rate of rejection that consequently renders the adaptation ineffective. The alternative is to use a componentwise independent MH approach, wherein only a single component (or a small block of components) of the parameter vector \bar{k} is proposed at a time, and the resulting parameter vector is immediately passed through

Algorithm 2 Model inference by the adaptive independence Metropolis-Hastings algorithm

- 1: **Given:** Data \mathcal{D} , prior density $p(\bar{k})$, likelihood function $p(\mathcal{D}|\bar{k})$, proposal distributions $q_s(k_i) = \lambda_i \tilde{q}(k_i; \tilde{\psi}_i) + (1 - \lambda_i)q(k_i; \psi_i)$, number of proposal updates N_{iter} , number of samples T between proposal updates
- 2: Initialize starting point \bar{k}^0 and proposal parameters $\bar{\psi}_0$
- 3: **for** $n = 1$ to N_{iter} **do**
- 4: **for** $t = 1$ to T **do**
- 5: Select the number of parameters $l \ll N$ to be updated.
- 6: Randomly select l parameter indices: $r_1, r_2, \dots, r_l < N$
- 7: Sample $u \sim \mathcal{U}_{[0,1]}$
- 8: **for** $p = 1$ to l **do**
- 9: Sample $k_{r_p}^* \sim q_s(k_{r_p}^*)$
- 10: **end for**
- 11: Set $k_{r_p}^* = k_{r_p}^{t-1}$ for $r_p \setminus \{r_1, r_2, \dots, r_l\}$
- 12: **if** $u < \mathcal{A}(\bar{k}^{t-1}, \bar{k}^*) = \min\{1, \frac{p(\bar{k}^*|\mathcal{D})q_s(\bar{k}^{t-1})}{p(\bar{k}^{t-1}|\mathcal{D})q_s(\bar{k}^*)}\}$ **then**
- 13: $\bar{k}^t = \bar{k}^*$
- 14: **else**
- 15: $\bar{k}^t = \bar{k}^{t-1}$
- 16: **end if**
- 17: **end for**
- 18: Update summary statistics $S_n^{O_{1:N,0:M}}, S_n^{P_{1:N,1:M}}, S_n^{Q_{1:N,1:M}}$, and $S_n^{R_{1:N,1:M}}$
- 19: Update proposal parameters $\bar{\psi}_n$: $b_{1:N,0:M}, \mu_{1:N,1:M}$, and $\sigma_{1:N,1:M}^2$
- 20: Store $\bar{k}^{1:T}$ and reset $\bar{k}^0 \leftarrow \bar{k}^T$
- 21: **end for**

an accept-reject step. We use a random-scan variant of the componentwise MH scheme which randomly selects a block of components to be updated. Updating only one or a few components at a time also provides the practical benefit of a local search, since this amounts to making small jumps in the model space. Small jumps in the model space, especially in problems where the parameters retain their meaning across different models, often result in significantly higher acceptance rates. Algorithm 2 summarizes the overall algorithm we use to generate samples from the posterior distribution. In implementing this algorithm for the following examples, we choose $T = 1000$ and $l = 1$ or 2.

5. Numerical demonstrations: catalytic reforming of methane

We demonstrate the approach formulated in the preceding sections on three example problems. In particular, we infer chemical kinetic models for steam and dry reforming of methane catalyzed by rhodium. The first problem uses synthetic data to demonstrate the consistency of the Bayesian model inference procedure, while the second and third examples use experimental data drawn from the literature. Methane reforming is an important process because it provides an effective route for the industrial production of syngas (CO+H₂). Catalytic reforming of methane has been studied previously, and a few kinetic models have been proposed (Deutschmann et al., 2001; Hartmann et al., 2010; McGuire et al., 2009,

2011). The development of these models has proceeded by collecting possible elementary reactions and making educated guesses about the appropriate pathways, with the selection of rate parameter values based on existing literature or fits to experimental data.

One of the most common experimental configurations for studying catalytic reactions is a stagnation flow reactor. Stagnation flow reactors provide favorable fluid-mechanical properties that enable measurement of the gas-phase boundary layer near the catalytic surface. Hence we use gas-phase measurements from stagnation flow reactors as data for our inference procedure. Recall that experimental data enters the Bayesian inference formulation through the likelihood function $p(\mathcal{D}|\bar{k})$ (Section 3). The likelihood function must also therefore incorporate a detailed numerical model of the stagnation flow reactor in order to compare the data with predictions based on any candidate kinetic model. We begin by discussing this reactor model.

5.1. Stagnation flow reactor model

The boundary layer flow equations in a stagnation flow reactor (schematic in Figure 2) can be modeled as a one-dimensional axisymmetric flow using similarity reduction (Kee et al., 2003). The stagnation-flow reactor boundary layer equations have been used by a number of authors in studies of catalytic surface reactions (Deutschmann et al., 2001; McGuire et al., 2009, 2011). The governing equations are:

$$\frac{d(\rho u)}{dz} + 2\rho V = 0 \quad (17)$$

$$\rho u \frac{dV}{dz} + \rho V^2 = -\Lambda_r + \frac{d}{dz} \left(\mu \frac{dV}{dz} \right) \quad (18)$$

$$\rho u c_p \frac{dT}{dz} = \frac{d}{dz} \left(\lambda \frac{dT}{dz} \right) - \sum_{\alpha=1}^{K_g} \rho Y_\alpha V_\alpha c_{p\alpha} \frac{dT}{dz} - \sum_{\alpha=1}^{K_g} h_\alpha W_\alpha \dot{\omega}_\alpha \quad (19)$$

$$\rho u \frac{dY_\alpha}{dz} = -\frac{d}{dz} (\rho Y_\alpha V_\alpha) + W_\alpha \dot{\omega}_\alpha, \quad (\alpha = 1 \dots K_g) \quad (20)$$

$$\dot{s}_\beta = 0, \quad (\beta = 1 \dots K_s) \quad (21)$$

$$p = \rho R T \sum_{\alpha=1}^{K_g} \frac{Y_\alpha}{W_\alpha} \quad (22)$$

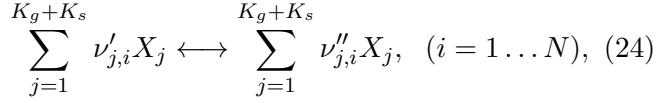
In the above equations, the axial spatial coordinate z is the independent variable, while the axial velocity u , the scaled radial velocity V , the fluid temperature T , and the species mass fractions Y_α are the dependent variables. The pressure-gradient eigenvalue is

$$\Lambda_r = \frac{1}{r} \frac{dp}{dr}. \quad (23)$$

The perfect gas equation (22) relates the pressure p to the temperature T , density ρ , and the species mass fractions at any point. In equations (17)–(21), μ is the fluid dynamic viscosity, λ is the thermal conductivity, c_p is the mixture specific heat, $c_{p\alpha}$ are species specific heats, h_α are species specific enthalpies, and W_α are the molecular weights of the species. $\dot{\omega}_\alpha$ denotes the molar production rate of the gas-phase species indexed by α , and \dot{s}_β the production rate of the surface species, indexed by β . There are K_g gas-phase species

and K_s surface species. A detailed chemical kinetic model is used to compute the species production rates $\dot{\omega}_\alpha$ and \dot{s}_β .

We assume that every candidate detailed chemical kinetic model involving N reactions among these species can be represented in the general form



where $\nu_{j,i}$ are integer stoichiometric coefficients and X_j is the chemical name of the j th species. The molar production rates $\dot{\omega}_\alpha$ and \dot{s}_β are summations over all reactions:

$$\dot{\omega}_\alpha = \sum_{i=1}^N \nu_{\alpha,i} q_i, \quad \dot{s}_\beta = \sum_{i=1}^N \nu_{\beta,i} q_i, \quad (25)$$

where

$$\nu_{\alpha,i} = \nu''_{\alpha,i} - \nu'_{\alpha,i}, \quad (26)$$

and similarly for $\nu_{\beta,i}$. The rate of progress q_i of the i th reaction, which is assumed to obey mass-action kinetics, is the difference between the forward and reverse reaction rates:

$$q_i = k_{i,f} \prod_{j=1}^{K_g+K_s} [X_j]^{\nu'_{j,i}} - k_{i,b} \prod_{j=1}^{K_g+K_s} [X_j]^{\nu''_{j,i}}. \quad (27)$$

The form of the concentrations $[X_j]$ in (27) depends on whether the species is in gas phase or on the surface. Also, it is known from earlier work (McGuire et al., 2011) that species production rates due to purely gas-phase reactions are negligible at normal operating conditions. Thus we omit purely gas-phase reactions when evaluating $\dot{\omega}_\alpha$ in our differential equation model.

The species diffusion velocities are computed using a multicomponent diffusion model as

$$V_\alpha = \frac{1}{X_\alpha \bar{W}} \sum_{j \neq \alpha}^{K_g} W_j D_{\alpha,j} \frac{dX_j}{dz} - \frac{D_\alpha^T}{\rho Y_\alpha T} \frac{1}{dz}. \quad (28)$$

Here X_α and X_j are the species mole fractions, \bar{W} is the mean molecular weight, $D_{\alpha,j}$ are multicomponent diffusion coefficients, and D_α^T are thermal diffusion coefficients. At the reactor inlet, boundary conditions are

$$u = U_{in}, \quad V = 0, \quad T = T_{in}, \quad Y_\alpha = Y_{\alpha,in}, \quad (29)$$

and at the catalytic stagnation surface, the boundary conditions are

$$u = 0, \quad V = 0, \quad T = T_s, \quad \rho Y_\alpha V_\alpha = F_{cg} \dot{\omega}_\alpha W_\alpha. \quad (30)$$

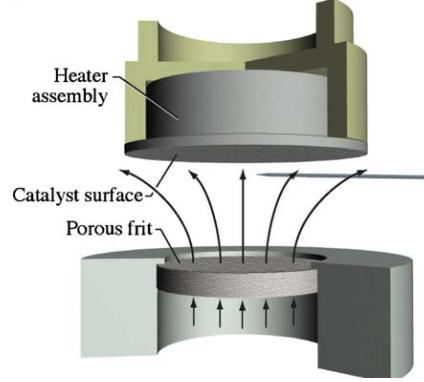


Figure 2: Stagnation flow reactor; figure reproduced from McGuire et al. (2009).

The boundary condition in (30) states that the gas-phase species diffusion flux at the stagnation surface is balanced by species consumption by catalytic reactions. The boundary condition also contains a parameter F_{cg} , which specifies the effective catalyst area. Since the catalyst particles are dispersed in a porous medium, the effective catalyst area $A_{catalyst}$ is much greater than the geometric area $A_{geometric}$ of the stagnation surface. The parameter F_{cg} is defined as

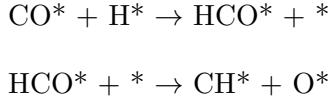
$$F_{cg} = \frac{A_{catalyst}}{A_{geometric}} \quad (31)$$

The steady-state stagnation flow axisymmetric boundary layer equations form a system of ordinary differential equations. These equations are discretized using a finite difference method and the resulting algebraic equations are solved using a combination of pseudo-time marching and Newton's method (Kee et al., 2003). We use the chemically reacting flow software package Cantera 2.0.2 (Goodwin et al., 2013) to compute species production rates and to solve the steady-state stagnation flow axisymmetric boundary layer equations.

5.2. Proposed elementary reactions

Beginning with the work of Hickman and Schmidt (1993), kinetic models for reactions of methane on rhodium have been developed via a combination of theoretical methods, fits to available experimental data, and previous analyses of related species. (Deutschmann et al., 2001; Hartmann et al., 2010; McGuire et al., 2011). Activation energies for surface reactions are often estimated using the semi-empirical unity bond index-quadratic exponential potential (UBI-QEP) method. The determination of pre-exponential factors, however, has largely relied on fits to observed data or the assignment of nominal values. The uncertainty associated with these rate determination techniques and limited understanding of the associated catalytic reaction pathways make this system a good candidate for Bayesian model inference.

The set of proposed elementary reactions we use in our inference demonstrations is taken from a comprehensive model proposed by McGuire et al. (2011) recently and is shown in Table 1. The reaction set consists of 42 irreversible elementary reactions involving 12 surface-adsorbed species and gas-phase species. We retain the rate parameters given by McGuire et al. (2011) as base values for all reactions, except the following two:



The pre-exponential factor of the first reaction above is changed from 5.0×10^{19} to 5.0×10^{18} , while that of the second reaction is changed from 3.7×10^{24} to 3.7×10^{23} . These changes yield minor improvements in agreement with data at the nominal parameter values. The pre-exponential factors were previously assigned nominal values. The activation energies were estimated by UBI-QEP method, which has an expected error of 1–3 kcal/mol (Hartmann et al., 2010).

The surface reactions shown in Table 1 are of two different types: adsorption/desorption of gas-phase species and reactions among surface intermediates. In this work, we do not consider the adsorption/desorption reactions (Reactions 31–42) to be uncertain; rather,

they are included in all models inferred. In the table, the adsorption/desorption reactions are shaded pink, while the surface reactions we consider to be uncertain are shaded green. In the three examples to be presented below, we treat the thermodynamic properties of the surface species as fixed (i.e., not uncertain). Although the thermodynamic properties are not precisely known, they are fixed indirectly through the individual forward and reverse rate constants. The base values of the forward and reverse rate constants, k_f and k_b , were originally established to satisfy approximate thermodynamic reversibility. Therefore, with the thermodynamic properties fixed, we need only to specify the prior distribution and apply the model inference framework on the forward reactions. The reverse rate constant of each reaction is then

$$k_b = \frac{k_f}{K_{eq}} = \frac{k_b^*}{k_f^*} k_f, \quad (32)$$

where K_{eq} , the equilibrium constant of the reaction, is a function of the thermodynamic properties of the species participating in the reaction. In the above equations, k_f and k_b are the perturbed rate constants of the reactions while k_f^* and k_b^* are the base rate constants.

5.3. Setup of the Bayesian model inference problem

Before discussing the three model inference examples individually, we describe the choices we make for the likelihood function and prior distribution in our Bayesian formulation. In the following, we use \tilde{k} to refer to the rate constants of the reactions that are treated as uncertain and \hat{k} to denote the rate constants of reactions that are kept fixed. By “fixed,” we mean that a particular reaction is always included in the model and that its rate constant is not a target of the inference procedure.

5.3.1. Likelihood function

As described in Section 3, evaluating the posterior probability in the Bayesian approach requires evaluating the likelihood function $p(\mathcal{D}|\bar{k})$, where \mathcal{D} are the data and $\bar{k} = (\tilde{k}, \hat{k})$ are the reaction parameters. We employ an i.i.d. additive Gaussian model for the difference between model predictions and observations; thus the data are represented as

$$\mathcal{D} = G(\tilde{k}, \hat{k}) + \epsilon_n, \quad (33)$$

where $\epsilon_n \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 I_n)$, n is the number of observations, I_n is an n -by- n identity matrix, and $G(\tilde{k}, \hat{k})$ is the prediction of the forward model at the given value of the reaction parameters. We let the noise standard deviation σ be 0.005. The deterministic predictions $G(\tilde{k}, \hat{k})$ are obtained with the stagnation flow reactor model explained in Section 5.1. The likelihood function is thus given by

$$\begin{aligned} p(\mathcal{D}|\bar{k}) &= \mathcal{N}_n(\mathcal{D}|G(\tilde{k}, \hat{k}), \sigma^2 I_n) \\ &= \prod_{t=1}^n \mathcal{N}(\mathcal{D}^t|G(\tilde{k}, \hat{k}), \sigma^2) \\ &= \prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d^t - G(\tilde{k}, \hat{k}))^2}{2\sigma^2}\right), \end{aligned} \quad (34)$$

where d^t are components of the data vector \mathcal{D} .

	Reaction	A	E_a	Uncertainty applied [†]
1	$H^* + O^* \rightarrow OH^* + *$	5.0×10^{22}	83.7	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
2	$H^* + OH^* \rightarrow H_2O^* + *$	3.0×10^{20}	33.5	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
3	$OH^* + OH^* \rightarrow H_2O^* + O^*$	3.0×10^{21}	100.8	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
4	$CO^* + O^* \rightarrow CO_2^* + *$	5.5×10^{18}	121.6	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
5	$CH_4^* + * \rightarrow CH_3^* + H^*$	3.7×10^{21}	61.0	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
6	$CH_3^* + * \rightarrow CH_2^* + H^*$	3.7×10^{24}	103.0	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
7	$CH_2^* + * \rightarrow CH^* + H^*$	3.7×10^{24}	100.0	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
8	$CH_4^* + O^* \rightarrow CH_3^* + OH^*$	1.7×10^{24}	80.34	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
9	$CH_3^* + O^* \rightarrow CH_2^* + OH^*$	3.7×10^{24}	120.31	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
10	$CH_2^* + O^* \rightarrow CH^* + OH^*$	3.7×10^{24}	114.5	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
11	$CH^* + * \rightarrow C^* + H^*$	3.7×10^{21}	21.0	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
12	$CH^* + O^* \rightarrow C^* + OH^*$	3.7×10^{21}	30.13	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
13	$C^* + O^* \rightarrow CO^* + *$	5.2×10^{23}	97.9	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
14	$CO^* + H^* \rightarrow HCO^* + *$	5.0×10^{18}	108.9	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
15	$HCO^* + * \rightarrow CH^* + O^*$	3.7×10^{23}	59.5	$\log_{10} k = \log_{10} k^* + \mathcal{U}[-2, 2]$
16	$OH^* + * \rightarrow H^* + O^*$	3.0×10^{20}	37.7	$k = k_{16}^* k_1/k_1^*$
17	$H_2O^* + * \rightarrow H^* + OH^*$	5.0×10^{22}	106.4	$k = k_{17}^* k_2/k_2^*$
18	$H_2O^* + O^* \rightarrow OH^* + OH^*$	3.0×10^{21}	171.8	$k = k_{18}^* k_3/k_3^*$
19	$CO_2^* + * \rightarrow CO^* + O^*$	3.0×10^{21}	115.3	$k = k_{19}^* k_4/k_4^*$
20	$CH_3^* + H^* \rightarrow CH_4^* + *$	3.7×10^{21}	51.0	$k = k_{20}^* k_5/k_5^*$
21	$CH_2^* + H^* \rightarrow CH_3^* + *$	3.7×10^{23}	44.1	$k = k_{21}^* k_6/k_6^*$
22	$CH^* + H^* \rightarrow CH_2^* + *$	3.7×10^{21}	68.0	$k = k_{22}^* k_7/k_7^*$
23	$CH_3^* + OH^* \rightarrow CH_4^* + O^*$	3.7×10^{21}	24.27	$k = k_{23}^* k_8/k_8^*$
24	$CH_2^* + OH^* \rightarrow CH_3^* + O^*$	3.7×10^{21}	15.06	$k = k_{24}^* k_9/k_9^*$
25	$CH^* + OH^* \rightarrow CH_2^* + O^*$	3.7×10^{21}	36.82	$k = k_{25}^* k_{10}/k_{10}^*$
26	$C^* + H^* \rightarrow CH^* + *$	3.7×10^{21}	172.8	$k = k_{26}^* k_{11}/k_{11}^*$
27	$C^* + OH^* \rightarrow CH^* + O^*$	3.7×10^{21}	136.0	$k = k_{27}^* k_{12}/k_{12}^*$
28	$CO^* + * \rightarrow C^* + O^*$	2.5×10^{21}	169.0	$k = k_{28}^* k_{13}/k_{13}^*$
29	$HCO^* + * \rightarrow CO^* + H^*$	3.7×10^{21}	0.0	$k = k_{29}^* k_{14}/k_{14}^*$
30	θ_{CO}^* $CH^* + O^* \rightarrow HCO^* + *$	3.7×10^{21}	167.5	$k = k_{30}^* k_{15}/k_{15}^*$
31	$H_2 + * + * \rightarrow H^* + H^*$	1.0×10^{-2a}	0.0	-
32	$O_2 + * + * \rightarrow O^* + O^*$	1.0×10^{-2a}	0.0	-
33	$CH_4 + * \rightarrow CH_4^*$	8.0×10^{-3a}	0.0	-
34	$H_2O + * \rightarrow H_2O^*$	1.0×10^{-1a}	0.0	-
35	$CO_2 + * \rightarrow CO_2^*$	4.8×10^{-2a}	0.0	-
36	$CO + * \rightarrow CO^*$	5.0×10^{-1a}	0.0	-
37	$H^* + H^* \rightarrow H_2 + * + *$	3.0×10^{21}	77.8	-
38	$O^* + O^* \rightarrow O_2 + * + *$	1.3×10^{22}	355.2	-
39	θ_O^* $CH_4^* \rightarrow CH_4 + *$	1.9×10^{14}	25.1	-
40	$H_2O^* \rightarrow H_2O + *$	3.0×10^{13}	45.0	-
41	$CO_2^* \rightarrow CO_2 + *$	4.1×10^{11}	18.0	-
42	$CO^* \rightarrow CO + *$	3.5×10^{13}	133.4	-
	θ_{CO}^*		-280.0 ^b	-
			-15.0 ^b	-

[†]Arrhenius rate expression for k^* (base value): $k^* = A \exp(-E_a/RT)$

^aSticking coefficient

^bCoverage-dependent activation energy

^bForward-backward reaction pair I consists of reactions I and I+15

Table 1: Proposed reactions for reforming of methane

5.3.2. Prior specification

The prior distribution in Bayesian analysis should encapsulate information about models and parameters that is available before assimilation of the data presently at hand. Often the priors come from known scientific principles and physical constraints on the parameters. In the context of chemical kinetics, the continuous component of the prior distribution may also be derived from previous investigations of the reactions. Furthermore, as described in Section 3, priors may also be shaped by expert elicitation (Garthwaite et al., 2005) or chosen to reflect relative ignorance about the rate parameter values. In our demonstrations we will choose relatively uninformative priors by allowing the rate constants \bar{k} to vary by two orders of magnitude above and below their base values. Other prior choices, e.g., an exponential distribution or a uniform distribution between zero and some positive upper bound, would also be reasonable. In the same way, prior information about model structure—applied in the form of prior weights on reaction inclusion or exclusion—can also be designed to reflect an investigator’s belief about the role or importance of particular reactions in the chemical process.

To illustrate the impact of the prior, we consider three different prior specifications in our numerical demonstrations:

- Prior 1: $p(k_{i,f}) = 0.2\delta(k_{i,f}) + 0.8\mathcal{C}(k_{i,f})$,
- Prior 2: $p(k_{i,f}) = 0.5\delta(k_{i,f}) + 0.5\mathcal{C}(k_{i,f})$,
- Prior 3: $p(k_{i,f}) = 0.8\delta(k_{i,f}) + 0.2\mathcal{C}(k_{i,f})$,

The prior distributions above are imposed identically on *each* reaction. Since reaction rate constants must be positive, while their uncertainties may multiple orders of magnitude, we take the continuous component of each prior distribution to be a bounded uniform distribution on the logarithm of the rate constant. Specifically, we set each $\mathcal{C}(k_{i,f})$ to

$$\mathcal{C}(k_{i,f}) : \log_{10} k_{i,f} \sim \mathcal{U}(\log_{10} k_{i,f}^* - 2, \log_{10} k_{i,f}^* + 2), \quad (35)$$

where each $k_{i,f}^*$ above is the base value of the i th forward rate constant. For simplicity, the priors used here are all of same family and width, but in general, one could certainly endow each of the 15 rate constants with distinct priors. One could even encode prior correlations among the rate constants.

The three prior specifications above reflect different prior beliefs in the size and sparsity of the reaction mechanism. Prior 1, with a weight of 0.8 on the continuous component, has a tendency to favor kinetic models with more reactions. Prior 2 is the *indifference* prior with no preference for inclusion or exclusion of reactions; it is equivalent to a uniform prior distribution on the space of all 2^N possible models, and thus allows the data to completely determine the most probable set of reactions. Prior 3 favors smaller models, and is an example of a *sparsity-promoting* prior. Such priors introduce additional parsimony in model structure, over and above the penalty on model complexity automatically imposed by the Bayesian Occam’s razor. By using priors that favor reaction exclusion, the posterior distribution over models is biased towards simpler reaction network structures; this has the potential of improving prediction accuracy over unobserved data (Hastie et al., 2009).

5.4. Example 1: Steam reforming of methane with synthetic data

In this first example, we infer kinetic models for steam reforming of methane from data generated using a known model. The goal of this example is to demonstrate the consistency of the Bayesian model inference process and to examine the impact of varying amounts of data. We create four synthetic (nested) data sets increasing in size from 10, 20, 40, to 60 points. The data are mole fractions of gas-phase species (H_2 , H_2O , CH_4 , CO , and CO_2) measured at different locations inside the stagnation flow reactor, at up to three different catalyst surface temperatures T_s . Data set 1 consists of mole fractions 0.1 mm and 0.7 mm from the catalyst surface, while data sets 2, 3, and 4 contain measurements performed 0.1 mm, 0.7 mm, 1.3 mm, and 2 mm from the catalyst surface. Further details on each data set are given in Table 2. We generate the data using a kinetic model that contains all the reactions shown in Table 1, except reaction pairs (4)–(19) and (6)–(21). Samples of Gaussian noise with mean zero and standard deviation $\sigma = 0.005$ are added to these model predictions to simulate noisy experimental observations. For the purpose of this example, we allow only four reaction pairs to have uncertain parameters and to be candidates for inclusion/exclusion. The other reactions are kept fixed at their base values for the likelihood calculation. The uncertain reaction pairs are shown in Table 3.

Data set	Number of data points	Catalyst temperatures
1	10	740°C
2	20	740°C
3	40	740°C, 790°C
4	60	740°C, 790°C, 840°C

Table 2: Synthetic data sets for Example 1.

Reaction pair [†]	Reaction
1 (1)–(16)	$\text{H}^* + \text{O}^* \leftrightarrow \text{OH}^* + *$
2 (4)–(19)	$\text{CO}^* + \text{O}^* \leftrightarrow \text{CO}_2^* + *$
3 (5)–(20)	$\text{CH}_4^* + * \leftrightarrow \text{CH}_3^* + \text{H}^*$
4 (6)–(21)	$\text{CH}_3^* + * \leftrightarrow \text{CH}_2^* + \text{H}^*$

[†]Reaction pair numbering in the leftmost column is specific to Example 1.

Table 3: Proposed reactions for inference in Example 1.

Because we consider only four reaction pairs to have inclusion/exclusion uncertainty, the number of possible models in the present example is $2^4 = 16$. We employ the indifference prior (Prior 2) described in the previous section. 200 000 samples are then simulated from the posterior distribution of \tilde{k} , using adaptive MCMC, for each of the four data sets. We begin adaptation after generating the first 20 000 samples and discard the next 20 000 samples as burn-in, while the proposal parameters adapt. The most probable models and their probabilities are computed using the remaining 160 000 samples; these are shown in Table 4.

Data set 1		Data set 2	
Reaction pairs included	Probability	Reaction pairs included	Probability
1, 3	0.281	1, 3	0.375
1, 4	0.256	1, 3, 4	0.197
1, 3, 4	0.165	1, 4	0.195
1	0.146	1	0.082
1, 2, 3	0.056	1, 2, 3	0.073
Data set 3		Data set 4	
Reaction pairs included	Probability	Reaction pairs included	Probability
1, 3	0.482	1, 3	0.525
1, 2, 3	0.316	1, 2, 3	0.253
1, 3, 4	0.122	1, 3, 4	0.152
1, 2, 3, 4	0.072	1, 2, 3, 4	0.070
1, 4	0.006	1, 4	0.001

Table 4: The five most probable models and their probabilities, from Example 1.

We see from the inference results that the data-generating model (i.e., the “true” model) is selected with highest posterior probability for every data set. Although it is possible in principle for the true model not to be assigned the highest posterior probability for finite data (Bishop, 2007), we notice here that the true model is always preferred and moreover that its probability increases with more data. This trend also demonstrates the diminishing impact of the prior distribution as more data are included. Indeed, Bayesian model inference is known to be asymptotically consistent, i.e., the posterior concentrates on the true model given infinite data (Berger and Pericchi, 2001), provided that true model is within the set of models being considered.

5.5. Example 2: Steam reforming of methane with real data

The second example considers inference of chemical kinetic models for steam reforming of methane using real experimental data from a stagnation flow reactor apparatus (McGuire et al., 2009). The operating conditions of the experiment are given in Table 5; further specifics on the experimental data set (e.g., species and measurement locations) can be found in McGuire et al. (2009).

In this example, we consider all three prior specifications (Section 5.3.2) imposed on all 15 of the uncertain reaction pairs in Table 1. Using the adaptive MCMC procedure of Section 3, we generate 200 000 samples from the posterior distribution of \tilde{k} using the adaptive MCMC procedure explained in Section 3. Again, we begin adaptation after generating the first 20 000 samples and discard the next 20 000 samples as burn-in. Posterior model probabilities are estimated from the remaining 160 000 samples.

Table 6 shows the ten most probable models for each prior specification, and their corresponding frequency (in a total population of 160 000 samples). As expected, the sparsity of the most probable models increases with the weight on the δ -component in the prior. In the case of Prior 1, the model that includes all the reactions is strongly preferred. For Prior 2, the most probable model includes all the reactions except pairs 6–21, 12–27, and 14–29. The exclusion of these three reaction pairs, particularly in the case of an indifference prior, is an example of the Bayesian Occam’s razor in action. Prior 3 results in extremely sparse

models. Reaction networks corresponding to the highest probability models for the three prior settings are shown in Figure 3.

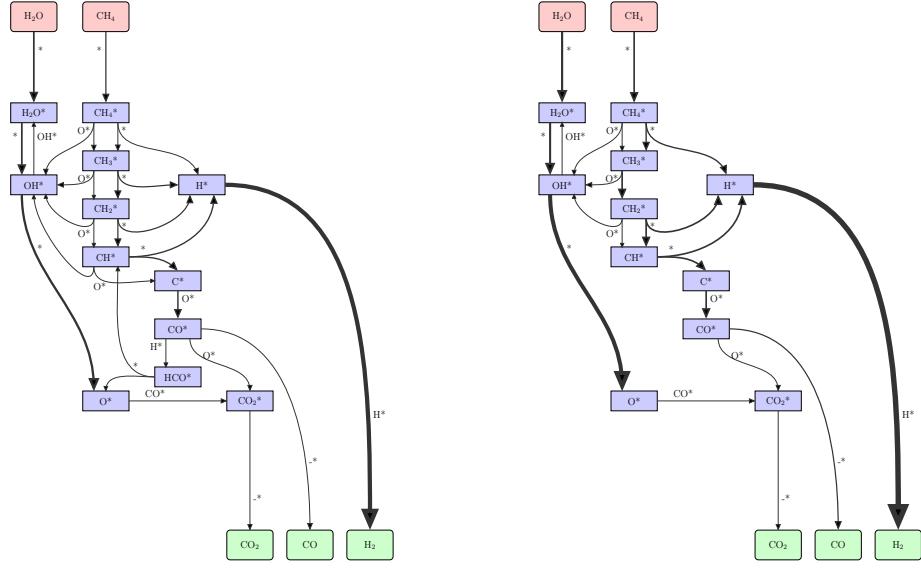
Condition	Value
Inlet composition (by mole fractions)	4.3% CH ₄ and 5.9% H ₂ O (balance Ar)
Inlet temperature	135°C
Catalyst surface temperature	740°C
Inlet velocity	1.3 m/s
Reactor pressure	300 Torr
F_{cg}	20

Table 5: Experimental operating conditions for Example 2, from McGuire et al. (2009).

We also show in Figure 4 the posterior marginal inclusion probability of each reaction pair. Since the marginal inclusion probability of a reaction is the average of its inclusion indicator over all possible models in the posterior, it provides a measure of how strongly an individual reaction is supported by the available data. In all three panels of Figure 4, we note that the posterior inclusion probability of reaction pair 3–18 is identical to its prior inclusion probability. Reaction pairs 14–29 and 15–30 (reactions involving species HCO*) also have a negligible difference between their prior and posterior inclusion probabilities. These results suggest that the data are not informative about these reactions; in other words, these reactions seem to have an insignificant effect on the level of agreement between model predictions and the available data. Invoking a further principle of parsimony, it may thus be prudent to exclude reaction pairs 3–18, 14–29, and 15–30 from the predictive model, or to reassess their importance by collecting more data.

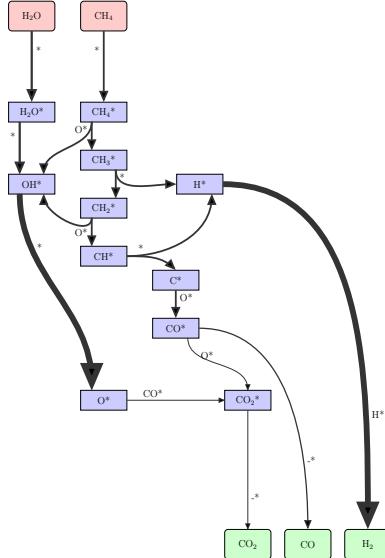
Figure 4 also shows that the posterior marginal inclusion probabilities of reaction pairs 1–16, 2–17, 4–19, and 13–28 remain close to one for each prior specification; these reactions are thus the most strongly supported by available data. The inclusion of reaction pair 4–19 with probability one in all the inferred models is confirmation that the inference procedure is working well, in that it does not exclude reactions that are absolutely necessary for production of the observed product species. Among all the uncertain reaction pairs, reactions 4 and 19 are the only pair containing CO₂^{*}, and their inclusion ensures that the model produces CO₂ as one of the products. In general, the near-unity posterior probabilities suggest that all four of these reaction pairs are critical to explaining the steam reforming behavior of methane.

It is important to note, however, that the marginal inclusion probabilities shown in Figure 4 do not capture correlations among the values of the reaction rate parameters or patterns of joint inclusion/exclusion. The joint posterior probability distribution of the rate parameters, which we have sampled using MCMC, in fact contains much more information. In particular, it contains information about combinations of reactions and how well particular combinations are supported by the current data.



$$(a) p(k_{i,f}) = 0.2\delta(k_{i,f}) + 0.8\mathcal{C}(k_{i,f})$$

$$(b) p(k_{i,f}) = 0.5\delta(k_{i,f}) + 0.5\mathcal{C}(k_{i,f})$$



$$(c) p(k_{i,f}) = 0.8\delta(k_{i,f}) + 0.2\mathcal{C}(k_{i,f})$$

Figure 3: Reaction networks of the highest posterior probability models for steam reforming of methane (Example 2), under different prior specifications. Edge thicknesses are proportional to reaction rates calculated using posterior mean values of the rate parameters.

Reaction prior probability $0.2\delta(k_{i,f}) + 0.8\mathcal{C}(k_{i,f})$		Reaction prior probability $0.5\delta(k_{i,f}) + 0.5\mathcal{C}(k_{i,f})$		Reaction prior probability $0.8\delta(k_{i,f}) + 0.2\mathcal{C}(k_{i,f})$	
Freq	Excluded pairs [†]	Freq	Excluded pairs [†]	Freq	Excluded pairs [†]
16157	—	521	6, 12, 14	4221	3, 5, 7, 9, 12, 14, 15
4894	7	496	6, 8, 10, 12, 14	3533	3, 5, 9, 10, 12, 14, 15
4845	6	493	7, 11	3447	3, 5, 6, 7, 11, 14, 15
4545	12	477	7, 9, 11, 14	3379	3, 5, 6, 10, 11, 14, 15
4326	9	464	3, 6, 12, 14	3340	3, 5, 7, 9, 11, 14, 15
4237	10	457	3, 6, 10, 11, 14	3318	3, 5, 6, 7, 12, 14, 15
4168	11	454	6, 10, 12, 14	3153	3, 6, 8, 10, 12, 14, 15
3954	3	451	6, 12	3033	3, 7, 8, 9, 12, 14, 15
3908	15	451	7, 11, 15	2878	3, 5, 9, 10, 11, 14, 15
3902	14	441	6, 10, 11, 15	2640	3, 7, 8, 9, 11, 14, 15

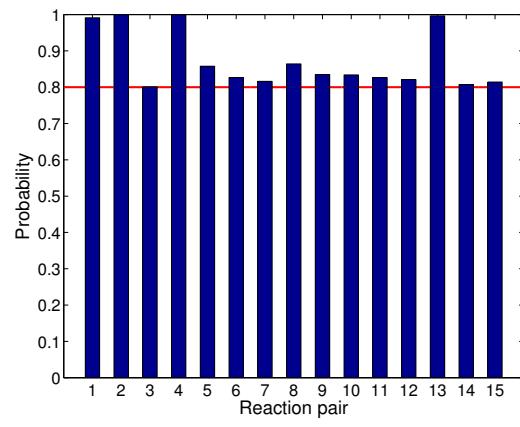
[†]Reaction pairs are denoted here by the number associated with the forward reaction

Table 6: The ten models with highest posterior probability in Example 2, for each choice of prior.

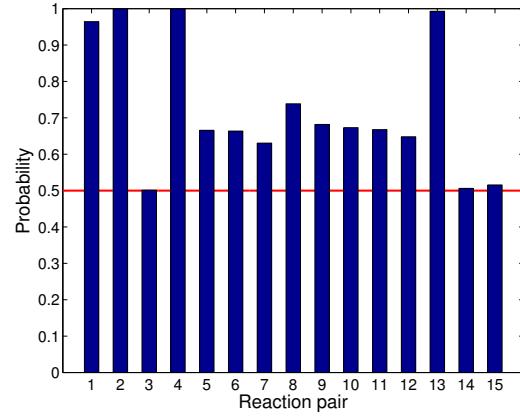
One way of interrogating the joint information embedded in the posterior distribution is to focus attention on particular *pathways* in the reaction network. Looking at the reaction network in Figure 3a (which contains *all* the proposed reactions), it is possible to discern three clear pathways for the conversion of reactants H₂O and CH₄ into products CO₂, CO, and H₂. The first pathway includes both C* and HCO* species, the second pathway excludes HCO* and retains C*, and the third pathway excludes C* but retains HCO*. The three possible reaction pathways are shown schematically in Figure 5. We use samples from the joint posterior distribution to quantify the degree to which each of these pathways is supported by available data. The posterior probability of each pathway is obtained by computing the fraction of posterior samples (i.e., candidate models) that contain the pathway. It is important to note that the probabilities obtained in this fashion correctly account for uncertainties in the other reactions (i.e., reactions not part of the pathway under consideration) by marginalizing over them. This contrasts with a method that simply compares three models, one corresponding to each pathway, while arbitrarily fixing or excluding all the other reactions. Given the data produced by the steam reforming experiments of McGuire et al. (2009), the estimated posterior probabilities of the three pathways are shown in Table 7. We observe that the dominant pathway is the C* pathway. The HCO* pathway has nearly zero probability. This conclusion supports the view commonly held in the literature that steam reforming of methane operates through the C* pathway and that the inclusion of HCO* in the kinetic model is superfluous (McGuire et al., 2009, 2011).

Pathway	Prior 1		Prior 2		Prior 3	
	Prior	Posterior	Prior	Posterior	Prior	Posterior
1 (both C* and HCO* present)	0.536	0.643	0.177	0.249	0.026	0.039
2 (only C* present)	0.302	0.343	0.529	0.733	0.633	0.961
3 (only HCO* present)	0.162	0.010	0.294	0.012	0.340	0.000

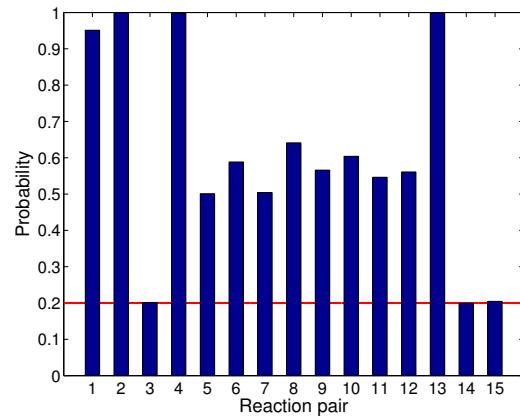
Table 7: Prior and posterior pathway probabilities for steam reforming of methane, Example 2.



(a) Prior: $0.2\delta(k_{i,f}) + 0.8\mathcal{C}(k_{i,f})$



(b) Prior: $0.5\delta(k_{i,f}) + 0.5\mathcal{C}(k_{i,f})$



(c) Prior: $0.8\delta(k_{i,f}) + 0.2\mathcal{C}(k_{i,f})$

Figure 4: Posterior reaction inclusion probabilities of all reactions for the three prior specifications in Example 2. The red line indicates the prior reaction inclusion probability.

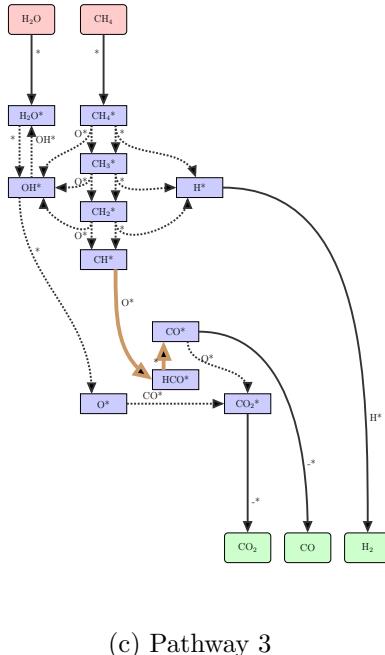
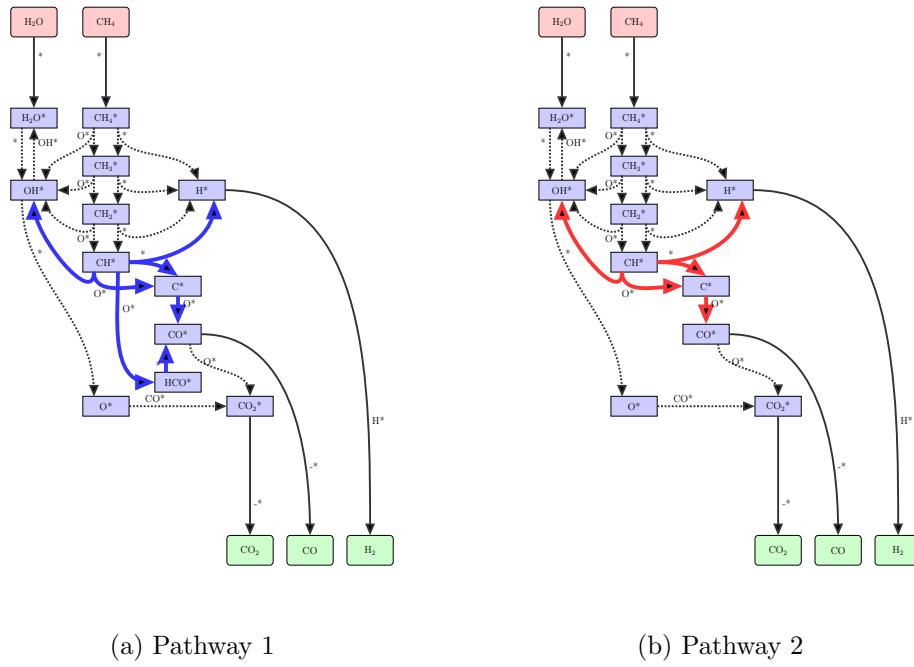


Figure 5: Reaction pathways for *steam reforming* of methane on rhodium (Example 2). Pathway 1 involves both species C^* and HCO^* , Pathway 2 excludes HCO^* , and Pathway 3 excludes C^* species. All other reactions that are treated as uncertain and that do not involve C^* and HCO^* are dotted. Reactions involving gas-phase species are shown as regular lines.

5.6. Example 3: Dry reforming of methane with real data

In the third application of our inference framework, we infer chemical kinetic models for *dry reforming* of methane using experimental data from a stagnation flow reactor reported in [McGuire et al. \(2011\)](#). Operating conditions for the experiment are given in Table 8, and further specifics on the experimental data set (e.g., measured species and their locations) can be found in [McGuire et al. \(2011\)](#). All three prior specifications discussed in Section 5.3.2 are again considered. As in the previous example, 200 000 posterior samples are simulated from a distribution encompassing all 15 uncertain reaction pairs given in Table 1, for each prior specification.

Table 9 shows the ten most probable models for each prior specification and their corresponding frequencies in 160 000 posterior samples. The highest posterior probability model obtained with Prior 1 includes all the reactions; as in the previous example, the weight specification of Prior 1 naturally favors larger models. With Prior 2, i.e., the indifference prior, the posterior excludes many reactions, slightly more than in Example 2. This reduction is again a demonstration of the penalty on model complexity built into evaluations of the marginal likelihood. The sparsity-promoting prior (Prior 3) pushes the posterior towards even smaller models, as seen the third column of Table 9. Reaction networks corresponding to the highest-frequency models for the three prior settings are illustrated in Figure 6.

Condition	Value
Inlet composition (by mole fractions)	10% CH ₄ and 15% CO ₂ (balance Ar)
Inlet temperature	25°C
Catalyst surface temperature	800°C
Inlet velocity	0.9 ms ⁻¹
Reactor pressure	300 Torr
F_{cg}	56

Table 8: Experimental operating conditions for Example 3 ([McGuire et al., 2011](#)).

Marginal posterior inclusion probabilities of all reaction pairs for the three prior specifications are shown in Figure 7. We see that the posterior inclusion probabilities of all the reactions deviate from their prior inclusion probabilities, in contrast to Example 2. This suggests that the experimental data used for this dry reforming example is influenced by—and thus contains information about—every single reaction pair. As in steam reforming, the inclusion of reaction pair 4–19 with probability one confirms that the inference procedure is working well.

Reaction prior probability $0.2\delta(k_{i,f}) + 0.8\mathcal{C}(k_{i,f})$		Reaction prior probability $0.5\delta(k_{i,f}) + 0.5\mathcal{C}(k_{i,f})$		Reaction prior probability $0.8\delta(k_{i,f}) + 0.2\mathcal{C}(k_{i,f})$	
Freq	Excluded pairs [†]	Freq	Excluded pairs [†]	Freq	Excluded pairs [†]
7616	—	537	1, 2, 5, 9, 11, 14	4635	1, 3, 5, 10, 11, 12, 13
4950	6	488	1, 10, 11, 12	3507	1, 2, 9, 10, 11, 14, 15
4800	5	484	1, 5, 10, 11, 14	3314	3, 5, 6, 10, 11, 12, 13
2892	7	445	1, 2, 5, 11, 14	3245	1, 2, 5, 9, 10, 11, 12, 13
2559	2, 6	401	1, 2, 5, 11	2608	1, 2, 9, 10, 11, 12, 13
2393	1	396	5, 6, 7, 14	2428	1, 2, 9, 10, 12, 14, 15
2107	12	392	5, 6, 7, 11, 14	2370	3, 5, 9, 10, 11, 12, 13
2023	15	391	5, 6, 10, 12, 15	1607	1, 3, 5, 7, 12, 14, 15
1889	9	369	5, 6, 7, 11	1511	3, 5, 6, 7, 12, 14, 15
1884	14	365	5, 6, 7, 14, 15	1448	1, 3, 7, 8, 12, 14, 15

[†]Reaction pairs are denoted here by the number associated with the forward reaction

Table 9: The ten models with highest posterior probability in Example 3, for each choice of prior.

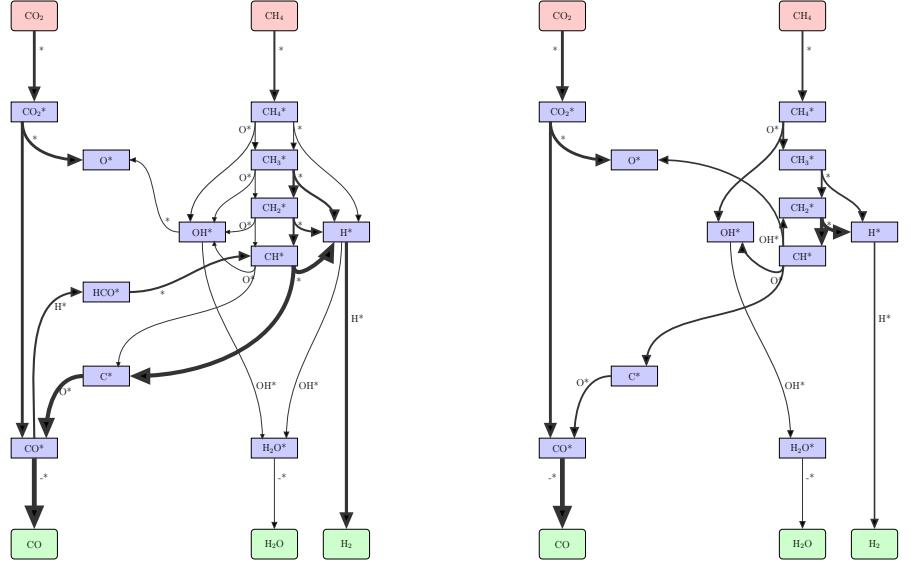
We also compute the posterior probabilities of the three distinct pathways shown in Figure 8. Pathway 1 includes both C* and HCO*, pathway 2 excludes HCO* and retains C*, and pathway 3 excludes C* and retains HCO*. The posterior probabilities of the three pathways are shown in Table 10. Compared to the corresponding results for steam reforming (Table 7), the present results suggest that the HCO* pathway is not unimportant to dry reforming. In other words, it is possible that dry reforming of methane is realized through the generation of HCO*. With an indifference prior, the HCO*-only pathway has a posterior probability of 8%. With a sparsity promoting prior, the posterior probability of pathway 1 decreases dramatically and the posterior places 23% of its mass on the HCO*-only route. That said, the C* pathway remains very much the dominant pathway given the current data. Pathway 1 also has strong support except in the case of the sparsity-promoting prior, which effectively forces the posterior to “choose” between the two more parsimonious options. A clearer conclusion can only result from collecting more data and repeating this analysis.

Pathway	Prior 1		Prior 2		Prior 3	
	Prior	Posterior	Prior	Posterior	Prior	Posterior
1 (both C* and HCO* present)	0.536	0.596	0.177	0.244	0.026	0.034
2 (only C* present)	0.302	0.347	0.529	0.677	0.633	0.732
3 (only HCO* present)	0.162	0.056	0.294	0.080	0.340	0.234

Table 10: Posterior pathway probabilities for dry reforming of methane, Example 3.

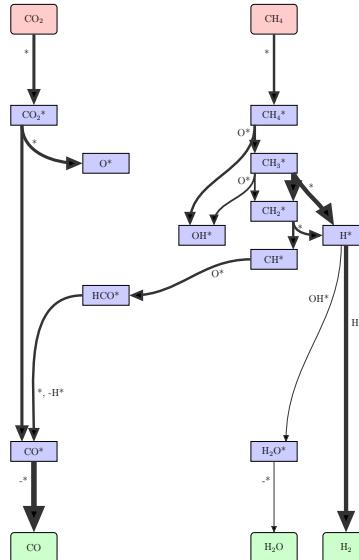
5.7. Efficiency of posterior sampling

To verify the numerical results reported in the preceding sections, we performed three independent MCMC runs for each example problem and each prior specification, with different initial guesses for the rate parameters \hat{k} in each case. Overall, the three replicate runs yielded very similar results; the independent chains were able to identify the high posterior



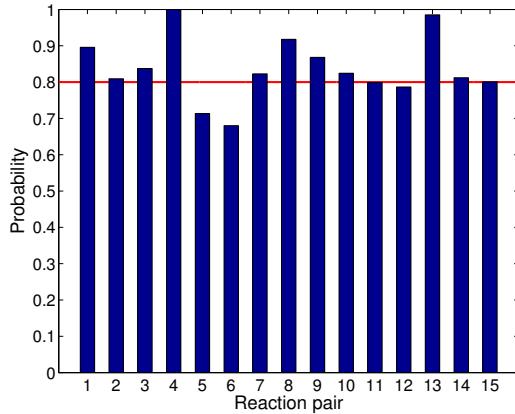
$$(a) p(k_{i,f}) = 0.2\delta(k_{i,f}) + 0.8\mathcal{C}(k_{i,f})$$

$$(b) p(k_{i,f}) = 0.5\delta(k_{i,f}) + 0.5\mathcal{C}(k_{i,f})$$

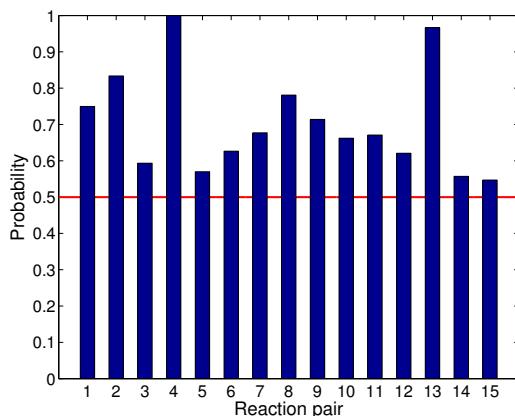


$$(c) p(k_{i,f}) = 0.8\delta(k_{i,f}) + 0.2\mathcal{C}(k_{i,f})$$

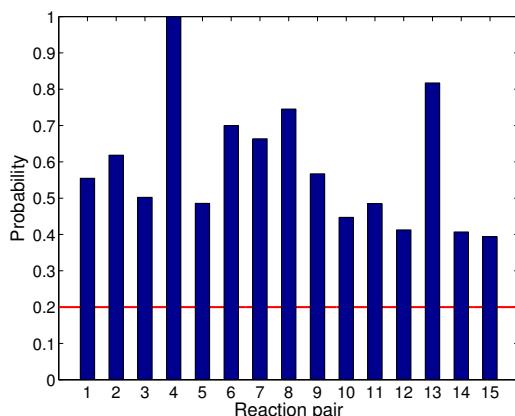
Figure 6: Reaction networks of the highest posterior probability models for dry reforming of methane (Example 3), under different prior specifications. Edge thicknesses are proportional to reaction rates calculated using posterior mean values of the rate parameters.



(a) Prior: $0.2\delta(k_{i,f}) + 0.8\mathcal{C}(k_{i,f})$

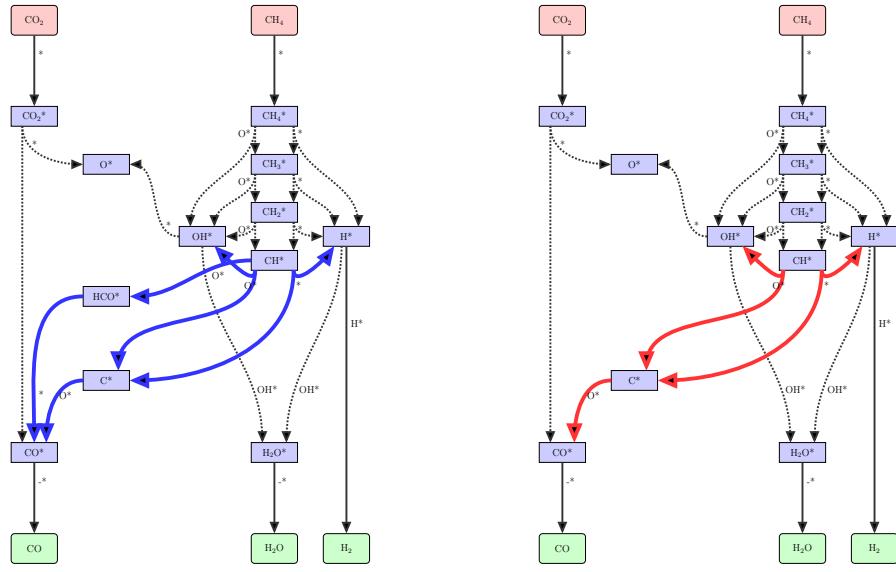


(b) Prior: $0.5\delta(k_{i,f}) + 0.5\mathcal{C}(k_{i,f})$



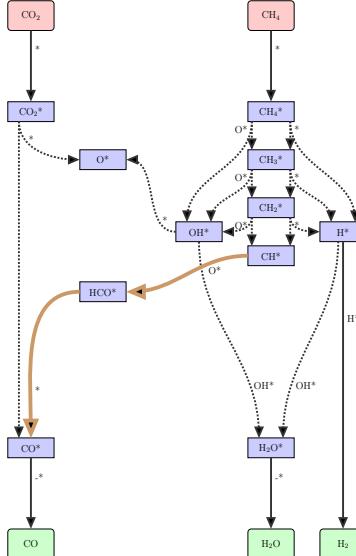
(c) Prior: $0.8\delta(k_{i,f}) + 0.2\mathcal{C}(k_{i,f})$

Figure 7: Posterior reaction inclusion probabilities of all reactions for the three prior specifications in Example 3. The red line indicates the prior reaction inclusion probability.



(a) Pathway 1

(b) Pathway 2



(c) Pathway 3

Figure 8: Reaction pathways for *dry reforming* of methane on rhodium (Example 3). Pathway 1 involves both species C^* and HCO^* , Pathway 2 excludes HCO^* , and Pathway 3 excludes C^* species. All other reactions that are treated as uncertain and that do not involve C^* and HCO^* are dotted. Reactions involving gas-phase species are shown as regular lines.

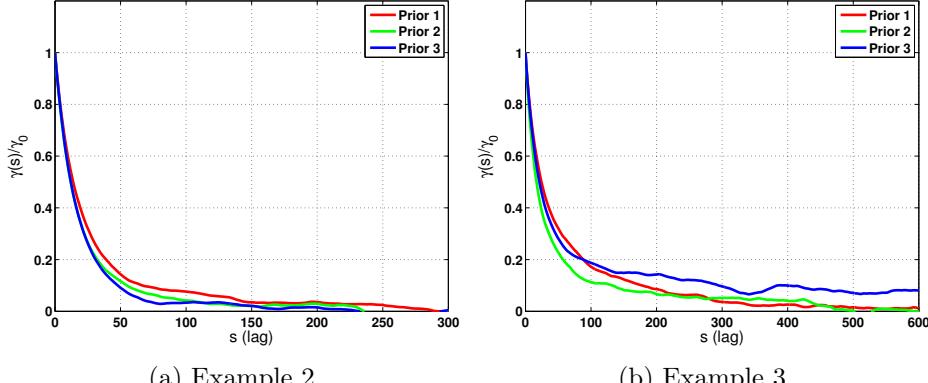


Figure 9: Autocorrelation at lag s of the log-posterior of the MCMC chains.

probability models and accurately reproduce their probabilities. Yet the quality of these posterior estimates, of course, depends on the number of posterior samples employed—i.e., the length of the MCMC chains. Because the forward models $G(\tilde{k}, \hat{k})$ in this setting are computationally expensive, it is practically important to limit the number of samples. As described in Section 4.1, the variance of a posterior estimate for a fixed number of samples depends on how well the chain is mixing (Andrieu et al., 2003; Robert and Casella, 2004). The adaptive MCMC scheme employed here has been shown to significantly improve mixing over non-adaptive schemes (Ji and Schmidler, 2013), but it is nonetheless important to assess the quality of its sampling.

A useful diagnostic for the quality of MCMC mixing is the empirical autocorrelation of the chain. In particular, we compute the correlation between samples as function of lag time. A steep decay in this autocorrelation means that successive samples are less correlated and more nearly independent. While one could compute the empirical autocorrelation for each reaction parameter individually, we instead summarize MCMC mixing by computing the autocorrelation of successive values of the log-posterior density. This is reported in Figure 9, for Examples 2 and 3 with all three prior specifications.

The decay of the autocorrelation is relatively good in both cases, though the MCMC chains mix more quickly in Example 2 than in Example 3. This difference can be ascribed to differences in posterior structure. In Example 2, the C* pathway is largely dominant, while in the dry reforming case of Example 3, both the C* and HCO* pathways have appreciable posterior probabilities. The MCMC chain in Example 3 thus has to switch between pathways more frequently, and each switch requires the inclusion and exclusion of multiple reactions. Even with the present adaptive proposal distribution, this coordinated inclusion and exclusion is a relatively “large” jump in the model space. Thus, while mixing is adequate in the current example, a more computationally efficient approach—i.e., one that could achieve similar results with fewer samples—might involve correlated proposal mechanisms that can learn not just the marginal structure of the posterior in each parameter direction but the joint structure of posterior. The design of such proposal mechanisms is a topic of ongoing research.

5.8. Posterior parameter uncertainties

Thus far, we have focused our analysis on the posterior description of uncertainties in *model structure*. But the across-model Bayesian inference framework also automatically produces a full description of uncertainties in rate *parameter values*. In other words, for every model in the posterior distribution, MCMC samples describe the joint probability distribution of the rate parameters that are included in that model (i.e., that are non-zero). Quantifying these parameter uncertainties is important when developing a rigorous assessment of uncertainties in model predictions.

Here we provide one example of the posterior parameter uncertainties obtained using our inference framework. Figure 10 shows 1-D and 2-D marginal distributions of the rate constants of the highest-probability model for steam reforming (Example 2), using Prior 3. This model includes 8 of the 15 possible reactions (as described in Table 6), and thus the continuous distribution over rate parameter values is supported on an eight-dimensional space. The diagonal of Figure 10 shows the marginal probability density function of one parameter at a time, while the off-diagonal elements show the joint probability density of each pair of parameters (marginalizing out the other six). The prior probability density was uniform over each box, and thus the more focused contours indicate a significant reduction in prior uncertainty. There is also some complex (and certainly non-Gaussian) structure in each pairwise marginal. This is the uncertainty given the experimental data at hand. Further reduction in parameter uncertainty would require the injection of additional data into the inference process.

6. Conclusions

This paper has introduced a new Bayesian framework for the inference of chemical kinetic models. This framework fuses experimental data with prior knowledge, incorporating a detailed simulation of the relationship between the underlying chemical mechanism and quantities that can be experimentally observed. Our approach quantifies uncertainty in the chemical kinetic model by exploring a probability distribution over both *model structures* (i.e., which reactions to include) and rate *parameter values*. This distribution results from a systematic approach to model inference, one that begins with a large number of proposed reactions and treats identification of the reaction network as a problem of nonlinear Bayesian variable selection. The approach thus inherits the benefits of Bayesian model selection, in that it avoids overfitting and includes an automatic Occam’s razor to penalize excessive model complexity. But it also offers a computationally efficient means of comparing a combinatorially large number of models, in contrast with other Bayesian approaches that directly compute marginal likelihoods for a small handful of chemical models.

We demonstrate our approach through the inference of kinetic models for steam and dry reforming of methane. A first example with synthetic data is used to show the consistency of the model inference procedure. Subsequent examples use real experimental data reported in the literature. We analyze the posterior distribution to quantify the degree to which individual reactions are supported by the data; furthermore, we show how the probabilities of particular multi-reaction pathways can be extracted from the posterior distribution. These pathways are competing hypotheses for how methane reforming proceeds through chains of intermediate species (e.g., C* or HCO*), and their posterior probabilities

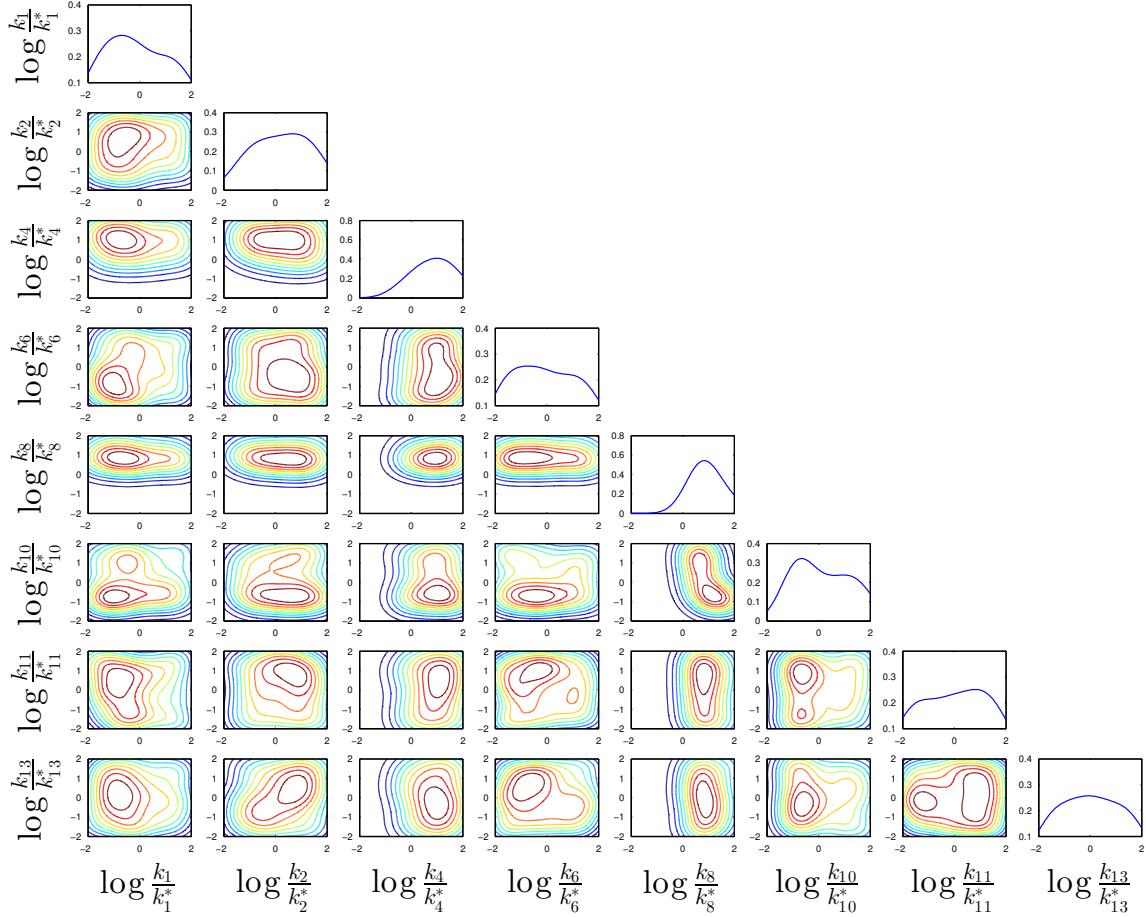


Figure 10: 1-D and 2-D posterior marginals of the rate constants of the highest-posterior-probability model for steam reforming (from Example 2), beginning with the prior $p(k_{i,f}) = 0.8\delta(k_{i,f}) + 0.2C(k_{i,f})$. The logarithms here are base 10.

quantify the extent to which each hypothesis is supported by the data. In all these analyses, we consider three different prior specifications. Prior probability distributions, both on the inclusion/exclusion of each reaction and its rate parameter value, provide a natural way of encapsulating current knowledge or preferences about the process under study. One interesting specification is the sparsity-promoting prior, which in the present examples yields smaller mechanisms than the uninformative “indifference” prior. In general, the prior specifications introduced here are intended to be illustrative; more complex or nuanced priors can easily be employed.

The key computational challenge raised by our Bayesian formulation is exploration of the posterior distribution via sampling. To make posterior sampling tractable, we employed an adaptive independence MCMC scheme, where adaptation of the point-mass mixture proposal distribution was guided by an online EM algorithm. This scheme offers significant gains in efficiency over non-adaptive samplers, but further improvements in the sampling scheme would be desirable—particularly to enable larger model comparisons, the use of more computationally intensive forward simulation models, and the inclusion of additional

uncertain parameters. Posterior distributions resulting from the across-model inference framework are intrinsically multi-modal, with complex correlations among the inclusion and exclusion of certain reactions. Further research is needed to design MCMC proposals that can quickly adapt to such posterior correlations and jump between well-separated posterior modes, thereby improving chain mixing. The present Bayesian model inference scheme can also be viewed as one component of a sequential experimental design procedure, where inference is used to identify the most uncertain reactions or to isolate particular groups of competing reaction pathways. New data, chosen via an experimental design criterion that focuses on these reactions, could be then collected and introduced into subsequent Bayesian model inference procedures.

7. Acknowledgements

This work was supported by the KAUST Global Research Partnership.

References

- H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- C. Andrieu and E. Moulines. On the ergodicity properties of some adaptive MCMC algorithms. *Annals of Applied Probability*, 16(3):1462–1505, 2006.
- C. Andrieu, N. deFreitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50:5–43, 2003.
- C. Andrieu, E. Moulines, and P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM Journal Control and Optimization*, 44(1):283–312, 2005.
- S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010.
- J. Berger and L. Pericchi. Objective Bayesian methods for model selection: introduction and comparison. *Model Selection (P.Lahiri, editor), IMS Lecture Notes – Monograph Series*, 38:135–207, 2001.
- J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley, 2000.
- B. Bhattacharjee, D. A. Schwer, P. I. Barton, and Jr. W. H. Green. Optimally-reduced kinetic models: reaction elimination in large-scale kinetic mechanisms. *Combustion and Flame*, 135(3):191–208, 2003.
- J. A. Bilmes. A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, 1998.
- C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2007.
- K. Braman, T. A. Oliver, and V. Raman. Bayesian analysis of syngas chemistry models. *Combustion Theory and Modelling*, 17(5):858–887, 2013.

- K. P. Burnham and D. R. Anderson. *Model Selection and Multimodal Inference: A Practical Information Theoretic Approach*. Springer, 2nd edition, 2002.
- B. P. Carlin and S. Chib. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(3):473–484, 1995.
- S. Chib and I. Jeliazkov. Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, 96(453):270–281, 2001.
- P. Dellaportas, J. J. Forster, and I. Ntzoufras. On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12:27–36, 2002.
- A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
- O. Deutschmann, R. Schwiedernoch, L. I. Maier, and D. Chatterjee. Comparison between calculated and experimentally determined selectivity and conversion for short-contact-time reactors using honeycomb monoliths. *Natural Gas Conversion VI, Studies in surface science and catalysis*, E. Iglesia, J.J. Spivey, T.H. Fleisch (eds.), Elsevier, 136:215–258, 2001.
- R. Feeley, M. Frenklach, M. Onsum, T. Russi, A. Arkin, and A. Packard. Model discrimination using data collaboration. *The Journal of Physical Chemistry A*, 110(21):6803–6813, 2006.
- N. Friedman and D. Koller. Being Bayesian about network structure. a Bayesian approach to structure discovery in Bayesian networks. *Machine Learning*, 50:95–125, 2003.
- N. Friel and A. N. Pettitt. Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B*, 70(3):589–607, 2008.
- P. H. Garthwaite, J. B. Kadane, and A. O'Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470):680–700, 2005.
- A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rudin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd edition, 2004.
- W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in Practice*. Chapman and Hall/ CRC, 1996.
- S. J. Godsill. On the relationship between Markov chain Monte Carlo methods for model uncertainty. *Journal of Computational and Graphical Statistics*, 10(2):230–248, 2001.
- D. Goodwin, N. Malalya, H. Moffat, and R. Speth. Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes, version 2.0.2. available at <https://code.google.com/p/cantera/>, 2013.
- P. Green and D. Hastie. Reversible jump MCMC. Available at <http://www.maths.bris.ac.uk/~mapjg/papers/>. Access date: 04/24/2014, 2009.

- P. J. Green. Reversible jump Markov chain Monte Carlo computation and model determination. *Annals of Stat*, 82(4):711–732, 1995.
- J. Hanna, W. Y. Lee, Y. Shi, and A. F. Ghoniem. Fundamentals of electro- and thermochemistry in the anode of solid-oxide fuel cells with hydrocarbon and syngas fuels. *Progress in Energy and Combustion Science*, 40:74–111, 2014.
- M. Hartmann, L. Maier, H. D. Minh, and O. Deutschmann. Catalytic partial oxidation of iso-octane over Rhodium catalysts: An experimental, modeling, and simulation study. *Combustion and Flame*, 157(9):1771–1782, 2010.
- T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition, 2009.
- D. Hickman and L. D. Schmidt. Steps in CH₄ oxidation on Pt and Rh surfaces: High-temperature reactor simulations. *AIChe J.*, 39(7):1164–1177, 1993.
- J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417, 1999.
- C. Ji and S. C. Schmidler. Adaptive Markov chain Monte Carlo for Bayesian variable selection. *J. Comp. and Graph. Stat.*, 22(3):708–728, 2013.
- R. E. Kass and A. E. Raftery. Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.
- R. J. Kee, M. E. Coltrin, and P. Glarborg. *Chemically Reacting Flow: Theory and Practice*. Wiley, 2003.
- H. J. Kushner and G. G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer, 2nd edition, 1997.
- D. J. C. Mackay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press UK, 2003.
- C. L. Mallows. Some comments on Cp. *Technometrics*, 15(4):661–675, 1973.
- N. E. McGuire, N. P. Sullivan, R. J. Kee, H. Zhu, A. J. Nabity, J. R. Engel, D. T. Wickham, and M. J. Kaufman. Catalytic steam reforming of methane using Rh supported on Sr-substituted hexaaluminate. *Chemical Engineering Science*, 64(24):5231–5239, 2009.
- N. E. McGuire, N. P. Sullivan, O. Deutschmann, H. Zhu, and R. J. Kee. Dry reforming of methane in a stagnation-flow reactor using Rh supported on strontium-substituted hexaaluminate. *Applied Catalysis A: General*, 394(1–2):257–265, 2011.
- T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.
- M. A. Newton and A. E. Raftery. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of Royal Statistical Society: Series B*, 56(1):3–48, 1994.

- C. J. Oates, B. T. Hennessy, Y. Lu, G. B. Mills, and S. Mukherjee. Network inference using steady-state data and goldbeter-koshland kinetics. *Bioinformatics*, 28(18):2342–2348, 2012.
- O. O. Oluwole, B. Bhattacharjee, J. E. Tolsma, P. I. Barton, and W. H. Green Jr. Rigorous valid ranges for optimally-reduced kinetic models. *Combustion and Flame*, 146(1–2):348–365, 2006.
- J. Prager, H. Najm, and J. Zador. Uncertainty quantification in the ab initio rate-coefficient calculation for the $\text{CH}_3\text{CH}(\text{OH})\text{CH}_3 + \text{OH} \rightarrow \text{CH}_3\text{C}(\text{OH})\text{CH}_3 + \text{H}_2\text{O}$ reaction. *Proceedings of the Combustion Institute*, 34(1):583–590, 2013.
- H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407, 1951.
- C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2nd edition, 2004.
- G. O. Roberts and J. S. Rosenthal. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of Applied Probability*, 44(1):458–475, 2007.
- K. Sachs, D. Gifford, T. Jaakkola, P. Sorger, and D. A. Lauffenburger. Bayesian network approach to cell signaling pathway modeling. *Science STKE*, 148:pe38, 2002.
- K. Sachs, O. Perez, D. Pe'er, D. Lauffenburger, and G. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.
- G. E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.
- D. S. Sivia. *Data Analysis: A Bayesian Tutorial*. Oxford University Press USA, 2nd edition, 2006.
- L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22(4):1701–1728, 1994.
- K. Vahteristo, A. Laari, H. Haario, and A. Solonen. Estimation of kinetic parameters in neopentyl glycol esterification with propionic acid. *Chemical Engineering Science*, 63(3):587–598, 2008.
- K. Vahteristo, A. Laari, and A. Solonen. Diels-alder reaction kinetics for production of Norbornene Monomers: Evaluation of parameters and model reliability using Markov chain. *Industrial and Engineering Chemistry Research*, 52(19):6357–6365, 2013.
- M. Vogler, A. Bieberle-Hütter, L. Gauckler, J. Warnatz, and W. G. Bessler. Modelling study of surface reactions, diffusion, and spillover at a Ni/YSZ patterned anode. *Journal of the Electrochemical Society*, 156(5):B663–B672, 2009.
- V. Vyshemirsky and M. A. Girolami. Bayesian ranking of biochemical system models. *Bioinformatics*, 24(6):833–839, 2008.

- V. Vyshemirsky, M. Girolami, A. Gormand, W. Kolch, and M. D. Houslay. A Bayesian analysis of the ERK signalling pathway, 2006.
- D. J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman and Hall/ CRC, 2011.
- C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.
- T.-R. Xu, V. Vyshemirsky, A. Gormand, A. von Kriegsheim, M. Girolami, G. S. Baillie, D. Ketley, A. J. Dunlop, G. Milligan, M. D. Houslay, and W. Kolch. Inferring signalling pathway topologies from multiple perturbation measurements of specific biochemical species. *Science Signaling*, 3(113):ra20, 2010.
- V. Yurkiv, D. Starukhin, H.-R. Volpp, and W. G. Bessler. Elementary reaction kinetics of the CO/CO₂/Ni/YSZ electrode. *Journal of the Electrochemical Society*, 158(1):B5–B10, 2011.
- C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods*, 11(2):197–202, 2014.

Appendix A. Online expectation-maximization for proposal adaptation

Here we present a derivation of the online EM algorithm applied to a general point mass mixture proposal:

$$q(k_i; \psi_i) = b_{i,0}\delta(k_i) + \sum_{m=1}^M b_{i,m}q_m(k_i; \theta_{i,m}). \quad (\text{A.1})$$

The marginal proposal distribution $q(k_i; \psi_i)$ shown in (A.1) can also be rewritten as

$$q(k_i; \psi_i) = \sum_{z_i} q(k_i, z_i; \theta_i), \quad (\text{A.2})$$

and taking $q(k_i; \psi_i)$ to be independent for each k_i , the joint proposal distribution for an N -dimensional problem follows:

$$q(\bar{k}; \bar{\psi}) = \sum_{\bar{z}} q(\bar{k}, \bar{z}; \bar{\psi}) = \prod_{i=1}^N \sum_{z_i} q(k_i, z_i; \theta_i). \quad (\text{A.3})$$

Here, z_i is a *latent variable* that takes one of $M + 1$ values corresponding to the $M + 1$ components that could generate the posterior sample. $q(\bar{k}, \bar{z})$ is the joint distribution of \bar{z} and \bar{k} and is referred to as the complete-data likelihood. Expanding (A.2) by the product rule of probability gives:

$$\begin{aligned} q(k_i; \psi_i) &= \sum_{z_i} q(k_i|z_i)q(z_i) \\ &= q(z_i = 0)\delta(k_i) + \sum_{m=1}^M q(z_i = m)q_m(k_i|z_i = m; \theta_{i,m}); \end{aligned} \quad (\text{A.4})$$

Comparing (A.4) to (A.1), we see that

$$q(z_i = 0) = b_{i,0} \quad \text{and} \quad q(z_i = m) = b_{i,m}. \quad (\text{A.5})$$

After the steps for a general point mass mixture proposal have been established, we will obtain specific expressions for the case when the continuous components of the above proposal distribution (12) are all Gaussian.

Appendix A.1. KL divergence minimization yields a maximum likelihood problem

Recall that our goal is to update the proposal distribution $q(\bar{k}; \bar{\psi})$ iteratively based on samples from the posterior distribution $p(\bar{k}|\mathcal{D})$ so as to minimize the KL divergence:

$$D_{KL}(p(\bar{k}|\mathcal{D}\|q(\bar{k}; \bar{\psi})) = \int p(\bar{k}|\mathcal{D}) \log \left(\frac{p(\bar{k}|\mathcal{D})}{q(\bar{k}; \bar{\psi})} \right) d\bar{k} \quad (\text{A.6})$$

w.r.t. the proposal parameters $\bar{\psi}$. Note that minimizing the KL divergence in (A.6) is equivalent to maximizing the cross entropy $\int p(\bar{k}|\mathcal{D})q(\bar{k}, \bar{\psi})d\bar{k}$. Thus the objective function can be rewritten as

$$\bar{\psi}^* = \arg \max_{\bar{\psi}} \int p(\bar{k}|\mathcal{D}) \log(q(\bar{k}; \bar{\psi})) d\bar{k}. \quad (\text{A.7})$$

The integral in (A.7) can be approximated by a Monte Carlo sum using T samples from the posterior distribution $p(\bar{k}|\mathcal{D})$ as

$$I = \frac{1}{T} \sum_{t=1}^T \log(q(\bar{k}^t; \bar{\psi})) = \frac{1}{T} \log \left(\prod_{t=1}^T q(\bar{k}^t; \bar{\psi}) \right). \quad (\text{A.8})$$

Now, if we think of $\bar{k}^{t=1:T}$ as pseudo-data and $q(\bar{k}^t; \theta)$ as a likelihood, cross entropy can be interpreted as a log-likelihood under infinite data and (A.7) as a maximum (log-)likelihood problem. Mathematically (A.7) can also be written as

$$\bar{\psi}^* = \arg \max_{\bar{\psi}} \lim_{T \rightarrow \infty} \frac{1}{T} \log \left(\prod_{t=1}^T q(\bar{k}^t; \bar{\psi}) \right) \quad (\text{A.9})$$

Appendix A.2. Classical EM algorithm

Suppose we are given T independent samples $(\bar{k}^1, \bar{k}^2, \dots, \bar{k}^T)$ distributed according to $p(\bar{k}|\mathcal{D})$. The solution of the maximum log-likelihood problem

$$\bar{\psi}^* = \arg \max_{\bar{\psi}} \frac{1}{T} \sum_{t=1}^T \log(q(\bar{k}^t; \bar{\psi})) \quad (\text{A.10})$$

can be obtained by taking the derivative of the log-likelihood and solving the resulting nonlinear equations. The nonlinear equations thus obtained seldom have a closed-form solution and thus are solved by numerical optimization.

An alternative known as expectation-maximization algorithm exists for the solution of the maximum log-likelihood problem (Bilmes, 1998; Dempster et al., 1977). The EM algorithm often results in simple analytical expressions and avoids the difficulties of gradient-based optimization approaches. The EM algorithm consists of two steps, known as the E-step and M-step, that are solved iteratively to obtain the optimal parameter values under mild regularity conditions (Wu, 1983). The two steps are given by

E-step

$$\begin{aligned} Q(\bar{\psi}, \bar{\psi}_{n-1}) &= \int \log \left(\prod_{t=1}^T q(\bar{k}^t, \bar{z}^t; \bar{\psi}) \right) \left(\prod_{t=1}^T q(\bar{z}^t | \bar{k}^t, \bar{\psi}_{n-1}) \right) d\bar{z}^1 \dots d\bar{z}^T \\ &= \mathbb{E}_{\bar{z}^1 \dots \bar{z}^T} \left[\log \left(\prod_{t=1}^T q(\bar{k}^t, \bar{z}^t; \bar{\psi}) \right) \right] \end{aligned} \quad (\text{A.11})$$

M-step

$$\bar{\psi}_n = \arg \max_{\bar{\psi}} Q(\bar{\psi}, \bar{\psi}_{n-1}) \quad (\text{A.12})$$

The E-step in the above equations evaluates the expectation of the logarithm of the complete-data likelihood, where the expectation is taken with respect to the latent variables conditioned on available (sampled) rate parameters. In the M-step, an updated set of parameter values are computed by maximizing the expected log-likelihood from the E-step. The EM algorithm as described in (A.11) and (A.12) is applicable if all the observed samples ($\bar{k}^{t=1:T}$) are available a priori and the samples are independent.

Our problem is different from the above case since we are generating samples from $p(\bar{k}|\mathcal{D})$ in batches. Moreover, the generated samples are not independent as they are coming from an MCMC scheme. Thus we use a sequential variant of the EM algorithm known as the online EM algorithm and specify conditions under which the resulting adaptive MCMC algorithm converges to the posterior distribution, $p(\bar{k}|\mathcal{D})$.

Appendix A.3. Online expectation maximization

We begin our discussion of the online EM algorithm by assuming that the proposal distribution $q(\bar{k}, \bar{z}; \bar{\psi})$ can be represented in the form

$$q(\bar{k}, \bar{z}; \bar{\psi}) = \exp(\langle s(\bar{k}, \bar{z}), \bar{\phi}(\bar{\psi}) \rangle - \bar{A}(\bar{\psi})). \quad (\text{A.13})$$

Distributions that can be cast in the above form are known to belong to the exponential family (Bernardo and Smith, 2000). Here, $s(\bar{k}, \bar{z})$ is a vector of sufficient statistics, $\bar{\phi}(\bar{\psi})$ refers to the natural parameters, and $\bar{A}(\bar{\psi})$ is the log base distribution. The operator $\langle \cdot \rangle$ is the standard inner product. Plugging the above expression for $q(\bar{k}, \bar{z})$ into (A.11) and (A.12), we get

E-step:

$$Q(\bar{\psi}, \bar{\psi}_{n-1}) = \int \prod_{t=1}^T q(\bar{z}^t | \bar{k}^t; \bar{\psi}_{n-1}) \sum_{t=1}^T (\langle s(\bar{k}^t, \bar{z}^t), \bar{\phi}(\bar{\psi}) \rangle - \bar{A}(\bar{\psi})) d\bar{z}^1 \dots d\bar{z}^T, \quad (\text{A.14})$$

M-step:

$$\bar{\psi}_n = \arg \max_{\bar{\psi}} Q(\bar{\psi}, \bar{\psi}_{n-1}). \quad (\text{A.15})$$

The above expectation and maximization steps can be recast in terms of sufficient statistics as

E-step:

$$S_n^T = \frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\bar{\psi}_{n-1}^T} [s(\bar{k}^t, \bar{z}^t) | \bar{k}^t], \quad (\text{A.16})$$

M-step:

$$\bar{\psi}_n^T = \Gamma\{S_n^T\}, \quad (\text{A.17})$$

where $\Gamma\{S_n^T\} = \arg \max_{\bar{\psi}} (\langle S_n^T, \bar{\phi}(\bar{\psi}) \rangle - \bar{A}(\bar{\psi}))$. Letting $T \rightarrow \infty$, the EM iterations are

E-step:

$$S_n = \mathbb{E}_{p(\bar{k}|\mathcal{D})} \left(\mathbb{E}_{\bar{\psi}_{n-1}} [s(\bar{k}, \bar{z}|\bar{k})] \right) \quad (\text{A.18})$$

M-step:

$$\bar{\psi}_n = \Gamma\{S_n\} \quad (\text{A.19})$$

Thus our overall goal of solving (A.7) is equivalent to locating the solutions of

$$\mathbb{E}_{p(\bar{k}|\mathcal{D})} (\mathbb{E}_{\Gamma\{S\}} [s(\bar{k}, \bar{z}|\bar{k})]) - S = 0 \quad (\text{A.20})$$

If we now take $\frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\Gamma\{S\}} [s(\bar{k}^t, \bar{z}^t) | \bar{k}^t]$ to be a noisy estimate of $\mathbb{E}_{p(\bar{k}|\mathcal{D})} (\mathbb{E}_{\Gamma\{S\}} [s(\bar{k}, \bar{z}|\bar{k})])$, application of the Robbins-Monro stochastic approximation algorithm results in the online EM algorithm (Andrieu and Moulines, 2006; Robbins and Monro, 1951). The online EM iterations are given by

E-step:

$$S_n = (1 - \eta_n) S_{n-1} + \eta_n \left(\frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\bar{\psi}_{n-1}} [s(\bar{k}_t, \bar{z}_t) | \bar{k}_t] \right) \quad (\text{A.21})$$

M-step:

$$\bar{\psi}_n = \Gamma\{S_n\} \quad (\text{A.22})$$

η_n here is a sequence of decreasing positive step sizes and satisfies the following two conditions:

$$\sum_{n=1}^{\infty} \eta_n = \infty \text{ and } \sum_{n=1}^{\infty} \eta_n^2 < \infty \quad (\text{A.23})$$

We take $\eta_n = 1/n$ in our work. We now return to the complete-data likelihood of the point-mass mixture proposal distribution ((A.2, A.3)). Assuming that the continuous parts of the proposal distribution for each rate parameter k_i are Gaussian distributions with arbitrary initial means and variances and recalling that the proposal for each k_i is independent, we obtain the complete-data log-likelihood as

$$\log q(\bar{k}, \bar{z}|\bar{\psi}) = \sum_{i=1}^N \sum_{m=0}^M z_{i,m} \log b_{i,m} + \sum_{i=1}^N \sum_{m=1}^M z_{i,m} \log \mathcal{N}(k_i; \mu_{i,m}, \sigma_{i,m}^2). \quad (\text{A.24})$$

It can be easily be shown that (A.24) can be cast in the form of (A.13) and that the corresponding sufficient statistics are given by:

For $i = 1$ to N

For $m = 0$ to M :

$$O_{i,m} = \frac{1}{T} \sum_{t=1}^T \gamma(z_{i,m}^t),$$

For $m = 1$ to M :

$$P_{i,m} = \frac{1}{T} \sum_{\substack{t=1 \\ k_i^t \neq 0}}^T \gamma(z_{i,m}^t) \quad Q_{i,m} = \frac{1}{T} \sum_{\substack{t=1 \\ k_i^t \neq 0}}^T \gamma(z_{i,m}^t) k_i^t \quad R_{i,m} = \frac{1}{T} \sum_{\substack{t=1 \\ k_i^t \neq 0}}^T \gamma(z_{i,m}^t) (k_i^t)^2$$

where $\gamma(z_{i,m}^t) = p(z_{i,m}^t | k_i^t; \psi_i)$ is given by

$$\gamma(z_{i,m}^t) = \begin{cases} 1 & \text{if } k_i^t = 0 \text{ and } m = 0 \\ 0 & \text{if } k_i^t = 0 \text{ and } m \neq 0 \\ 0 & \text{if } k_i^t \neq 0 \text{ and } m = 0 \\ \frac{b_{i,m} \mathcal{N}(k_i^t; \mu_{i,m}, \sigma_{i,m}^2)}{\sum_{m'=1}^M b_{i,m'} \mathcal{N}(k_i^t; \mu_{i,m'}, \sigma_{i,m'}^2)} & \text{if } k_i^t \neq 0 \text{ and } m \neq 0. \end{cases} \quad (\text{A.25})$$

Thus the online EM iterations consist of the following two steps

E-step:

$$\begin{aligned} S_n^{O_{i,m}} &= S_{n-1}^{O_{i,m}} + \eta_n (O_{i,m} - S_{n-1}^{O_{i,m}}) \\ S_n^{P_{i,m}} &= S_{n-1}^{P_{i,m}} + \eta_n (P_{i,m} - S_{n-1}^{P_{i,m}}) \\ S_n^{Q_{i,m}} &= S_{n-1}^{Q_{i,m}} + \eta_n (Q_{i,m} - S_{n-1}^{Q_{i,m}}) \\ S_n^{R_{i,m}} &= S_{n-1}^{R_{i,m}} + \eta_n (R_{i,m} - S_{n-1}^{R_{i,m}}) \end{aligned} \quad (\text{A.26})$$

M-step:

$$\begin{aligned}
b_{i,m} &= \frac{S_n^{O_{i,m}}}{\sum_{m'=0}^M S_n^{O_{i,m'}}} \\
\mu_{i,m} &= \frac{S_n^{Q_{i,m}}}{S_n^{P_{i,m}}} \\
\sigma_{i,m}^2 &= \frac{\mu_{i,m}^2 S_n^{P_{i,m}} - 2\mu_{i,m} S_n^{Q_{i,m}} + S_n^{R_{i,m}}}{S_n^{P_{i,m}}}
\end{aligned} \tag{A.27}$$

We have thus arrived at the steps of an adaptive MCMC algorithm that involves simulating a batch of T samples from the posterior distribution in each iteration and updating the proposal parameters based on (A.26) and (A.27). Because online EM adjusts proposal parameters based on all the past samples, standard proofs that guarantee asymptotic convergence of non-adaptive MCMC methods do not apply here. [Andrieu and Moulines \(2006\)](#) provide rigorous technical conditions that guarantee a law of large numbers and a central limit theorem for the online EM algorithm. These conditions also require that one include a non-adaptive fixed component in the proposal distribution; we do so in our simulations in the form of a multi-dimensional Gaussian with fixed parameters $\tilde{q}(\bar{k}; \tilde{\psi})$. [Roberts and Rosenthal \(2007\)](#), in contrast, develop simpler conditions that ensure convergence of the adaptive MCMC scheme to the target distribution and provide a law of large numbers. The first is known as diminishing adaptation, which requires that the magnitude of adaptation is continuously decreasing. The online EM-based adaptive MCMC approach described above satisfies this condition since the step size $\eta_n \rightarrow 0$. The second condition, known as bounded convergence, is satisfied as long as the non-adaptive component $\tilde{q}(\bar{k}; \tilde{\psi})$ has sufficiently heavy tails or the support of \bar{k} is compact ([Ji and Schmidler, 2013](#)).