BNL-226494-2025-JAAM

Particle Markov Chain Monte Carlo Approach to Inference in Transient Surface Kinetics

M. Iloska, J. Boscoboinik

To be published in "Journal of Chemical Theory and Computation"

January 2025

Center for Functional Nanomaterials

Brookhaven National Laboratory

U.S. Department of Energy

USDOE Office of Science (SC), Basic Energy Sciences (BES). Scientific User Facilities (SUF)

DISCLAIMER

A Particle Markov Chain Monte Carlo approach to inference in transient surface kinetics

Marija Iloska,[*,†] J. Anibal Boscoboinik,[‡] Qin Wu,[‡] and Mónica F. Bugallo[†]

†*Department of Electrical & Computer Engineering,*

Stony Brook University, Stony Brook, NY 11794, United States

‡*Center of Functional Nanomaterials,*

Brookhaven National Lab, Upton, NY 11973, United States

E-mail: marija.iloska@stonybrook.edu

Abstract

In this work, we develop a novel Bayesian approach to study the adsorption and desorption of CO onto a Pd(111) surface, a process of high importance in natural sciences. The motivation for this work comes from the recent availability of time-resolved infrared spectroscopy data, and the need of model interpretability and uncertainty quantification in chemical processes. The objective is to learn the relevant parameters which characterize the process: coverage with time, rate constants, activation energies, and pre-exponential factors. Our approach consists of three main schemes: i) a problem design and probabilistic model for the whole system, ii) a particle Markov chain Monte Carlo sampler to learn the hidden coverages and rate constant parameters, and iii) two Bayesian formulations to infer the activation energies and pre-exponential factors. The flexibility of the Bayesian framework allows for uncertainty quantification where possible, and integration of mathematical constraints in the model to reflect the system physically. We found that our results for the activation energies and pre-exponential factor are in agreement with those reported in experimental literature, independently,

and we provide discussions on the advantages and disadvantages as well as applicability to other systems.

Introduction

Chemical reactions and processes are stochastic by nature. Learning and understanding their rates and mechanisms requires probabilistic approaches[1] to account for uncertainties in a principled way. A probabilistic framework that is particularly well-suited for uncertainty quantification is Bayesian inference. In practice, most Bayesian approaches cannot be solved analytically and rely on approximate inference such as using Monte Carlo (MC) algorithms[2,3] or variational inference.[4] We specifically refer the reader unfamiliar with Bayesian inference to the work in,[5] which offers great intuition behind Bayesian reasoning within the scope of chemical applications.

There has been more push in the literature for Bayesian modeling in kinetics.[6–8] For example, the authors in[9] use an adaptive Markov chain Monte carlo (MCMC) method with expectation-maximization (EM)[10] to learn the rate parameters and model dimensions given stagnation flow reactor data. Further, the important role of Bayesian inference in kinetics modeling was highlighted in the comprehensive review in.[11] The authors provide in-dept discussions of all the possible sources of error in kinetics, and illustrative examples of Bayesian parameter estimation in their context. Other works (e.g.,[5]) use MCMC toolboxes such as STAN[12] or JAGS,[13] which require only a specified data model and prior as an input from the user. In an effort to generalize a Bayesian toolbox for chemical kinetics, an encompassing Python package for kinetics abbreviated CheKiPEUQ was proposed by.[14] The authors demonstrate its use on several examples including temperature programmed reactions and further show that it can also be used as aid in experimental design.[15] Soon after, another open source Python toolbox called CKBIT[16] was proposed, which similarly allows practitioners to learn their kinetic model from concentration data, or the activation energies given

the reaction rates. The authors discuss the pros and cons of using different toolboxes and coding platforms, as well as the limitations on parameter domains when choosing priors and models. The development of toolboxes such as CKBIT and CheKiPEUQ is of utmost importance to allow practitioners in kinetics to reap the benefits of the Bayesian paradigm.

Recently, there has also been use of sequential Monte Carlo or particle filtering (PF)[3] in kinetic applications. Specifically,[17] use PF track the concentration dynamics of a dissolution-precipitation process from measurements of an intermediate only, and pair it with EM to learn the rate constant parameters. Another work[18] uses PF to study a CO_2 adsorption process in a hollow fiber sorbent, and quantifies uncertainty in cases of "residual variability". Many problem formulations and modeling choices depend heavily on the data/experiment available for the system in question, and for this reason, we recommend the related works[15,19,20] for configuring optimal experimental designs.

In this paper, we propose a novel Bayesian framework for studying the adsorption and desorption of CO onto Pd(111) from infrared (IR) spectroscopy time-series. The adsorption of CO on Pd(111) is a well-studied process of particular importance in environmental settings (green house gas emissions[21]), industrial (catalytic converters design[22]), and for general understanding of surface kinetics.[23] The amount of CO adsorbed onto the Pd(111) is referred to as the coverage of the system. There is a plethora of both experimental,[24,25] and computational studies,[26,27] bringing forth much discussion about the underlying mechanism. Many complications arise because CO can adsorb onto a-top, bridge, and hollow sites, obscuring the measurable signals. The objective is to find a physically meaningful model, and to learn the coverage dynamics, the kinetic parameters, and the measurement noise for a range of temperatures, given IR data. Additionally, we want to use the learned parameters to estimate the activation energies and pre-exponential factors involved. The advancement of novel experimental design and data acquisition techniques such as temporal IR spectroscopy[28] provides new opportunities to elucidate the mechanisms of molecular processes. An immediate distinction from the above-mentioned literature is that we are learning

from time-resolved IR data.

We summarize our contributions as follows:

1. A partition of the system into 4 regions, and interpretable probabilistic models, which incorporate physical constraints and maintain continuity between the regions; this includes models of the coverage dynamics, the rate constants, and the observation noise,

2. A particle Markov chain Monte Carlo (pMCMC) sampler to infer the parameters mentioned in 1).

3. Two Bayesian approaches for inference of the activation energies and pre-exponential factors and an analysis of the uncertainty quantification. In the first approach, we obtain estimates as a direct consequence of the pMCMC results, and in the second approach we propose a Bayesian model that achieves conjugacy.

Unlike MCMC, pMCMC[29] methods are not standardized in toolboxes such as STAN, JAGS or other Bayesian kinetic toolboxes. Several groups[30–32] have demonstrated success of pMCMC algorithms for biological kinetic systems (e.g., prokaryotic auto-regulatory networks, predator-prey models). As they are designed to track population of species from noisy population data, their formulations and targets are not directly applicable to the system we are considering. To the best of our knowledge, there are so far no existing Bayesian methods that learn from specifically time-series IR reflection-absorption spectroscopy (IRRAS) data. Further, we are not familiar with any works with the proposed probabilistic modeling for the coverages, the rate constant parameters, as well as the conjugate models derived for the activation energies and pre-exponential factors. We validate our results reported in extensive experimental literature,[33–35] and provide a discussion on challenges when working with real data.

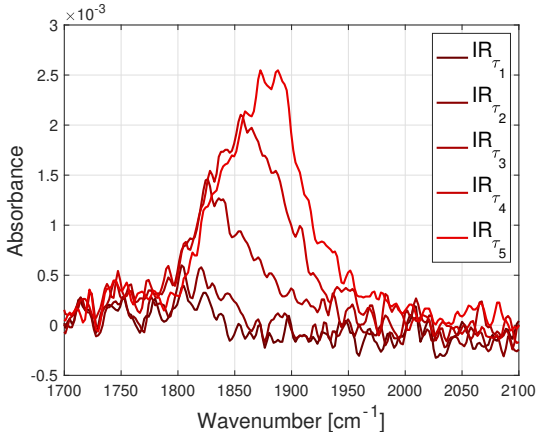
Problem Description

Experiment Experiments are carried out in an ultra-high vacuum system with a base pressure of $3 \times 10^9 mbar$. The system has two interconnected chambers. One is a preparation chamber used for sample cleaning, and the second one is used for transient kinetics IRRAS.

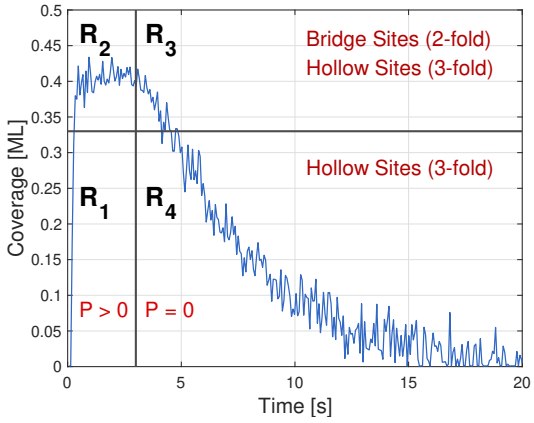
Sample cleaning A Pd(111) single crystal was cleaned by Ar ion sputtering and annealing cycles, as described in reference.[36] Briefly, the Pd(111) crystal was cleaned by Ar+ sputtering at $1.1keV$ at a pressure of $1 \times 10^{-5} mbar$. The sample was then annealed to $1073K$ for 5 min and cooled to $973K$ in O_2 atmosphere at $1 \times 10^7 mbar$. This cleaning step is repeated 3 times, followed by flash annealing to $1073K$ and transferring to the IRRAS chamber.

IR scanning The sample is then stabilized at temperature T and an ultra-fast valve with a capillary facing the sample is open producing an influx of CO gas ($P > 0$) for a set amount of time $\tau = \tau_p$. This time is 3 seconds for all experiments described here, which is enough to reach an equilibrium coverage, as evident by the constant position and area of the CO IRRAS peak. At this point ($\tau = \tau_p$), the CO valve is closed ($P = 0$), and CO starts to desorb. IRRA spectra is obtained throughout this cycle of opening and closing the valve (note that the valve has a time resolution of 200 microseconds). The IRRAS measurements can detect the presence of CO on the Pd surface based on the vibrational frequency of the molecule.[37] The measurements are collected throughout the entire duration of the experiment, i.e., until CO is completely desorbed ($\tau = \tau_F$), and no CO presence is detected. An illustration of CO coverage over the course of the experiment is shown in Fig. 1b.

Data IR spectroscopy is specially suited to track CO adsorption, and it is sensitive to the type of site on which the molecule sticks on the metal surface. However, it is a space averaging method, which requires inference to obtain kinetic and mechanistic insights about the process at the atomic scale. The raw dataset for an individual temperature T is a time-series IR spectra of absorbance vs wavenumber (see Fig. 1a). For our purposes, we are interested in the area under the curve of the IR spectra, as it is proportional to the coverage.[38] Ulti-



(a) IR spectra for the first 5 time instants.



(b) Expected coverage dynamics and problem partitioning. In regions $R_{1\&2}$ we have adsorption + desorption. In regions $R_{3\&4}$ we have desoprtion.

Figure 1: Visualization of the IR data and expected coverage for a single temperature.

mately, we have available the area $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}_{\tau_1:\tau_N}$, where the notation $\tau_1 : \tau_N$ indicates a vector of all the points at times $\tau_1, \tau_2, \ldots, \tau_N$. An area point is obtained as $\mathcal{A}_{\tau_n} = \int_{\nu_0}^{\nu} \mathcal{IR}_{\tau_n}$, where $\mathcal{IR}_{\tau_n}$ denotes the full IR spectrum at time τ_n. The information about CO in an IR spectrum lies within the wavenumber interval $[\nu_0, \nu]$, which in our case is $[1700cm^{-1}, 2100cm^{-1}]$. The experiment was repeated for $T \in \{450, 460, 470, 475, 480, 490\}K$ and a visualization of the obtained area data $\boldsymbol{\mathcal{A}}$ for each temperature is provided in the Supporting Information (SI).

Physical Constraints Formally, coverage θ is defined as the number of CO molecules over number of Pd surface atoms in units - monolayers (ML). Within the temperature range of our experiments $450K \leq T \leq 490K$, CO can adsorb on two possible sites: hollow (3-fold) and bridge (2-fold).[39,40] There are several prior findings that are crucial to allow for our modeling, at the experimental conditions used.

1. The maximum coverage (for a surface packed with CO atoms) is $0.5ML$. Hence, we have the constraint that $0 \leq \theta \leq \frac{1}{2}$.

2. There is a "phase" change around $\frac{1}{3}ML$. For low coverages when $\theta < \frac{1}{3}$, CO adsorbs only on hollow sites, and for $\frac{1}{3} \leq \theta \leq \frac{1}{2}$ it adsorbs on both hollow and bridge sites. However, there is much debate[24,26,40] about the proportions of hollow and bridge

adsorptions.

3. Estimates of the coverage at equilibrium θ^{EQ} for different temperatures are available from.[41]

We use the phase and pressure changes to divide the problem in 4 regions (see Fig. 1b), which we denote R_r, with $r \in \{1, 2, 3, 4\}$.

Modeling

The chemical equation that describes adsorption and desorption of CO onto a Pd surface from gas phase is:

$$\mathrm{CO} \underset{k_{D_\mathrm{r}}}{\overset{k_{A_\mathrm{r}}}{\rightleftharpoons}} \mathrm{CO^{Pd}}, \tag{1}$$

where CO is CO in the gas phase, and CO-Pd is CO adsorbed on the Pd surface. The process is characterized by k_{Ar} and k_{Dr}, the rate constants of adsorption and desorption in region r, respectively. Then, we can write the rate of the process with the following differential equation:

$$\frac{\partial \theta_\tau}{\partial \tau} = \overbrace{k_{Ar} P(\theta^{MAX} - \theta_\tau)}^{\text{adsorption term}} \overbrace{-k_{Dr}\theta_\tau}^{\text{desorption term}} \quad , \tag{2}$$

where $\theta^{MAX} = \frac{1}{2}$ is the absolute maximum possible coverage. We discretize equation (2) to obtain:

$$\theta_{\tau_n} = (1 - k_{Dr}\Delta\tau)\theta_{\tau_{n-1}} + k_{Ar}\Delta\tau P(\theta^{MAX} - \theta_{\tau_{n-1}}), \tag{3}$$

where $P = 0$ for the desorption regions $R_{3\&4}$.

Given the phase change at $\theta = \frac{1}{3}$, we take that the same desorption process happens in the regions R_1 and R_4, as well as the regions R_2 and R_3, and set the parameters $k_{D1} = k_{D4}$, $k_{D2} = k_{D3}$. We acknowledge that a more complicated model involving adsoroption on both hollow sites and bridge sites, and their molecular interaction is required in the regions R_2 and R_3 to capture a more realistic depiction of the process. Our purpose within the scope of this work is to attempt to model and estimate the net effects in R_{23} and provide uncertainty.

The goal is to first estimate the latent states $\boldsymbol{\theta} = [\theta_{\tau_1}, \ldots, \theta_{\tau_N}]^\top$, and the rate constants $\mathbf{k} = [k_{A1}, k_{A2}, k_{D3}, k_{D4}]$, given the area data $\boldsymbol{\mathcal{A}}$ for each temperature T_t, then to use these results to find the activation energies $\mathbf{E}_a = [E_{a_1}, E_{a_2}, E_{a_3}, E_{a_4}]$ and pre-exponential factors $\mathbf{A} = [A_1, A_2, A_3, A_4]$ for all regions.

Methods

Preliminaries

In this work, we will be using state-space models (SSMs) which are suitable for dynamical systems. SSMs are constructed of two equations: one that describes the evolution of unobserved states in a system (hidden states), and another which describes how our collected data probabilistically relates to the hidden states. In our case, the coverage over time $\boldsymbol{\theta} = \boldsymbol{\theta}_{\tau_0:\tau_N}$ represents the hidden states, and the time-series of the areas $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}_{\tau_0:\tau_N}$ are the collected data. The evolution of the coverage $\boldsymbol{\theta}$ is governed by unknown rate constants, which we denote with the vector $\mathbf{k}$. Our goal is to estimate the hidden states $\boldsymbol{\theta}$ and model parameters $\mathbf{k}$ given the data $\boldsymbol{\mathcal{A}}$. A complete picture of the unknowns given the data $\boldsymbol{\mathcal{A}}$ would be summarized in the joint posterior distribution $p(\boldsymbol{\theta}, \mathbf{k}|\boldsymbol{\mathcal{A}})$. This entails probable values of the parameters, point estimates, and uncertainty quantification. In practice however, this posterior is rarely analytically tractable, and one can turn to MC techniques for approximations. For this work, the MC techniques we use are MCMC and PF together. Next, we describe the idea behind each, and how they are suitable to our system.

Given the time-series $\boldsymbol{\mathcal{A}}$, a standard MCMC sampler to infer the unknowns $\boldsymbol{\theta}$ and $\mathbf{k}$ would be a Gibbs sampler (GS) which iteratively samples from the conditional posteriors as

follows:

$$\begin{aligned}\boldsymbol{\theta}^{(i)} &\sim p(\boldsymbol{\theta}|\mathbf{k}^{(i-1)}, \boldsymbol{\mathcal{A}}),\\ \mathbf{k}^{(i)} &\sim p(\mathbf{k}|\boldsymbol{\theta}^{(i)}, \boldsymbol{\mathcal{A}}),\end{aligned} \tag{4}$$

for $i = 1, \ldots, I$. The distribution $p(\boldsymbol{\theta}|\mathbf{k}, \boldsymbol{\mathcal{A}})$ assumes that the parameter $\mathbf{k}$ is known, and similarly, the distribution $p(\mathbf{k}|\boldsymbol{\theta}, \boldsymbol{\mathcal{A}})$ assumes that the entire trajectory of the coverage $\boldsymbol{\theta}$ is known. With enough iterations I, the samples converge to samples of the joint posterior distribution $p(\boldsymbol{\theta}, \mathbf{k}|\boldsymbol{\mathcal{A}})$, and the resulting chain of samples $\{\boldsymbol{\theta}^{(i)}, \mathbf{k}^{(i)}\}_{i=1}^{I}$ is called a Markov chain. In cases where direct sampling from the conditional posteriors in eq. (4) is also not available, one can resort to PF[42] for sampling hidden states $\boldsymbol{\theta}$, and Metropolis-Hastings $(\text{MH})^2$ for sampling model parameters $\mathbf{k}$.

Particle Filtering PF is a numerical technique for Bayesian inference used to track a changing system. Here, the term particle stands for a numerical sample or possible value for the hidden states, i.e., the coverages $\boldsymbol{\theta}$. The workflow per time instant is based on two main steps: i) propose particles, and ii) compute their importance. The particles are generated from a proposal probability distribution that typically admits information physically meaningful of the system. For our system, this will be information from the ODE hypothesized to describe the adsorption and desorption process. Then, the "goodness" of the proposed particles is evaluated using the measured data $\boldsymbol{\mathcal{A}}$ and law of probability, resulting in an importance weight $w_{\tau_n}^{(m)}$ for each particle $\theta_{\tau_n}^{(m)}$, for $m = 1, \ldots, M$. Each weight reflects the probability that the corresponding particle is the best guess for the true coverage . The final estimate per time step, $\hat{\theta}_{\tau_n}$ is a weighted average of all the proposed particles:

$$\hat{\theta}_{\tau_n} = \sum_{m=1} w_{\tau_n}^{(m)} \theta_{\tau_n}^{(m)}.$$

In the Bayesian framework, the obtained set of particles and weights $\{\theta_{\tau_n}^{(m)}, w_{\tau_n}^{(m)}\}_{m=1}^{M}$ is a discrete approximation of the conditional posterior $p(\theta_{\tau_n}|\mathbf{k}, \boldsymbol{\mathcal{A}}_{\tau_0:\tau_n})$. This approximated pos-

terior together with the ODE will serve as a prior (proposal distribution) in the next time step τ_{n+1}. Then, the newly proposed particles $\{\theta_{\tau_{n+1}}\}_{m=1}^{M}$ are "filtered" using the likelihood $p(\mathcal{A}_{\tau_{n+1}}|\theta_{\tau_{n+1}}^{(m)}, \mathbf{k})$, which answers "how well can the value $\theta_{\tau_{n+1}}^{(m)}$ reproduce the observed data $\mathcal{A}_{\tau_{n+1}}$?". And thus, we obtain our new set of particles and weights $\{\theta_{\tau_{n+1}}^{(m)}, w_{\tau_{n+1}}^{(m)}\}_{m=1}^{M}$. Note that we are specifically describing the bootstrap PF in which the weights simplify to evaluation of the likelihood, however, there are many different types of PF (see chapter 5 in[43]). In a later section, we provide the details of the probabilistic modeling, and explicit implementation steps for PF in the SI as well as code available at `https://github.com/marija-iloska/pmcmc_CO_Pd`.

Metropolis-Hastings MH is a MC algorithm used to learn the static (non-varying) model parameters, which in our example are the rate constants $\mathbf{k}$. In essence, a sample is proposed from a proposal distribution, and it is accepted or rejected as part of the Markov chain based on some metric of the data. Let $*$ denote a newly proposed sample, and (old) denote the most recent accepted sample. The proposed sample $\mathbf{k}^*$ is generated from a proposal $q(\mathbf{k}|\mathbf{k}^{(old)})$ that depends on the last accepted sample $\mathbf{k}^{(old)}$. Then, it is accepted or rejected in the chain with acceptance probability $\tilde{\alpha} = min\{1, \frac{p(\mathbf{k}^*|\boldsymbol{\theta},\boldsymbol{\mathcal{A}})q(\mathbf{k}^{(old)}|\mathbf{k}^*)}{q(\mathbf{k}^*|\mathbf{k}^{(old)})p(\mathbf{k}^{(old)}|\boldsymbol{\theta},\boldsymbol{\mathcal{A}})}\}$, derived from the detailed balance equation.[2] In our problem, we opt to incorporate a single iteration of the MH algorithm as a sampling step within a GS, known as the Metropolis-within-Gibbs sampler.

Ultimately, the proposed algorithm is a pMCMC[29] that obtains samples of the coverages $\boldsymbol{\theta}$ with PF and of the rate constants $\mathbf{k}$ with MH as:

$$\boldsymbol{\theta}^{(i)} \sim p_{PF}(\boldsymbol{\theta}|\mathbf{k}^{(i-1)}, \boldsymbol{\mathcal{A}}),$$
$$\mathbf{k}^{(i)} \sim p_{MH}(\mathbf{k}|\boldsymbol{\theta}^{(i)}, \boldsymbol{\mathcal{A}}),$$

for $i = 1, \ldots, I$. It is named pMCMC to specify the use of PF to sample one or more of the unknowns. As standard in MCMC algorithms, a burn-in period I_0 is applied resulting the

chain of samples $\{\boldsymbol{\theta}^{(i)}, \mathbf{k}^{(i)}\}_{i=I_0+1}^{I}$.

Probabilistic Modeling

The goal in our modeling is to find distributions that will reflect the physical reality of the system. The most popular and practical distribution is the Gaussian distribution defined as $\mathcal{N}(x|\mu, \sigma^2)$ for some arbitrary random variable x, mean parameter μ and variance σ^2. However, it is unbounded $-\infty < x < \infty$ and not suitable in all cases. Another particularly useful distribution is the Beta distribution, which is defined as $\mathcal{B}(x|\alpha, \beta)$ with shape parameters α and β. The Beta bounds the support of x to $0 \leq x \leq 1$, and it has mean (expected value) $\mathbb{E}[x] = \frac{\alpha}{\alpha+\beta}$. The shape of the distribution can be used to incorporate information from our system. Following the constraint 1) for the coverage in the problem description, we propose the following SSM:

$$\begin{aligned} 2\theta_{\tau_n} &\sim \mathcal{B}(2\theta_{\tau_n}|\alpha, \beta_{\tau_n}), \\ \mathcal{A}_{\tau_n} &\sim \mathcal{N}(\mathcal{A}_{\tau_n}|\epsilon_{\tau_n}\theta_{\tau_n}, \sigma^2), \quad \tau_n \in R_r. \end{aligned} \tag{5}$$

By choosing to model $x = 2\theta_{\tau_n}$, we enforce the constraint $0 \leq \theta_{\tau_n} \leq 0.5$. Then, to incorporate information from the ODE in our model, we choose the mean $\mathbb{E}[2\theta_{\tau_n}]$ to be $2\mu_{\tau_n}$, where μ_{τ_n} follows the kinetic model in equation (3), i.e.,

$$\mu_{\tau_n} = (1 - k_{Dr}\Delta\tau)\theta_{\tau_{n-1}} + k_{Ar}\Delta\tau P(\theta^{MAX} - \theta_{\tau_{n-1}}).$$

We can ensure that the distribution $\mathcal{B}(2\theta_{\tau_n}|\alpha, \beta_{\tau_n})$ has mean $2\mu_{\tau_n}$ by keeping α fixed and setting $\beta_{\tau_n} = \alpha\frac{1-2\mu_{\tau_n}}{2\mu_{\tau_n}}$. In the second equation, the constant ϵ_{τ_n} is the molar absorptivity, which for practical purposes we fix to prior obtained estimations in[28] that use information from the location of the peak in the IR spectra. We present the details and values for obtaining ϵ_{τ_n} in the SI. Finally, σ^2 represents the measurement noise variance, which is also unknown.

Mathematical Constraints In regions R_3 and R_4, where $P = 0$ and we have desorption only, we know that $\theta_{\tau_n} < \theta_{\tau_{n-1}}$, and thus we have the following constraints on the desorption rate parameters:

$$0 < 1 - k_{Dr}\Delta\tau = \frac{\theta_{\tau_n}}{\theta_{\tau_{n-1}}} < 1 \quad \rightarrow \quad 0 < 1 - k_{Dr}\Delta\tau < 1 \quad \rightarrow \quad 0 < k_{Dr}\Delta\tau < 1. \tag{6}$$

In regions R_1 and R_2, we find a lower limit LL for the adsorption parameters conditional on the desoprtion ones:

$$\begin{aligned} &k_{Ar}\Delta\tau P(\theta^{MAX} - \theta_{\tau_n}) - k_{Dr}\Delta\tau\theta_{\tau_n} > 0, \\ &k_{Ar} > k_{Dr}\frac{\theta^{MAX_r}}{P(\theta^{MAX} - \theta^{MAX_r})} = LL_r, \end{aligned} \tag{7}$$

where θ^{MAX_r} is the highest coverage in region R_r.

Particle Markov chain Monte Carlo sampler

The objective here is to construct a pMCMC sampler to infer the hidden coverage evolution $\boldsymbol{\theta}$, the rate constant parameters $\mathbf{k}$, and the noise measurement σ^2 given the time-series area data $\boldsymbol{\mathcal{A}}$. The global structure of the pMCMC algorithm takes the form presented as Algorithm 1 in Fig. 2. Next, we describe in detail the sampling strategies that we propose for each parameter.

Sampling $\boldsymbol{\theta}$ As described earlier, we use the bootstrap PF to sample the hidden states $\boldsymbol{\theta}$ from the distribution $p_{PF}(\boldsymbol{\theta}|\mathbf{k}, \sigma^2, \boldsymbol{\mathcal{A}}) \approx p(\boldsymbol{\theta}|\mathbf{k}, \sigma^2, \boldsymbol{\mathcal{A}})$. We choose our proposal to be the transition probability given in the SSM in equation (5) with the Beta distribution. Specifically, we propose the particles by sampling $2\theta_{\tau_n}^{(m)} \sim \mathcal{B}(2\theta_{\tau_n}|\alpha, \beta_{\tau_n})$, for $m = 1, \ldots, M$. Then, we compute the importance weights by evaluating likelihood i.e., $w_{\tau_n}^{(m)} = \mathcal{N}(\mathcal{A}_{\tau_n}|\epsilon_{\tau_n}\theta_{\tau_n}^{(m)}, \sigma^2)$, creating the posterior set of samples $\{\boldsymbol{\theta}_{\tau_n}^{(m)}, w_{\tau_n}^{(m)}\}_{m=1}^{M}$. At the last time step τ_N we obtain

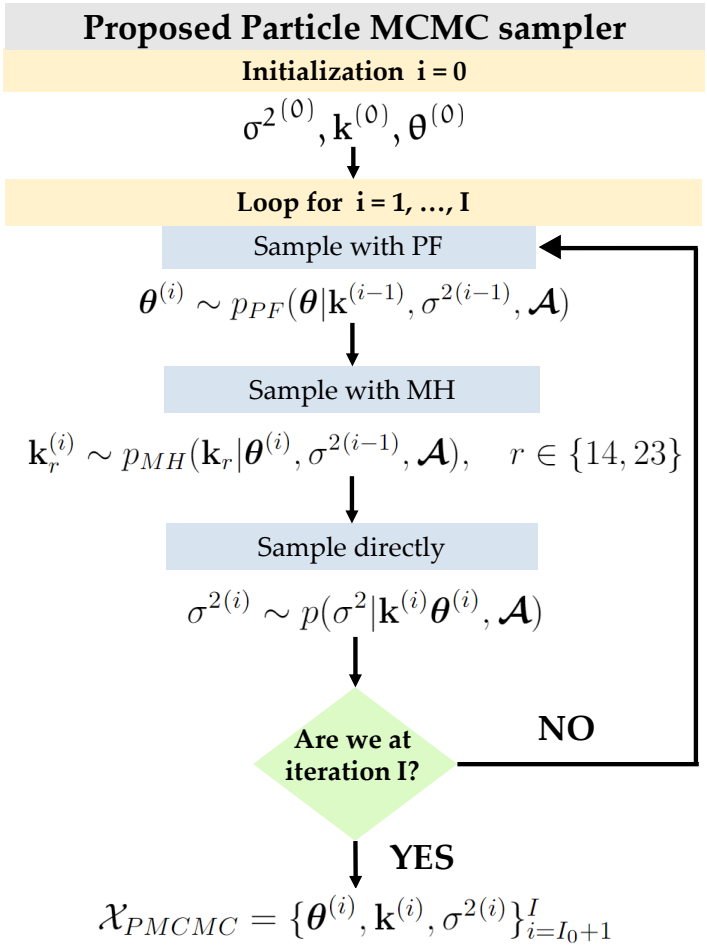


Figure 2: Flowchart of Algorithm 1. Proposed Particle MCMC sampler.

samples of the entire trajectory of the coverage $\{\boldsymbol{\theta}_{\tau_0:\tau_N}^{(m)}\}_{m=1}^{M}$ from which we uniformly sample one for our PMCMC iteration i, and discard the rest.

Sampling k Let us denote $\mathbf{k}_{14}$ and $\mathbf{k}_{23}$ as the rate constants associated with regions R_{14} and R_{23}, respectively. Since R_{14} is independent from R_{23} in terms of model parameters, we can rewrite the conditional posterior of $\mathbf{k}$ as a product of the regions:

$$p(\mathbf{k}|\boldsymbol{\theta}, \sigma^2, \boldsymbol{\mathcal{A}}) = p(\mathbf{k}_{14}|\boldsymbol{\theta}, \sigma^2, \boldsymbol{\mathcal{A}})p(\mathbf{k}_{23}|\boldsymbol{\theta}, \sigma^2, \boldsymbol{\mathcal{A}}). \tag{8}$$

We will focus on deriving the sampling strategy for regions R_{14} only, as sampling for regions

R_{23} follows analogous steps. Then, the conditional posterior of the model parameters for region R_{14} can be expressed as:

$$\begin{aligned} & p(\mathbf{k}_{14}|\boldsymbol{\theta}, \sigma^2, \boldsymbol{\mathcal{A}}) \\ & \quad \propto p(\boldsymbol{\theta}_{14}|\mathbf{k}_{14}) p(\mathbf{k}_{14}) \\ & \quad = p(k_{A1}|k_{D4}) p(k_{D4}) \prod_{\tau_n \in R_{14}} p(\theta_{\tau_n}|\theta_{\tau_{n-1}}, \mathbf{k}_{14}), \end{aligned} \tag{9}$$

where the likelihood now is a product of beta distributions given in equation (5). To carry out the sampling step above, we layout the details of the MH algorithm. Let (new) denote new accepted samples. We choose proposals whose means are the previous sample. Given the constraints of the domains in (6) and (7), we assign a Beta prior $p(k_{D4}\Delta\tau) = \mathcal{B}(k_{D4}\Delta\tau|\alpha_4, \beta_4)$ for desorption and a Gamma prior $p(z_1) = \mathcal{G}(z_1|\alpha_1, \beta_1)$, with the transformation $k_{A1} = LL_1 + z_1$ for adsorption. For reference, the Gamma distribution for an arbitrary random variable x is defined as $\mathcal{G}(x|\alpha, \beta)$ with shape parameter α and scale parameter β. The mean is $\mathbb{E}[x] = \alpha\beta$ and the support is semi-bounded, i.e., $x > 0$. By adding the lower bound LL_1 to the Gamma sample z_1, we enforce the mathematical constraint to k_{A1}. To preserve continuity in the regions, we must sample k_{D4} and k_{A1} jointly. Thus, we first sample k_{D4} independently, and k_{A1} conditional on k_{D4}. Then, the proposed samples are jointly accepted or rejected with acceptance probability, computed according to:

$$\tilde{\alpha}_{14} = \frac{\mathcal{B}(k_{D4}^{(old)}\Delta\tau|\alpha_4, \beta_4^*) \prod_{\tau \in R_{14}} p(\theta_{\tau_n}|\theta_{\tau_{n-1}}, \mathbf{k}_{14}^*),}{\mathcal{B}(k_{D4}^*\Delta\tau|\alpha_4, \beta_4^{(old)}) \prod_{\tau \in R_{14}} p(\theta_{\tau_n}|\theta_{\tau_{n-1}}, \mathbf{k}_{14}^{(old)})}, \tag{10}$$

and the parameters $\beta_4^{(old)}$ and β_4^* are defined as $\beta_4 = \alpha_4 \frac{1-k_{D4}}{k_{D4}}$, evaluating $k_{D4}^{(old)}$ and k_{D4}^*, respectively. The step-by-step implementation of the MH algorithm is provided in the SI.

Sampling σ^2 The noise variance is the simplest to obtain as we choose the inverse Gamma distribution as prior $p(\sigma^2) = \mathcal{IG}(\sigma^2|\alpha_0, \beta_0)$, which is conjugate to the likelihood $p(\boldsymbol{\mathcal{A}}|\boldsymbol{\theta}, \sigma^2) = \prod_{\tau_n \in R} \mathcal{N}(\mathcal{A}_{\tau_n}|\epsilon_{\tau_n}\theta_{\tau_n}, \sigma^2)$. That means we can directly sample from the exact

posterior distribution

$$\begin{aligned} p(\sigma^2|\mathbf{k},\boldsymbol{\theta},\boldsymbol{\mathcal{A}}) &\propto p(\boldsymbol{\mathcal{A}}|\boldsymbol{\theta},\sigma^2)p(\sigma^2) \\ &\propto \mathcal{IG}(\sigma^2|\alpha_{\sigma^2},\beta_{\sigma^2}), \end{aligned} \tag{11}$$

with posterior parameters $\alpha_{\sigma^2} = \alpha_0 + \frac{N}{2}$, and $\beta_{\sigma^2} = \beta_0 + \frac{1}{2}\sum_{\tau_n \in R}(\mathcal{A}_{\tau_n} - \epsilon_{\tau_n}\theta_{\tau_n})^2$. For convenience, we have included in the SI a list of the distributions and their characteristics, along with some helpful notes and definitions.

Bayesian Estimation of the Activation Energy and Pre-exponential Factor

Having available rate constants $\mathbf{k}$ of a system for d_T number of temperatures T, the activation energy E_a and pre-exponential factor A can be obtained through the Arrhenius equation:[44]

$$k = Ae^{-\frac{E_a}{\mathcal{R}T}} \quad \rightarrow \quad lnk = -E_a\frac{1}{\mathcal{R}T} + lnA \tag{12}$$

where $\mathcal{R}$ is the ideal gas constant. Let the notation $k_{r,t}$ indicate the rate constant for region R_r and temperature T_t, and $\mathbf{k}_{r,T} = \{k_{r,t}\}_{t=1}^{d_T}$. The standard deterministic approach is to do ordinary least-squares[45] to fit the line and obtain E_{a_r} and lnA_r as the slope and intercept of that line. The solution to least squares is analytically tractable, so we can express our variables as functions of the rate constants:

$$\begin{aligned} E_{a_r} &= \frac{\sum_{t=1}^{d_T} lnk_{r,t}h_t}{\sum_{t=1}^{d_T} lnk_{r,t}^2} = f_E(\mathbf{k}_{r,T}), \\ lnA_r &= \frac{\sum_{t=1}^{t_F} lnk_{r,t}\sum_{t=1}^{d_T} h_t^2 - \sum_{t=1}^{d_T} h_t \sum_{t=1}^{d_T} lnk_{r,t}h_t}{t_F\sum_{t=1}^{d_T} h_t^2 - (\sum_{t=1}^{d_T} h_t)^2} = f_A(\mathbf{k}_{r,T}), \end{aligned}$$

where $h_t = -\frac{1}{\mathcal{R}T_t}$. The simplest method would be obtaining an estimate of the rate constants by taking the mean of the chain samples as $\hat{\mathbf{k}}_{r,T} = \frac{1}{I-I0}\sum_{i=I0+1}^{I}\mathbf{k}_{r,T}^{(i)}$, to get the single point

estimates $\widehat{E}_{a_r} = f_E(\hat{\mathbf{k}}_{r,T})$ and $\widehat{lnA}_r = f_A(\hat{\mathbf{k}}_{r,T})$. To make use of the MCMC samples for uncertainty quantification, we will consider two Bayesian approaches.

Approach 1

MC techniques are based on the fundamental MC approximation of expected values. For an arbitrary random variable x with distribution $p(x)$ and expected value $\mathbb{E}_p[x] = \int xp(x)dx$, the integral is approximated as:

$$\mathbb{E}_p[x] \approx \frac{1}{S}\sum_{s=1}^{S} x^{(s)}, \tag{13}$$

where $\{x^{(s)}\}_{s=1}^{S}$ are samples from $p(x)$. The approximation generalizes to any arbitrary function $h(\cdot)$ as $\mathbb{E}_p[h(x)] \approx \frac{1}{S}\sum_{s=1}^{S} h(x^{(s)})$ and will be useful in this section.

Let us denote $\pi(\mathbf{k}_{r,T}) = \prod_{t=1}^{d_T} p(\mathbf{k}_{r,t}|\boldsymbol{\theta}, \sigma^2, \boldsymbol{\mathcal{A}})$, for $r \in R$. By evaluating $f_E(\cdot)$ and $f_A(\cdot)$ for every sample $\mathbf{k}_{r,T}^{(i)}$ we implicitly obtain empirical distributions for E_{a_r} and lnA_r. Then, using MC approximation, we can compute quantities of interest, such as the mean and variance as expectations with respect to the posterior $\pi(\mathbf{k}_{r,T})$. Here, we approximate μ_{E_r}, the mean of E_{a_r} as

$$\begin{aligned} \mu_{E_r} &= \mathbb{E}_\pi[f_E(\mathbf{k}_{r,T})] \\ &\approx \frac{1}{I - I_0}\sum_{i=I_0+1}^{I} f_E(\mathbf{k}_{r,T}^{(i)}), \end{aligned} \tag{14}$$

and the variance $\sigma_{E_r}^2$ as

$$\begin{aligned} \sigma_{E_r}^2 &= Var[f_E(\mathbf{k}_{r,T})] \\ &\approx \frac{1}{I - I_0}\sum_{i=I_0+1}^{I} f_E(\mathbf{k}_{r,T}^{(i)})^2 - \mu_{E_r}^2, \end{aligned} \tag{15}$$

with $\mathbf{k}_{r,T}^{(i)} \sim \pi(\mathbf{k}_{r,T})$, for $i = I_0 + 1, \ldots, I$. An analogous procedure allows to obtain μ_{A_r} and

$\sigma^2_{A_r}$, the mean and variance of lnA_r:

$$\mu_{A_r} \approx \frac{1}{I - I_0} \sum_{i=I_0+1}^{I} f_A(\mathbf{k}^{(i)}_{r,T}), \qquad \sigma^2_{A_r} \approx \frac{1}{I - I_0} \sum_{i=I_0+1}^{I} f_A(\mathbf{k}^{(i)}_{r,T})^2 - \mu^2_{A_r}. \tag{16}$$

Approach 2

Since the probabilistic assumption of regression for the linear model in equation (12) is that $\mathbb{E}[lnk] = 0$, one concern with Approach 1 is the domain of the variables. We propose a fully Bayesian approach which takes into consideration the domain constraints. The goal is to find the posterior distributions $p(E_{a_r}|\mathbf{k}_{r,T})$ and $p(lnA_r|\mathbf{k}_{r,T})$. Given that we have a chain of samples, we can obtain $I - I_0$ local posteriors $\left\{p(E_{a_r}|\mathbf{k}^{(i)}_{r,T}), p(lnA_r|\mathbf{k}^{(i)}_{r,T})\right\}_{i=I_0+1}^{I}$, and fuse their information to obtain the global posterior distributions, respectively.

First, we propose likelihood models that respect the mathematical constraints for lnk. Specifically, for the regions $r \in \{1, 2\}$, we propose:

$$\begin{aligned} lnk_{r,t} &\sim \mathcal{G}(lnk_{r,t}|a_{k_{r,t}}, b_k), \\ a_{k_{r,t}} &= E_{a_r} h_t + lnA_r, \end{aligned} \tag{17}$$

and for the regions $r \in \{3, 4\}$:

$$\begin{aligned} -lnk_{r,t} &\sim \mathcal{G}(-lnk_{r,t}|a_{k_{r,t}}, b_k), \\ a_{k_{r,t}} &= -E_{a_r} h_t - lnA_r, \end{aligned} \tag{18}$$

with scale parameter $b_k = 1$. In this way, the means of our distributions follow the Arrhenius equation. Next, we have domain knowledge that $E_{a_r} > 0$ and $lnA_r > 0$, so we propose the respective priors $p(E_{a_r}) = \mathcal{G}(E_{a_r}|a_{E_0}, b_{E_0})$ and $p(lnA_r) = \mathcal{G}(lnA_r|a_{A_0}, b_{A_0})$, $\forall r$. With some algebraic manipulations, the posterior distributions for the activation energies can be

expressed as:

$$\begin{aligned} p(E_{a_r}|\mathbf{k}_{r,T}^{(i)}) &\propto p(\mathbf{k}_{r,T}^{(i)}|E_{a_r})p(E_{a_r}) \\ &= \mathcal{G}(E_{a_r}|a_{E_0}, b_{E_0}) \prod_{t=1}^{d_T} \mathcal{G}(lnk_{r,t}^{(i)}|a_{k_{r,t}}, b_k) \\ &\propto \mathcal{G}(E_{a_r}|a_{E_0}, b_{E_r}). \end{aligned} \tag{19}$$

Similarly, we obtain the posterior distribution for lnA_r, $\mathcal{G}(lnA_r|a_{A_0}, b_{A_r})$. Let us define $g_{r,t} = lnk_{r,t}$. Explicitly, the posterior parameters we obtained are:

$$b_{E_r} = \left(\frac{1}{b_{E_0}} + \sum_{t=1}^{d_T} \frac{ln(g_{r,t})}{\mathcal{R}T_t}\right)^{-1} \qquad b_{A_r} = \left(\frac{1}{b_{A_0}} - \sum_{t=1}^{d_T} ln(g_{r,t})\right)^{-1} \qquad r \in \{1, 2\} \tag{20}$$

$$b_{E_r} = \left(\frac{1}{b_{E_0}} - \sum_{t=1}^{d_T} \frac{ln(-g_{r,t})}{\mathcal{R}T_t}\right)^{-1} \qquad b_{A_r} = \left(\frac{1}{b_{A_0}} + \sum_{t=1}^{d_T} ln(-g_{r,t})\right)^{-1} \qquad r \in \{3, 4\} \tag{21}$$

Finally, we propose to use linear pooling[46] to fuse the local posteriors, and represent our global posterior as a mixture distribution:

$$p(E_{a_r}|\mathbf{k}_{r,T}) = \sum_{i=I_0+1}^{I} \frac{1}{I - I_0} p(E_{a_r}|\mathbf{k}_{r,T}^{(i)}),$$

where the weights $\frac{1}{I-I_0}$ imply that we sample from any ith distribution with equal probability. The assumption for linear pooling is that our sources are at most weakly correlated. MCMC samples have autocorrelation which can be weakened by using a thining-rate, i.e., skipping every few samples when preserving the final Markov chain. Then, we can compute $\tilde{\mu}_{E_r}$, the mean of E_{a_r} as an expectation with respect to its own posterior

$$\tilde{\mu}_{E_r} = \int E_{a_r} p(E_{a_r}|\mathbf{k}_{r,T}) dE_{a_r} \approx \frac{1}{L} \sum_{\ell=1}^{L} E_{a_r}^{(\ell)}, \tag{22}$$

and the variance $\tilde{\sigma^2}_{E_r}$ as

$$\tilde{\sigma^2}_{E_r} = \mathbb{E}[E_{a_r}^2] - \tilde{\mu}_{E_r}^2 = \frac{1}{L}\sum_{\ell=1}^{L} E_{a_r}^{2\ (\ell)} - \tilde{\mu}_{E_r}^2, \tag{23}$$

with $E_{a_r}^{(\ell)} \sim p(E_{a_r}|\mathbf{k}_{r,T})$, for $\ell = 1, \ldots, L$. Analogous steps are done to obtain $\tilde{\mu}_{A_r}$ and $\tilde{\sigma^2}_{A_r}$, the mean and variance of lnA_r, respectively.

Results and discussion

We have obtained the area from collected IR time-series datasets with constant pressure $P = 10^{-3}\ Pa$ at temperatures $T \in \{450, 460, 470, 475, 480, 490\}K$, and scanning resolution of $\Delta\tau = 0.067s$. The proposed pMCMC sampler was run with $I = 7000$ iterations and burn-in $I_0 = 2000$ for each temperature. The noise variance was initialized (σ_0^2) to the sample variance of the last 5 data points of the area time-series, where we assume we have complete desorption. The prior hyperparameters were set as the following: $k_{Dr}\Delta\tau \sim \mathcal{B}(k_{Dr}\Delta\tau|3,3)$, $z_r \sim \mathcal{G}(z_r|1000,2)$, and $\sigma^2 \sim \mathcal{IG}(\sigma^2|1,\sigma_0^2)$.

In Fig. 3, we show the estimated time series of the coverages for all temperatures together, obtained as the mean of their Markov chain samples. We also show the converged Markov chains in Fig. 4a for the rate constants in all regions for $T = 460$K only. The rest are similar and provided in the SI. In Fig. 4b we show the corresponding Arrhenius equation fitting using the estimates of the rate constants $\hat{\mathbf{k}}$. A careful comparison of the scale in these plots reveals that the fitting for regions $R_{3\&4}$ is more sensitive and may indicate higher uncertainty in the measurements for low coverage regions. We presume this to be due to instrumentation and experimental errors from temperature to temperature datasets. The fittings in $R_{1\&2}$ are practically straight lines. This is in good agreement with the theory that the adsorption of CO onto Pd(111) is a non-activated process, which implies its activation energy $E_{a_1} \approx 0$.

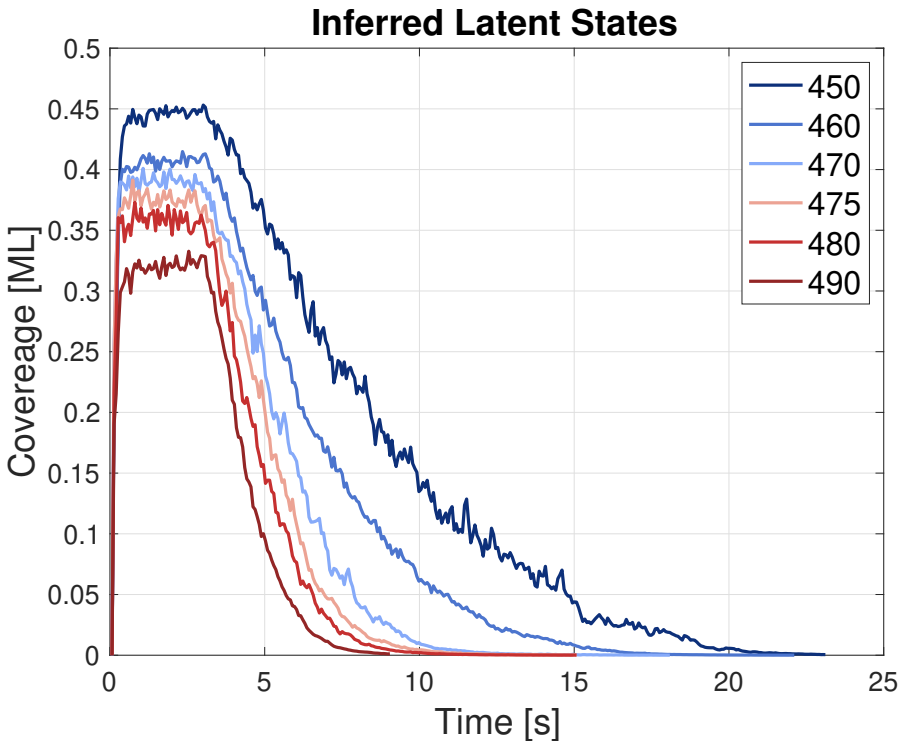
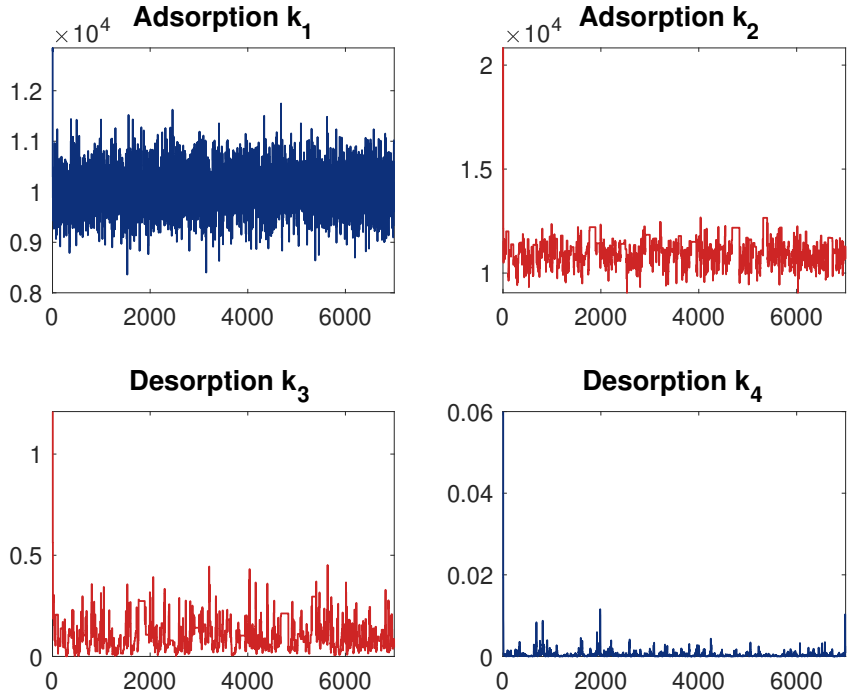


Figure 3: Estimated coverages for all temperatures.

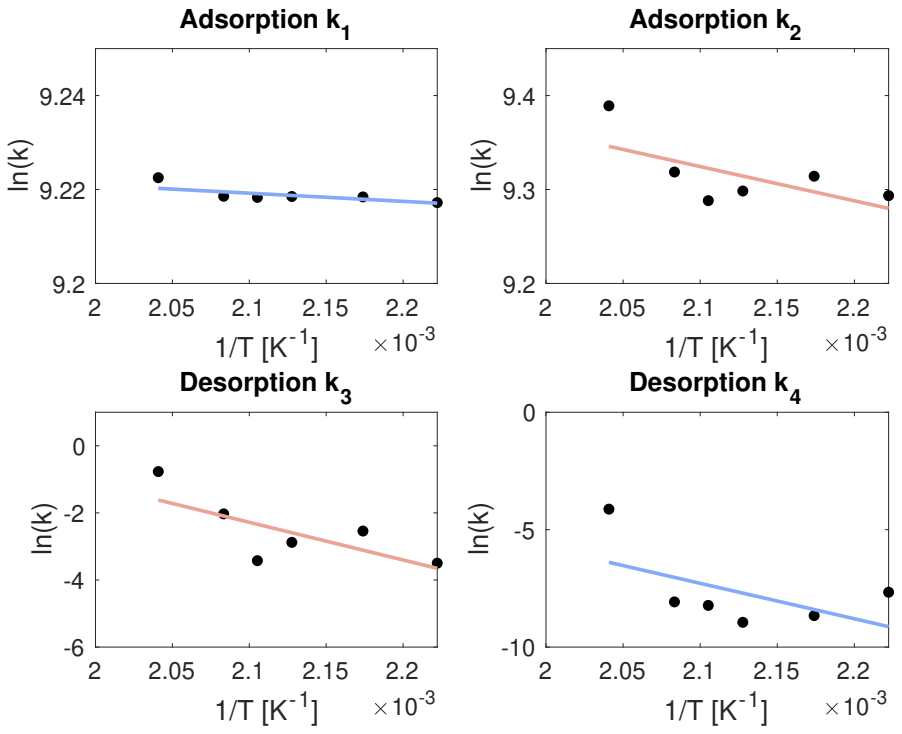
When estimating the activation energies and pre-exponential factors with Approach 2, we considered 3 cases to demonstrate the sensitivity of the priors when choosing hyperparameters: case 1) where we take the means of our priors to be the point estimates we obtained in Approach 1 and we have low variance $\sigma^2_{E_0} = \sigma^2_{A_0} = 1$; case 2) same as case 1) but with high variance of $\sigma^2_{E_0} = \sigma^2_{A_0} = 50$; and case 3) where we keep high variance of 50, but we shift the means by a factor of 2. We ensured the means and variances of our priors by moment matching, i.e.,

$$a_{E_0} = \frac{\mu^2_{E_0}}{\sigma^2_{E_0}} \quad b_{E_0} = \frac{\sigma^2_{E_0}}{\mu_{E_0}}$$

and similarly for lnA. Bayesian inference is ideal when we have prior information available. In this study, we are able to use estimates from Approach 1 (and even literature), however, we do not always have this luxury. In situations where we do not have prior information, we must rely on having enough data for quality results. We plot the empirical distributions

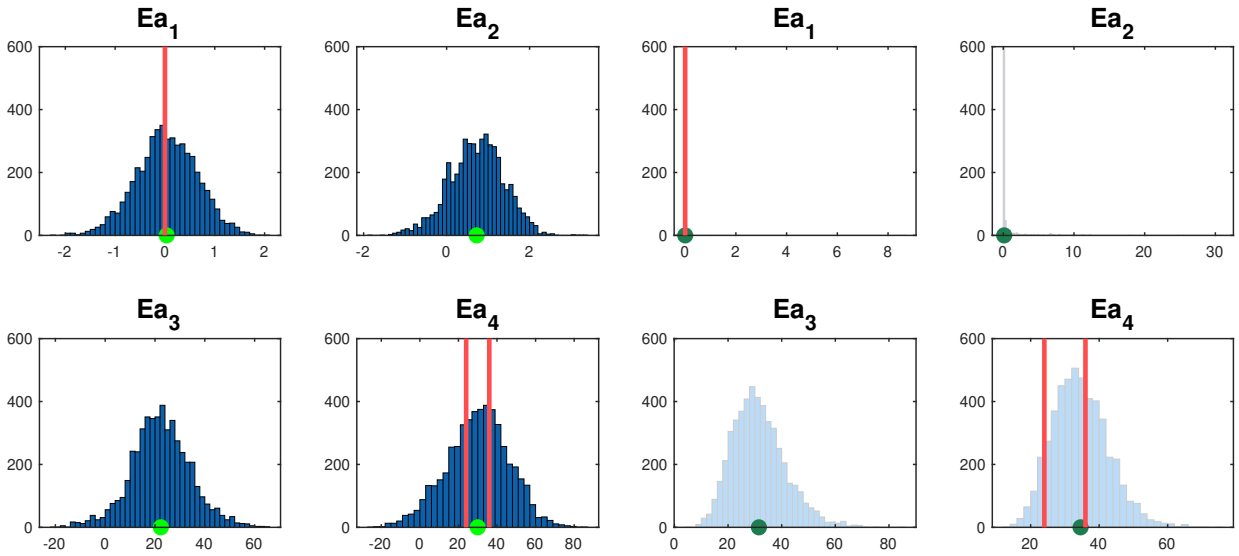


(a) Converged Markov chains for $T = 460$K.

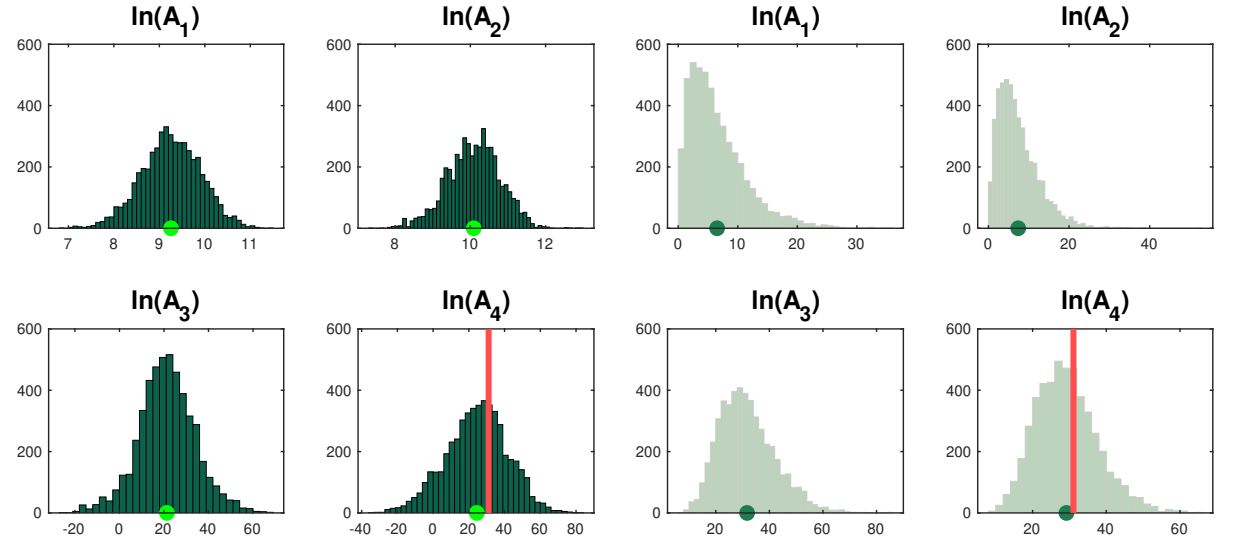


(b) Fitting of Arrhenius equation with averaged rate constants.

Figure 4: Results of the proposed pMCMC sampler. The unit of the rate constants for desorption is $[s^{-1}]$, and for for adsorption $[(s \cdot Pa)^{-1}]$.



(a) Histogram results of the activation energy for each region. Units $[kcal/mol]$.



(b) Histogram results of the pre-exponential factor for each region. Units correspond to those of the rate constants.

Figure 5: Estimation of activation energy and pre-exponential factor. The dark histograms (left) result from Approach 1. The light histograms (right) result from Approach 2. The green dot in each histogram is the mean of the samples corresponding to equation (16) and equation (22), respectively. The red frames indicate ranges reported in literature.[33–35] We would like to point out the difference in scales of the x-axes when comparing the range and width of the histograms.

Table 1: Literature reported values and mean estimations from Approach 1 and 2.

Regions	R_1		R_2		R_2&R_3		R_1&R_4	
Process	Adsorption		Adsorption		Desorption		Desorption	
Parameters	E_a $\left[\frac{kcal}{mol}\right]$	lnA	E_a $\left[\frac{kcal}{mol}\right]$	lnA	E_a $\left[\frac{kcal}{mol}\right]$	lnA	E_a $\left[\frac{kcal}{mol}\right]$	lnA
Approach 1	0.0345	9.2557	0.7242	10.0897	22.3280	21.3154	30.0046	24.4237
Approach 2	7.323×10^{-1}	6.6505	0.1227	7.4017	31.4516	31.7077	34.4057	29.4816
Literature	0						$24 - 36$[33–35]	31.0849[33]

over the activation energies and pre-exponential factors for all regions, obtained through the pMCMC samples, with Approach 1 (left) and case 2 of Approach 2 (right) side by side in Fig. 5. The red vertical lines on top of the histograms mark values from literature. We also provide a Table 1 for numerical representation of our estimates. With Approach 1, we can see that the means of the energies in R_4 and R_1 and the log of the pre-exponential factor in R_4 seem to match with the values from literature. For R_2 and R_3 there are no reported values yet, however, the relative values obtained $\mu_{E_1} < \mu_{E_2}$ and $\mu_{E_3} < \mu_{E_4}$ would physically imply that adsorption is more difficult in R_2 (where it is more crowded), and that desorption happens faster in R_3 (again, where it is more crowded). The uncertainty is reflected visually with the width of the histograms, however, we see that the samples extend to values that are not physically plausible. One possibility is to discard the impossible values and obtain a truncated distribution with a meaningful range. With Approach 2, we have addressed the support of the variables and we observe mostly sharper distributions. The means of the distributions have shifted, but remain comparable to those from Approach 1. The reason we present case 2) is to use a slightly weaker prior, and allow more influence of the data. However, the number of our data points ($d_T = 6$) is too low, and we cannot obtain good posteriors without a moderately strong prior (i.e., in this case, having some knowledge on the prior means). We presented the results of case 1) and case 3) in the SI.

Uncertainty and Modeling A challenging aspect when working on applications with real data is accounting for (and acknowledging) all the sources and layers of uncertainty. In the problem considered, we quantified aleatoric uncertainty due to the inherent randomness

of the system by proposing the probabilistic model in equation (5). However, aleatoric uncertainty due to the inevitable experimental challenges (clean Pd(111) sample, instrumental precision, system perturbations) remains, and can only be reduced by repeating the experiment for a single temperature several times. The sources of epistemic uncertainty come from parameter estimation and model simplification. In our study with the proposed sampler, we can quantify the epistemic uncertainty due to rate constant parameter estimation using our pMCMC posteriors given our SSM and probabilistic models of the rest of the parameters. However, the choice of our models propagates unclear layers of uncertainty, especially in regions R_{23}, where we are only modeling the net outcome. A complete model that explicitly represents the details between the types of occupied sites in R_{23} requires careful thought. We would need to consider both finding representative equations that can capture the conversions between hollow and bridge adsorption sites, and assigning sensible priors on the rate constant parameters involved. This becomes a problem of model selection, where one can explore the presence of a process, for example, hollow to bridge conversion or the reverse. While it may seem straightforward to model hollow and bridge sites as individual species in the SSM, depending on the model proposed and experimental data available, one may run into the model non-identifiability issue.[47,48] This implies that there is no unique solution to the problem, i.e., different sets of parameters can produce the same observed data. In MCMC, the symptoms of model non-identifiability are high sensitivity to priors, parameters converging to different values in different runs, and/or complete lack of convergence.[19] We provide details of our explicit models considered for bridge and hollow in regions $R_{2,3}$ in the SI. Obtaining parameter constraints for the parameters characterizing these explicit models proved challenging. Additionally, the parameters must be sampled jointly because of the dependencies between them, and even one sampled out of bounds can lead to either low acceptance probability in MH, misleading results, or numerical errors in the PF. To address model non-identifiability, one must impose enough constraints by either reformulating the model, running experiments to collect additional and likely different type of data, or mak-

ing strong assumptions.[47] For these reasons, we resorted to modeling the net effects as a step towards the eventual goal of fully capturing the mechanistic details about the sites of adsorption.

Regarding our study of the activation energies and pre-exponential factors, we have reduced epistemic uncertainty moving from Approach 1 to 2, by formulating a model that situates our parameters in a realistic domain. When it comes to uncertainty quantification, the concern has been raised that linearization of the Arrhenius equation obscures the error distribution for E_a and lnA.[49] The proposed alternative is to perform direct nonlinear fitting, using the fitting results of the linearized version as initial guesses. We have ran nonlinear regression using the estimated rate constants from the PMCMC chains and report the details in the SI as well as the implementation on Github. The results have shown that using initial guesses from the linearized fitting can be helpful, but it does not guarantee convergence and brute-force search is still required. Further, manually searching for the best initial guesses proved challenging to find a reasonable fitting for R_4, as the optimization algorithm faces numerical overflow handling the large exponential values of A. To truly obtain the error distribution of Ea and A that will reflect the errors estimated in the rate constants from the PMCMC and the assumed model, one would need to run nonlinear regression for all the samples in the rate constants chain, which would be more expensive than Approach 1, comparable in cost to Approach 2, but sensitive to numerical overflow and to initial guesses for every individual sample from the chain. A final constraint in Approach 2 is that it is the linearized form of the Arrhenius equation that allows for conjugacy in the probabilistic modeling, i.e., possibility to directly sample from the posterior distributions of E_a and lnA. The benefit of our Bayesian approach is that the uncertainty quantification and the estimated values reflect the uncertainty from the experiments, uncertainty from the models assumed, as well as uncertainty propagated from parameter to parameter.

A final challenge in estimation for both Approaches 1 and 2 comes from having limited number of temperature datasets (only $d_T = 6$). Having a much larger d_T would significantly

improve the fittings in Fig. 4b, and reduce the sensitivity to the priors in Approach 2.

Cost and Limitations Like any MCMC method, pMCMC methods can be expensive as they require thousands of iterations to ensure convergence. Informally, we report as a semi-quantitive reference that the proposed PMCMC for all 6 temperatures with $I = 7000$ iterations and $M = 40$ particles took ~ 17 mins in total to run (without any parallelization). Like in many cases, in the present problem 7000 iterations were run out of caution, but even 4000 would suffice. For additional reference, we report that running with $I = 4000$ took ~ 10 mins. However, these quantities depend on the power of the computer as well as the efficiency in the implementation of the code. More formally, for a system with d_θ hidden states, M particles, and τ_N time steps, the cost of PF scales as $O(d_\theta M \tau_N)$. The cost of an MH step is dominated by the computation of the acceptance probability, which requires an evaluation of a product of likelihoods, in our case, $O(\tau_N)$ for each temperature. Then, the overall cost of PMCMC with one PF and two MH steps becomes $O(I \times (d_\theta M \tau_N + d_p \tau_N))$, where d_p is the number of MH steps required. One of the limitations of PF is systems with high number of states d_θ (e.g., $d_\theta > 20$). For such high dimensional systems with many species, one can use multiple PF (MPF), which partitions the state vector into K partitions of size d_k and tracks each with M_k particles. A benefit of MPF is also a reduction in cost to an impressive $O(\tau_N \sum_{k=1}^{K} M_k d_k)$ and parallel implementation. We refer the interested reader to[50] for details.

An additional limitation of PF is the possibility of unknown static parameters (in our case, the rate constants $\mathbf{k}$. If the values used for $\mathbf{k}$ when running the PF are unrealistic in a system that is mathematically bounded, then the filtering may result in either poor estimates or numerical errors. For this reason, it is especially important in the PMCMC to ensure that the MH step produces meaningful samples. In the MH algorithm, constructing an inadequate proposal distribution can generate samples from spaces of low probability, leading to high rejection rates, slow convergence, or even no convergence. We can best

leverage their power for systems where we are able to design meaningful proposals using mathematical constraints and domain knowledge.

The benefit of using pMCMC approaches for chemical systems is that they provide an interpretable and complete description of the system, including static and dynamic parameters with uncertainty quantification. This makes them suitable to evaluate and advance what we know about kinetic systems, simultaneously learning the change in concentration or coverage of chemical species and the kinetic parameters from noisy time-series data.

Conclusion

In this work we proposed a novel modeling scheme for the adsorption and desorption of CO onto Pd(111) surface that divides the problem in 4 regions, and a Bayesian approach to infer its kinetic parameters from IR time-series data. Specifically, we constructed a particle Markov chain Monte Carlo method to learn the hidden process, and the rate constant parameters governing this process. We then used the resulting samples to formulate two Bayesian approaches to obtain distributions over the activation energies and pre-exponential factors of each region. The flexibility of the Bayesian framework allowed us to work in the correct domain of the variables, to quantify several layers of uncertainty, and to provide physically meaningful interpretations of the results. Our findings match with values reported in literature, specifically activation energies in the first and last region, and pre-exponential factor in the last region. The rest of the parameters have not been reported, however, the relative values seem to reflect sound physical implications.

Acknowledgement

We extend special thanks to Zubin Darbari for providing the IR data used in this work, and Talin Avensian for theoretical contribution to the problem. We further express special thanks to the reviewers whose constructive feedback helped greatly improve the clarity and

quality of the paper.

Supporting Information Available

In the SI, we provide visualization of the processed data (the area time-series), description on the estimation of the molar absorptivity, step-by-step algorithms for implementation, the results of the Markov chains of the rest of the temperature datasets, the derivations of the posterior distributions from Approach 2, and the prior sensitivity when estimating the activation energies and pre-exponential factors.

References

(1) Gillespie, D. T. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **2007**, *58*, 35–55.

(2) Chib, S.; Greenberg, E. Understanding the Metropolis-Hastings algorithm. *The American Statistician* **1995**, *49*, 327–335.

(3) Doucet, A.; De Freitas, N.; Gordon, N. J.; others *Sequential Monte Carlo methods in practice*; Springer, 2001; Vol. 1.

(4) Blei, D. M.; Kucukelbir, A.; McAuliffe, J. D. Variational inference: A review for statisticians. *Journal of the American Statistical Association* **2017**, *112*, 859–877.

(5) Van Boekel, M. On the pros and cons of Bayesian kinetic modeling in food science. *Trends in Food Science & Technology* **2020**, *99*, 181–193.

(6) Wang, J.; Zhou, Z.; Lin, K.; Law, C. K.; Yang, B. Facilitating Bayesian analysis of combustion kinetic models with artificial neural network. *Combustion and Flame* **2020**, *213*, 87–97.

(7) Baltussen, M. G.; van de Wiel, J.; Fernandez Regueiro, C. L.; Jakstaite, M.; Huck, W. T. A Bayesian approach to extracting kinetic information from artificial enzymatic networks. *Analytical Chemistry* **2022**, *94*, 7311–7318.

(8) Iapteff, L.; Jacques, J.; Celse, B.; Costa, V. Reducing the number of experimental points to fit kinetic models: A Bayesian approach. *Industrial & Engineering Chemistry Research* **2023**, *62*, 10903–10914.

(9) Galagali, N.; Marzouk, Y. M. Bayesian inference of chemical kinetic models from proposed reactions. *Chemical Engineering Science* **2015**, *123*, 170–190.

(10) Moon, T. K. The expectation-maximization algorithm. *IEEE Signal Processing Magazine* **1996**, *13*, 47–60.

(11) Matera, S.; Schneider, W. F.; Heyden, A.; Savara, A. Progress in accurate chemical kinetic modeling, simulations, and parameter estimation for heterogeneous catalysis. *Acs Catalysis* **2019**, *9*, 6624–6647.

(12) Carpenter, B.; Gelman, A.; Hoffman, M. D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M. A.; Guo, J.; Li, P.; Riddell, A. Stan: A probabilistic programming language. *Journal of Statistical Software* **2017**, *76*.

(13) Plummer, M.; others JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd international workshop on distributed statistical computing. 2003; pp 1–10.

(14) Savara, A.; Walker, E. A. CheKiPEUQ intro 1: Bayesian parameter estimation considering uncertainty or error from both experiments and theory. *ChemCatChem* **2020**, *12*, 5385–5400.

(15) Walker, E. A.; Ravisankar, K.; Savara, A. CheKiPEUQ intro 2: Harnessing uncertain-

ties from data sets, Bayesian design of experiments in chemical kinetics. *ChemCatChem* **2020**, *12*, 5401–5410.

(16) Cohen, M.; Vlachos, D. G. Chemical kinetics Bayesian inference toolbox (CKBIT). *Computer Physics Communications* **2021**, *265*, 107989.

(17) Omori, T.; Kuwatani, T.; Okamoto, A.; Hukushima, K. Bayesian inversion analysis of nonlinear dynamics in surface heterogeneous reactions. *Physical Review E* **2016**, *94*, 033305.

(18) Kalyanaraman, J.; Kawajiri, Y.; Lively, R. P.; Realff, M. J. Uncertainty quantification via Bayesian inference using sequential Monte Carlo methods for CO2 adsorption process. *AIChE Journal* **2016**, *62*, 3352–3368.

(19) Raue, A.; Kreutz, C.; Theis, F. J.; Timmer, J. Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2013**, *371*, 20110544.

(20) Kalyanaraman, J.; Fan, Y.; Labreche, Y.; Lively, R. P.; Kawajiri, Y.; Realff, M. J. Bayesian estimation of parametric uncertainties, quantification and reduction using optimal design of experiments for CO2 adsorption on amine sorbents. *Computers & chemical engineering* **2015**, *81*, 376–388.

(21) Feyzbar-Khalkhali-Nejad, F.; Hassani, E.; Rashti, A.; Oh, T.-S. Adsorption-based CO removal: Principles and materials. *Journal of Environmental Chemical Engineering* **2021**, *9*, 105317.

(22) Ozensoy, E.; Hess, C.; Goodman, D. W. Understanding the catalytic conversion of automobile exhaust emissions using model catalysts: CO+ NO reaction on Pd(111). *Topics in Catalysis* **2004**, *28*, 13–23.

(23) Somorjai, G. A.; Li, Y. *Introduction to Surface Chemistry and catalysis*; John Wiley & Sons, 2010.

(24) Conrad, H.; Ertl, G.; Küppers, J. Interactions between oxygen and carbon monoxide on a Pd(111) surface. *Surface Science* **1978**, *76*, 323–342.

(25) Fischer-Wolfarth, J.-H.; Farmer, J. A.; Flores-Camacho, J. M.; Genest, A.; Yudanov, I. V.; Rösch, N.; Campbell, C. T.; Schauermann, S.; Freund, H.-J. Particle-size dependent heats of adsorption of CO on supported Pd nanoparticles as measured with a single-crystal microcalorimeter. *Physical Review B* **2010**, *81*, 241416.

(26) Hooshmand, Z.; Le, D.; Rahman, T. S. CO adsorption on Pd (111) at 0.5 ML: A first principles study. *Surface Science* **2017**, *655*, 7–11.

(27) Martin, N.; Van den Bossche, M.; Gronbeck, H.; Hakanoglu, C.; Zhang, F.; Li, T.; Gustafson, J.; Weaver, J.; Lundgren, E. CO adsorption on clean and oxidized Pd (111). *The Journal of Physical Chemistry C* **2014**, *118*, 1118–1128.

(28) Darbari, Z.; Iloska, M.; Bugallo, M. F.; Boscoboinik, J. A. Addition of transient kinetics capabilities to an infrared reflection absorption spectroscopy system through synchronized gas pulsing and data acquisition. *Catalysis Today* **2023**, *417*, 113762.

(29) Andrieu, C.; Doucet, A.; Holenstein, R. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2010**, *72*, 269–342.

(30) Golightly, A.; Wilkinson, D. J. Bayesian parameter inference for stochastic biochemical network models using Particle Markov chain Monte Carlo. *Interface Focus* **2011**, *1*, 807–820.

(31) Koblents, E.; Míguez, J. A comparison of nonlinear population Monte Carlo and Par-

ticle Markov chain Monte Carlo algorithms for Bayesian inference in stochastic kinetic models. *arXiv preprint arXiv:1404.5218* **2014**,

(32) Golightly, A.; Henderson, D. A.; Sherlock, C. Delayed acceptance Particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing* **2015**, *25*, 1039–1055.

(33) Guo, X.; Yates Jr, J. T. Dependence of effective desorption kinetic parameters on surface coverage and adsorption temperature: CO on Pd(111). *The Journal of Chemical Physics* **1989**, *90*, 6761–6766.

(34) Kuhn, W. K.; Szanyi, J.; Goodman, D. W. CO adsorption on Pd(111): the effects of temperature and pressure. *Surface Science* **1992**, *274*, L611–L618.

(35) Engel, T.; Ertl, G. Surface residence times and reaction mechanism in the catalytic oxidation of CO on Pd(111). *Chemical Physics Letters* **1978**, *54*, 95–98.

(36) Avanesian, T.; Darbari, Z.; Iloska, M.; Boscoboinik, J. A.; Wu, Q. Investigation of coverage dependence of the stretching frequency of CO adsorbed on Pd surfaces at low coverage limits. *Surface Science* **2024**, 122534.

(37) Savara, A.; Weitz, E. Elucidation of intermediates and mechanisms in heterogeneous catalysis using infrared spectroscopy. *Annual review of physical chemistry* **2014**, *65*, 249–273.

(38) Swinehart, D. F. The beer-lambert law. *Journal of Chemical Education* **1962**, *39*, 333.

(39) Loffreda, D.; Simon, D.; Sautet, P. Dependence of stretching frequency on surface coverage and adsorbate–adsorbate interactions: a density-functional theory approach of CO on Pd(111). *Surface Science* **1999**, *425*, 68–80.

(40) Bradshaw, A.; Hoffmann, F. The chemisorption of carbon monoxide on palladium single

crystal surfaces: IR spectroscopic evidence for localised site adsorption. *Surface Science* **1978**, *72*, 513–535.

(41) Ertl, G.; Koch, J. Adsorption von CO auf einer Palladium (111)-Oberfläche. *Zeitschrift für Naturforschung A* **1970**, *25*, 1906–1912.

(42) Särkkä, S. *Bayesian filtering and smoothing*; Cambridge University Press, 2013.

(43) Adali, T.; Haykin, S. *Adaptive signal processing: next generation solutions*; John Wiley & Sons, 2010.

(44) Jensen, F. Activation energies and the Arrhenius equation. *Quality and Reliability Engineering International* **1985**, *1*, 13–17.

(45) Theodoridis, S. *Machine Learning: a Bayesian and Optimization Perspective*; Academic press, 2015.

(46) Koliander, G.; El-Laham, Y.; Djurić, P. M.; Hlawatsch, F. Fusion of probability density functions. *Proceedings of the IEEE* **2022**, *110*, 404–453.

(47) Park, S.; Himmelblau, D. Parameter estimation and unique identifiability. *The Chemical Engineering Journal* **1982**, *25*, 163–174.

(48) Cole, D. *Parameter redundancy and identifiability*; Chapman and Hall/CRC, 2020.

(49) Brauner, N.; Shacham, M. Statistical analysis of linear and nonlinear correlation of the Arrhenius equation constants. *Chemical Engineering and Processing: Process Intensification* **1997**, *36*, 243–249.

(50) Djuric, P. M.; Lu, T.; Bugallo, M. F. Multiple Particle filtering. 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07. 2007; pp III–1181.