Received: 17 June 2021 | Revised: 3 September 2021 | Accepted: 1 October 2021

DOI: 10.1002/jcc.26770

FULL PAPER

Journal of COMPUTATIONAL CHEMISTRY WILEY

Estimating reaction parameters in mechanism-enabled population balance models of nanoparticle size distributions: A Bayesian inverse problem approach

Danny K. Long[1] | Wolfgang Bangerth[1,2] | Derek R. Handwerk[3] | Christopher B. Whitehead[3,4] | Patrick D. Shipman[1] | Richard G. Finke[3]

[1]Department of Mathematics, Colorado State University, Fort Collins, Colorado, USA

[2]Department of Geosciences, Colorado State University, Fort Collins, Colorado, USA

[3]Department of Chemistry, Colorado State University, Fort Collins, Colorado, USA

[4]Department of Chemistry, University of Basel, Basel, Switzerland

Correspondence
Wolfgang Bangerth, Department of Mathematics, Colorado State University, Fort Collins, CO 80523, USA.
Email: bangerth@colostate.edu

Funding information
National Science Foundation, Grant/Award Numbers: DMS-1814941, DMS-1821210, EAR-1925595, OAC-1835673; U.S. Department of Energy, Grant/Award Number: SE-FG402-02ER15453

Abstract

In order to quantitatively predict nano- as well as other particle-size distributions, one needs to have both a mathematical model and estimates of the parameters that appear in these models. Here, we show how one can use Bayesian inversion to obtain statistical estimates for the parameters that appear in recently derived mechanism-enabled population balance models (ME-PBM) of nanoparticle growth. The Bayesian approach addresses the question of "how well do we know our parameters, along with their uncertainties?." The results reveal that Bayesian inversion statistical analysis on an example, prototype $Ir(0)_n$ nanoparticle formation system allows one to estimate not just the most likely rate constants and other parameter values, but also their SDs, confidence intervals, and other statistical information. Moreover, knowing the reliability of the mechanistic model's parameters in turn helps inform one about the reliability of the proposed mechanism, as well as the reliability of its predictions. The paper can also be seen as a tutorial with the additional goal of achieving a "Gold Standard" Bayesian inversion ME-PBM benchmark that others can use as a control to check their own use of this methodology for other systems of interest throughout nature. Overall, the results provide strong support for the hypothesis that there is substantial value in using a Bayesian inversion methodology for parameter estimation in particle formation systems.

KEYWORDS
Bayesian inversion, kinetics and mechanism, nanoparticles, nucleation and growth, particle size distribution, population balance modeling

1 | INTRODUCTION

Nanoparticles are widely used in homogeneous and heterogeneous catalysis,[1–4] as light-emitting diodes,[5–8] and in solar cells,[9–12] to mention just a few among many more important examples.[13] In nearly every case, the desired application is dependent on the particle size and particle-size distribution (PSD).[14,15] It is perhaps not surprising, then, that attempts at developing conceptual and mathematical models for nanoparticle PSDs date back at least 56 and likely more than 100 years.[16,17]

Fundamentally, control of nanoparticle size and PSDs requires two basic items: (i) an understanding of particle formation mechanisms that is, the mechanisms of nucleation, growth and agglomeration and, then, (ii) a mathematical model that allows computation of PSDs in a way that is consistent with that experimentally derived mechanistic knowledge. Presently, the first requirement is satisfied by the availability of five classes of disproof-based,[18] hence more reliable, deliberately minimum particle-formation mechanisms developed over the past 25 years expressed in composite, pseudo-elementary step

form.[19–24] (For more information on and the use of pseudo-elementary steps [PESteps], the reader is directed to Watzky and Finke,[19] Field and Noyes[25]). Little appreciated at present is that a total of ≥96 separate PEStep mechanisms are now available if one also includes important work on the role of nanoparticle ligands from Karim's group.[26] The second requirement of a mathematical way to compute PSDs has been available for some time and is known as a "population balance model" (PBM). The general approach of constructing a PBM is based on its successful use in chemical engineering and statistical physics any time a distribution of objects is present.[17,27–30]

However, what has not been available until recently was the *combination* of the experimentally based, minimum mechanisms for particle formation with the PBM methodology, what we have recently labeled as a "mechanism-enabled" population-based model and modeling (ME-PBM and ME-PB modeling).[31–34] ME-PBMs are derived using the law of mass action and faithfully represent a particular, experimentally determined, pseudo-elementary step mechanism. ME-PBMs yield realistic PSDs, including the shape of the PSD.[31–33] These models have parameters, such as reaction rate constants and a particle-size-distinguishing ("cut-off") parameter, that are a priori unknown and need to be determined with a sufficient degree of certainty to ensure predictions of useful accuracy. Determining if the parameters are well defined can also serve as a critical test of the proposed mechanism. It is determining these crucial parameters of the ME-PBM, and what they in turn tell about the input mechanism, that are the focus of the present contribution.

A prototypical $Ir(0)_n$ nanoparticle formation system[19,31,32] is where we first developed a ME-PBM able to describe how the concentrations of nanoparticles evolve over time, specifically how the PSD, not just the *average* particle size, evolves over time. We showed that for the $Ir(0)_n$ system, only a "4-step mechanism"[24] and a newly discovered "3-step mechanism"[31] are capable of producing PSDs that reasonably matched experimental data. Eleven other hypothesized mechanisms were tested but found unable to match the experimental PSD, thereby disproving their applicability to the $Ir(0)_n$ system. A key finding in our prior work is that it was possible to predict the observed narrow PSDs when (i) inputting the experimental finding that nucleation is continuous, and (ii) when the growth rates of the small particles are faster than those of larger particles.[31–33] Given this success with the iridium nanoparticle model system, our efforts in the rest of this paper will be based on this prototype, $Ir(0)_n$ system.

In our previous work,[31–33] we proceeded by asking what set of parameters yields predictions that are *as close as possible* to actual measured PSDs? This is an optimization problem which is readily solved once we define exactly what we mean by "as close as possible." Noteworthy here is that finding a "best-fit set of parameters" does not tell us anything about how certain we can actually be about those parameters. It is possible that the data allow us to determine parameter values to just a few percent accuracy; but it is also possible that a wide range of values would all have yielded essentially the same predictions. In these latter cases, the data are *uninformative* about the values of certain parameters because the (observed part of the) model is *insensitive* to variations in their values. Yet, even though a particular set of parameters might have had little effect on the quantities we measured, it is also possible that these parameters are important in other situations (say, at higher or lower reaction temperatures, or if one were to run the reaction for substantially longer times). Being able to accurately predict reaction outcomes for a broad range of conditions requires us to at least know which parameters we know well and which are poorly determined.

Herein we therefore address this critical question of *how well do we know our parameter estimates, along with their error estimates*? We adapt techniques from the statistical and mathematical sciences that help us seek a probability distribution in parameter space that does not just identify the one *most likely* set of parameter values or that which gives the *best fit*, but instead a probability distribution that tells us *how likely* different parameter values are. In other words, we do not just want to know the most likely parameter values, but also their SDs, confidence intervals, and other statistical properties such as the presence of long tails when desired. In particular, we want to *quantify the uncertainty* in our parameter estimates. The approach we will discuss below is typically called a "probabilistic" or "Bayesian inverse" problem. Bayesian inversion is a key statistical way to enforce Ockham's razor,[18] a critical component of disproof-based, hence more rigorous construction, development, and refinement of a minimum, more reliable chemical mechanism. Bayesian inversion such as what we present below has proven to be very useful over the last twenty or so years in other disciplines, including the geosciences,[35] hydrology,[36] and astronomy.[37] As a consequence there are now introductory textbooks[38,39] as well as tutorial-style articles available to interested readers who wish to know more about Bayesian inversion methodology.[40–42] A literature review[43–49] teaches that the Bayesian inversion framework has been used in chemistry, specifically in combustion chemistry and associated mechanisms as well as in chemometrics, forensic sciences, medical testing, microbiology/DNA analysis, chromatography and mass spectrometry, environmental chemistry, and occupational health and safety, among other areas as detailed by Hibbert and Armstrong in their highly recommended reviews.[50,51] However, the Bayesian inversion approach is still not widely employed in mechanistic chemistry in general and there is little to no use in nanoparticle chemistry and mechanisms to the best of our knowledge.*

Herein we apply the Bayesian inversion method to the analysis of the experimental kinetics and PSD data according to the previously developed ME-PBM models.[31–33] The key points we demonstrate in this paper are the following:

1. The use of the Bayesian inversion method with ME-PBM models. In particular, we demonstrate that we can not only infer model parameters quantitatively, but also provide parameter uncertainties—and hence parameter reliability estimates—while performing a global analysis and fitting all of the available particle formation kinetics and PSD data.
2. The use of the resultant parameters from the Bayesian inversion ME-PBM method, specifically their use to judge the apparent

reliability/correctness of the input mechanistic model. Our results are consistent with and strongly supportive of the notion that Bayesian inversion methods are critical for judging whether or not one has a model with too many, poorly determined parameters (e.g., if overfitting is an issue) or if one has the desired minimalistic, "Ockham's-razor-obeying" model according to the Bayesian inversion statistical method, and also if significant parameter correlations exist. In short, the results which follow demonstrate that Bayesian inversion methodology is a significant help in selecting the "correct" parameters as well as the "correct" underlying mechanism from among those considered.

3. The approach we develop below to estimating the model's parameters is more general and equally applicable to other systems as well, systems that are a focus of additional work in progress that will be published separately in due course.
4. As such, the present work also goes far toward the important goal of achieving a "Gold Standard" benchmark (i.e., a "Ground Truth") in ME-PBM fortified by Bayesian statistical methods that others can, therefore, use as a control to check on their own use of the Bayesian methodology.

Our results strongly support the hypothesis that Bayesian inversion expands significantly on the use of ME-PBM for nanoparticle formulation and one's ability to examine in a critical and reliable way the many plausible models and their parameters.

Outline of this paper. First, in Section 2, we will briefly explain the chemical background of the formation of iridium nanoparticles, a corresponding mathematical model, and the kind of experimental data we have for the nanoparticle formation, especially PSD versus time data. This section also briefly discusses the parameters that appear in the mathematical model. Section 3 will cover the techniques we use to infer both the most likely values for these parameters, as well as statistical uncertainties. Section 4 will then show what we find with these techniques and then discuss how the results can be interpreted to meaningfully relate to the iridium nanoparticle system. Finally, Section 5 contains our conclusions of how these statistical results provide evidence that a 3-step mechanism is the minimal chemical mechanism required to predict the PSDs observed in at least the prototype iridium nanoparticle system. Information available in Appendix: First, in Appendix A we discuss an alternative formulation of the so-called "likelihood" described in Section 3.2.1; Appendix B contains a detailed account of how our Bayesian inversion analysis was conducted.

2 | IR(0)$_N$ NANOPARTICLE FORMATION

Let us start our discussions by outlining the chemical basis for our mathematical models of nanoparticle formation. Specifically, we examine a nanoparticle system in which $\{(1,5-\mathrm{COD})\mathrm{Ir}^{\mathrm{I}}\cdot\mathrm{POM}\}^{8-}$ is reduced under H_2. (Here and below, POM = **p**oly**ox**o**m**etalate, $P_2W_{15}Nb_3O_{62}{}^{9-}$ and $1,5-\mathrm{COD}$ is 1,5-cyclooctadiene, C_8H_{12}.) Our ME-PBM approach to accurately accounting for the PSD entails breaking the reaction down into pseudo-elementary steps describing nucleation, growth, and any agglomeration of the nanoparticles. We will address these in the following two subsections along with a concrete, mathematical ME-PBM model followed by a discussion about the data we have available to estimate the model's parameters.

2.1 | Nucleation of nanoparticles

Using the experimental finding of *continuous* nucleation[19]—as opposed to the now disproved[65–67] theory of classical nucleation theory as employed for so-called "burst nucleation models"[68]—we are able to start the particle formation process off with the correct critical first steps at a nearly elementary step level, notably via the experimentally identified "alternative termolecular" nucleation mechanism.[69] We can then use that known nucleation mechanism with different combinations of pseudo-elementary steps for growth and any agglomeration.[32] The previously deduced minimal mechanism, dubbed the "alternative 3-step" mechanism, was previously shown to be capable of matching the shape of experimental PSDs.[31] The nucleation mechanism of Ir$(0)_n$ so identified is illustrated by the following reactions, consisting of a fast, prior equilibrium step, and continuous nucleation that is overall third-order in iridium:

$$2[(\mathrm{COD})\mathrm{Ir}\cdot\mathrm{POM}]^{8-}+4\mathrm{solv}\underset{k_b}{\overset{k_f}{\rightleftharpoons}}2\left[(\mathrm{COD})\mathrm{Ir}(\mathrm{solv})_2\right]^{+}+2\mathrm{POM}^{9-},$$

$$2\left[(\mathrm{COD})\mathrm{Ir}(\mathrm{solv})_2\right]^{+}+[(\mathrm{COD})\mathrm{Ir}\cdot\mathrm{POM}]^{8-}+7.5\mathrm{H}_2\overset{k_1}{\rightarrow}\mathrm{Ir}(0)_3+\mathrm{POM}^{9-}+3\mathrm{H}^{+}+3\mathrm{cyclooctane}+4\mathrm{solv}.$$

For brevity, we will in the following denote the precursor $[(\mathrm{COD})\mathrm{Ir}\cdot\mathrm{POM}]^{8-}$ by "A," the solvated complex, $\left[(\mathrm{COD})\mathrm{Ir}(\mathrm{solv})_2\right]^{+}$, as $\mathrm{A}(\mathrm{solv})_2$, the ligand, POM^{9-}, as "L," and (small) particles consisting of iridium(0) atoms generally as "B." The nucleation mechanism above can, then, equivalently be written as

$$\begin{aligned}&2\left\{\mathrm{A}\cdot\mathrm{L}+2\mathrm{solv}\underset{k_b}{\overset{k_f}{\rightleftharpoons}}\mathrm{A}(\mathrm{solv})_2+\mathrm{L}\right\},\\&2\mathrm{A}(\mathrm{solv})_2+\mathrm{A}\cdot\mathrm{L}\overset{k_1}{\rightarrow}\mathrm{B}_3+\mathrm{L}+4\mathrm{solv}.\end{aligned}\tag{1}$$

In the formulas above, the symbols k_* indicate reaction rate constants. It will be important for the following to note that k_f and k_b are not independent; rather, their ratio can be determined by measuring the equilibrium concentrations of the quantities involved in the reaction, and has been determined experimentally to be $k_f/k_b\approx 5\times 10^{-7}\,\mathrm{mol\,L^{-1}}$ (25°C); see reference 70 of Handwerk et al.[32]

2.2 | Nanoparticle growth and agglomeration mechanisms

The nucleation mechanism described in the previous section provides an experimentally based model for the critical question of *how does a particle form in the first place?* But, we also need to model what happens to each and every particle after it is formed. Here we assume, consistently with

experimental evidence, that growth and possibly agglomeration (i.e., aggregation) can occur, but that the reverse reactions do not occur. Our ME-PBM model is then formed from the following processes:

1. *Growth*: A precursor reacts with a particle of size n (that is, consisting of n precursor molecules), resulting in a particle of size $n+1$.
2. *Agglomeration*: A particle of size n reacts with a particle of size m, resulting in a particle of size $n+m$.

Among the many possibilities of combining these steps, we will in the following consider only what we will refer to as the "3-step" and "4-step" mechanisms—see References 31–33 for the evidence that these are the only models from among twelve considered that can accurately fit the measured data.

Specifically, the 3-step model is based on our earlier, critical finding that *different size particles grow at different rates*.[31–33] For this, let us denote by "B" a "small" particle (which we define as having at most M atoms, including the ones that result from the nucleation mechanism discussed above) and by "C" a "large" particle (with more than M atoms). The growth mechanism that augments (1) then reads as follows:

$$\begin{aligned} A+B &\xrightarrow{k_2} C+L \\ A+C &\xrightarrow{k_3} 1.5C+L. \end{aligned} \tag{2}$$

The 3-step model above is a simplified, 1-step less, condensed version of a more complex, 4-step model in which *agglomeration* due to small particles does have a substantial influence on the resulting PSD. This augmented model would then add the following reaction to (1) and (2)

$$B+B \xrightarrow{k_4} C. \tag{3}$$

In both of these mechanisms, it is important to remember that both B and C represent particles of different sizes via a sharp, delta-function at the size-cutoff parameter, M, an admittedly zeroth-order approximation of the true growth kernel. The parameter M and its use will become clearer in the following section where we derive a mathematical description of these models. For a comprehensive chemical account of these mechanisms, the reader is directed to the References 19–24, 31–33, 69, 70.

2.3 | The mathematical model

As discussed in the Introduction, an important component of being able to predict nanoparticle properties resulting from a system of reactions is having a mathematical model that faithfully obeys the experimental mechanistic evidence. Based on the conceptual model described in the previous subsection, let us next outline a mathematical transcription of these ideas. More details about this transcription can again be found in References 31–33.

The models for both the 3- and 4-step mechanisms are formulated by describing the number (or concentration) of particles of size j with a function $n_j(t)$ and asking how these functions $n_j(t)$ evolve over time. In the data we will use (see Section 2.4 and Figure 1 below), we observe no particles larger than 4 nm, and consequently only consider variables n_j with $j \le J = 2500$. Using the conversion function $\text{diameter}(j) = 0.3000805 j^{1/3}$, based on published size data for $\text{Ir}(0)_j$ particles (see figure S1 in Handwerk et al.[31]), a particle size of $J = 2500$ corresponds to a particle diameter of about 4.0 nm.

To describe all of the species in Equations (1)–(3), let us denote by n_1 the concentration of the precursor A, n_s the concentration of the disassociated precursor $A(solv)_2$, p the concentration of the ligand L, n_j the concentration of particles of size j, s the concentration of solvent (solv), used in the reaction, and $r_i = 2.677 i^{0.72}/i$, a function which limits interactions to the surface atoms of a particle.[72] Our models then lead to the following set of ordinary differential equations for the 3-step mechanism (1)–(2):

$$\begin{aligned} \frac{dn_1}{dt} &= -k_f n_1 s^2 + k_b n_s p - k_1 n_1 n_s^2 - k_2 n_1 \sum_{i=3}^{M} i r_i n_i - k_3 n_1 \sum_{i=M+1}^{J} i r_i n_i, \\ \frac{dn_s}{dt} &= k_f n_1 s^2 - k_b n_s p - 2k_1 n_1 n_s^2, \\ \frac{dp}{dt} &= k_f n_1 n_s^2 - k_b n_s p + k_1 n_1 n_s^2 + k_2 n_1 \sum_{i=3}^{M} i r_i n_i + k_3 n_1 \sum_{i=M+1}^{J} i r_i n_i, \\ \frac{dn_3}{dt} &= k_1 n_1 n_s^2 - 3k_2 n_1 r_3 n_3, \\ \frac{dn_j}{dt} &= k_2 n_1 \{ r_{j-1}(j-1) n_{j-1} - r_j j n_j \}, \qquad 4 \le j \le M, \\ \frac{dn_{M+1}}{dt} &= k_2 n_1 M r_M n_M - k_3 n_1 (M+1) r_{M+1} n_{M+1}, \\ \frac{dn_j}{dt} &= k_3 n_1 \{ r_{j-1}(j-1) n_{j-1} - r_j j n_j \}, \qquad M+2 \le j \le J. \end{aligned} \tag{4}$$

Similarly, for the 4-step mechanism in Equations (1)–(3) the equations are

$$\begin{aligned} \frac{dn_1}{dt} &= -k_f n_1 s^2 + k_b n_s p - k_1 n_1 n_s^2 - k_2 n_1 \sum_{i=3}^{M} i r_i n_i - k_3 n_1 \sum_{i=M+1}^{J} i r_i n_i, \\ \frac{dn_s}{dt} &= k_f n_1 s^2 - k_b n_s p - 2k_1 n_1 n_s^2, \\ \frac{dp}{dt} &= k_f n_1 n_s^2 - k_b n_s p + k_1 n_1 n_s^2 + k_2 n_1 \sum_{i=3}^{M} i r_i n_i + k_3 n_1 \sum_{i=M+1}^{J} i r_i n_i, \\ \frac{dn_3}{dt} &= k_1 n_1 n_s^2 - 3k_2 n_1 r_3 n_3 - 3k_4 r_3 n_3 \sum_{k=3}^{M} r_k k n_k - k_4 (3 r_3 n_3)^2, \\ \frac{dn_j}{dt} &= k_2 n_1 \{ r_{j-1}(j-1) n_{j-1} - r_j j n_j \} - k_4 r_j j n_j \sum_{k=3}^{M} r_k k n_k - k_4 (r_j j n_j)^2, \quad j = 4,5, \\ \frac{dn_j}{dt} &= k_2 n_1 \{ r_{j-1}(j-1) n_{j-1} - r_j j n_j \} - k_4 r_j j n_j \sum_{k=3}^{M} r_k k n_k - k_4 (r_j j n_j)^2 \\ &\quad + k_4 \sum_{\mu+\nu=j} r_\mu \mu n_\mu r_\nu \nu n_\nu, \qquad 6 \le j \le M, \\ \frac{dn_{M+1}}{dt} &= k_2 n_1 M r_M n_M - k_3 n_1 (M+1) r_{M+1} n_{M+1} + k_4 \sum_{\substack{\mu+\nu=M+1 \\ \mu \ge 3, \nu \ge 3}} r_\mu \mu n_\mu r_\nu \nu n_\nu, \\ \frac{dn_j}{dt} &= k_3 n_1 \{ r_{j-1}(j-1) n_{j-1} - r_j j n_j \} + k_4 \sum_{\substack{\mu+\nu=j \\ \mu \ge 3, \nu \ge 3}} r_\mu \mu n_\mu r_\nu \nu n_\nu, \qquad M+2 \le j \le J. \end{aligned} \tag{5}$$

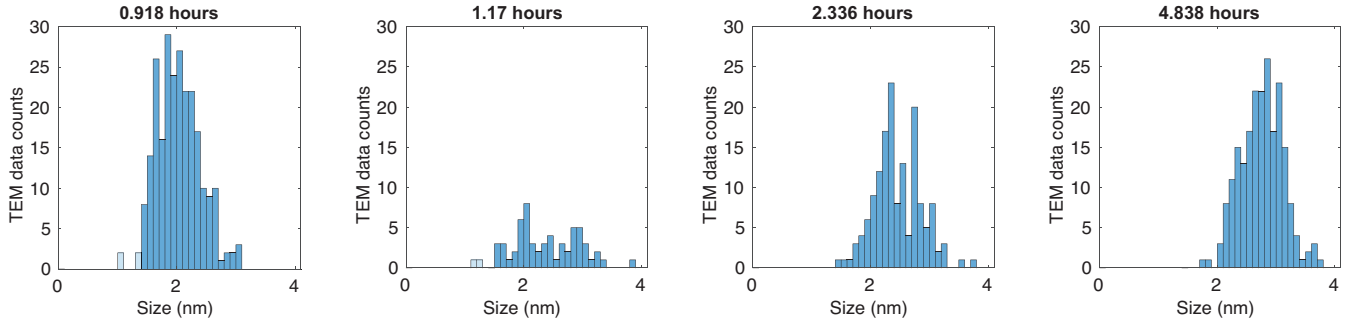

FIGURE 1 Histograms showing the number of $Ir(0)_n$ particles measured via transmission electron microscopy at different time points of the reaction (based on Watzky et al.[71]). Particle counts for particles smaller than 1.4 nm are unreliable and are shown in a lighter color. For comparison, nanoparticles composed of $J = 2500$ iridium atoms have a size of 4.0 nm

Both Equations (4) and (5) are accompanied by the known initial conditions of $n_1(0) = 0.0012\,\text{mol}\,\text{L}^{-1}$, $n_s(0) = p(0) = n_3(0) = \cdots = n_J(0) = 0$, matching our initial reaction state in which only precursor, but no nanoparticles, are present.

The question we would like to answer is what are the values of the parameters k_b, k_1, k_2, k_3, k_4 (reaction rates) and M (the size cut-off between "small" and "large" particles)? (The solvent concentration s is a known quantity in this analysis and k_f is determined by the fixed ratio $k_f/k_b \approx 5 \times 10^{-7}\,\text{L}\,\text{mol}^{-1}$ (25°C), see reference 70 in Handwerk et al.[32]) The purpose of this paper is, then and as noted earlier, to employ Bayesian inversion techniques for determining the values and the uncertainties in the unknown parameters.

To simplify notation, we will group all of these parameters into sets $K_{3-\text{step}} = \{k_b, k_1, k_2, k_3, M\}$ and $K_{4-\text{step}} = \{k_b, k_1, k_2, k_3, k_4, M\}$, and use the vector K to represent a set of parameters when we are talking about a generic mechanism. In the following, whenever we report concrete numbers for these parameters, we will imply the following units: $[k_b] = \text{L}^2\,\text{mol}^{-2}\,\text{h}^{-1}$, $[k_1] = \text{L}^2\,\text{mol}^{-2}\,\text{h}^{-1}$, $[k_2] = \text{L}^1\,\text{mol}^{-1}\,\text{h}^{-1}$, $[k_3] = \text{L}^1\,\text{mol}^{-1}\,\text{h}^{-1}$, $[k_4] = \text{L}^1\,\text{mol}^{-1}\,\text{h}^{-1}$, $[M] = 1$. The units here are the same as published previously.[31–33] From here forward all rate and equilibrium constants will be given without these (known, implied) units for the sake of simplicity.

2.4 | Experimentally determined nanoparticle size distributions

Figure 1 shows measured size distributions for the $Ir(0)_n$ system described above, for samples taken at different times during the reaction. These data come from the experimental work first provided elsewhere.[71] The figure shows how many particles fall into size bins of 1×10^{-1} nm width, based on measuring the sizes of a sample of particles as seen in transmission electron microscopy (TEM) images.

The key feature of the data is that the final size distribution is unimodal and surprisingly narrow, especially if nucleation is continuously occurring and thousands of steps are involved in particle formation. Understanding the origins of this narrowness continues to be an important driver of past research, as it would enable many applications of nanoparticles if their sizes could be predicted and controlled.[14,15,67,73]

2.5 | How reliable are these data?

In the end, we will want to use the data shown in Figure 1 to infer the values of the parameters associated with the reactions discussed in the previous subsections. At the same time, it is clear that we will not be able to determine parameters more accurately than the accuracy of the data used for this purpose.

The errors associated with the size distributions shown in the figure fall into at least three categories:

1. *Sampling error*: The data shown in the figure represent a random sample of particles for which transmission electron micrographs were used to determine their sizes. However, the number of measured particles is relatively small (246 at 0.918 h, 61 at 1.17 h, 150 at 2.336 h, and 213 at 4.838 h) and it is clear that there is stochastic noise associated with having such a limited number of particles. Consequently, the PSDs may not adequately represent the overall, "true" size distribution. In practice, this error manifests in "jumpy" (as opposed to smooth) histograms, most notably for time 1.17 h where there is limited data to represent what the true size distribution likely looks like.
2. *TEM size threshold*: Particles that are too small are not visible in TEM images. The fact that such particles seem absent in Figure 1 does not imply that they don't exist. A realistic assessment of the measurement methodology used for the data in Figure 1 suggests that we can only accurately see particles with sizes greater than 1.4 nm (equivalent to about 100 atoms). As a consequence, our methodology to match models against measurements will only include particles of size greater than this threshold, ignoring any predicted particles below the size threshold that appear absent in the data.
3. *Inaccurate size determination*: We measured particle sizes by assessing their diameter in TEM images. But these images can be fuzzy, and electron beams may also not be absorbed sufficiently in

the outer parts of a nanoparticle to clearly delineate the particle edge. We therefore believe that our determinations of diameters are only accurate to within $\pm 1 \times 10^{-1}$ nm accuracy at the very best.

As a consequence of this consideration, all figures in this manuscript use histograms of particle counts that have bin widths of 1×10^{-1} nm, starting at the threshold of 1.4 nm mentioned above.

We will take all of these sources of error into account in the statistical framework for parameter identification that we will present below.

3 | THE INVERSE PROBLEM

Section 2.3 introduced the mathematical model describing nanoparticle nucleation and growth. It contained reaction rate and cut-off parameters K which we would like to know accurately so that we can *predict* and *control* the system. Determining these parameters is called an "inverse problem" or "parameter estimation problem," and in the following we will first briefly introduce the traditional "deterministic" approach to estimating parameters, and then the more general "Bayesian" perspective we want to follow herein.

3.1 | The "deterministic" inverse problem

Having a model and having data allows us to ask what the values of the parameters $K_{3-\text{step}}$ and $K_{4-\text{step}}$ in Equation (4) or (5) might be. Traditionally, this parameter estimation problem is formulated as a "deterministic inverse problem"—namely, by asking for that set of parameters that minimizes a function that is often chosen as the least-squares "misfit":

$$\Phi(K) = \left\| \text{data}_{\text{predicted}}(K) - \text{data}_{\text{measured}} \right\|^2. \tag{6}$$

Here, $\text{data}_{\text{predicted}}(K)$ involves solving the forward model for one of (4) or (5) for the PSD given a set of parameters, and then computing from it what we would measure—in our case, how many particles we would find in each size bin.

We have followed this paradigm in Handwerk et al.[32] and found that for the data shown in Figure 1, the best-fit parameters are[†]

$$\begin{aligned} K^*_{3\text{-step}} &= \{k_b^* = 7.27 \times 10^4,\ k_1^* = 6.55 \times 10^4,\ k_2^* = 1.65 \times 10^4, \\ &\qquad k_3^* = 5.63 \times 10^3,\ M^* = 274\} \\ K^*_{4\text{-step}} &= \{k_b^* = 7.27 \times 10^4,\ k_1^* = 6.40 \times 10^4,\ k_2^* = 1.61 \times 10^4, \\ &\qquad k_3^* = 5.45 \times 10^3,\ k_4^* = 1.20 \times 10^1,\ M^* = 265\}. \end{aligned} \tag{7}$$

These "optimal" parameters correspond to the "best fit" of the model to the data, but this does not mean that the fit is actually "good." Indeed, in Handwerk et al.[32] we visually determined that only the 3- and 4-step mechanisms used here can adequately describe the data, whereas for other proposed mechanisms, the best fit was so poor that the mechanism could not be considered realistic.

That said, even in cases where we can visually determine that a given set of optimal parameters leads to a good match between prediction and measurements, we still do not know *how accurately we know these parameters*. We will address this in the following section.

3.2 | The "Bayesian" inverse problem

An alternative perspective on the inverse problem is the so-called "Bayesian approach."[38–41] In it, we seek a probability distribution $p(K|\text{data})$—that is, the probability that K are the correct parameters *given the measured data*. This approach makes intuitive sense given that the data themselves are fundamentally stochastic: for example, the size bin data shown in Figure 1 is based on a *randomly* chosen subset of particles whose sizes we have then measured, with the size measurement itself subject to measurement uncertainties. In other words, if we repeated measurements we would get different data, and we need to transform this uncertainty in data space into uncertainty in parameter space.

In order to compute the probability distribution $p(K|\text{data})$, we make use of Bayes' theorem that states that (see References 38 and 39)

$$p(K|\text{data}) \propto \underbrace{p_L(\text{data}|K)}_{\text{"likelihood"}} \underbrace{p_{\text{pr}}(K)}_{\text{"prior"}}, \tag{8}$$

where the "likelihood" describes how likely it would be to observe the measured data if K were the "true" set of parameters, and the "prior" encodes what we know a priori about the parameters. The probability distribution $p(K|\text{data})$ is typically called the "posterior probability" since it is informed by our measurements, as opposed to the prior. Since our data is static, we will simplify our notation for the likelihood function to be

$$L(K) = p_L(\text{data}|K).$$

The posterior in Equation (8) is defined as a proportionality rather than an equality—that is, the computable right-hand side is a non-normalized probability density. In practice, this proportionality is sufficient: it allows for comparisons of relative probability density—that is, is $p(K_1|\text{data}) > p(K_2|\text{data})$ or $p(K_1|\text{data}) < p(K_2|\text{data})$—and that is all our algorithms will need.

In the following subsections, we will discuss the construction of the likelihood and prior, and then how one can use $p(K|\text{data})$ to make inferences about parameter values.

3.2.1 | The likelihood

Computing the "likelihood" $L(K)$ implies solving the forward model (4) or (5) with K, and then comparing its predictions with the measured data. In the current context, we do this as follows:

1. Given a particular K, we can solve the forward model (4) or (5) numerically using a standard ODE integrator to obtain values $n_j^{\text{pred}}(t_i;K)$ for the predicted concentrations of nanoparticles of size j at $t_1 = 0.918$, $t_2 = 1.170$, $t_3 = 2.336$, and $t_4 = 4.838$ h, where we choose a notation that makes it explicit that n_j^{pred} depends on the set of parameters K used to run the forward simulation.
2. From these predicted concentrations $n_j^{\text{pred}}(t_i;K)$ at time t_i, we can infer the concentrations $b_{i,\ell}^{\text{pred}}(K)$ of particles at time t_i that fall within the size range of the ℓth bin, where we use the size bins defined in Section 2.5 above. Using these binned concentrations, we can easily calculate the fraction of particles in the ℓth bin:

$$p_{i,\ell}(K) = \left(\sum_{\ell=1}^{N_{\text{bins}}} b_{i,\ell}^{\text{pred}}(K)\right)^{-1} b_{i,\ell}^{\text{pred}}(K). \tag{9}$$

   where N_{bins} is the number of bins used to group particle sizes.
3. For the likelihood, we then need to determine how likely it is that a given measurement of particle sizes results from these relative probabilities $p_{i,\ell}$. If a measurement consists of N_i particles' sizes grouped into bins (as in Figure 1), then this process can be understood in the same way as drawing N_i balls of different colors from an urn with a very large number of balls with known color distribution (corresponding to $p_{i,\ell}$).
   Our measured data is a set of values $\beta_{i,\ell}^{\text{measured}}$. The question for us to answer is: How likely is it to get $\beta_{i,\ell}^{\text{measured}}$ particles in bin ℓ at time i if the probabilities of particles being in these bins are given by $p_{i,\ell}$? This likelihood can be computed by the analogy to drawing balls from an urn, and is given by

$$L_i(K) = \underbrace{\left(\frac{N_{\text{total}}!}{(N_{\text{total}} - N_i)!} \prod_{\ell=1}^{N_{\text{bins}}} \frac{1}{\beta_{i,\ell}^{\text{measured}}!}\right)}_{\text{normalization factor}} \underbrace{\prod_{\ell=1}^{N_{\text{bins}}} \left(p_{i,\ell}(K)\right)^{\beta_{i,\ell}^{\text{measured}}}}_{\text{computable}} \tag{10}$$

   where N_{total} is the total number of particles in the chemical solution. The first term cannot be computed because we do not know N_{total}, but is independent of K. Akin to the discussion following Equation (8), we can ignore this normalization factor, and we are left with

$$L_i(K) \propto \prod_{\ell=1}^{N_{\text{bins}}} \left(p_{i,\ell}(K)\right)^{\beta_{i,\ell}^{\text{measured}}}. \tag{11}$$

4. Finally, we model the likelihood $L(K)$ as the product of the probabilities for finding bins as measured at each of the times as measured. That is:

$$L(K) \propto \prod_{i=1}^{4} L_i(K). \tag{12}$$

Underlying this product structure is the assumption that the measurements at different times and their errors are statistically independent. This assumption is justified given that the data obtained at each time point resulted from the removal of a small amount of reaction products from the ongoing reaction, and subsequent independent analysis of these samples.

This likelihood can be computed for a given set of parameters K with a bit of effort, but in a relatively straightforward way. It requires solving the system of differential Equations (4) or (5), plus the statistical evaluations in Equations (11) and (12).

We end this section by noting that one could have defined the likelihood also in ways that do not use binning. We explore this possibility in Appendix A.

3.2.2 | The prior probability

The prior probability p_{pr} in Equation (8) encodes what we know a priori about the parameters. This is often very little. In the current context, all we really know is that all of these parameters must be non-negative, and that $M \geq 3$. We describe this as follows:

$$p_{\text{pr}}(K_{3-\text{step}}) = \begin{cases} 1 & \text{if } 0 \leq k_b \leq k_{b,\max} \text{ and } 0 \leq k_1 \leq k_{1,\max} \\ & \quad \text{and } 0 \leq k_2 \leq k_{2,\max} \text{ and } 0 \leq k_3 \leq k_{3,\max} \\ & \quad \text{and } 3 \leq M \leq M_{\max}, \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

and similar for $K_{4\text{-step}}$. $k_{b,\max}, k_{2,\max}, k_{2,\max}, k_{3,\max}$, and $M_{\max}$ are chosen large enough that their concrete value does not affect results. In fact, they could be chosen as infinity.

The prior $p_{\text{pr}}(K)$ is also not normalized, that is, the integral over all parameter values does not add up to one—and would not even be finite if the maximal values are chosen as infinity. However, as discussed previously, the missing normalization constant is of no consequence.

3.2.3 | Evaluating the posterior probability

The function $p(K|\text{data})$ is, in general, difficult, high dimensional, and without a closed-form expression (because it involves the solution of a differential equation). As a consequence, we cannot easily evaluate quantities of interest such as the mean value and SD of each parameter in K. For example, the mean value $\bar{k}_b$ in $K_{3\text{-step}}$ is

$$\bar{k}_b = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k_b \; p(k_b,k_1,k_2,k_3,M|\text{data}) \; dk_b\, dk_1\, dk_2\, dk_3\, dM, \tag{14}$$

but this integral cannot be evaluated for lack of a closed-form expression for $p(K|\text{data})$.

Since such integrals cannot be computed exactly, we must *approximate* them. The typical approach to do this is through *sampling* using Markov Chain Monte Carlo methods such as the Metropolis–Hastings sampler or variations thereof. All of these methods fundamentally start at a point $K^{(0)}$ and then repeatedly perform the following steps:

1. Propose a trial sample K^{trial}, typically chosen near the current sample K^{current}.
2. Evaluate the ratio of probabilities,

$$\frac{p\left(K^{\text{trial}}|\text{data}\right)}{p\left(K^{\text{current}}|\text{data}\right)}$$

and based on this ratio and other information, either "accept" or "reject" K^{trial}. If accepted, it becomes K^{current}, if rejected the previous K^{current} is kept. In both cases, K^{current} is appended to the list of samples in the chain.

A concise definition of how the Metropolis-Hastings and other samplers define trial samples, and when they accept them, can be found in references[39]; we will provide an outline of specific choices we made for the implementation in Appendix B. In any case, because only the *ratio* of probabilities is used, it is now clear why the normalization constants in Equations (10) and (13) do not matter, and why in the definition of (8) it was sufficient to state a proportionality, rather than an equality.

The end result is a chain of samples, $\left\{K^{(0)},K^{(1)},K^{(2)},K^{(3)},\ldots\right\}$ that is constructed in such a way that there are many samples where $p(K|\text{data})$ is large, and few samples where $p(K|\text{data})$ is small. Through this process we have a representation of the approximate posterior distribution. It can then be shown that we can approximate

$$\overline{k}_b \approx \frac{1}{P}\sum_{p=1}^{P} k_b^{(p)}\; p\left(K^{(p)}|\text{data}\right) \tag{15}$$

using P samples, with similar approximations for the mean values of the other parameters, as well as for the SDs or other statistical quantities.

In practice, the approximation gets better the more samples P one has. We will often want to use many thousands or millions, despite the fact that the creation of a sample requires the evaluation of the ratio of probabilities which in turn requires the solution of the forward model and some statistical evaluations, as discussed in Sections 3.2.1–3.2.2. In the examples below, we have used several million samples, each of which required in the range of 1–5 s to compute. The overall computational cost of these evaluations is therefore on the order of a few CPU years, though one can run many computations in parallel.

4 | RESULTS

Using the data presented in Section 2 and the formalism of the previous section, we can represent the posterior probability distribution via a large number of samples in a number of scenarios that we will discuss in the following. In particular, we will first use this approach to identify the parameters in the "3-step" mechanism discussed in Section 2.3, followed by a discussion of corresponding results for the "4-step" mechanism.

4.1 | Inversion based on individual time points

The experiments of Watzky et al.[71] and summarized in Section 2 provided particle counts at four different time points of the reaction, at $t_1 = 0.918, t_2 = 1.170, t_3 = 2.336, t_4 = 4.838$ h. Yet, our previous exploration of the reaction mechanism determined "best fit" parameters for $K_{3-\text{step}}$ and $K_{4-\text{step}}$ was based only on the last of these time points, reasoning that all of these parameters must surely affect the outcome at the last time point to the same or a larger degree as the first three time points.

But one can question this: maybe one of the reactions is fast, and the effect of its reaction rate would be visible in one of the earlier times but its value is no longer important to explain the results at later times. This reasoning suggests that each of the time points could provide *complementary* information that, taken together, would yield a better picture of the true parameter values than just considering one time point.

The Bayesian approach allows us to test this. Instead of defining the likelihood function in Equation (12) as the product of likelihoods from all four time points, we can take into account only one time point *i*. The first four columns of Figure 1 show the one-dimensional marginal probability densities derived from the posterior probability density based on just the data from one time point each.

Indeed, reading each of the rows of the figure left to right shows that the probability distributions we obtain using data from different time points are substantially different both from each other, and from the probability distribution obtained from all data together (shown in the last column). In other words, each data set contains different, independent information.

Also, the graphs of the second column (considering only the data for t_2) show marginal probability distributions for most parameters that are so flat that they do not have substantial mass within the horizontal range shown in the figure. This is easily explained since, as discussed in Section 2 (see also Figure 1), we have very little data at t_2 compared to the sampling noise, and this is apparently not enough to substantially constrain parameter values. As a consequence, using only this time point, we cannot infer parameter values to any kind of certainty. In contrast, the probability distributions obtained from time points t_1 and t_4, for which we have the most data, are narrow and therefore provide estimates for the parameters with relatively small uncertainties.

4.2 | Inversion based on all time points jointly

If, as indeed shown above, the data from different time points provides *complementary* information, it makes sense to use *all* of this information to determine the parameters in our models. The last column of Figure 2 shows the probability distributions using all time points jointly by utilizing the likelihood function as originally described in Equation (12).[‡]

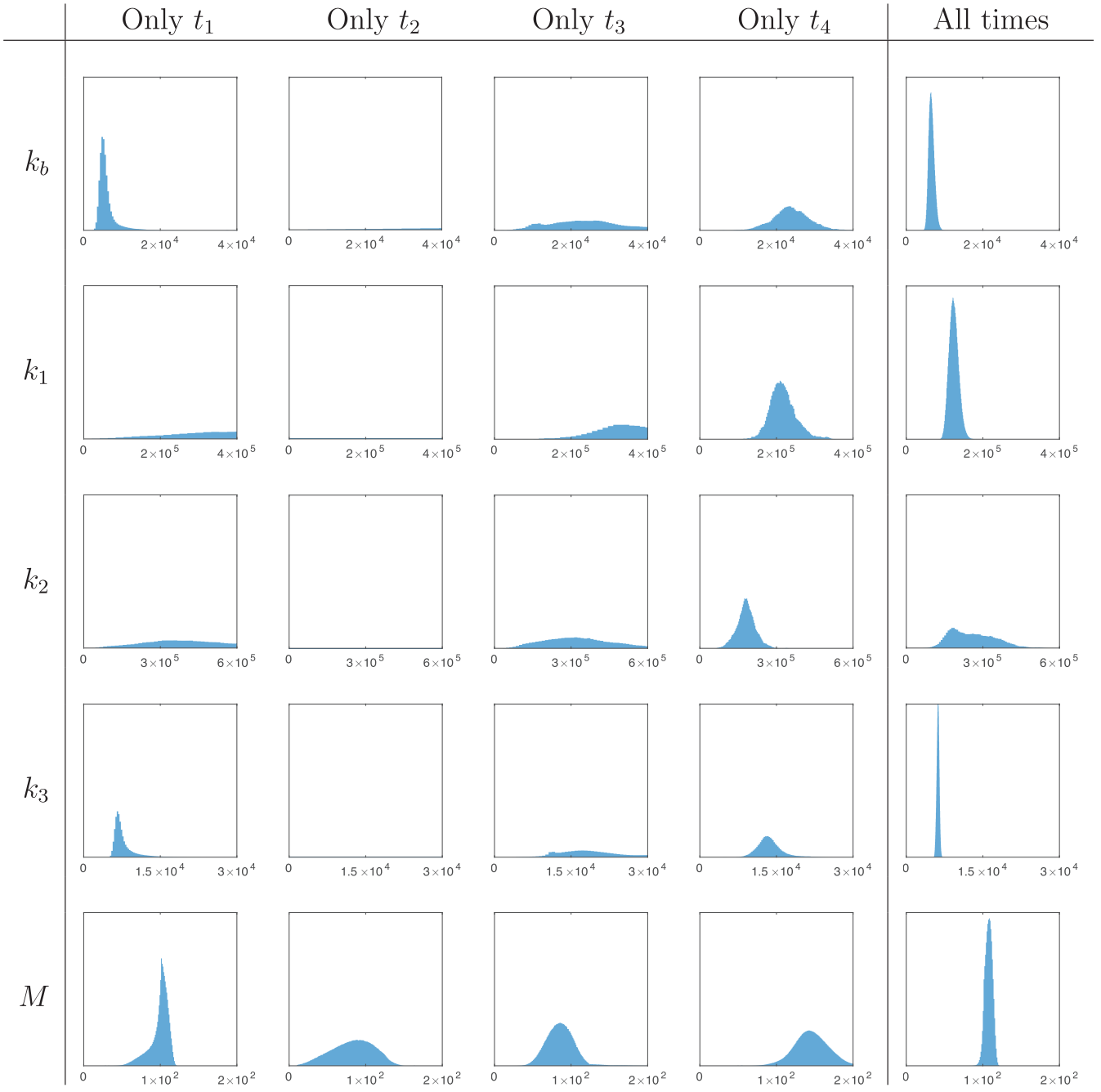

FIGURE 2 One-dimensional marginal probability densities for the parameters in the 3-step mechanism. The first four columns show probability densities computed using only measured data from one of the four time points each ($t_1 = 0.918$h, $t_2 = 1.170$h, $t_3 = 2.336$h, and $t_4 = 4.838$h). The last column uses all data jointly. Each column is computed from simulations using 4.408×10^6 samples. All plots in a row use the same vertical and horizontal scales; plots that look empty simply have a small and very broad probability distribution that may extend beyond the left and right edges of the plot. (Appendix D and Figure S3 show alternate ways of visualizing the data that underlies this figure.)

The fact that the combined probability distributions are generally narrower than the ones obtained from the measurements at individual time points illustrates that using more data helps narrow down the uncertainties in how well we know each of the parameters. Moreover, from the information shown in the figures, we can provide not only improved estimates of the parameters (to be compared to those originally reported in Handwerk et al.[32] and reproduced in Equation (7) in Section 3.1), but importantly also their uncertainties:

$$\begin{aligned} K^*_{\text{3-step}} = \{ & k_b^* = (6.62 \pm 0.75) \times 10^3, \;\; k_1^* = (1.24 \pm 0.12) \times 10^5, \\ & k_2^* = (2.60 \pm 0.86) \times 10^5, \;\; k_3^* = (6.22 \pm 0.25) \times 10^3, \\ & M^* = 107 \pm 5 \}. \end{aligned} \tag{16}$$

These values are the mean value and SD of the probability distributions shown in the rightmost column of Figure 1. We note that for all of the parameters besides k_b and k_2, the SD is an order of magnitude lower than the mean value. In other words, the data we have allows us to infer the parameters with confidence to one digit of accuracy.

A key conclusion from our previous work is that for the observed narrow PSDs to form, smaller particles must grow more quickly than larger particles—that is, $k_2 > k_3$. The best fit of the 3-step model to the t_4 data in Handwerk et al.[32] (see Equation (7)) gave $k_2 = 1.65 \times 10^4$ and $k_3 = 5.63 \times 10^3$, so that $\frac{k_2}{k_3} = 2.93$. The Bayesian approach supports the conclusion that $k_2 > k_3$. Indeed, for the values shown in Equation (16) above, using all four time steps, the lower bound of the confidence interval for k_2, namely 1.74×10^5, is larger than the upper bound of the confidence interval for k_3, namely 6.47×10^3, and the ratio of these parameters' mean values is $\frac{k_2}{k_3} = 26.89$.

We can compare the previous data also against what the Bayesian approach would yield when using only the fourth time step. In that case, we obtain

$$K^*_{3\text{-step,only } t_4} = \{k_b^* = (2.38 \pm 0.48) \times 10^4,\ k_1^* = (2.17 \pm 0.34) \times 10^5,\ k_2^* = (1.81 \pm 0.37) \times 10^5,\ k_3^* = (1.36 \pm 0.25) \times 10^4,\ M^* = 145 \pm 23\}.$$

Again, the lower bound of the confidence interval for k_2, namely 1.44×10^5, is larger than the upper bound of the confidence interval for k_3, namely 1.61×10^4, and the ratio of these values is $\frac{k_2}{k_3} = 8.94$, close to the previously reported ratio.

4.3 | Assessment for the 4-step mechanism

We can repeat the same process for the 4-step mechanism provided previously. Figure 3 shows the results of inverting for parameter values using all time points jointly, for the six parameters in Equation (5). As before, we can compute mean and SDs for these parameters (again to be compared to those originally reported in Handwerk et al.[32] and reproduced in Equation (7)):

$$K^*_{4\text{-step}} = \{k_b^* = (1.37 \pm 0.13) \times 10^5,\ k_1^* = (7.69 \pm 0.86) \times 10^4,\ k_2^* = (1.40 \pm 0.10) \times 10^4,\ k_3^* = (7.15 \pm 0.32) \times 10^3,\ k_4^* = (1.74 \pm 0.78) \times 10^3,\ M^* = 111 \pm 14\} \quad (17)$$

Comparing Figure 3 to the rightmost column of Figure 2, we see many similarities in the probability distributions between the 3- and 4-step models. In particular, the distributions for k_b,k_1,k_3 are all quite narrow, suggesting both models are sensitive to these parameters. Conversely, parameters k_2 and M have qualitatively different probability distributions between the 3- and 4-step models. Indeed, we see that the 4-step model is more sensitive to k_2 but less sensitive to M. In addition to the visual representation of uncertainty, the SDs in Equations (16) and (17) show similar levels of accuracy for parameters k_b and k_3—the 4-step model's k_1 has an inflated SD due to its heavy tail. Similarly, the SDs for k_2 and M reflect the differences in parameter sensitivity between the 3- and 4-step models.

Finally, we note that the uncertainty in k_4 is relatively large, and that k_4 is associated with the agglomeration reaction that distinguishes the 3- and 4-step models. The fact that we have a large uncertainty in this reaction's rate suggests that this additional reaction is not particularly important in describing the observed PSD. One can interpret this as suggesting that the 3-step mechanism is the minimal mechanism that can explain the data, and that the 4-step mechanism is an unnecessary complication for the present iridium nanoparticle system.[§]

4.4 | Assessing accuracy and uncertainty in model predictions

It is conceivable that one can obtain narrow parameter distributions yet the model with these parameters does not reproduce the

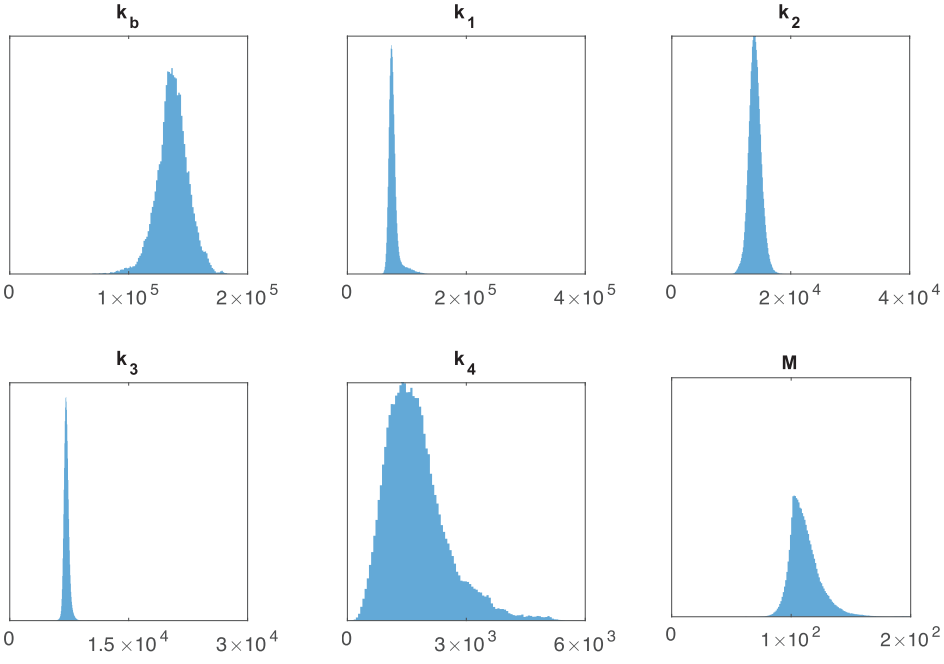

FIGURE 3 One-dimensional marginal probabilities for the parameters in the 4-step mechanism, using measured data at all four time points. These probability distributions are computed using 4.408×10^6 samples

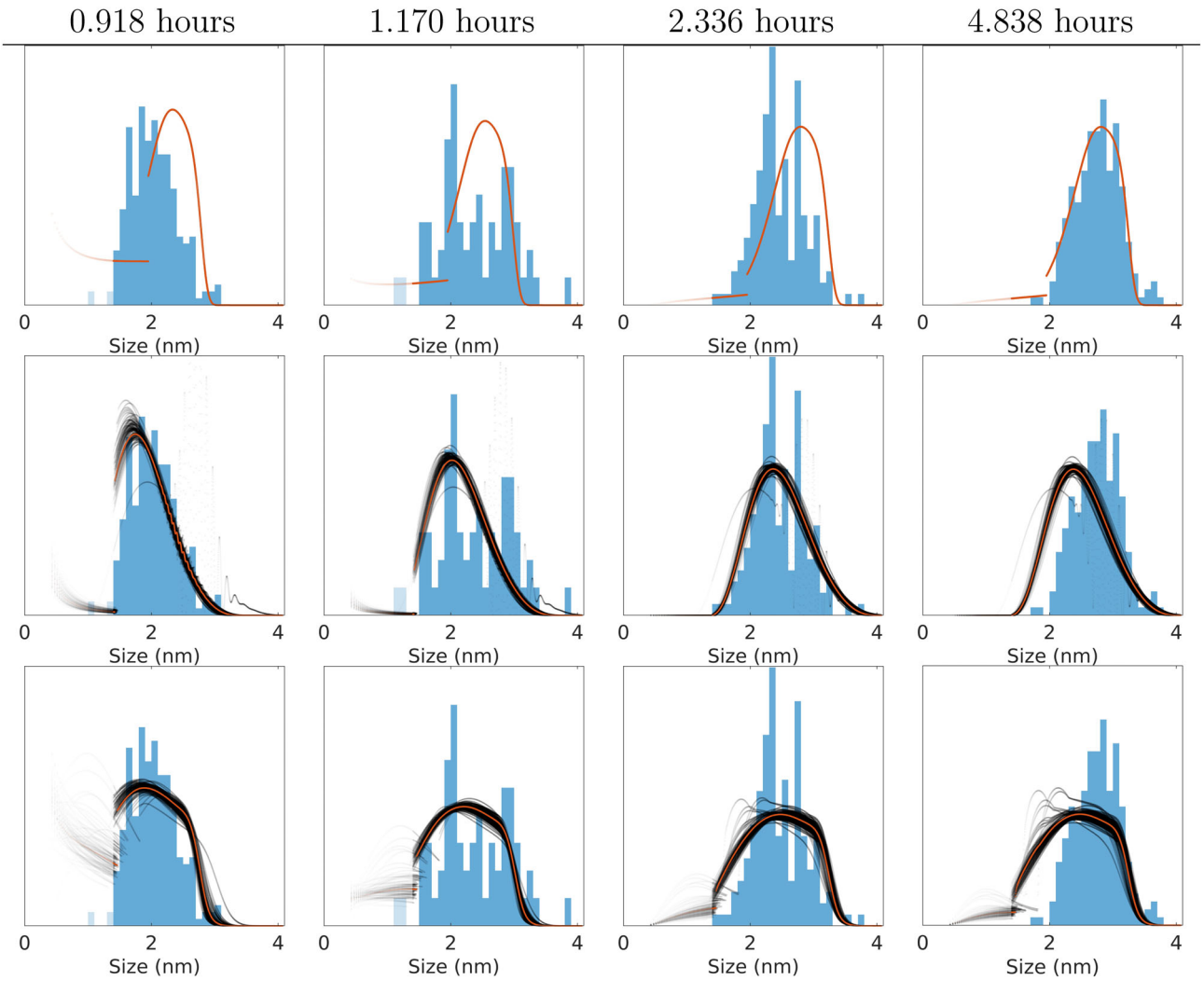

FIGURE 4 Comparison of simulation results using (top) the 3-step mechanism with previously published parameters determined using the deterministic approach (7), (middle) the 3-step mechanism with parameters determined using the Bayesian approach, and (bottom) the 4-step mechanism with parameters determined using the Bayesian approach. Recall that the previously published values shown in Equation (7) only considered t_4, and so it is no surprise that the top-right figure shows an excellent fit, with worse fits for the other time points. For the bottom two rows, the orange curve represents particle size predictions obtained using the mean parameters provided in Equations (16) and (17), respectively. The many black curves represent predictions using 100 parameters K randomly chosen from the probability distribution $p(K|\text{data})$; this visualization provides an indication of the uncertainty in predictions. As in Figure 1, particles smaller than 1.4 nm are shown in a lighter color

observed PSD—this would be the case if the model is simply unable to reproduce reality even with the "best" values for the coefficients. Therefore, it is important to test this possibility. In Figure 4, we compare the simulated PSD using the previously published values for $K^*_{3-\text{step}}$ (see Equation 7) to the simulated PSD using the mean values obtained in our Bayesian analysis for the 3- and 4-step mechanisms (see Equations 16 and 17).

Our predictions are all reasonably close to the measured data, and reasonably accurately represent the observed, narrow nanoparticle PSD. Visually, it is difficult to assess whether the 3- or 4-step mechanisms provide better matches for the data, supporting our conclusion above that the 4-step mechanism may be an unnecessary complication of the 3-step model. At the same time, from the lack of fit in the first row for times other than t_4, it is clear that it is important to use all of available data in the fitting procedure.[**]

The figure also shows, in thin black lines, predicted PSDs computed with a random subset of 100 parameter values chosen from our Markov chains. These lines provide a visualization of the spread of predictions, corresponding to the spread in parameter values, and illustrate that the parameter distributions result in relatively uniform predictions.

5 | CONCLUSIONS

In this contribution, we have discussed the use of a Bayesian methodology for the estimation of parameters in a mathematical model of

iridium nanoparticle formation; that is, we have performed "Bayesian inversion assisted, mechanism-enabled population-balance modeling" (BIA-ME-PBM).

The Bayesian inversion approach we employ is more complicated and substantially more computationally expensive than our previous deterministic approach described in Section 3.1 and previously reported in References 31–33. At the same time, the BIA-ME-PBM approach provides valuable insights not available from the deterministic method. Specifically, the deterministic inverse problem does not provide us with a way to quantify the uncertainty in parameter estimates. Through our discussion in Section 4, we have seen that the Bayesian approach provides this missing knowledge. Furthermore, the method not only provides a best-fit value for the parameter k_4 that appears in the equation that distinguishes the 3- and 4-step mechanisms, but also a large uncertainty for it. In other words, the details of the agglomeration reaction (3, vide supra) do not seem to matter much for fitting the available data; hence, we can interpret this as further evidence that the 3-step mechanism, rather than the 4-step mechanism, is the *minimal mechanism* able to describe the formation of $\mathrm{Ir}(0)_n$ nanoparticles. As a consequence, the Bayesian approach can also be used as a tool in *model selection* here mechanistic model selection.

Quantitatively, the probability distributions for all parameters that appear in the model, taking into account all available data (see the rightmost column of Figure 2), are narrow enough to determine all parameters to approximately one digit of accuracy. Furthermore, our assessment of the uncertainty in predictions of the model using uncertain parameters in Figure 4 shows that the parameter ranges we have identified for all components of K all yield relatively similar predictions. Together, this allows us to draw two important conclusions that one cannot obtain from the deterministic inverse problem alone:

1. While the experimental nanoparticle size data shown in Figure 1 contain substantial noise, it is not entirely inadequate for a reasonable determination of parameter values. Clearly, one always wishes for better data, in particular more measurements at time t_2; better data generally leads to smaller uncertainties in parameter estimates, and would allow us to determine them to more than around one digit of accuracy. At the same time, the data shown in Figure 1 allows for reasonably accurate estimates of reaction parameters.
2. The uncertainty in parameters leads to relatively small variability in predictions, and this enables the important application of *optimizing the reaction conditions for specific outcomes*. More specifically, optimization of initial conditions or reaction temperatures is only meaningful if model predictions are relatively stable with regard to the uncertainty in estimated parameters—as is indeed the case here, based on the results shown in Figure 4.

These conclusions show that there is substantial value in using a Bayesian methodology to parameter estimation. To the best of our knowledge, Bayesian methods similar to the one we presented are not widely used in this part of the chemistry community. Indeed, the Bayesian framework is a highly generalizable approach that can be used in many parameter estimation problems. The construction of the likelihood function in Section 3.2.1 and the prior distribution in Section 3.2.2 will differ between problems, but the interpretations we make of the resulting probability distributions are the same insights one would seek for a general parameter estimation problem: narrow, unimodal distributions are indicative of an adequate mathematical model and sufficient data. We hope to see more use of the Bayesian approach within the chemistry community. We will report our own efforts expanding the BIA-ME-PBM methodology to a second iridium system where more data is present and where the mechanism of especially nucleation is currently not 100% clear, as well as in semiconductor and other nanoparticle systems.

ACKNOWLEDGMENTS

D. K. L.'s research time was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Catalysis Science Program, via Award SE-FG402-02ER15453 to R. F. G. W. B.'s research was partially supported by the National Science Foundation under award OAC-1835673 as part of the Cyberinfrastructure for Sustained Scientific Innovation (CSSI) program; by award DMS-1821210; by award EAR-1925595. P. D. S.'s research was partially supported by the National Science Foundation Division of Mathematical Sciences, via NSF grant DMS-1814941.

DATA AVAILABILITY STATEMENT

The software that implements the statistical analyses described herein is available at https://github.com/dklong-csu/mepbm under an open source license, along with documentation that explains its use.[76] This repository also contains all of the data—Specifically the measured PSDs—Used as input for our computations.

ORCID

Wolfgang Bangerth https://orcid.org/0000-0003-2311-9402

ENDNOTES

[*] In our literature search, we found a number of uses of Bayesian inversion in chemistry.[18,43–49,52–64] Armstrong and Hibbert[50,51] also provide a comprehensive, albeit now decade-old survey of the uses of Bayesian methods in chemistry. However, our attempt at a comprehensive literature search revealed no uses of Bayesian methods for nanoparticle mechanistic chemistry nor evidence for its deserved, more extensive use in mechanistic chemistry in general.

[†] The actual methodology used in References 31–33 replaced the measured data (shown in Figure 1) by a smoothed version to mitigate the problem of sampling error mentioned in Section 2.5 at least to some degree. The method there also tried to fit the entire particle size distribution, rather than only for those particles with a diameter of more than 1.4 nm as explained above. This is equivalent to assuming that measurements simply found no small particles. Finally, our methodology used only the final PSD data at t_4, deliberately ignoring the data at t_1, t_2, t_3 in that initial ME-PBM effort.

[‡] Indeed, looking at the form of (12), we recognize that the combined likelihood is the product of the likelihoods obtained from the four time points individually. Furthermore, the specific form of the prior in Equation (13) then implies that the *combined* posterior probability density $p(K|\text{data})$ is simply the product of the posterior probability densities we

get if we only consider a single time point—that is, the plots shown in the last column of Figure 2 would simply depict the product of the probability densities for individual time points, shown in the preceding four columns, if we had infinitely many samples.

[§] The suggestion that the 3-step model is sufficient can be formally tested with the help of "Bayes factors"[74] to determine whether the 4-step model yields a benefit that outweighs its greater complexity.

[**] One can find quantitative ways to assess whether one model fits data better than another model. If one were to simply fit parameters of a model to a data set, then the *R*-squared criterion is often used. In Bayesian models, this notion needs to be generalized and is often referred to as "posterior predictive assessment," see, for example, Gelman et al.[75] Visual inspection of the plots in the second and third row of Figure 4 suggests that the 3- and 4-step models yield fits that are *not qualitatively different* in their goodness-of-fit, and we consequently decided not to go into the details of posterior predictive assessment.

REFERENCES

[1] I. E. Beck, V. I. Bukhtiyarov, I. Y. Pakharukov, V. I. Zaikovsky, V. V. Kriventsov, V. N. Parmon, *J. Catal.* **2009**, *268*, 60. https://doi.org/10.1016/j.jcat.2009.09.001

[2] M. R. Axet, K. Philippot, *Chem. Rev.* **2020**, *120*, 1085. https://doi.org/10.1021/acs.chemrev.9b00434

[3] I. Favier, D. Pla, M. Gómez, *Chem. Rev.* **2019**, *120*, 1146. https://doi.org/10.1021/acs.chemrev.9b00204

[4] F. P. da Silva, J. L. Fiorio, L. M. Rossi, *ACS Omega* **2017**, *2*, 6014. https://doi.org/10.1021/acsomega.7b00836

[5] D. C. Gary, M. W. Terban, S. J. L. Billinge, B. M. Cossairt, *Chem. Mater.* **2015**, *27*, 1432. https://doi.org/10.1021/acs.chemmater.5b00286

[6] J. W. Stouwdam, R. A. J. Janssen, *J. Mater. Chem.* **2008**, *18*, 1889. https://doi.org/10.1039/b800028j

[7] V. L. Colvin, M. C. Schlamp, A. P. Alivisatos, *Nature* **1994**, *370*, 354. https://doi.org/10.1038/370354a0

[8] Y. Shirasaki, G. J. Supran, M. G. Bawendi, V. Bulović, *Nat. Photonics* **2012**, *7*, 13. https://doi.org/10.1038/nphoton.2012.328

[9] R. D. Schaller, V. I. Klimov, *Phys. Rev. Lett.* **2004**, *92*, 186601. https://doi.org/10.1103/physrevlett.92.186601

[10] A. G. Pattantyus-Abraham, I. J. Kramer, A. R. Barkhouse, X. Wang, G. Konstantatos, R. Debnath, L. Levina, I. Raabe, M. K. Nazeeruddin, M. Grätzel, E. H. Sargent, *ACS Nano* **2010**, *4*, 3374. https://doi.org/10.1021/nn100335g

[11] L. Protesescu, S. Yakunin, M. I. Bodnarchuk, F. Krieg, R. Caputo, C. H. Hendon, R. X. Yang, A. Walsh, M. V. Kovalenko, *Nano Lett.* **2015**, *15*, 3692. https://doi.org/10.1021/nl5048779

[12] I. Spanopoulos, I. Hadar, W. Ke, P. Guo, E. M. Mozur, E. Morgan, S. Wang, D. Zheng, S. Padgaonkar, G. N. Manjunatha Reddy, E. A. Weiss, M. C. Hersam, R. Seshadri, R. D. Schaller, M. G. Kanatzidis, *J. Am. Chem. Soc.* **2021**, *143*, 7069. https://doi.org/10.1021/jacs.1c01727

[13] D. Gielen, F. Boshell, D. Saygin, *Nat. Mater.* **2016**, *15*, 117. https://doi.org/10.1038/nmat4545

[14] Q. N. Nguyen, R. Chen, Z. Lyu, Y. Xia, *Inorg. Chem.* **2021**, *60*, 4182. https://doi.org/10.1021/acs.inorgchem.0c03576

[15] D. V. Talapin, J.-S. Lee, M. V. Kovalenko, E. V. Shevchenko, *Chem. Rev.* **2009**, *110*, 389. https://doi.org/10.1021/cr900137k

[16] M. Smoluchowski, *Z. Phys. Chem.* **1917**, *92*, 129.

[17] H. M. Hulburt, S. Katz, *Chem. Eng. Sci.* **1964**, *19*, 555.

[18] R. Hoffmann, V. I. Minkin, B. K. Carpenter, *Bull. Soc. Chim. Fr.* **1996**, *133*, 117.

[19] M. A. Watzky, R. G. Finke, *J. Am. Chem. Soc.* **1997**, *119*, 10382.

[20] C. Besson, E. E. Finney, R. G. Finke, *J. Am. Chem. Soc.* **2005a**, *127*, 8179. https://doi.org/10.1021/ja0504439

[21] C. Besson, E. E. Finney, R. G. Finke, *Chem. Mater.* **2005b**, *17*, 4925. https://doi.org/10.1021/cm050207x

[22] B. J. Hornstein, R. G. Finke, *Chem. Mater.* **2003**, *16*, 139. https://doi.org/10.1021/cm034585i

[23] E. E. Finney, R. G. Finke, *Chem. Mater.* **2008**, *20*, 1956. https://doi.org/10.1021/cm071088j

[24] P. D. Kent, J. E. Mondloch, R. G. Finke, *J. Am. Chem. Soc.* **2014**, *136*, 1930.

[25] R. J. Field, R. M. Noyes, *Acc. Chem. Res.* **1977**, *10*, 214. https://doi.org/10.1021/ar50114a004

[26] S. Mozaffari, W. Li, C. Thompson, S. Ivanov, S. Seifert, B. Lee, L. Kovarik, A. M. Karim, *Nanoscale* **2017**, *9*, 13772. https://doi.org/10.1039/c7nr04101b

[27] D. Ramkrishna, *Population Balances: Theory and Applications to Particulate Systems in Engineering*, Elsevier, **2000**. https://doi.org/10.1016/B978-0-12-576970-9.X5000-0.

[28] F. Sporleder, Z. Borka, J. Solsvik, H. A. Jakobsen, *Rev. Chem. Eng.* **2012**, *28*, 149.

[29] D. Ramkrishna, M. R. Singh, *Annu. Rev. Chem. Biomol. Eng.* **2014**, *5*, 123.

[30] R. I. Jeldres, P. D. Fawell, B. J. Florio, *Powder Technol.* **2018**, *326*, 190.

[31] D. R. Handwerk, P. D. Shipman, C. B. Whitehead, S. Zkar, R. G. Finke, *J. Am. Chem. Soc.* **2019**, *141*, 15827. https://doi.org/10.1021/jacs.9b06364

[32] D. R. Handwerk, P. D. Shipman, C. B. Whitehead, S. Özkar, R. G. Finke, *J. Phys. Chem. C* **2020**, *124*, 4852. https://doi.org/10.1021/acs.jpcc.9b11239

[33] D. Handwerk, PhD Thesis, Colorado State University **2019**.

[34] C. B. Whitehead, D. R. Handwerk, P. D. Shipman, Y. Li, A. I. Frenkel, B. Ingham, N. M. Kirby, R. G. Finke, *J. Phys. Chem. C* **2021**, 13449. https://doi.org/10.1021/acs.jpcc.1c03475

[35] A. Tarantola, *Inverse Problem Theory*, Elsevier, Amsterdam **1987**.

[36] Y. Jiang, A. D. Woodbury, *Geophys. J. Int.* **2006**, *167*, 1501. https://doi.org/10.1111/j.1365-246x.2006.03145.x

[37] I. Craig, J. Brown, in *Bayesian Astrophysics* (Eds: A. A. Ramos, I. Arregui), Cambridge University Press, **1986**, p. 31. https://doi.org/10.1017/9781316182406.003

[38] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, **2005**. https://doi.org/10.1137/1.9780898717921

[39] J. Kaipio, E. Somersalo, *Statistical and Computational Inverse Problems*, Springer-Verlag, New York, NY, **2005**. https://doi.org/10.1007/b138659

[40] M. Allmaras, W. Bangerth, J. M. Linhart, J. Polanco, F. Wang, K. Wang, J. Webster, S. Zedler, *SIAM Rev.* **2013**, *55*, 149.

[41] O. Aguilar, M. Allmaras, W. Bangerth, L. Tenorio, *SIAM Rev.* **2015**, *57*, 131.

[42] M. Dashti, A. M. Stuart, *The Bayesian Approach to Inverse Problems*. **2015**.

[43] J. Prager, H. N. Najm, K. Sargsyan, C. Safta, W. J. Pitz, *Combust. Flame* **2013**, *160*, 1583. https://doi.org/10.1016/j.combustflame.2013.01.008

[44] L. Cai, H. Pitsch, S. Y. Mohamed, V. Raman, J. Bugler, H. Curran, S. M. Sarathy, *Combust. Flame* **2016**, *173*, 468. https://doi.org/10.1016/j.combustflame.2016.04.022

[45] L. Hakim, G. Lacaze, M. Khalil, H. N. Najm, J. C. Oefelein, *J. Eng. Gas Turbine. Power* **2016**, *138*, 112806. https://doi.org/10.1115/1.4033502

[46] E. Cisneros-Garibay, C. Pantano, J. B. Freund, *Combust. Flame* **2019**, *208*, 219. https://doi.org/10.1016/j.combustflame.2019.06.028

[47] K. Braman, T. A. Oliver, V. Raman, *Combust. Theory Model.* **2013**, *17*, 858. https://doi.org/10.1080/13647830.2013.811541

[48] L. Hakim, G. Lacaze, M. Khalil, K. Sargsyan, H. Najm, J. Oefelein, *Combust. Theory Model.* **2018**, *22*, 446. https://doi.org/10.1080/13647830.2017.1403653

[49] J. Zádor, I. G. Zsély, T. Turányi, M. Ratto, S. Tarantola, A. Saltelli, *J. Phys. Chem. A* **2005**, *109*, 9795. https://doi.org/10.1021/jp053270i
[50] N. Armstrong, D. Hibbert, *Chemom. Intell. Lab. Syst.* **2009**, *97*, 194. https://doi.org/10.1016/j.chemolab.2009.04.001
[51] D. Hibbert, N. Armstrong, *Chemom. Intell. Lab. Syst.* **2009**, *97*, 211. https://doi.org/10.1016/j.chemolab.2009.03.009
[52] W. Shao, X. Tian, *Chem. Eng. Res. Des.* **2015**, *95*, 113. https://doi.org/10.1016/j.cherd.2015.01.006
[53] M.-Y. Fan, Y.-L. Zhang, Y.-C. Lin, J. Li, H. Cheng, N. An, Y. Sun, Y. Qiu, F. Cao, P. Fu, *Environ. Sci. Technol. Lett.* **2020**, *7*, 883. https://doi.org/10.1021/acs.estlett.0c00623
[54] E. Gallicchio, M. Andrec, A. K. Felts, R. M. Levy, *J. Phys. Chem. B* **2005**, *109*, 6722. https://doi.org/10.1021/jp045294f
[55] H. O. Lloyd-Laney, N. D. J. Yates, M. J. Robinson, A. R. Hewson, J. D. Firth, D. M. Elton, J. Zhang, A. M. Bond, A. Parkin, D. J. Gavaghan, *Anal. Chem.* **2021**, *93*, 2062. https://doi.org/10.1021/acs.analchem.0c03774
[56] A. Puliyanda, K. Sivaramakrishnan, Z. Li, A. de Klerk, V. Prasad, *React. Chem. Eng.* **2020**, *5*, 1719. https://doi.org/10.1039/d0re00147c
[57] R. H. Johnstone, E. T. Chang, R. Bardenet, T. P. de Boer, D. J. Gavaghan, P. Pathmanathan, R. H. Clayton, G. R. Mirams, *J. Mol. Cell. Cardiol.* **2016**, *96*, 49. https://doi.org/10.1016/j.yjmcc.2015.11.018
[58] D. J. Gavaghan, J. Cooper, A. C. Daly, C. Gill, K. Gillow, M. Robinson, A. N. Simonov, J. Zhang, A. M. Bond, *ChemElectroChem* **2017**, *5*, 917. https://doi.org/10.1002/celc.201700678
[59] N. Galagali, Y. M. Marzouk, *Chem. Eng. Sci.* **2015**, *123*, 170. https://doi.org/10.1016/j.ces.2014.10.030
[60] R. D. Berry, H. N. Najm, B. J. Debusschere, Y. M. Marzouk, H. Adalsteinsson, *J. Comput. Phys.* **2012**, *231*, 2180. https://doi.org/10.1016/j.jcp.2011.10.031
[61] F. T. Bölle, A. E. G. Mikkelsen, K. S. Thygesen, T. Vegge, I. E. Castelli, *NPJ Comput. Mater.* **2021**, *7*. 41. https://doi.org/10.1038/s41524-021-00505-9
[62] H. Kaneko, K. Funatsu, *Chemom. Intell. Lab. Syst.* **2014**, *137*, 57. https://doi.org/10.1016/j.chemolab.2014.06.008
[63] N. Sun, R. J. Carroll, H. Zhao, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7988. https://doi.org/10.1073/pnas.0600164103
[64] K. Sargsyan, H. N. Najm, R. Ghanem, *Int. J. Chem. Kinet.* **2015**, *47*, 246. https://doi.org/10.1002/kin.20906
[65] C. B. Whitehead, S. Özkar, R. G. Finke, *Chem. Mater.* **2019**, *31*, 7116. https://doi.org/10.1021/acs.chemmater.9b01273
[66] C. B. Whitehead, S. Özkar, R. G. Finke, *Mater. Adv.* **2021**, *2*, 186. https://doi.org/10.1039/d0ma00439a
[67] C. B. Whitehead, M. A. Watzky, R. G. Finke, *J. Phys. Chem. C* **2020**, *124*, 24543. https://doi.org/10.1021/acs.jpcc.0c06875
[68] V. K. LaMer, R. H. Dinegar, *J. Am. Chem. Soc.* **1950**, *72*, 4847.
[69] S. Özkar, R. G. Finke, *J. Am. Chem. Soc.* **2017**, *139*, 5444. https://doi.org/10.1021/jacs.7b00958
[70] W. W. Laxson, R. G. Finke, *J. Am. Chem. Soc.* **2014**, *136*, 17601. https://doi.org/10.1021/ja510263s
[71] M. A. Watzky, E. E. Finney, R. G. Finke, *J. Am. Chem. Soc.* **2008**, *130*, 11959. https://doi.org/10.1021/ja8017412
[72] A. F. Schmidt, V. V. Smirnov, *Top. Catal.* **2005**, *32*, 71. https://doi.org/10.1007/s11244-005-9261-4
[73] J. M. Lee, R. C. Miller, L. J. Moloney, A. L. Prieto, *J. Solid State Chem.* **2019**, *273*, 243. https://doi.org/10.1016/j.jssc.2018.12.053
[74] R. E. Kass, A. E. Raftery, *J. Am. Stat. Assoc.* **1995**, *90*, 773. https://doi.org/10.1080/01621459.1995.10476572
[75] A. Gelman, X.-L. Meng, H. Stern, *Stat. Sin.* **1996**, *6*, 733.
[76] D. Long, W. Bangerth, dklong-csu/mepbm: ME-PBM version 1.0 **2021**, https://zenodo.org/record/4970247.
[77] A. Gelman, G. O. Roberts, W. R. Gilks, *Bayesian Stat.* **1996**, *5*, 599.
[78] G. O. Roberts, J. S. Rosenthal, *Stat. Sci.* **2001**, *16*, 351. https://doi.org/10.1214/ss/1015346320
[79] A. Sokal, *Functional Integration*, Springer, Boston, MA **1997**, p. 131.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: D. K. Long, W. Bangerth, D. R. Handwerk, C. B. Whitehead, P. D. Shipman, R. G. Finke, *J. Comput. Chem.* **2021**, 1. https://doi.org/10.1002/jcc.26770