On the Statistical Calibration of Physical Models

K. SARGSYAN,[1] H. N. NAJM,[1] R. GHANEM[2]

[1]*Sandia National Laboratories, Livermore*, CA 94550

[2]*University of Southern California, Los Angeles*, CA 90089

Received 23 August 2014; revised 22 January 2015; accepted 22 January 2015

DOI 10.1002/kin.20906
Published online 23 February 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: We introduce a novel statistical calibration framework for physical models, relying on probabilistic embedding of model discrepancy error within the model. For clarity of illustration, we take the measurement errors out of consideration, calibrating a chemical model of interest with respect to a more detailed model, considered as "truth" for the present purpose. We employ Bayesian statistical methods for such model-to-model calibration and demonstrate their capabilities on simple synthetic models, leading to a well-defined parameter estimation problem that employs approximate Bayesian computation. The method is then demonstrated on two case studies for calibration of kinetic rate parameters for methane air chemistry, where ignition time information from a detailed elementary-step kinetic model is used to estimate rate coefficients of a simple chemical mechanism. We show that the calibrated model predictions fit the data and that uncertainty in these predictions is consistent in a mean-square sense with the discrepancy from the detailed model data. © 2015 Wiley Periodicals, Inc. Int J Chem Kinet 4: 246–276, 2015

INTRODUCTION

Chemical rate coefficients are most commonly estimated from experimental measurements, e.g., using shock tube tests [1–3], or via analytic computations using established data sets [4–6]. Values for these coefficients are typically estimated using least-squares (LS) fitting of model output observables to corresponding experimental measurements [7,8]. Together with the best fit value, the LS formalism provides an estimate of the second-order error statistics and a corresponding Gaussian probability measure [9]. However, resulting uncertainty estimates rely on underlying assumptions, such as Gaussian noise and linear model dependence on parameters. Moreover, these methods provide at best ad hoc means of handling nuisance parameters,e.g., uncertain parameters available from prior measurements that are required for the data fitting but not otherwise of interest in the measurement. On the other hand, Bayesian inference methods [10–13] provide a more general framework for model calibration and parameter estimation. They also provide a robust framework for handling multiple sources of information, different data types, sequential learning and updating of prior information, as well as handling of nuisance parameters. Besides, the established deterministic methods for parameter estimation can be cast equivalently within a Bayesian formulation by treating the deterministic

Correspondence to: Khachik Sargsyan; e-mail: ksargsy@sandia.gov.

objective function as a basis for a Bayesian likelihood. With these evident capabilities, Bayesian methods have become increasingly popular in the recent chemical kinetic literature [2,3,14–24].

In the context of parameter estimation, the Bayesian framework permits the update of existing probabilistic knowledge about the parameters as new knowledge is acquired. This update is carried out in a manner that is consistent with the axioms of probability and does not necessitate an extrinsic interpretation of optimality such as required by LS methods, for instance. More specifically, and starting with prior probability density function (PDF), $\pi(\boldsymbol{\lambda})$, on the parameters of a model, together with the likelihood of the data, $\mathcal{D}$, given a value for $\boldsymbol{\lambda}$, a Bayes rule yields the joint posterior PDF on $\boldsymbol{\lambda}$. The likelihood function, $L(\boldsymbol{\lambda}) = p(\mathcal{D}|\boldsymbol{\lambda})$, is an expression of the conditional likelihood of the data given the fit-model and a specific value of $\boldsymbol{\lambda}$. As opposed to classical optimization approaches that typically provide deterministic results for parameters $\boldsymbol{\lambda}$, Bayesian methods lead to a probabilistic description, encapsulated in a joint posterior PDF on $\boldsymbol{\lambda}$ that takes into account uncertainties due to *lack of knowledge* or insufficient data. Employment of Bayesian inference typically requires many model evaluations, but the expense can be somewhat alleviated by the use of precomputed surrogate models [25] that approximate the responses of the model over parameter ranges of interest and effectively replace the model under consideration in search-based algorithms, similar to the solution mapping technique [26], which has been used for chemical model calibration in the context of deterministic LS-type objective functions.

Furthermore, it is well documented [3,7,14,17,20, 21,24,27–30] thatb manyb parameter estimation studies implicitly assume that the model is *perfect*, i.e. it exactly replicates the true behavior of an observable of interest. The Bayesian framework allows for model uncertainty assessment for subsequent model selection, e.g., via evidence computation [21,31], model class plausibility [20,32,33], or information criteria [24,34]. To a lesser extent, this is also true for deterministic parameter estimation methods, e.g., using model fitness [27], or again information criteria such as Akaike Information Criterion [35,36]. Also, the data collaboration framework [27,28] deals with the data-driven inference of model parameters in a deterministic, interval-prediction context and accepts the prediction bounds as a measure of model quality [37,38]. However, all these methods do not explicitly quantify the key driver behind model uncertainty assessment and model selection that is the *model error*, i.e. the discrepancy of the model from the reality, nor do they provide predictive uncertainty that is representative of the extent of model error. In a deterministic LS parameter estimation context, systematic offsets that mimic model error are dealt with in [7] to fit multiple kinetics data sets within the same model. In [2,17,18], the model error is recognized but effectively folded into the experimental error term. It is indeed difficult to explicitly quantify model errors as one would need to know the "truth" to evaluate the discrepancies present in models. The work of Kennedy and O'Hagan [39] comes closest, building a Bayesian calibration machinery by explicitly modeling the discrepancy term as a Gaussian process. However, this explicit additive term may break physical/chemical laws, gets entangled with the measurement error, and does not improve the predictive utility of a model generally as it is finetuned toward a specific observable of interest only.

We note here that, while there are many scenarios where model error is small, such that the conventional strategy of ignoring it is valid to some extent, we are particularly concerned here with highly simplified/reduced chemical kinetic models for hydrocarbon fuels that are typically necessary, for computational tractability, in large-scale turbulent combustion computations [40,41]. In this context, it is understood that agreement with the detailed underlying physics, as provided e.g., by calibration with respect to a "validated" detailed chemical kinetic mechanism, can only be ensured on a small subset of observables and only to within some coarse degree of accuracy. More broadly, simplified/skeletal kinetic models are typically produced employing some strategy, e.g., [42–45], whereby they are compared to databases of detailed model solutions over a range of operating/initial conditions, employing a specific subset of target species, and with chosen error tolerances. It is clear that agreement with *all* detailed model outputs is not guaranteed nor is accurate prediction of *chosen* targets guaranteed beyond chosen error thresholds.

In this work, we introduce a Bayesian calibration framework that is tuned toward improving the predictive capabilities of a model by burdening specific model parameterizations with model error terms, rather than assuming an explicit additive component. This framework does not break any physical constraints by construction, has the potential for disambiguation of model error from the errors associated with measurement noise and, is well-suited for model-to-model calibration studies as demonstrated in this paper.

Consider the *truth* function $y = g(x)$, where x and y are real scalars, with x being the independent variable (e.g., a spatial coordinate), and y an observable dependent variable. The choice of scalars, rather than vectors, is for convenience, and has no impact on the generality of the discussion. Furthermore, consider a

parameterized model for representing the (x, y) dependence given by $y_m = f(x;\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a real parameter vector, and the m subscript indicates the prediction of the observable y by the model $f()$. Let $\{(x_i, y_i), i = 1, \ldots, N\}$ be the set of N measurement pairs. We can write, for purposes of illustration, and following Kennedy and O'Hagan [39],

$$y_i = \underbrace{f(x_i;\boldsymbol{\lambda}) + \varepsilon_{mi}}_{g(x_i)} + \varepsilon_{di} \tag{1}$$

where $\{\varepsilon_{mi}\}_{i=1}^N$ is the sequence of model error values, where model error is the discrepancy between the model prediction and the truth, and $\{\varepsilon_{di}\}_{i=1}^N$ is the sequence of data error values, where data error is the discrepancy between the measurement and the truth. It is typical to refer to both ε_d and ε_m, each a vector of length N, as error terms and to endow each of them with a multivariate statistical structure. For example, for an instrument with Gaussian noise, one may model $\varepsilon_d \sim N(\mu_d, \Sigma_d)$, where μ_d is the bias and Σ_d is the covariance matrix, which captures the dependence of the noise structure on the measurement instrument details. The construction of an accurate ε_d statistical model can be challenging. However, with sufficient availability of measurement data, with a discernible noise structure, at least some aspects of the statistical structure of ε_d may be observable. In contrast, the assignment of a statistical structure for model error ε_m is faced with the difficulty that there is no underlying associated observable noise structure to justify any specific choice of distribution or correlation structure. Accordingly, a significant degree of modeling is involved. Furthermore, and particularly when considering physically based models with associated constraints, the addition of the model discrepancy error term as a statistical noise on model predictions is challenging. Consider, for example, a physical system such as incompressible flow, whose output y is required to satisfy implicitly a divergence free constraint $\nabla\bullet\, y = 0$. In this case, presuming that $y_m = f(x, \boldsymbol{\lambda})$ is divergence-free by construction, declaring $y = g(x) = f(x;\boldsymbol{\lambda}) + \varepsilon_m(x)$ requires that the statistical model $\varepsilon_m(x)$ equally satisfies the requisite constraint, which can be challenging. Analogous arguments can be considered for initial and boundary conditions. These limitations have been recognized in [46], where various challenges with the statistical additive error term are spelled out, and Bayesian constraint-driven priors are suggested as a potential remedy to respect physics-based constraints. The construction of such priors may not always be feasible, however, in general. Another issue with this ε_m construction is that, in fact the calibrated model is $f(x, \boldsymbol{\lambda}) + \varepsilon_m(x)$, and not simply the original governing equations $f(x, \boldsymbol{\lambda})$ with the posterior $p(\boldsymbol{\lambda}|\mathcal{D})$. This, from a practical context, is problematic, because it is altering the model in a somewhat arbitrary fashion. Given the investment scientists and engineers make in developing and justifying physical model structure, it is not likely that this approach would be acceptable. Finally, this approach provides a model discrepancy correction to only the one observable being measured. In the case of models intended for predicting other, nonobservable quantities of interest, there is no provision for an associated discrepancy correction. This set of difficulties argues very strongly for alternate means of employing model discrepancy error in the calibration of models of physical systems.

One such alternate approach recognizes that, for physical models, it may be more appropriate to embed model discrepancy error terms in elements of $f()$ where phenomenological or similarly approximate constructions are used. As a step in this direction, Berliner and co-workers [47] employed a Bayesian statistical approach with a model for ice glacier transport, where the model was calibrated with specific embedded proposed alterations exploring potential phenomenologies. However, these were embedded deterministic parameters and not random model discrepancy error terms per the present discussion. Along the same lines, but more generally, embedding model discrepancy error terms in the model pursues alteration of model structure in a manner that is both physically acceptable and informative. For example, calibration with different proposed embeddings of ε_m in select submodels or phenomenological model elements enables the identification of the model element where discrepancy/model-update can potentially explain and correct for error in predicted observables. Furthermore, the proposed model changes are made in the context of updating underlying assumptions, rather than simply additional statistical terms on output predictions. Finally, this approach also immediately impacts all model outputs. Recent work by Strong et al. [48,49] has recognized the need for zooming into discrepancies driven by component submodels. However, these studies rely on expert opinion and are strongly tied to the specific context of health-economic decision models. Others [50–52] also recognized the need to deal with model error "internally" via exploring physics improvements or embedding stochasticity into the model. Dwight et al. [53] implemented a Bayesian calibration framework in the simplest, x-independent setting, where they calibrate for an existing *aleatoric* model parameter. In other words, the primary target was not model error assessment but the evaluation of a parameter that already has a structural variability dictated

by physics. Similarly, a density estimation framework was employed in [33], albeit restricted to five PDF classes, employing Bayesian machinery to infer PDF parameters and find the best class of PDFs.

In the present context, we explore the utility of such embedded model discrepancy error representations in the context of calibration of chemical systems and estimation of reaction rate coefficients. More specifically, and to focus on model errors, we consider the context of no-noise calibration of one (simple) model $f(x, \boldsymbol{\lambda})$ against computational data produced from some more detailed model $g(x)$. The latter is considered as the "truth" function in the present context. In this case, ε_m is the only error term to be considered, as there is no measurement noise, i.e. $\varepsilon_d = 0$ in Eq. (1). In principle, ε_m can be embedded anywhere in $f(x, \boldsymbol{\lambda})$, according to the proposed choice of the modeler. Thus, e.g., a model with altered phenomenology could be proposed employing the model error term in specific model elements of interest, $y = \tilde{f}(x, \boldsymbol{\lambda}, \varepsilon_m)$. Alternately, and for a simpler context, one can consider the construction $y = f(x, \boldsymbol{\lambda} + \varepsilon_m)$. In this case, $\boldsymbol{\lambda} + \varepsilon_m$ is a random variable arrived at by a $\boldsymbol{\lambda}$-shift in the mean of ε_m. This is equivalent to the original model $y = f(x, \boldsymbol{\lambda})$, but where $\boldsymbol{\lambda}$ is a random variable. To fix ideas, and to focus on a simple context, we consider this specific choice here, where the model error is embedded in the parameters $\boldsymbol{\lambda}$. This is not a requirement for the proposed construction, but is one choice selected for illustration.b It is well understood that this embedding of ε_m in existing model parameters may not be a sufficiently flexible approach in general. For example, the model $f(x, \boldsymbol{\lambda})$ may not be consistent with data from $g(x)$ in any meaningful sense for any value of $\boldsymbol{\lambda}$. In this case, the more general proposed embedding in new phenomenology/physics above would be more appropriate. Still, the simplicity of the present approach is advantageous in introducing and analyzing the method and is to be understood as a special case where the model has sufficient flexibility to be in some sense consistent with the data. Note that, in this context, $\boldsymbol{\lambda}$ is no longer a vector of unknown parameters to be estimated. Rather, it is a random vector, whose characteristics (density, moments, etc.) are to be estimated. Effectively, then the problem is transformed to a density estimation problem, where we seek to estimate parameters α that define the density of $\boldsymbol{\lambda}$, denoted by $\pi_{\boldsymbol{\lambda}}(\cdot; \alpha)$, such that the pushed forward posterior (PFP) distributions $p(\boldsymbol{\lambda}|\mathcal{D})$, and, consequently, $p(f(x, \boldsymbol{\lambda})|\mathcal{D})$ are in some sense consistent with the $g(x)$ data $\mathcal{D}$. *

The definition of this consistency is of some relevance. In the Bayesian context, $y = f(x, \boldsymbol{\lambda})$ is the model for the data, whose density is the likelihood $p(f(x, \boldsymbol{\lambda})|\mathcal{D})$ employed in Bayes formula. We will discuss below challenges with Bayesian implementations employing alternate forms/approximations of this likelihood. Given these challenges, we outline in the following an approach that targets-specific constraints of interest on the resulting calibrated model, implemented in an approximate Bayesian computational context.

From a practical perspective, we consider a particular objective behind calibration of the simple model, namely to enable its use for prediction with uncertainty, accounting for model error, with requisite consistency between its uncertain predictions and those from the detailed (truth) model.† Obviously, if $f(x, \boldsymbol{\lambda})$ had the exact structure of $g(x)$, then the ideal result of calibration is the *true* values of the $\boldsymbol{\lambda}$ parameters of $f()$. In this context, $\varepsilon_m \equiv 0$ and $p(\boldsymbol{\lambda}|\mathcal{D})$ is a delta function centered on the true values of the parameters. Accordingly, $p(f(x, \boldsymbol{\lambda})|\mathcal{D})$ would be a delta function centered on $g(x)$, for any x. On the other hand, when model discrepancy is finite, it is natural to seek a $p(f(x, \boldsymbol{\lambda})|\mathcal{D})$ that is, in a LS sense, centered on $g(x)$, and that has a width that is consistent with the discrepancy between $g(x)$ and the mean prediction $\mu(x) \equiv \mathrm{E}_{\boldsymbol{\lambda}}[f(x, \boldsymbol{\lambda})]$, again in a LS sense. In this manner, uncertain model predictions are consistent in the mean with the truth and their degree of uncertainty is consistent with the discrepancy from the truth. These requirements, which will be made more precise below, provide the key guidance behind the proposed density estimation construction.

While we allow for general parameterization of $p(\boldsymbol{\lambda}|\mathcal{D})$, in this work we assume $\boldsymbol{\lambda}$ obeys a multivariate normal distribution with parameters α that uniquely define the mean vector and the covariance matrix of $\boldsymbol{\lambda}$. This is a special case of spectral polynomial chaos expansion (PCE) as a general representation of any finite-variance random vector [54,55]. Spectral

*To be clear, we note that, in the conventional use of Bayesian inference for model calibration and parameter estimation, while one estimates a posterior density $p(\boldsymbol{\lambda}|\mathcal{D})$ on the parameters $\boldsymbol{\lambda}$ of interest, one is strictly inferring the *parameter* and not its density. Thus, presuming a consistent Bayesian implementation, as more data is taken, the posterior density width shrinks, tending to a delta function at the true parameter value $\boldsymbol{\lambda}^*$ in the infinite data limit. On the other hand, in the present context of *density estimation*, the object of inference is the parameters α defining the density $\pi_{\boldsymbol{\lambda}}(\cdot; \alpha)$, and the posterior density is on α. Thus, additional data would provide better information on α, with the infinite-data limit providing a delta function posterior centered on the true value α^*. Therefore, with infinite data, one has perfect information on α; however, $\pi_{\boldsymbol{\lambda}}(\cdot; \alpha)$ is not a delta function; it is rather a density whose parameters are perfectly known.

†Note that, if the objective is specifically physical parameter estimation, with no particular regard for model error and uncertain predictive utility, then the present approach, encumbering model parameters with model error, may not be the best choice.

expansions have been extensively used for forward propagation of uncertainties in chemical kinetic models [22,23,56–71] and have recently been employed in the calibration studies via the the MUM-PCE method (Method of Uncertainty Minimization) [29,30,72–75]. It can be shown [21] that MUM-PCE is a few special assumptions away from a general Bayesian formulation. More specifically, we note that the objective function in MUM-PCE is similar to the likelihood presented in this work and aims at finding the best spectral expansion for model parameters by matching moments of predictive uncertainties of the model with moments of the observational errors. The present development differs in a number of key respects. To begin with, it is fully Bayesian, in the sense that we employ an (approximate) likelihood function, rather than a deterministic objective function. Second, we do not explicitly seek agreement of predictive uncertainties with *observational* errors. Rather, we focus on model error explicitly as a motivation for casting deterministic input parameters as uncertain with the express purpose of matching the predictive uncertainties and capturing the consequences of model error in predictions.

Given a parameterization of $p(\boldsymbol{\lambda}|\mathcal{D})$, as $\pi_\lambda(\cdot;\alpha)$, specifically via $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\cdot;\alpha)$, where α is a parameter vector, the object of the calibration procedure becomes the estimation of α. This, of course, can be done in a LS fitting framework, seeking the best fit α. We propose, rather, to remain in the Bayesian framework, seeking estimation of $p(\alpha|\mathcal{D})$. This provides an estimate of uncertainty in α whose reduction with additional data utilization can be examined to provide guidance on the quality of the α estimation. It also allows us to take advantage of the Bayesian formalism for purposes of model comparison/selection, and for future extensions to contexts where both model discrepancy and data noise are present.

In the third section, we outline the proposed construction in detail and demonstrate it in a setting where we calibrate a "simple" model $f(x, \boldsymbol{\lambda})$ against noise-free data generated by a "detailed" model $g(x)$. Note that there are a number of challenges. From a fundamental perspective, for example, given the lack of noise in the data, there is clearly no local information on the PDF structure at any given x. Furthermore, the full likelihood is both very expensive to evaluate, and, more critically, degenerate, as will be shown below and also described in a different context in [76]. Moreover, a well-defined alternative, the marginalized likelihood is shown to be not that useful without prior regularizations. We will outline a number of optional means of solving the proposed density estimation problem, discussing associated challenges, arriving eventually at an effective construction that makes use of approximate Bayesian computation (ABC) methods. We will detail the method, followed by illustrations of the performance of the construction with synthetic examples, before demonstrating the algorithm in the context of calibration of a simple chemical model for methane oxidation, relative to a detailed model, in an ignition problem. We rely on ignition time data for a homogeneous methane-air mixture, computed using the detailed model, over a range of initial temperature and equivalence ratio, at atmospheric pressure. These data are then used to calibrate the simple chemical model and to compare resulting uncertain predictions with those from the detailed model, outlining the utility of the present construction.

The paper is organized as follows. In the next section, we briefly describe the conventional calibration mechanism with explicit Gaussian representation of the model-to-data discrepancy. Next, we introduce our calibration approach that embeds that discrepancy within model parameters probabilistically leading to a density estimation problem. The appendix summarizes the predictive quantities associated with this calibration approach and derives their moments. Furthermore, we demonstrate the approach on synthetic cases that help dissect the behavior of the new method. Then, the calibration mechanism is applied to a chemical kinetic model with ignition time data. The discussion and conclusions are presented in the last two sections.

CONVENTIONAL LIKELIHOOD CONSTRUCTION

In the Bayesian inverse modeling framework, consider data that is the set of detailed ("truth") model simulations y_i at different conditions x_i, i.e., $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$. Let us first introduce the conventional calibration mechanism, in which one assumes a statistical model error term, with some endowed structure, superposed on the model output. For illustration purposes, and with no loss of generality of the resulting conclusions, we consider here an additive Gaussian discrepancy on the collected output data. Thus, we assume a statistical discrepancy, between the truth model $g(x)$ and the model $f(x;\boldsymbol{\lambda})$ that is being calibrated using data $\mathcal{D}$, as follows:

$$g(x_i) = f(x_i;\boldsymbol{\lambda}) + \varepsilon_m(x_i) \tag{2}$$

where ε_m is generally endowed with some correlation structure. Here again, for illustration purposes, let the $\varepsilon_m(x_i)$ be independent identically distributed (i.i.d.) Gaussian errors with fixed variance σ^2. Often, there

are no direct indicators suggesting a good value for σ. Therefore, it is natural to add it as a hyperparameter to the set of inferred parameters $\boldsymbol{\lambda}$. In fact, to ensure positivity, and to better regularize the inference problem, it is common to employ $\ln\sigma$ as a hyperparameter instead of σ. In a Bayesian formulation, this leads to a posterior distribution for $(\boldsymbol{\lambda}, \ln\sigma)$. Presuming independent priors $p(\boldsymbol{\lambda}), p(\ln\sigma)$, we have

$$\underbrace{p(\boldsymbol{\lambda}, \ln\sigma|\mathcal{D})}_{\text{posterior}} \propto \underbrace{L_{\mathcal{D}}(\boldsymbol{\lambda}, \ln\sigma)}_{\text{likelihood}} \underbrace{p(\boldsymbol{\lambda})p(\ln\sigma)}_{\text{prior}} \tag{3}$$

Bayes rule (3) relates the prior probability distribution on $\boldsymbol{\lambda}$ and $\ln\sigma$ to the posterior distribution in light of data $\mathcal{D}$, via the likelihood function $L(\boldsymbol{\lambda}, \ln\sigma) = p(\mathcal{D}|\boldsymbol{\lambda}, \ln\sigma)$. The latter is defined according to the assumptions made for the error ε_m, i.e.,

$$L_{\mathcal{D}}(\boldsymbol{\lambda}, \ln\sigma) = p(\mathcal{D}|\boldsymbol{\lambda}, \ln\sigma) = \frac{1}{(2\pi)^{N/2}\sigma^N} \prod_{i=1}^{N} \times \exp\left(-\frac{(g(x_i) - f(x_i;\boldsymbol{\lambda}))^2}{2\sigma^2}\right) \tag{4}$$

Note thatb the logarithm of this likelihoodb is—within an additive shift and a multiplicative factor—the LS objective function that is very common for deterministic parameter estimation studies [7–9,15,26]. The prior then corresponds to regularization terms that may be used in the deterministic optimization. This suggests the maximum value of the posterior is exactly the same as the maximum of the LS objective function with a regularization term. We will employ uniform priors on both $\boldsymbol{\lambda}$ and $\ln\sigma$. With the posterior density in hand, the marginal posterior distribution $p(\boldsymbol{\lambda}|\mathcal{D}) = \int_{-\infty}^{\infty} p(\boldsymbol{\lambda}, \ln\sigma|\mathcal{D})d\ln\sigma$ can then be pushed through the model $f(x_i;\boldsymbol{\lambda})$ to produce probabilistic predictions of the model outputs. However, in many practical situations, e.g., when highly simplified/reduced models are necessary for computational tractability in turbulent combustion computations [40,41], the model is clearly not perfect and it would not replicate the "truth", e.g., detailed kinetic model predictions, to a high degree of accuracy no matter what the chosen values for the parameters $\boldsymbol{\lambda}$ are. In such cases, one expects the posterior predictive (PP) distribution [77], being the push-forward of $p(\boldsymbol{\lambda}, \ln\sigma|\mathcal{D})$ through the data model $y = f(x, \boldsymbol{\lambda}) + \varepsilon_m(x)$, to exhibit a degree of uncertainty that is consistent with the discrepancy from the data. However, the PFP, being the push-forward of $p(\boldsymbol{\lambda}, \ln\sigma)$ through the model $y = f(x, \boldsymbol{\lambda})$, which we claim is more relevant from a physical modeling perspective, is generally inconsistent with the data discrepancy. Thus, in this context, not only are the mean PFP predictions off because of model error, but their uncertainty estimate is misleading and does not account for the model's inability to match the truth, i.e. model error. This deficiency of the PFP prediction based on conventional calibration mechanisms will be illustrated further in the context of a chemical ignition model, together with the benefits of the proposed novel calibration approach described in the next section.

METHOD

Again, assume that the "truth" values of a quantity of interest $g(x)$ are modeled with a simpler model $f(x;\boldsymbol{\lambda})$ with parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ using data set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, where $y_i = g(x_i)$. We seek an approximate relation

$$g(x) \approx f(x;\boldsymbol{\lambda}) \tag{5}$$

by casting $\boldsymbol{\lambda}$ as an uncertain quantity with PDF $\pi_{\boldsymbol{\lambda}}(\cdot;\alpha)$ and turning the problem into one of *density estimation*,[‡], i.e. finding parameters α that define the PDF of $\boldsymbol{\lambda}$. We parameterize the random variable vector $\boldsymbol{\lambda}$ (or its PDF) with parameters α in polynomial chaos (PC) spectral expansion form, that provide a common framework [54,55] for representing a finite-variance random variable ζ as a polynomial expansion

$$\zeta = \sum_k c_k \psi_k(\xi) \tag{6}$$

with respect to a standard random variable *germ* ξ and associated polynomials $\psi_k(\xi)$ that are orthogonal with respect to the PDF of ξ, i.e.,

$$\langle \psi_i \psi_j \rangle \equiv \int_{\Xi} \psi_i(\xi)\psi_j(\xi)\pi(\xi)d\xi = \delta_{ij}\langle \psi_i^2 \rangle \tag{7}$$

The most commonly used PC expansions (PCEs) employ either Legendre-uniform (LU) or Hermite-Gauss (HG) polynomial-germ pairs. The PCE (6) can be generalized to represent a random vector $\boldsymbol{\zeta}$ as an expansion of multivariate polynomials of standard i.i.d.variables ξ_i that form the vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$,

$$\zeta_i = \sum_{\boldsymbol{k}} c_{\boldsymbol{k}}^{(i)} \Psi_{\boldsymbol{k}}(\xi_1, \ldots, \xi_d) \tag{8}$$

[‡]Note that here and throughout the paper, we denote the PDF of a random quantity X by $\pi_X(\cdot;\beta)$, where β denotes the set of parameters defining the PDF or the random variable X.

where $\boldsymbol{k} = (k_1, \ldots, k_d)$ is the multiindex corresponding to the multivariate polynomial $\Psi_{\boldsymbol{k}}(\xi_1, \ldots, \xi_d) = \psi_{k_1}(\xi_1)\psi_{k_2}(\xi_2)\ldots\psi_{k_d}(\xi_d)$. The PCE is essentially a parameterization of the random quantity $\boldsymbol{\zeta}$ by a set of deterministic coefficients $c_{\boldsymbol{k}}^{(i)}$. The orthogonality of the basis polynomials $\Psi_{\boldsymbol{k}}$ allows easy extraction of the first two moments of ζ_i for all $i = 1, \ldots, d$, leading to formulae for means and variances:

$$\mathrm{E}[\zeta_i] = c^{(i)}_{\substack{\rightarrow \\ 0}} \quad \text{and} \quad \mathrm{V}[\zeta_i] = \sum_{\boldsymbol{k} \neq \substack{\rightarrow \\ 0}} \left(c_{\boldsymbol{k}}^{(i)}\right)^2 \langle \Psi_{\boldsymbol{k}}^2 \rangle \tag{9}$$

where, as defined above, $\langle\cdot\rangle$ denotes the average with respect to the PDF of $\boldsymbol{\xi}$.

For the demonstrations in this work, we will use first-order HG PCEs for the components of $\boldsymbol{\lambda}$, effectively assuming a multivariate normal distribution for $\boldsymbol{\lambda}$:

$$\begin{aligned}
\lambda_1 &= a_{10} + a_{11}\xi_1 \\
\lambda_2 &= a_{20} + a_{21}\xi_1 + a_{22}\xi_2 \\
&\vdots \\
\lambda_d &= a_{d0} + a_{d1}\xi_1 + a_{d2}\xi_2 + \ldots + a_{dd}\xi_d
\end{aligned} \tag{10}$$

Note that, to avoid rotational invariance, we used a "triangular" form in the PC description. The parameters $\alpha = \{a_{jk}\}_{j=1,\ldots,d}^{k=0,\ldots,j}$ define a random vector $\boldsymbol{\lambda}$ with multivariate PDF $\pi_{\boldsymbol{\lambda}}(\cdot;\alpha)$. As a density estimation problem, the parameters α will be the objects of inference in the Bayesian context. Namely, Bayes' formula in this case reads

$$\underbrace{p(\alpha|\mathcal{D})}_{\text{posterior}} \propto \underbrace{p(\mathcal{D}|\alpha)}_{\text{likelihood}} \underbrace{p(\alpha)}_{\text{prior}} \tag{11}$$

We will employ i.i.d. uniform priors for components of α, i.e.

$$p(\alpha) = \prod_{j=1}^{d}\prod_{k=0}^{j} p(a_{jk}) = \text{ const.} \tag{12}$$

Note that in Eq. (10) simultaneous flips of signs in parameters $a_{kk}, a_{k+1,k}, \ldots, a_{dk}$, $\forall k > 0$ lead to the same joint distribution for $\boldsymbol{\lambda}$. This invariance leads to multimodal posteriors that make the posterior sampling harder to tune. Therefore, for convenience, we use positive priors for the nonleading PC coefficients of the last element λ_d, i.e., $p(a_{dk}) = 0$ for $a_{dk} < 0$ and $k > 0$, to ensure unimodal posteriors. The prior (12) is improper as its integral is not finite. However, this does not affect the feasibility of the Bayesian approach. An appropriate choice of priors is essential for Bayesian machinery. The above symmetry-driven restrictions make use of the very structure by which α affects the model. More informative priors can in principle be employed, driven primarily by expert knowledge about, say, moments of the parameters λ_i, thus constraining the PC coefficients of λ_i in Eq. (10), i.e. providing further information on the vector α. The key component of the Bayesian framework—and the one we will focus on in this work—is the likelihood as a function of parameters α given the current data set $\mathcal{D}$ of the "truth" model values. The data $\mathcal{D}$ consist of values of the *deterministic*, "truth" model $g(x)$, whereas the likelihood function provides a means of comparing these output values of $g(x)$ with the output of the simpler model-to-be-calibrated $f(x, \boldsymbol{\lambda})$ with embedded probabilistic $\boldsymbol{\lambda}$. With both prior and likelihood functions in place, we then employ an adaptive Markov chain Monte Carlo [78] algorithm to sample values of α according to the posterior distribution $p(\alpha|\mathcal{D})$. We will then interrogate and utilize the posterior distribution in a few different ways, described in the "Posterior Predictions" section. Three commonly used summaries of the posterior distribution are the mean α_{mean}, standard deviation α_{std} and the maximum a posteriori (MAP) value $\alpha_{\text{MAP}} = \text{argmax}_{\alpha}\, p(\alpha|\mathcal{D})$. Often, when only the latter is of interest, one can proceed to posterior maximization via standard optimization methods, without invoking Markov chain Monte Carlo (MCMC). To this end, we note that, when uniform priors are employed, the MAP estimate coincides with commonly used maximum likelihood estimate (MLE).

Our resulting proposed calibration framework is illustrated in Fig. 1.

Note that when σ is fixed and not inferred, the conventional calibration procedure described earlier (in the second section) can be viewed as a degenerate simplification of our proposed calibration framework.[§]

Likelihood Construction for Density Estimation

In this section, we will detail several options for the likelihood construction, starting from its definition and

[§]Indeed, consider a "probabilistic" model for $\boldsymbol{\lambda}$ that has zero variance, e.g., a PCE of zeroth order. Then, as outlined later below, the ABC likelihood with no variance constraints from Eq. (20) is proportional to the additive-error likelihood in Eq. (4), and σ plays the role of the ABC tolerance parameter ε.

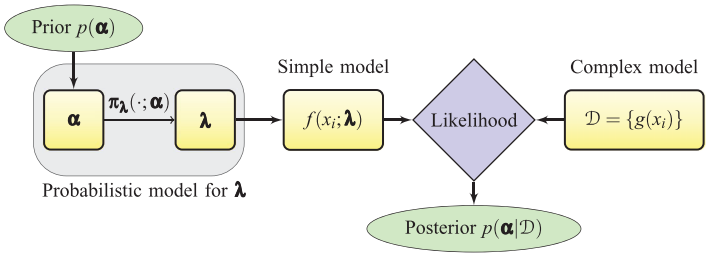

Figure 1 Schematic of the proposed calibration framework.

Table I A Summary of Potential Options for Likelihood Construction, Including the Corresponding Subsection and Equation Number

Section	Likelihood Type	Formula, $L(\alpha)$	Equation
Full Likelihood	Full likelihood	$\pi_{\boldsymbol{f}}(\boldsymbol{y};\alpha)$	(13)
Marginalized Likelihood	Marginalized likelihood	$\prod_{i=1}^{N} \pi_{f_i}(y_i;\alpha)$	(16)
Gaussian Approximation to Marginalized Likelihood	Gaussian approximation	$\frac{1}{(2\pi)^{N/2}} \prod_{i=1}^{N} \frac{1}{\sigma_i(\alpha)} \exp\left(-\frac{(\mu_i(\alpha)-y_i)^2}{2\sigma_i(\alpha)^2}\right)$	(17)
Approximate Bayesian Computation	ABC; mean only	$\frac{1}{\varepsilon\sqrt{2\pi}} \prod_{i=1}^{N} \exp\left(-\frac{(\mu_i(\alpha)-y_i)^2}{2\varepsilon^2}\right)$	(20)
	ABC; mean and standard deviation	$\frac{1}{\varepsilon\sqrt{2\pi}} \prod_{i=1}^{N} \exp\left(-\frac{(\mu_i(\alpha)-y_i)^2+(\sigma_i(\alpha)-\gamma\lvert\mu_i(\alpha)-y_i\rvert)^2}{2\varepsilon^2}\right)$	(24)

drawing several approximations or surrogate forms for it, summarized in Table I.

Full Likelihood. The random vector $\boldsymbol{\lambda}$ induces a stochastic process $f(x;\boldsymbol{\lambda})$; therefore, by definition, the exact likelihood is

$$L(\alpha) = p(\mathcal{D}|\alpha) = \pi_{\boldsymbol{f}}(\boldsymbol{y};\alpha) \tag{13}$$

where $\pi_{\boldsymbol{f}}(\cdot;\alpha)$ is a N-variate PDF on the vector $\boldsymbol{f} = (f_1, \ldots, f_N)$ with $f_i = f(x_i;\boldsymbol{\lambda})$, parameterized by α, at the given x values. From this viewpoint, we have a single data point $\boldsymbol{y}$ in the N-dimensional space. However, the distribution $\pi_{\boldsymbol{f}}(\cdot)$ is degenerate in general, when $N > d$, where d is the number of parameters in the vector $\boldsymbol{\lambda}$. Consider a simple example $f(x,\lambda) = \lambda x$, i.e., $d = 1$ parameter, cast as a normal random variable $\lambda \in \mathcal{N}(\mu, \sigma^2)$, where $\alpha = (\mu, \sigma)$ are the parameters defining the random variable λ. Assume there are $N > 1$ data points $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$. Now, each realization of λ invokes a line $f(x,\lambda) = \lambda x$ going through the origin. The likelihood can then be nonzero only if there are values of λ such that all data points are aligned in a line going through the origin, i.e., satisfy the constraints $y_i = \lambda x_i$ for all $i = 1, \ldots, N$. The likelihood can be written as

$$\begin{aligned} p(\mathcal{D}|\alpha) &= \int_{-\infty}^{+\infty} \underbrace{p(\mathcal{D}|\lambda')}_{\mathbb{I}_\Omega} \underbrace{p(\lambda'|\alpha)}_{\pi_\lambda(\lambda';\alpha)} d\lambda' \\ &= \int_\Omega \pi_\lambda(\lambda';\alpha)\, d\lambda' = P(\lambda \in \Omega) \end{aligned} \tag{14}$$

where $\Omega = \{\lambda : y_i = \lambda x_i, \text{ for all } i = 1, \ldots, N\}$ is the set of values of λ that satisfy all constraints $y_i = \lambda x_i$. Indeed, by definition,

$$p(\mathcal{D}|\lambda') = \begin{cases} 1, & \text{if } \lambda' = \frac{y_i}{x_i} \text{ for all } i = 1, \ldots, N, \\ & \qquad \text{i.e. } \lambda' \in \Omega \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

thus inducing the constraints $y_i = \lambda x_i$. Generically, if the points (x_i, y_i) are not aligned in a line going through the origin, then Ω is empty. In a degenerate case, when all points are aligned along a single line $y = \lambda_0 x$ going through the origin, the set $\Omega = \{\lambda_0\}$ consists of one point and its likelihood still vanishes. In more complicated cases, for general $f(x;\boldsymbol{\lambda})$ with higher number of variables $\boldsymbol{\lambda}$, i.e. with more degrees of freedom to satisfy the N constraints $y_i = f(x_i;\boldsymbol{\lambda})$, one may arrive at a set Ω that includes more than a single value. However, the likelihood in (14) will still vanish as Ω is a lower dimensional manifold satisfying the

N constraints $y_i = f(x_i; \boldsymbol{\lambda})$ in a d-dimensional space. Note that this is only possible for $N \leq d$ or in degenerately distributed data case with $N > d$. Clearly, whether Ω is empty (most often) or it is a lower dimensional structure in d-dimensional space, the likelihood in (14) vanishes. In other words, one usually cannot tune the simpler model $f(x, \boldsymbol{\lambda})$ to exactly match the complex model $g(x)$ at all locations of interest x_i for $i = 1, \ldots, N$. This degeneracy is also discovered and detailed, in the context of Bayesian identification of random variables, in [76].

Marginalized Likelihood. We can instead construct a likelihood function that looks at individual, marginalized fits at each x_i. From such a viewpoint, there are N data points y_i and the likelihood is written as

$$L_M(\alpha) = \prod_{i=1}^{N} \pi_{f_i}(y_i; \alpha) \tag{16}$$

where $\pi_{f_i}(\cdot)$ is a univariate PDF, parameterized by α, corresponding to the variability in $\boldsymbol{\lambda}$ for the model output $f(x_i; \boldsymbol{\lambda})$ at fixed x_i.

To compute the marginal PDFs $\pi_{f_i}(\cdot; \alpha)$, one generally relies on sampling the uncertain parameter $\boldsymbol{\lambda}$ for a fixed value of α and evaluating the model at each sample to arrive at the sample set for $f_i = f(x_i; \boldsymbol{\lambda})$. This will require kernel density estimation (KDE) and can be prohibitively expensive as this computation is requested at each MCMC step. To avoid sampling and potentially expensive KDE calculations, one can use a Gaussian approximation for marginalized likelihoods, described in the next subsection. This approximation, however, will reveal another inherent shortcoming of the marginalized approximation. Namely, the posterior has multiple singularities corresponding to α values leading to vanishing marginal variances at each data location x_i. This will be clarified at the end of the next subsection and demonstrated in the section "Deficiencies in the Marginalized Likelihood Construction" on a simple example.

Gaussian Approximation to Marginalized Likelihood. Instead of computing the marginal PDFs in Eq. (16), we can compute the first two moments of the model output and use a Gaussian approximation of the marginal PDFs. Denote by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)$ and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)$ the mean and standard deviation vectors of the model f evaluated at all locations $(x_1, \ldots, x_N)$, respectively. Both, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ depend on the PDF parameters α. Then, the likelihood takes the form

$$L_G(\alpha) = \frac{1}{(2\pi)^{N/2}} \prod_{i=1}^{N} \frac{1}{\sigma_i(\alpha)} \exp\left(-\frac{(\mu_i(\alpha) - y_i)^2}{2\sigma_i(\alpha)^2}\right) \tag{17}$$

Using marginalized likelihoods, whether in PDF form (16) or its Gaussian approximation (17), leads to inappropriate results as tests in the "Deficiencies in the Marginalized Likelihood Construction" section show. There are multiple posterior singularities, corresponding for vanishing $\sigma_i(\alpha)$ for each i, and posterior maximization always finds one of these singularities, fitting exactly one point perfectly, while completely misfitting the rest of the points.

Taking into consideration the challenges associated with the density estimation framework with true KDE-based likelihood or its marginalized approximations, whether via KDE or Gaussian approximations, we argue for and employ ABC, or likelihood-free methods that replace the likelihood with a measure of discrepancy between statistics of the model outputs f_i and those of the data y_i, for all i [79–81].

Approximate Bayesian Computation. ABC, or likelihood-free, methods are useful when a proper likelihood is not known or is prohibitively expensive to evaluate. In this framework, one uses a synthetic likelihood or fit function that measures the discrepancy between a chosen set of statistics of model outputs and the corresponding estimates from the data. In this work, we will primarily adopt the mean and standard deviation of the uncertain model output as statistics of interest.

The ABC formulation relies on a likelihood constructed as a kernel function evaluated on a chosen distance measure $\rho(S_{\mathcal{M}}, S_{\mathcal{D}})$ between the estimates of the statistic of interest on model prediction ($\mathcal{M}$) and data ($\mathcal{D}$):

$$L_{ABC}(\alpha) = \frac{1}{\varepsilon} K\left(\frac{\rho(S_{\mathcal{M}}, S_{\mathcal{D}})}{\varepsilon}\right) \tag{18}$$

By construction, this density penalizes/favors model parameters for which the model and data statistics match/mismatch. The scaling factor ε plays the role of a tolerance parameter. The discussion on the choice of its value will be relegated to the "Posterior Distribution: Choice of N, ε and γ" section. We will work with the most commonly used, Gaussian, kernel function $K(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ (see [79]).

In the simplest setting, focusing only on matching the means, we can write a distance function between the model mean $\boldsymbol{\mu}$ and the data $\boldsymbol{y}$ as

$$\rho(S_{\mathcal{M}}, S_{\mathcal{D}}) = ||\boldsymbol{\mu}(\alpha) - \boldsymbol{y}||_2, \tag{19}$$

implying the likelihood form

$$L_{\text{ABCM}}(\alpha) = \frac{1}{\varepsilon\sqrt{2\pi}} \prod_{i=1}^{N} \exp\left(-\frac{(\mu_i(\alpha) - y_i)^2}{2\varepsilon^2}\right) \tag{20}$$

However, this does not constrain $\sigma_i(\alpha)$ at all, and the optimization problem is not well-defined for terms that contribute to $\sigma_i(\alpha)$ (ie non-leading coefficients in the PCE of $\boldsymbol{\lambda}$). Nevertheless, this form of the likelihood has parallels with standard least-squares methods where a Gaussian additive discrepancy term is assumed. Indeed, if one considers the limit case of 0-th order PCE (10) for $\boldsymbol{\lambda}$, then having no constraints on $\sigma_i(\alpha)$ is not a handicap, as $\boldsymbol{\lambda}$ is cast as a deterministic quantity, and the typical least-squares likelihood is equivalent to (20).

Beyond this, one can require the discrepancy test between model and data to include the standard deviation as well, *i.e.* require the standard deviation of the model uncertain prediction to be *consistent* with the data spread around the mean model prediction. That is, besides requesting the mean-match

$$\mu_i(\alpha) \approx y_i, \tag{21}$$

we apply an additional constraint

$$\sigma_i(\alpha) \approx \gamma|\mu_i(\alpha) - y_i| \tag{22}$$

for all $i = 1, \ldots, N$ and for some positive γ that is $\mathcal{O}(1)$. The latter constraint is chosen to ensure that the predictive uncertainty in the simple model is, to some extent, comparable to the residual error made at the same location i. By default, we set $\gamma = 1$, relegating the discussion of its relevance to the "Posterior Distribution: Choice of N, ε and γ" section. Inducing ℓ_2 norms that correspond to the approximate equalities in Eqs. (21) and (22), e.g.,

$$\rho(S_{\mathcal{M}}, S_{\mathcal{D}})^2 = \sum_{i=1}^{N}((\mu_i(\alpha) - y_i)^2 + (\sigma_i(\alpha) - \gamma|\mu_i(\alpha) - y_i|\,)^2) \tag{23}$$

one arrives at a likelihood

$$L_{ABC}(\alpha) = \frac{1}{\varepsilon\sqrt{2\pi}} \prod_{i=1}^{N} \exp\left(-\frac{(\mu_i(\alpha) - y_i)^2 + (\sigma_i(\alpha) - \gamma|\mu_i(\alpha) - y_i|)^2}{2\varepsilon^2}\right) \tag{24}$$

This form of the likelihood will be the main focus of the studies in this paper. We note that this ABC likelihood is similar to the moment-matching likelihood in the MUM-PCE method [29] where parameters defining spectral expansions of uncertain model inputs are being sought for. In the next subsection, we proceed in this direction by first outlining an uncertainty propagation mechanism to estimate the means $\mu_i(\alpha)$ and standard deviations $\sigma_i(\alpha)$ of the calibrated uncertain model prediction for a given parameter value α.

Forward Propagation of Uncertainty

Both the Gaussian approximation (17) and the ABC likelihood (24) require the knowledge of the mean prediction $\boldsymbol{\mu}(\alpha)$ and the predictive standard deviation $\boldsymbol{\sigma}(\alpha)$ only, without having to construct the full PDF $\pi_f(\cdot;\alpha)$ or its marginals $\pi_{f_i}(\cdot;\alpha)$. In such cases, there are efficient alternatives to sampling methods that accurately estimate the first two moments of $f(x_i, \boldsymbol{\lambda})$ for each $i = 1, \ldots, N$, given the PC representation (10) of the uncertain quantity $\boldsymbol{\lambda}$. To this end, we will employ nonintrusive spectral projection (NISP) [63,82–84] to obtain a PC representation of the model outputs $f_i = f(x_i;\boldsymbol{\lambda})$. We seek a PCE

$$f(x_i;\boldsymbol{\lambda}(\boldsymbol{\xi};\alpha)) = \sum_{k=0}^{K-1} c_{ik}(\alpha)\Psi_k(\boldsymbol{\xi}) \tag{25}$$

at each data location $i = 1, \ldots, N$, *given* the linear PC representation (10) for each $\lambda_j(\boldsymbol{\xi};\alpha) = a_{j0} + \sum_{k=1}^{j} a_{jk}\xi_k$, where $\alpha = \{a_{jk}\}_{j=1,\ldots,d}^{k=0,\ldots,j}$. The projection technique uses the orthogonality of the PC bases $\Psi_k(\boldsymbol{\xi})$ to obtain formulae for the coefficients

$$c_{ik} = \frac{\langle f_i\Psi_k\rangle}{\langle\Psi_k^2\rangle} \tag{26}$$

where $\langle\cdot\rangle$ denotes expectation with respect to the PDF of $\boldsymbol{\xi}$, defined as $\langle h(\boldsymbol{\xi})\rangle = \int_{\xi} h(\boldsymbol{\xi})\pi(\boldsymbol{\xi})\,d\boldsymbol{\xi}$. While the denominator in (26) can be computed and stored in advance, the numerator requires a computation of an

integral tackled by Gaussian quadrature

$$\langle f_i \Psi_k \rangle = \int_{\xi} f(x_i; \boldsymbol{\lambda}(\boldsymbol{\xi}; \alpha)) \Psi_k(\boldsymbol{\xi}) \pi(\boldsymbol{\xi})\, d\boldsymbol{\xi}$$

$$\approx \sum_{q=1}^{Q} f(x_i; \boldsymbol{\lambda}(\boldsymbol{\xi}_q; \alpha)) \Psi_k(\boldsymbol{\xi}_q) w_q \qquad (27)$$

for a set of quadrature point-weight pairs $\{(\boldsymbol{\xi}_q, w_q)\}_{q=1}^{Q}$. Note that in the quadrature summation above, the integration PDF $\pi(\boldsymbol{\xi}_q)$ is absent as it is "folded" into the weights w_q by the definition of Gaussian quadrature. Furthermore, note that when $f(\cdot)$ is a polynomial surrogate model in $\boldsymbol{\xi}$, as in the problem described in the fifth section, or any polynomial model in $\boldsymbol{\xi}$ in general, then a quadrature formula with high enough order leads to exact integration. Namely, if f is an order-p polynomial and the sought-after PCE is of the same order p, and $\boldsymbol{\lambda}(\boldsymbol{\xi}; \alpha)$ is only linear in $\boldsymbol{\xi}$, a Gaussian product-rule quadrature of $(p+1)$ points per dimension is sufficient for exact integration in (27), leading to $K = (d+p)!/(d!\,p!)$ PC terms and $Q = (p+1)^d$ quadrature points. In principle, one can drastically reduce the number of required quadrature points by employing sparse quadrature rules. However, the focus of this work is the calibration method, and we deal with low values of d for which product rules are computationally feasible. The evaluation of the PC coefficients $c_{ik}(\alpha)$ for a fixed α, therefore, requires Q evaluations of the linear PCE $\boldsymbol{\lambda}(\boldsymbol{\xi}_q; \alpha)$ and the forward model $f(x_i; \boldsymbol{\lambda})$ for the requisite values of $\boldsymbol{\lambda}$, across all $i = 1, \ldots, N$. It is worth noting that in cases when the forward model $f(x_i; \boldsymbol{\lambda})$ is a polynomial or other simple algebraic relation, one can also rely on intrusive spectral projection (ISP) to obtain the PC representation (25) thus avoiding any sampled forward model simulations. However, to preserve the general feasibility, we will employ NISP for all the cases described in this work.

After the PCE (25) for the model output is constructed, one can arrive at the PDF of $f(\cdot; \alpha)$ by sampling the PC followed by a KDE. Although sampling is generally computationally inexpensive, one still needs many samples for smooth and accurate density estimates. Also, this computation is carried out for each value of α invoked by the MCMC or optimization algorithms. However, the ABC methods and the Gaussian approximation described in the "Likelihood Construction for Density Estimation" section only require the knowledge of the first two moments of the output, and the PC forward propagation described in this section offers easy access to the first two moments by

$$\mu_i(\alpha) = \mathrm{E}_{\xi}[f(x_i; \boldsymbol{\lambda}(\boldsymbol{\xi}; \alpha)] = c_{i0}(\alpha) \qquad (28)$$

and

$$\sigma_i(\alpha)^2 = \mathrm{V}_{\xi}[f(x_i; \boldsymbol{\lambda}(\boldsymbol{\xi}; \alpha)] = \sum_{k=1}^{K-1} c_{ik}(\alpha)^2 \langle \Psi_k^2 \rangle \qquad (29)$$

With the first two moments of the forward model available, the likelihood function (24) can be computed, leading to the full posterior $p(\alpha|\mathcal{D})$, in the Bayesian context, or simply to its maximum $\alpha_{MAP} = \mathrm{argmax}_{\alpha} \mathrm{p}(\alpha|\mathcal{D})$, if one only employs optimization with the posterior serving as the objective function. The next subsection details potential predictive options after the posterior or its maximum become available.

Posterior Predictions

Let us denote by $Z(x; \alpha) = f(x; \boldsymbol{\lambda}(\boldsymbol{\xi}; \alpha))$ the probabilistic prediction of the model f for fixed value of α. We use the capital Z notation to emphasize the probabilistic nature of this quantity associated with the variability of the random variable $\boldsymbol{\lambda}$, or, in the PC setting, $\boldsymbol{\xi}$.

The random variable $Z^{\mathrm{MP}}(x) = Z(x; \alpha_{\mathrm{MAP}})$ associated with the best value of α, *i*.e. α_{MAP}, follows the *MAP predictive* distribution. Its moments are

$$Z^{\mathrm{MP}}_{\mathrm{mean}}(x) = \mu(\alpha_{\mathrm{MAP}}; x) \qquad \text{and}$$

$$Z^{\mathrm{MP}}_{\mathrm{var}}(x) = \sigma(\alpha_{\mathrm{MAP}}; x)^2 \qquad (30)$$

where $\mu(\alpha; x) = \mathrm{E}_{\xi}[f(x; \boldsymbol{\lambda}(\boldsymbol{\xi}; \alpha))]$ and $\sigma(\alpha; x)^2 = \mathrm{V}_{\xi}[f(x; \boldsymbol{\lambda}(\boldsymbol{\xi}; \alpha))]$ are the forward-propagated mean and variance of the function $f(x; \boldsymbol{\lambda}(\boldsymbol{\xi}; \alpha))$ at location x for given α.

Furthermore, when the full posterior distribution becomes available, one can inspect the *PP* random variable $Z^{\mathrm{PP}}(x) = \mathrm{E}_{\alpha}[Z(x; \alpha)]$ whose distribution is given by

$$\pi_{Z^{\mathrm{PP}}(x)}(z) = p(z|\mathcal{D}) = \int_{\alpha} p(z|\alpha) p(\alpha|\mathcal{D})\, d\alpha$$
$$= \mathrm{E}_{\alpha}[p(z|\alpha)] \qquad (31)$$

which is the probability distribution of the model at any given location x, i.e. $f(x; \boldsymbol{\lambda}(\boldsymbol{\xi}; \alpha))$ integrated over the posterior distribution of α, $p(\alpha|\mathcal{D})$. The notation E_{α} denotes expectation with respect to the posterior distribution of α. It needs to be emphasized that both

$Z^{\mathrm{MP}}(x)$ and $Z^{\mathrm{PP}}(x)$ are random variables: While α-variability is fixed or integrated over, they both still include the variability built into $\boldsymbol{\lambda}$, which in the PC setting is characterized by $\boldsymbol{\xi}$.

One can find the mean and the variance of the PP random variable

$$Z^{\mathrm{PP}}_{\mathrm{mean}}(x) = \mathrm{E}_\alpha[\mu(\alpha; x)] \text{ and}$$
$$Z^{\mathrm{PP}}_{\mathrm{var}}(x) = \mathrm{E}_\alpha[\sigma(\alpha; x)^2] + \mathrm{V}_\alpha[\mu(\alpha; x)] \quad (32)$$

where V_α denotes variance with respect to the posterior distribution of α.

The detailed derivations of these predictive quantities, as well as their relationships, are presented in the appendix. For synthetic studies, demonstrated the next section, we will analyze the ABC method by separately focusing on the MAP predictive distribution, as well as clearly identifying the character of the posterior distribution.

SYNTHETIC STUDIES

In this section, we apply the calibration method described above on test functions, to demonstrate the key features of the algorithm. The task is to calibrate parameters $\boldsymbol{\lambda}$ of a simpler model $f(x;\boldsymbol{\lambda})$ with respect to simulation results obtained from a more complicated model, $g(x)$.

Consider a cubic function as a complex model or "truth,"

$$g(x) = 2 + 2x + 3x^2 - 5x^3 \quad (33)$$

defined on $x \in [-1, 1]$. We will calibrate three different models $f(x;\boldsymbol{\lambda})$ with respect to data gathered from $g(x)$ at N equidistant locations across the x-axis, for $N = 11$ and $N = 51$. The chosen "simpler" models for calibration are

$$\begin{aligned} f(x;\boldsymbol{\lambda}) &= \lambda_0 + \lambda_1 L_1(x) && \text{Linear} \\ f(x;\boldsymbol{\lambda}) &= \lambda_0 + \lambda_1 L_1(x) + \lambda_2 L_2(x) && \text{Quadratic} \quad (34) \\ f(x;\boldsymbol{\lambda}) &= \lambda_0 + \lambda_1 L_1(x) + \lambda_2 L_2(x) \\ &\quad + \lambda_3 L_3(x) && \text{Cubic} \end{aligned}$$

where $L_k(x)$ is the Legendre polynomial of degree k. The parameters λ_j will be cast as first-order Hermite–Gauss PC random variable as described in Eq. (10).

For the initial tests, we employ the ABC likelihood (24) with fixed $\varepsilon = 0.1$ and $\gamma = 1$, and focus on the MAP estimate of α only. As Fig. 2 illustrates, the simpler model uncertain prediction "covers" the complex model results well enough given the constraints of the simpler model. Clearly, the higher order model fits better, and when one employs the most accurate, third-order polynomial as f, the uncertainty on model output vanishes, as there is no uncertainty attributed to model error. Also, note that increasing the number of data points from $N = 11$ to $N = 51$ does not improve the uncertain MAP prediction results, by construction. This is intuitively clear, because the MAP prediction standard deviation $\sqrt{Z^{MP}_{\mathrm{var}}(x)} = \sigma(\alpha_{\mathrm{MAP}}; x)$ corresponds to the discrepancy between the simpler and more complex model and does not relate to the amount of data gathered from the complex model $g(x)$. It is worth noting, however, that increasing N does reduce the posterior width for α, such that with increasing data volume there is little residual change of the model output uncertainty with additional data. This will be demonstrated in the next subsection, while explaining the posterior distribution structure and dependence on parameters of the method.

Posterior Distribution: Choice of N, ε, and γ

In this subsection, we recall the ABC likelihood, also given in Eq. (24),

$$L_{\mathrm{ABC}}(\alpha) = \frac{1}{\varepsilon\sqrt{2\pi}} \prod_{i=1}^{N} \exp\left(-\frac{(\mu_i(\alpha) - y_i)^2 + (\sigma_i(\alpha) - \gamma|\mu_i(\alpha) - y_i|)^2}{2\varepsilon^2} \right) \quad (35)$$

and study its dependence on values of N, ε and γ for the cubic complex function (33) and linear simple model from (34), i.e., $f(x;\boldsymbol{\lambda}) = \lambda_0 + \lambda_1 L_1(x)$, where λ_0 and λ_1 are cast as random variables with a "triangular" PCE

$$\lambda_0 = a_{00} + a_{01}\xi_1$$
$$\lambda_1 = a_{10} + a_{11}\xi_1 + a_{12}\xi_2 \quad (36)$$

It is important to recognize that the likelihood replacement (35) leads to a posterior distribution $p(\alpha|\mathcal{D})$ that corresponds to the allowed variability in α that still satisfies the constraints enforced by the ABC likelihood. The parameter ε in ABC, while directly affecting the posterior distribution width, merely plays the role of a tolerance parameter, i.e. it specifies how much one

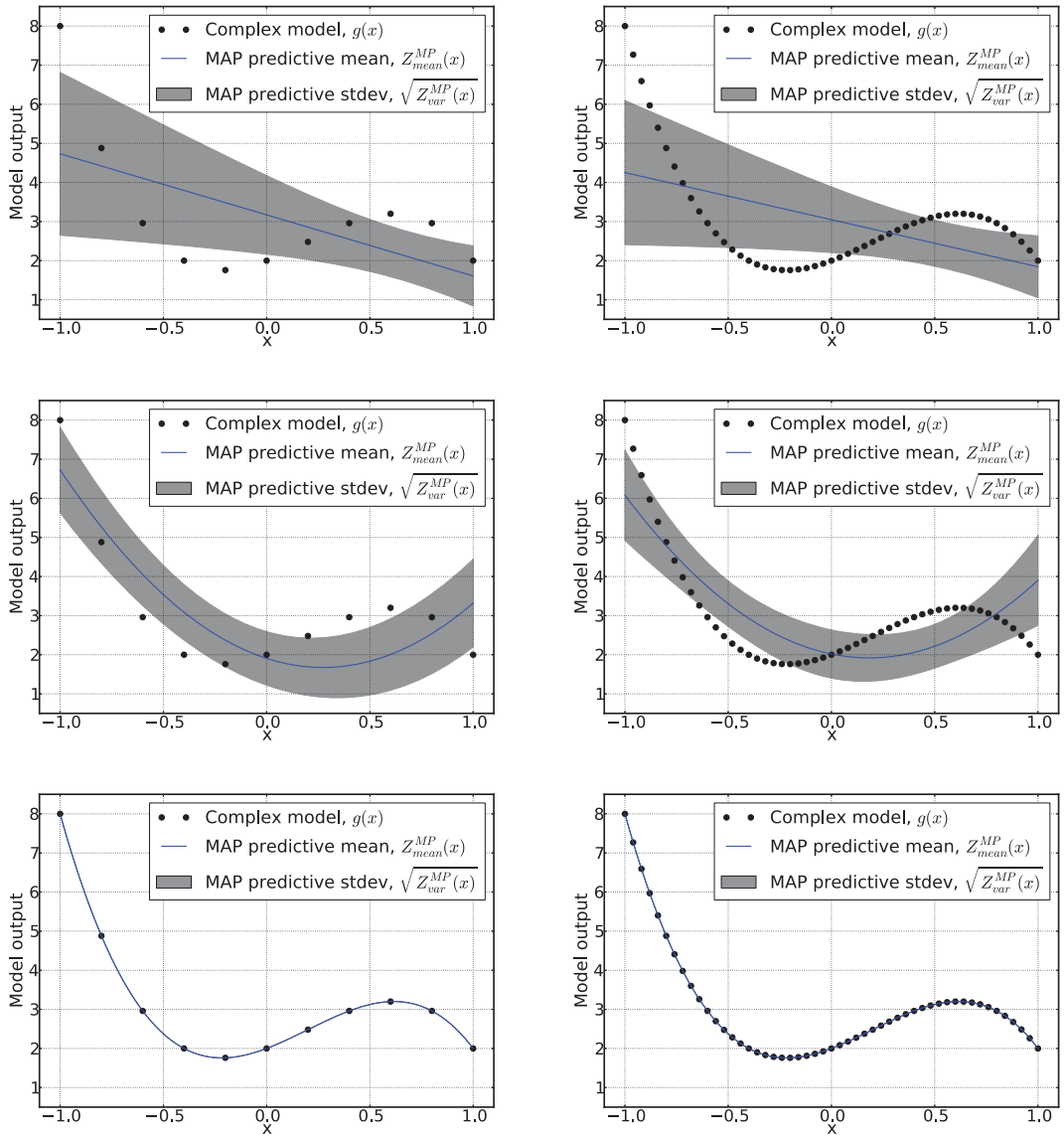

Figure 2 The results of ABC calibration for a test case with cubic complex function $g(x)$ from Eq. (33). The simpler function $f(x;\boldsymbol{\lambda})$ is set to linear, quadratic, and cubic, from top to bottom. The left column corresponds to $N = 11$ evaluations of the complex function, whereas the right column corresponds to $N = 51$. Plotted are calibrated model predictions using the MAP value of α, i.e., $Z^{MP}_{\text{mean}}(x) = \mu(\alpha_{\text{MAP}}; x)$ and $\sqrt{Z^{MP}_{\text{var}}(x)} = \sigma(\alpha_{\text{MAP}}; x)$. The ABC parameters in Eq. (24) are set to $\gamma = 1$ and $\varepsilon = 0.1$.

should penalize for the mismatch in the chosen statistics of the simpler versus the complex-model/data. Figure 3 demonstrates the dependence of the posterior width on both ε and N. Clearly—and this can be seen from the likelihood (35) itself—increasing ε linearly increases the posterior width, while keeping its MAP value intact. Furthermore, increasing the amount of data N also reduces the posterior width, albeit not in a linear fashion. Moreover, a new data set also affects the MAP value of the parameters, although, as can be seen from the top row of Fig. 2, the change does not drastically affect the ability of the simple model uncertain output to predict the range of discrepancy from the complex model.

Next, we study the dependence on γ. In the ABC likelihood (35), γ relates the simple model output variability $\sigma_i(\alpha)$ to the discrepancy $|\mu_i(\alpha) - y_i|$ between the simple model mean prediction and the complex model. Larger γ allows more variability in the embedded model parameter $\boldsymbol{\lambda}$. Figure 4 illustrates that effect.

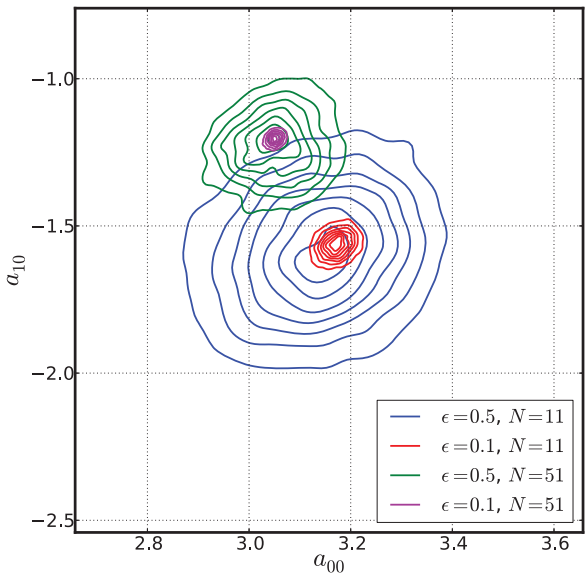

Figure 3 Joint posterior distribution for the leading terms in the PCE (10), i.e. for (a_{00}, a_{10}), for four different cases with varying N and ε. The posterior distribution is computed from 20, 000 chain samples, using the ABC likelihood (35) with $\gamma = 1$, for the cubic complex function $g(x)$ and linear simple function $f(x;\boldsymbol{\lambda})$ introduced in Eqs. (33) and (34), respectively.

It shows the posterior distributions for two pairs of parameters for the default, $\gamma = 1$ case and the $\gamma = 0.5$ case, where the linear model is calibrated against the cubic model. While the MAP values for the leading terms in the PCE (10), shown on the left plot, are not altered much, the higher order terms, shown on the right plot, are affected by the magnitude of γ and therefore, allow for larger variability in the MAP predictions of the calibrated model output as demonstrated in Fig. 5. Also note that the posterior width itself is only slightly affected by the values of γ across all components of α.

Deficiencies in the Marginalized Likelihood Construction

In this section, we demonstrate the challenges associated with the marginalized likelihood form (17). As the likelihood form suggests, if any one of the $\sigma_i(\alpha)$ nearly vanishes, the posterior becomes very large. This suggests the presence of many "spiky" local maxima in the posterior distribution. Figure 6 illustrates this observation, as we started with three different values of α, and found three different MAP values, each of them corresponding to exactly one $\sigma_i(\alpha)$ being zero.

This is an inherent deficiency of the marginalized likelihood construction, whether it is with the Gaussian assumption or not. That is, the issue persists for the sampling-based marginalized likelihood (16). In principle, one can employ more informative priors on α to ensure that none of the pushed-forward standard deviations $\sigma_i(\alpha)$ vanishes, e.g., by penalizing against the small values of the nonconstant terms of the PCE (36). Such a synthetic penalization can lead to a specific, goal-oriented likelihood that is effectively very similar to the one (24) employed in this work.

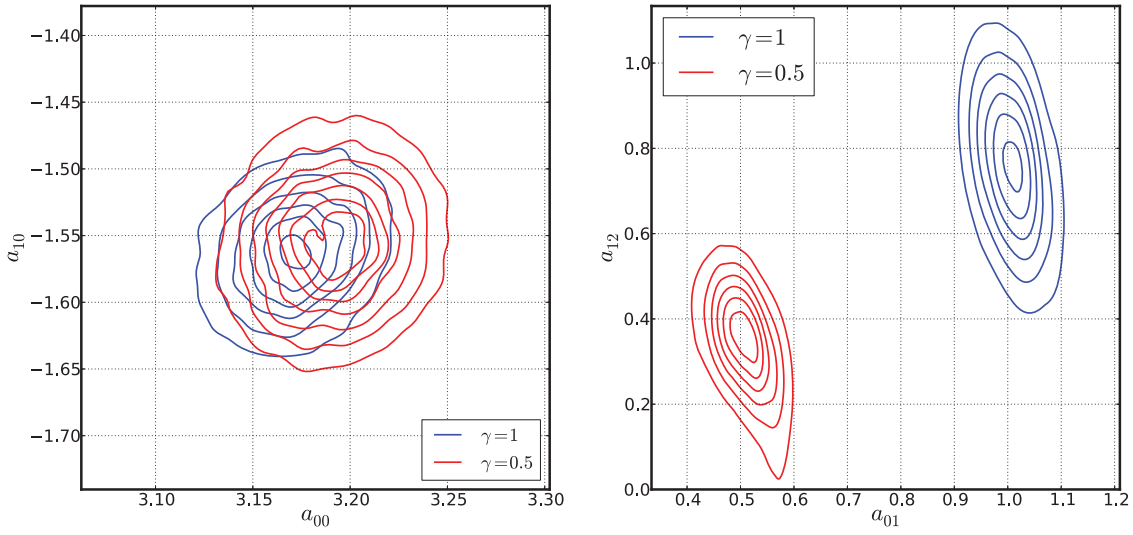

Figure 4 Joint posterior distributions for parameter pairs (a_{00}, a_{10}) and (a_{01}, a_{12}) from the PC model (36) for two values of γ. The posterior distribution is computed from 30, 000 chain samples, using the ABC likelihood (35) with $\varepsilon = 0.1$, $N = 11$, for the cubic complex function $g(x)$ and linear simple function $f(x;\boldsymbol{\lambda})$ introduced in Eqs. (33) and (34), respectively.

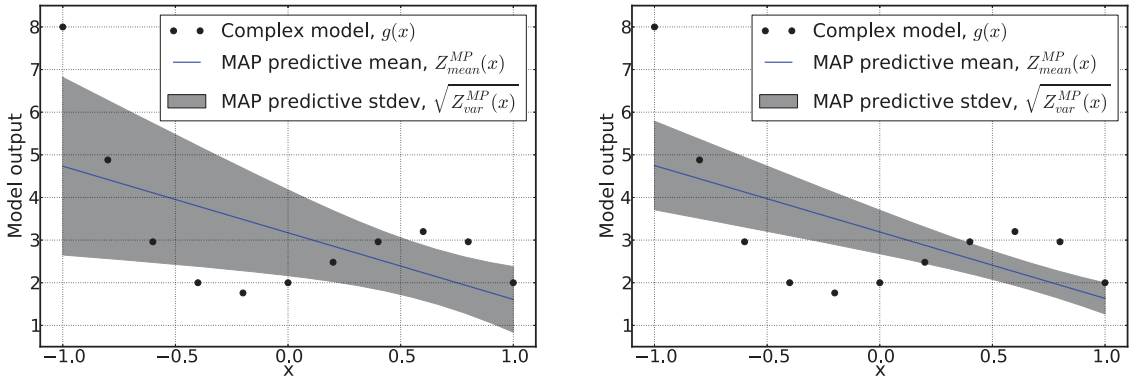

Figure 5 Predictive distribution of the simple linear model implied by the MAP value of α, for two different values of γ, (left) $\gamma = 1$ and (right) $\gamma = 0.5$.

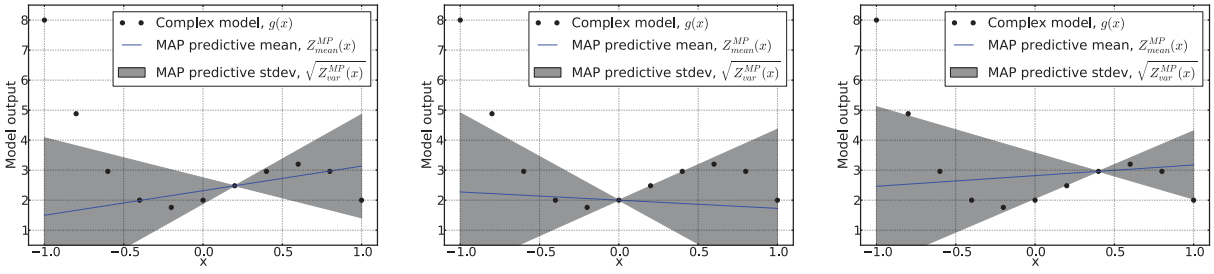

Figure 6 Demonstration of results based on the likelihood implied by a marginalized Gaussian approximation (17). The posterior exhibits various local maxima, corresponding to $\sigma_i(\alpha) \approx 0$ for each of the locations i. Depending on the starting point (three different options are shown), an optimization or MCMC technique finds one of the global maxima, completely failing to represent the model discrepancy at the rest of the locations.

CALIBRATION OF A CHEMICAL KINETIC MODEL

In the following, we consider the calibration of a simple chemical model with respect to a detailed "truth" model, using a homogeneous ignition problem. We examine two cases, one where we consider ignition over a range of initial temperature, but with a stoichiometric mixture, and the other where we vary both initial temperature and stoichiometry.

Case Study 1: Ignition Time versus Initial Temperature

Consider a preheated stoichiometric methane-air mixture, at atmospheric pressure. We compute the constant-pressure homogeneous ignition of this system using GRImech3.0 [85] (the "detailed" model in the present context), over a range of initial temperature $T^0 \in [1000, 1300]$ K, with N values uniformly spaced as $\{T_i^0 = 1000 + 300(i-1)/(N-1)\}_{i=1}^N$. The system evolves through a preheat phase during which a radical pool develops, and the mixture slowly heats up, before going through a fast ignition characterized by a rapid rise in temperature, consumption of reactants, and formation of combustion products. Defining the ignition delay time τ as the time instant at which the temperature reaches 1500 K, we observe the expected decrease in τ with increasing T^0, as shown in Fig. 7, for two cases, with $N = 11$ or $N = 51$.

In this context, the data are the set of ignition time values from the detailed GRImech3.0 model at different initial temperatures, $\mathcal{D} = \{(T_i^0, \ln \tau_i)\}_{i=1}^N$. We seek to calibrate a simpler model, specifically a global single-step irreversible methane-air mechanism,

$$CH_4 + 2O_2 \xrightarrow{k} CO_2 + 2H_2O \qquad (37)$$

for prediction of the quantity of interest $\ln \tau(T^0)$, with the forward rate of progress[‖]

[‖]Note that, if the reaction in Eq. (37) were an elementary reaction step, its rate of progress would involve the square of $[O_2]$. However, with the present *global* reaction the exponents on $[CH_4]$ and $[O_2]$ are largely a convenient modeling choice and are chosen here to be unity.

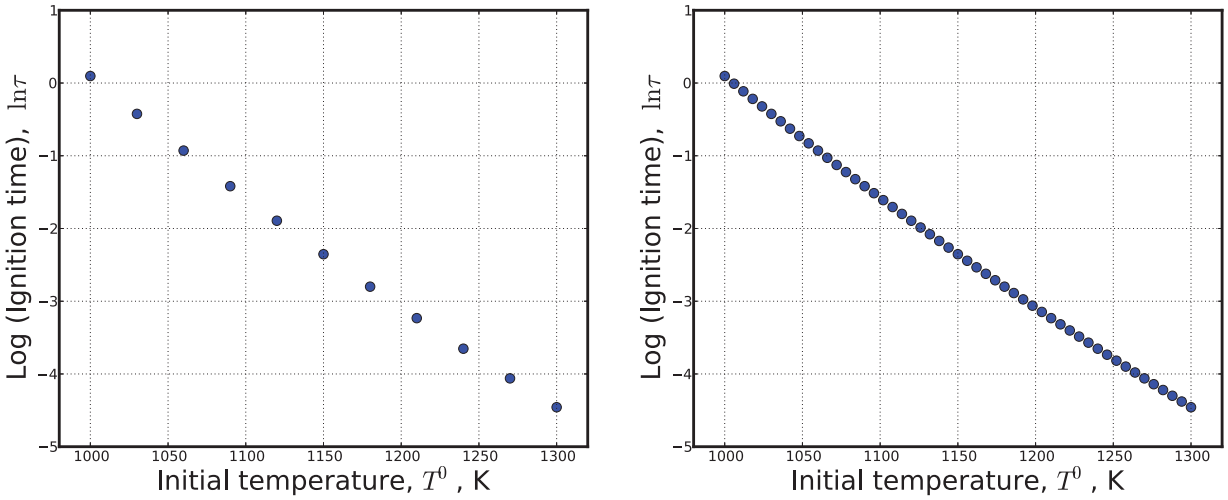

Figure 7 Ignition time data collected from the detailed methane-air ignition computations using GRImech3.0, given a stoichiometric homogeneous reactants mixture, for two cases $N = 11$ (left) and $N = 51$ (right).

$$\mathcal{R} = [CH_4][O_2]k(T) \tag{38}$$

where $[\cdot]$ is the concentration (mol cm^{-3}). $k(T)$ is the forward rate, modeled using the Arrhenius rate expression

$$k(T) = A\exp(-E/RT) \tag{39}$$

where $R = 1.9872041$ cal mol^{-1}K^{-1} is the universal gas constant, T is the temperature, E is the activation energy (cal mol^{-1}), and A is the positive pre-exponential constant (cm^3 mol^{-1} s^{-1}). The goal is to calibrate the parameters $\boldsymbol{\lambda} = (\ln A, E)$ to match the ignition-time data (shown in Fig. 7) produced by the detailed model. Note that, to enforce positivity of A, we use $\ln A$ as a parameter of interest, instead of A. Furthermore, following the above framework, we cast $\boldsymbol{\lambda}$ as a random variable and assume a multivariate normal distribution for its components. It then can be written as a first-order joint Hermite–Gauss PCE with respect to standard normal random variables ξ_1 and ξ_2,

$$\begin{aligned}\ln\tilde{A} &= a_0 + a_1\xi_1\\ \tilde{E} &= b_0 + b_1\xi_1 + b_2\xi_2\end{aligned} \tag{40}$$

where, to make the MCMC calibration easier to tune, we used linearly rescaled quantities $\ln\tilde{A}, \tilde{E} \in [-1, 1]$ that correspond to the physical values in the domains

$$\begin{aligned}\ln A &\in [\ln A_0 - \Delta\ln A, \ln A_0 + \Delta\ln A]\\ &\equiv [28.5, 36.0]\\ E &\in [E_0 - \Delta E, E_0 + \Delta E] \equiv [36315.5, 56954.0]\end{aligned}$$

As described in the "Method" section, we set priors on b_1 and b_2 to have positive support only, i.e. enforce $b_1 > 0$ and $b_2 > 0$ to avoid the sign-flip invariance. The five parameters $\alpha = (a_0, a_1, b_0, b_1, b_2)$ define a joint density for $\boldsymbol{\lambda}$, $\pi_{\boldsymbol{\lambda}}(\cdot;\alpha)$. These parameters will be the objects of inference in the Bayesian context seeking an approximate equality $\ln\tau(T^0)|_{\text{GRI}} \approx f(T^0;\boldsymbol{\lambda})|_{\text{simple-model}}$.

Furthermore, to avoid the cost of prohibitively many ignition model evaluations during optimization or MCMC, we construct a polynomial *surrogate* for the model $f(T^0;(\ln A, E))$ using a sixth order, three-variate Legendre polynomial basis with respect to parameter triple $(T^0, \ln A, E)$,

$$\begin{aligned}f(T^0;\ln A, E) &\approx f_s(T^0;\ln A, E)\\ &= \sum_{|\boldsymbol{k}|\le 6} s_{\boldsymbol{k}} L_{\boldsymbol{k}}(\eta_1, \eta_2, \eta_3)\end{aligned} \tag{41}$$

The surrogate is constructed as an approximation to the model f over a parameter domain

$$\begin{aligned}T^0 &\in [1000, 1300]\\ \ln A &\in [28.5, 36.0]\\ E &\in [36315.5, 56954.0]\end{aligned} \tag{42}$$

implying the linear scaling $T^0 \to \eta_1$, $\ln A \to \eta_2$, and $E \to \eta_3$ to have $\eta_i \in [-1, 1]$. The orthogonality of Legendre polynomials enables projection formulae

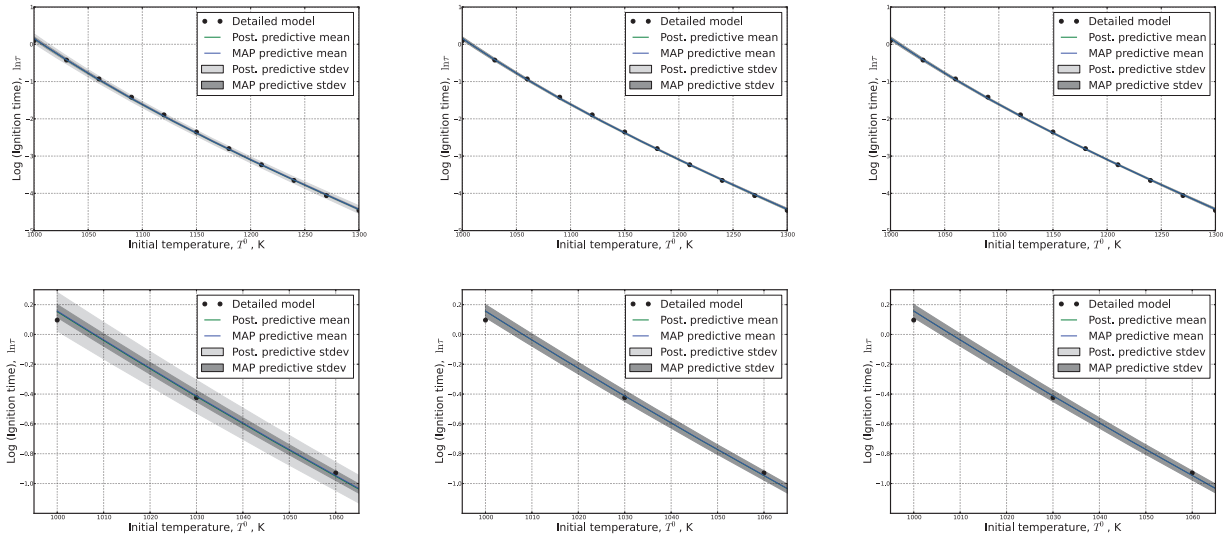

Figure 8 Calibration of the simple single-step model from Eq. (37) using the GRImech3.0 detailed model data with $N = 11$, for stoichiometric conditions. Two predictive distributions are illustrated, the MAP predictive $Z^{\mathrm{MP}}(T^0)$ (grey) and the PP $Z^{\mathrm{PP}}(T^0)$ (dark grey). The parameters in the ABC likelihood (24) are set to $\gamma = 1$ and varying ε, from left to right, $\varepsilon = 0.1$, $\varepsilon = 0.01$, and $\varepsilon = 0.001$. The bottom plots simply illustrate the error bars better by zooming in the top left corner of the corresponding graphs.

that lead to the surrogate coefficients

$$s_{\boldsymbol{k}} = \frac{1}{\langle L_{\boldsymbol{k}}^2 \rangle} \int_{\boldsymbol{\eta} \in [-1,1]^3} f(T^0(\eta_1); \ln A(\eta_2), E(\eta_3)) L_{\boldsymbol{k}}(\boldsymbol{\eta}) \frac{1}{2^3} d\boldsymbol{\eta} \qquad (43)$$

The integral above is evaluated using Gauss–Legendre quadrature. After the coefficients $s_{\boldsymbol{k}}$ are obtained, we replace the calibrated model $f(T^0; \ln A, E)$ with its accurate surrogate representation $f_s(T^0; \ln A, E)$ for both calibration and postprocessing steps. We have chosen first-order PC expansion for the computation of the first two moments of the forward model output within the ABC likelihood evaluation.

Figures 8 and 9 illustrate both the MAP predictive distribution and PP distribution for the $N = 11$ and $N = 51$ GRImech3.0 data, respectively, for different values of ε. The bottom plots are simply the zoomed-in versions to illustrate the error bar coverage better. Clearly, as ε is reduced the effect of posterior variability is diminished compared to the variability due to $\boldsymbol{\xi}$, i.e, built-in variability of $(\ln A, E)$ that accounts for model discrepancy. This is more clear from Fig. 10 where only the predictive standard deviations are shown, together with the error between the MAP predictive mean and the actual detailed model data. In principle, the choice of ε can be dictated by such an analysis—if one strives to obtain predictions that are independent of the posterior variability, one should strive to use smaller values of ε. In this case, clearly $\varepsilon = 0.001$ is small enough to guarantee that the posterior width, i.e. the uncertainty in α, does not appreciably affect the predictive performance, for both $N = 11$ and $N = 51$ regimes. Moreover, for the large data size case, $N = 51$, Fig. 10 indicates that the reduction in the PP uncertainty is stronger with ε as compared to the $N = 11$ case. This results from the reduced uncertainty in α given the enhanced information available from the additional data. Thus, for the $N = 51$ case, one can use a larger value of ε, e.g., $\varepsilon = 0.01$ without significantly affecting the predictive variance.

We emphasize that while reducing ε clearly reduces the prediction variance and then saturates, the α-posterior width itself keeps reducing. This is evident from the ABC likelihood definition (18), as well as from Fig. 11, which illustrates the posterior width dependence on N and ε. The effect of saturating predictive variance with reducing ε can also be seen from the PP formula (32), derived in the appendix. Indeed, with smaller ε, the second component of the variance in (32) reduces, whereas the first one does not. Clearly, for smaller values of ε, the posterior narrows down, making MAP and posterior predictions coincide, as seen in Fig. 10.

Furthermore, as a key outcome of the calibration, we illustrate the predictive joint PDF of $\boldsymbol{\lambda} =$

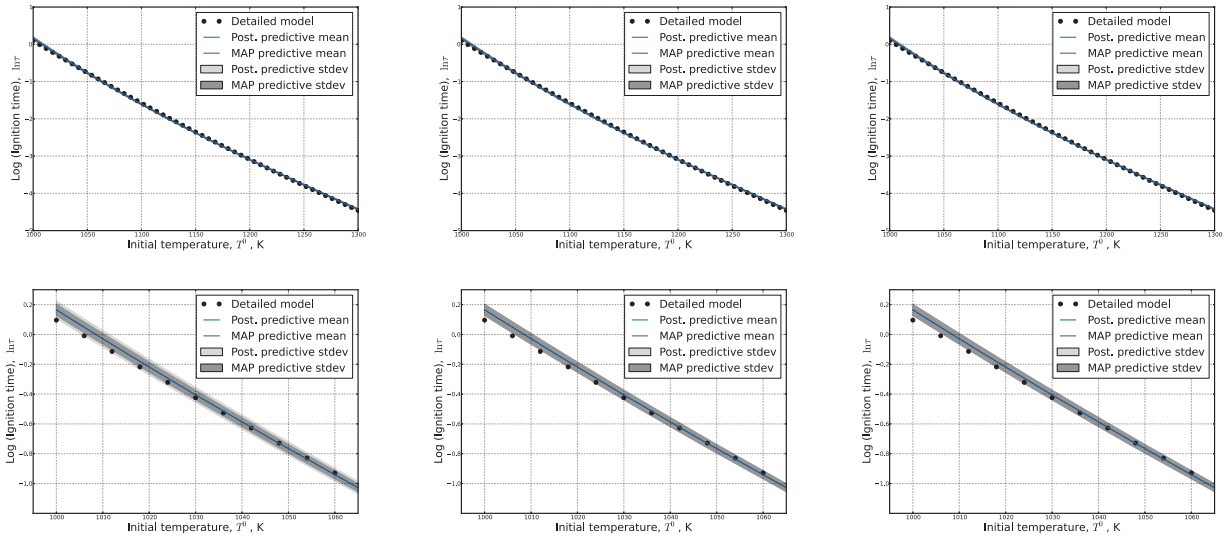

Figure 9 Calibration of the simple single-step model from Eq. (37) using the GRImech3.0 detailed model data with $N = 51$, under stoichiometric conditions. Two predictive distributions are illustrated, the MAP predictive $Z^{\mathrm{MP}}(T^0)$ (grey) and the PP $Z^{\mathrm{PP}}(T^0)$ (dark grey). The parameters in the ABC likelihood (24) are set to $\gamma = 1$ and varying ε, from left to right, $\varepsilon = 0.1$, $\varepsilon = 0.01$, and $\varepsilon = 0.001$. The bottom plots simply illustrate the error bars better by zooming in the top left corner of the corresponding graphs.

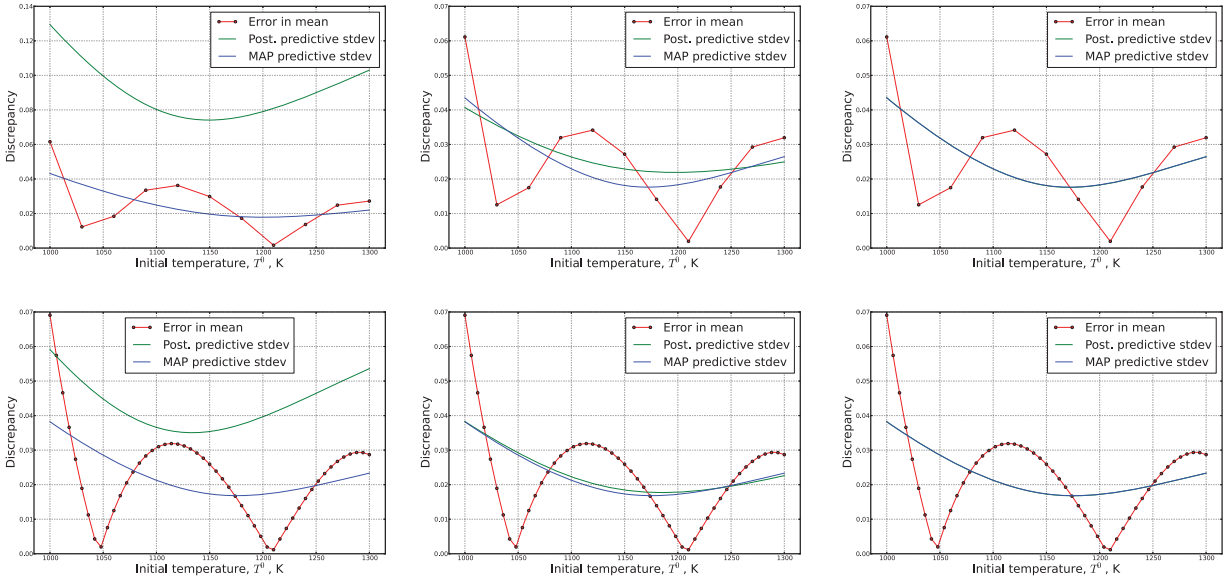

Figure 10 Detailed view of the predictive standard deviations and the error from the map predictive mean to the data. Calibration of simple single-step model from Eq. (37) using the GRImech3.0 detailed model data, for stoichiometric initial mixture conditions, with $N = 11$ (top row) and $N = 51$ (bottom row). Two predictive distributions are illustrated, the MAP predictive $Z^{\mathrm{MP}}(T^0)$ (blue) and the PP $Z^{\mathrm{PP}}(T^0)$ (green). The parameters in the ABC likelihood (24) are set to $\gamma = 1$ and varying ε, from left to right, $\varepsilon = 0.1$, $\varepsilon = 0.01$, and $\varepsilon = 0.001$.

$(\ln A, E)$ in Fig. 12, for various values of ε. The MAP predictive distribution of $\boldsymbol{\lambda} = (\ln A, E)$ is defined as $\pi_{\lambda}(\boldsymbol{\lambda}; \alpha_{\mathrm{MAP}})$, whereas the PP is the average of the PDF model $\pi_{\lambda}(\boldsymbol{\lambda}; \alpha)$ over the posterior, i.e.

$$p(\boldsymbol{\lambda}|\mathcal{D}) = \mathrm{E}_{\alpha}[\pi_{\lambda}(\boldsymbol{\lambda}; \alpha)] = \int_{\alpha} \pi_{\lambda}(\boldsymbol{\lambda}; \alpha) p(\alpha|\mathcal{D})\, d\alpha. \tag{44}$$

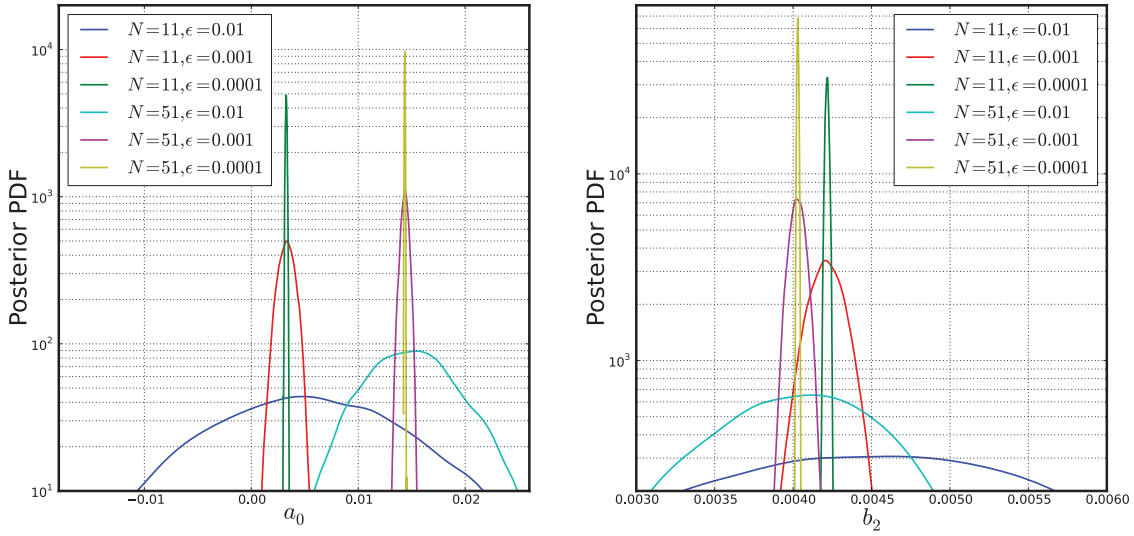

Figure 11 Marginal posterior distributions for parameters a_0 and b_2, demonstrating the dependence on N and ε. These posteriors are evaluated using 100, 000 MCMC samples. Note that logarithmic y-axis is used to facilitate comparing the PDFs visually.

As Fig. 12 shows, the posterior predictions of $(\ln A, E)$ coincide with the MAP predictive distribution for small enough values of ε, i.e. when the posterior distribution of $\alpha = (a_0, a_1, b_0, b_1, b_2)$ is narrow enough. Note that, in principle, it is not necessary that the PP variance is larger than the MAP predictive variance, as can be seen from the $\varepsilon = 0.01$ case.

Finally, Fig. 13 illustrates the structure of the two-dimensional (2D) and one-dimensional (1D) marginals based on the full joint posterior distribution using samples of $\alpha = (a_0, a_1, b_0, b_1, b_2)$ generated by the MCMC algorithm, for the case with $N = 11$, $\varepsilon = 0.001$, and $\gamma = 1$. The MAP values are highlighted on the 1D marginal plots on the diagonal of the figure. As improper uniform priors were employed for all parameters, it is clear that the data constrain the parameters well leading to unimodal posteriors of reasonably small widths. Also, all marginal posteriors seem to have somewhat comparable widths, a_0 and b_0 being the narrowest, i.e., the parameters constrained most. The strong correlation between the mean components a_0 and b_0 mimics the strong correlation in the pair $(\ln A, E)$ seen in [69] as well as in the conventional calibration with i.i.d. additive noise, described in the "Conventional Calibration Approach for both Case Studies" section. Also, note that the positive correlation between components a_1 and b_1 indicates strong constraining of the major axis tilt in the $(\ln A, E)$ density illustrated in Fig. 12. There is also somewhat strong negative correlation between b_2 and b_1 (and, consequently, between b_2 and a_1).

Overall, this case study shows that the simple reaction model (37) captures the ignition-time GRIMech3.0 data very well. The results indicate that very small—albeit correlated—embedded uncertainty in the parameter pair $(\ln A, E)$ explains the model discrepancy between the data from the simple reaction model and GRImech3.0. In the next subsection, we will consider a situation where the collected data reveals the inability of the simple model to match the results of the detailed model, and we will strive to properly characterize the associated model errors that are not small anymore.

Case Study 2: Ignition Time versus Initial Temperature and Equivalence Ratio

Let us now consider a situation that challenges the simple reaction model (37). Specifically, we collect data from the detailed methane-air mixture model not only across a set of initial temperatures but also across $N_\Phi = 7$ instances of the equivalence ratio Φ.[#] The domain of variation of equivalence ratio is chosen as $\Phi \in [0.6, 1.8]$, with the seven equidistant instances separated by increments of 0.2, whereas initial temperature is still set to $N_T = 11$ equidistant values across $T^0 \in [1000, 1300]$. The detailed model data are shown in Fig. 14. Now, we seek an approximate

[#]The equivalence ratio is defined as the fuel-to-air ratio, divided by the stoichiometric fuel-to-air ratio.

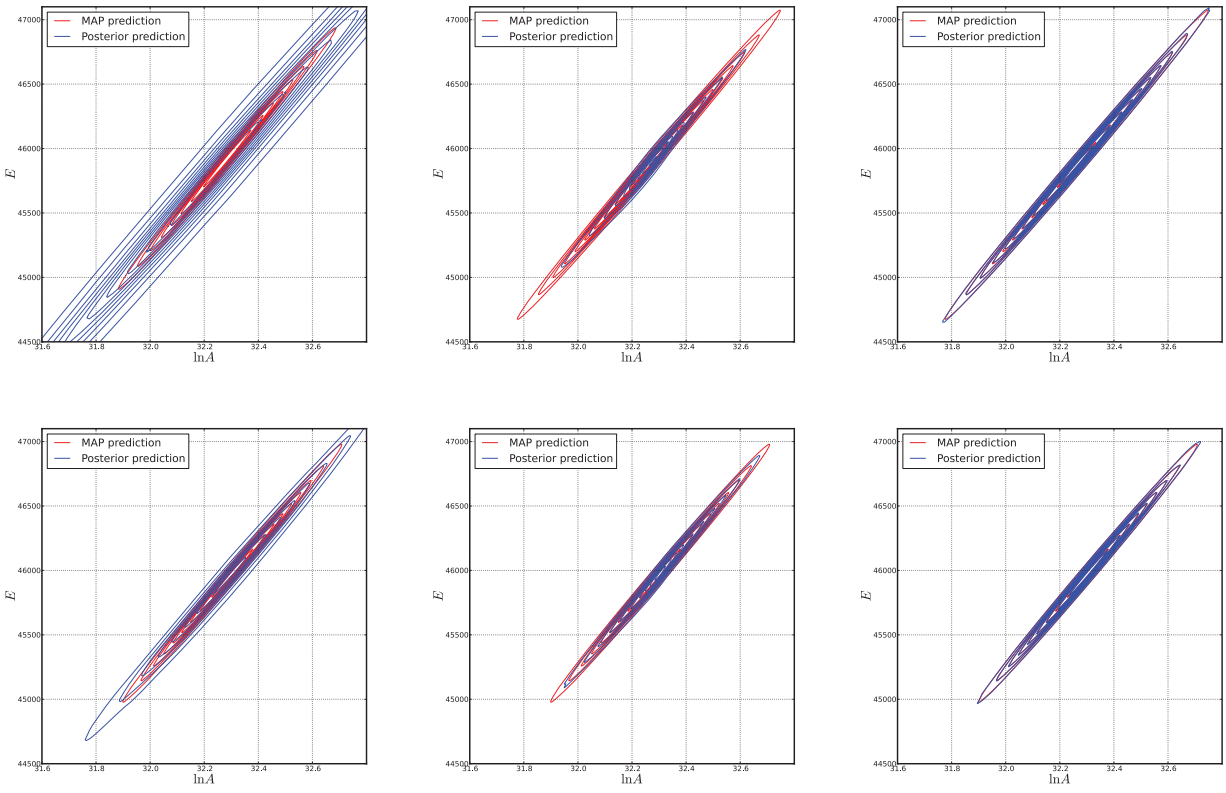

Figure 12 PP (blue) and MAP (red) predictive densities of $\boldsymbol{\lambda} = (\ln A, E)$ for $\varepsilon = 0.1, 0.01, 0.001$, from left to right, for two cases, $N = 11$ (top) and $N = 51$ (bottom), for stoichiometric reactants. The contour plots for the PP are based on $10^4 \times 10^3$ samples, i.e. 10^4 samples of $\boldsymbol{\lambda}$ according to $\pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}; \alpha)$, for each of 10^3 posterior samples of α. All plots are intentionally shown on the same axes scales. Clearly, for small enough ε, the MAP and PP contours coincide. Also, somewhat counterintuitively, for the $\varepsilon = 0.01$ case, the PP is narrower than the MAP predictive.

equality $\ln \tau(T^0, \Phi)|_{\text{GRI}} \approx f(T^0, \Phi; \boldsymbol{\lambda})|_{\text{simple-model}}$, by casting $\boldsymbol{\lambda} = (\ln A, E)$ as a bivariate normal variable, the scaled version of which is described by the first-order Hermite–Gauss PC (40). The uncertainty propagation from the PCE of the inputs to the model output is again performed via NISP, using a sixth order surrogate as in the first case study, but in a four-dimensional space of $(T^0, \Phi, \ln A, E)$.

Figure 15 shows the mean prediction surface, together with the data. Clearly, the simple model fails to capture the ignition time behavior of the detailed model across the range of T^0 and Φ. While the model cannot simultaneously capture all the GRImech3.0 data, it matches well near the central regime of $\Phi \approx 1.0$. The associated prediction uncertainties are illustrated in Fig. 16, in which the PP distribution is shown, for all seven instances of Φ. These plots show that while the simple model has structural deficiencies that prevent it from capturing the full range of GRImech3.0 data, it has predictive error bars that capture these model errors sufficiently well, given the present calibration procedure. Furthermore, Fig. 17 shows the predictive standard deviations across all values of T^0, as well as the error in the mean compared to the detailed model, for all seven instances of Φ. Similar to the first case study, we have performed tests across various values of the ABC likelihood parameter ε, discovering that $\varepsilon = 0.001$ is a "good" value where the ABC constraints are satisfied with sufficient accuracy such that the PP distributions are nearly equivalent to the MAP predictive distributions. In fact, for $\varepsilon = 10^{-4}$, the two distributions are effectively indistinguishable. Indeed, Fig. 18 demonstrates both posterior and MAP predictions of the density of $\boldsymbol{\lambda} = (\ln A, E)$ for sequentially reducing values of ε. Interestingly, it suggests that the best—according to the objective function that is the ABC likelihood—multivariate normal density of $(\ln A, E)$ is the uncorrelated one. For larger values of ε, the posterior variability is dominant leading to a positively correlated density. Note the clear difference in these figures compared to the same ones for the first case study in Fig. 12, in which the detailed model data are only collected across a range of initial temperatures T^0. Finally, Fig. 19 demonstrates the marginal

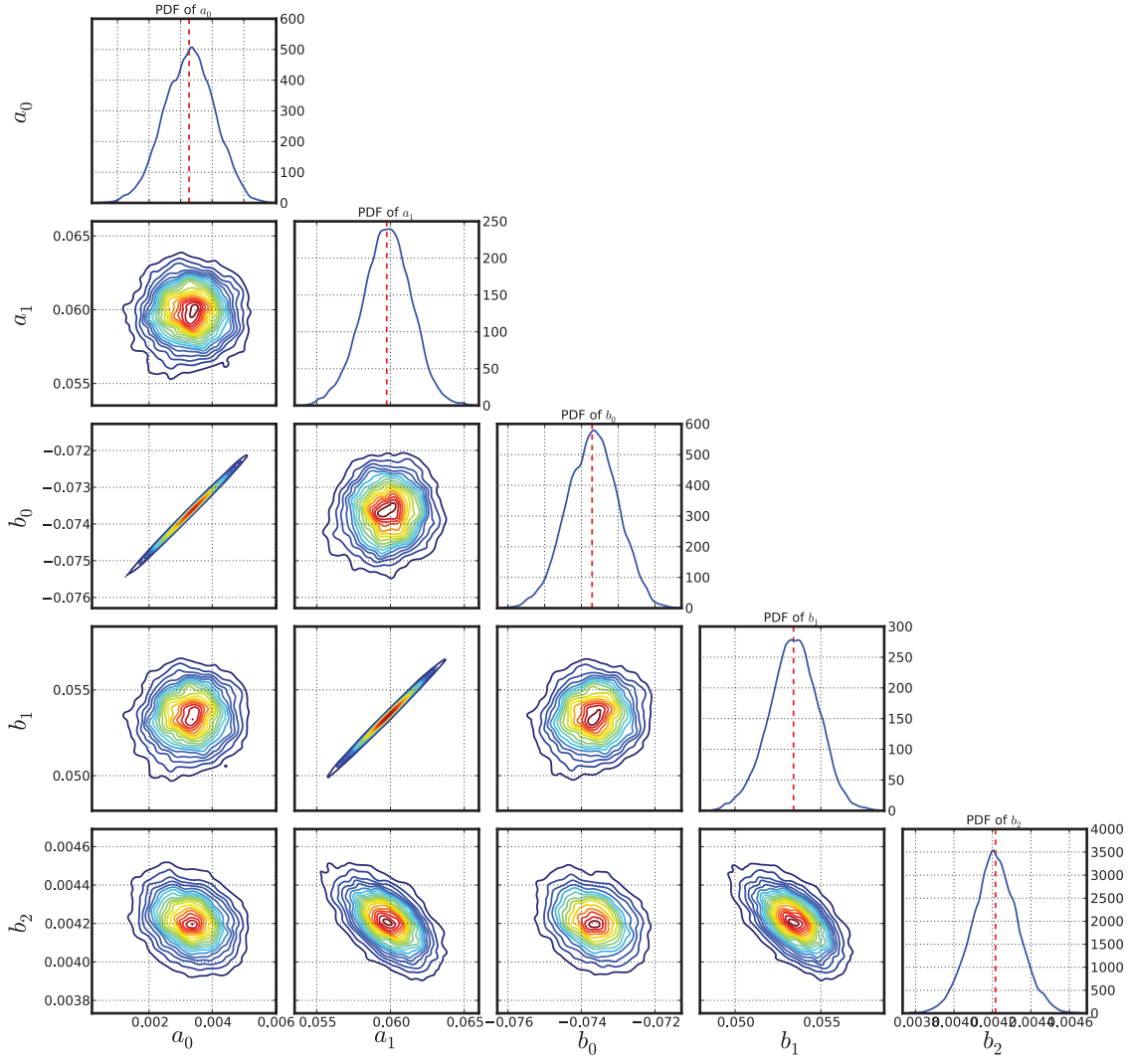

Figure 13 Marginal and joint posterior densities of the parameters $\alpha = (a_0, a_1, b_0, b_1, b_2)$ for the case with $N = 11$. Densities are computed via KDE using 10^5 MCMC samples. The parameters of the ABC likelihood are set to $\varepsilon = 0.001$ and $\gamma = 1$. The components of the MAP value α_{MAP} are highlighted by vertical red dashed lines on the 1D marginal plots.

and joint posterior densities of the parameters $\alpha = (a_0, a_1, b_0, b_1, b_2)$ for the default case of $\varepsilon = 0.001$. As in the first case study, there is a clear posterior correlation between the mean components of the parameters, i.e. between a_0 and b_0. The MAP value of the parameter b_1 is near 0, which indeed leads to uncorrelated joint density in $(\ln A, E)$ according to their probabilistic description (40). The narrowest marginal distribution also belongs to b_1, suggesting that the detailed model data strongly constraints the correlations built-in the density of $(\ln A, E)$ to vanish. Furthermore, we note a comparatively wide marginal posterior in b_2, which suggests that the variability in the vertical, E-direction, in the density of parameters (shown in Fig. 18) is less constrained by the data than the variations in the horizontal, $\ln A$-direction.

Conventional Calibration Approach for Both Case Studies

In this subsection, we show results of calibration with the conventional method, described in the second section, for both case studies, illustrating how it leads to biased predictions while not accounting for model-to-truth errors.

Figure 20 shows the posterior distribution of $\boldsymbol{\lambda}$ as well as the ID marginal posterior distribution of σ for both case studies, using the conventional calibration mechanism with Gaussian additive likelihood. The first two rows illustrate the first case study with $N = 11$ and $N = 51$, respectively. It is evident that more data imply a narrower posterior. At the same time, the MAP value shifts slightly from the $N = 11$ case to the $N = 51$

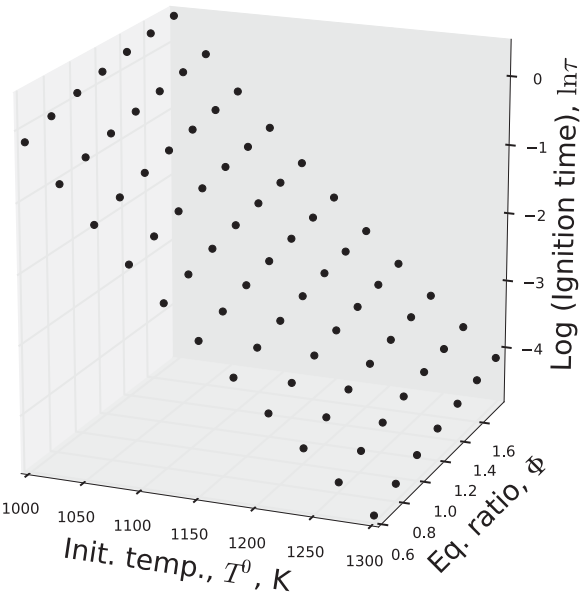

Figure 14 Ignition time data collected from the detailed methane-air mixture model using GRImech3.0, for $N_T = 11$ and $N_\Phi = 7$.

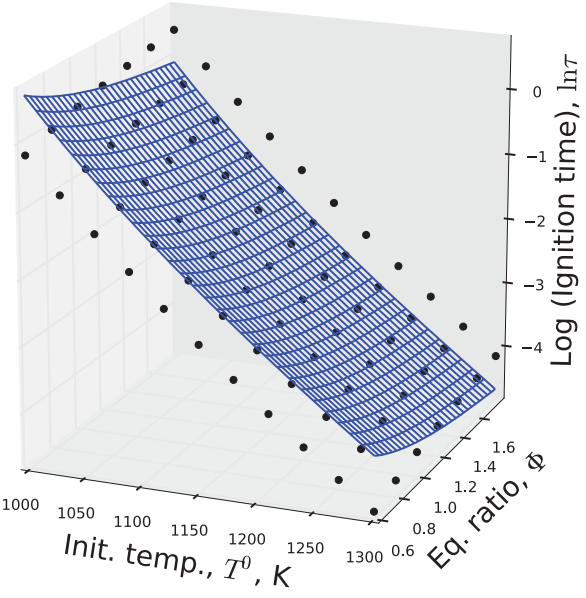

Figure 15 Ignition time data collected from the detailed methane-air mixture model using GRImech3.0, together with the mean prediction from the simple model (37).

case. For the second case study, i.e., the bottom row in Fig. 20, the posterior distribution is much wider, together with a much-larger best value of σ. It is of course expected, as we already know that the model to-be-calibrated does not capture well the detailed model data across the full range of variability of Φ.

Figure 21 illustrates the PFP uncertainties for the log-ignition time based on this calibration analysis. The first two rows correspond to the first case study. Increasing the amount of data from $N = 11$ to $N = 51$ reduces the posterior width, and therefore the PFP variance is reduced, clearly indicating that the associated predictive variance does not adequately capture the data. Similarly, for the second case study, illustrated in the bottom row of Fig. 21, the PFP again does not capture the data variability, leading to predictions that are inconsistent with the data.

Both cases demonstrate the deficiency of the conventional approach, particularly when there is significant model discrepancy, for arriving at uncertain predictions that are consistent with the discrepancy from the (noise-free) data. Of course, one can use the PP distribution, instead of the PFP, thereby accounting for ε_m, to be consistent with the data. However, this external artificial addition of a statistical mismatch, be it a Gaussian i.i.d. noise term or any other explicitly additive representation, is not the part of the predictive model, and its inclusion in model predictions has a number of shortcomings, as already discussed in the introduction.

Prediction of the Forward Reaction Rate

Given the calibrated PDF of input parameter $\boldsymbol{\lambda} = (\ln A, E)$ that is found with the ABC-based mechanism introduced in this paper, one can proceed to predictions of other quantities of interest relevant to the calibrated model. For example, the logarithm of the forward reaction rate can be written as

$$\ln k(T) = \ln A - \frac{1}{RT} E \tag{45}$$

which is a linear function with respect to components of $\boldsymbol{\lambda}$, i.e. calibrated input parameter pair $(\ln A, E)$. For the MAP predictions, one can then proceed analytically,

$$\ln k(T) = \ln A_0 + \Delta \ln A(a_0 + a_1\xi_1) - \frac{E_0 + \Delta E(b_0 + b_1\xi_1 + b_2\xi_2)}{RT} \tag{46}$$

Therefore, the MAP prediction mean and variance can be written as

$$\text{Mean}_{\ln k(T)} = \ln A_0 - \frac{E_0}{RT} + a_0 \Delta \ln A - b_0 \frac{\Delta E}{RT} \tag{47}$$

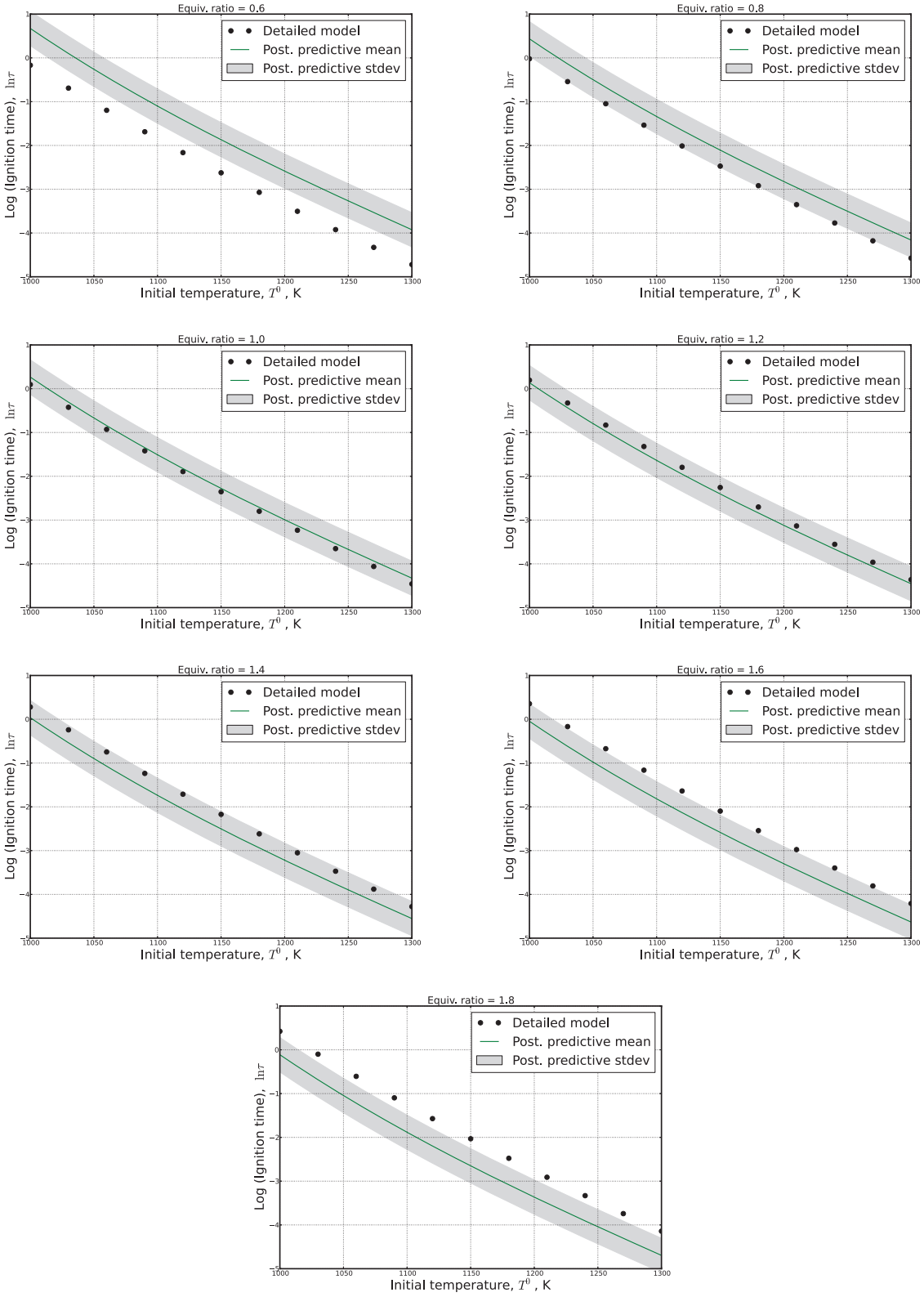

Figure 16 The results of calibration of simple single-step model from Eq. (37) using the GRImech3.0 detailed model data, for various values of the equivalence ratio Φ. The posterior predictive standard deviations are illustrated with grey bands. The parameters in the ABC likelihood (24) are set to $\varepsilon = 0.001$ and $\gamma = 1$. For the chosen value of ε, the MAP predictive mean and standard deviation (not shown) nearly coincide with the posterior predictive ones.

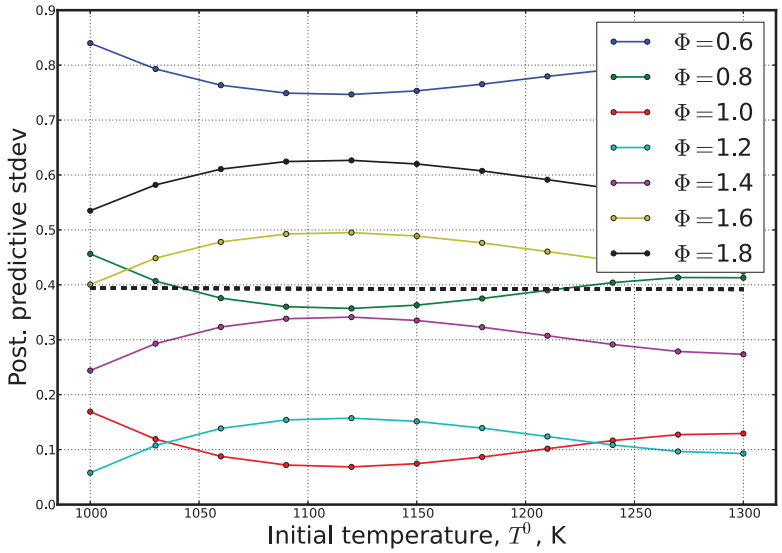

Figure 17 The error in the mean predictions for all instances of Φ. The MAP and PP standard deviations nearly coincide for this $\varepsilon = 0.001$ case. Also, they are almost uniform across the initial temperature range and are illustrated with a dashed black line.

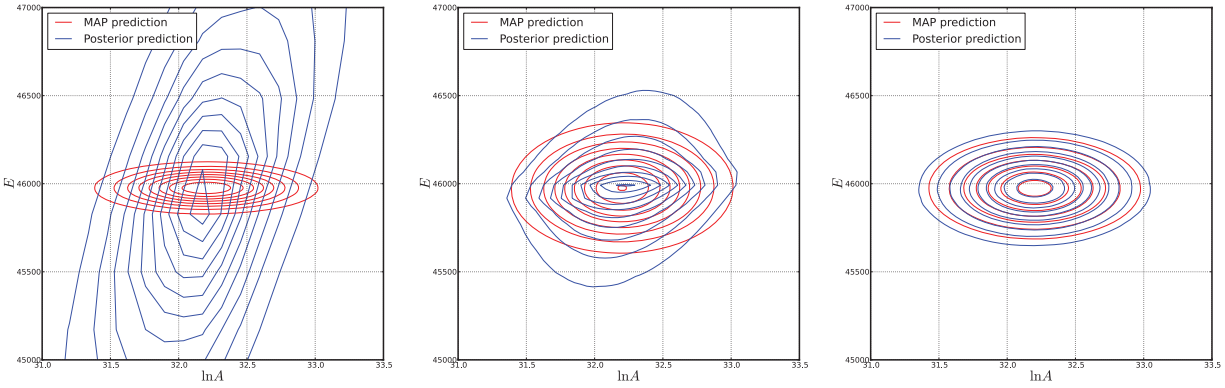

Figure 18 PP (blue) and MAP predictive (red) densities of $\boldsymbol{\lambda} = (\ln A, E)$ for four different values of ε: from left to right, $\varepsilon = 0.1$, $\varepsilon = 0.01$, and $\varepsilon = 0.001$. The contour plot for the PP is based on $10^4 \times 10^3$ samples, i.e. 10^4 samples of $\boldsymbol{\lambda}$ according to $\pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}; \alpha)$, for each of 10^3 posterior samples of α.

and

$$\mathrm{Var}_{\ln k(T)} = \left(a_1 \Delta \ln A - b_1 \frac{\Delta E}{RT}\right)^2 + \left(b_2 \frac{\Delta E}{RT}\right)^2 \tag{48}$$

where $(a_0, a_1, b_0, b_1, b_2)$ are the MAP parameters. With decreasing ε, the posterior distribution narrows down, and the PP variance reduces to the MAP predictive one, similar to the prediction of ignition time observed in Case Studies 1 and 2. Figure 22 demonstrates the predictions for the forward rate and the predictive standard deviations for the two case studies with $\varepsilon = 0.001$. The linear dependence of the log forward rate on the inverse temperature is evident by definition. In the first case study, the calibrated model performed very well in terms of matching the detailed model data leading to smaller predictive error bars compared to those for the second case study, which indicates the model's deficiencies when the equivalence ratio is also considered, thus increasing the predictive error bars and implying reduced predictive fidelity of the model.

DISCUSSION

The above illustrations, including the cubic function case, as well as the two ignition cases with/without

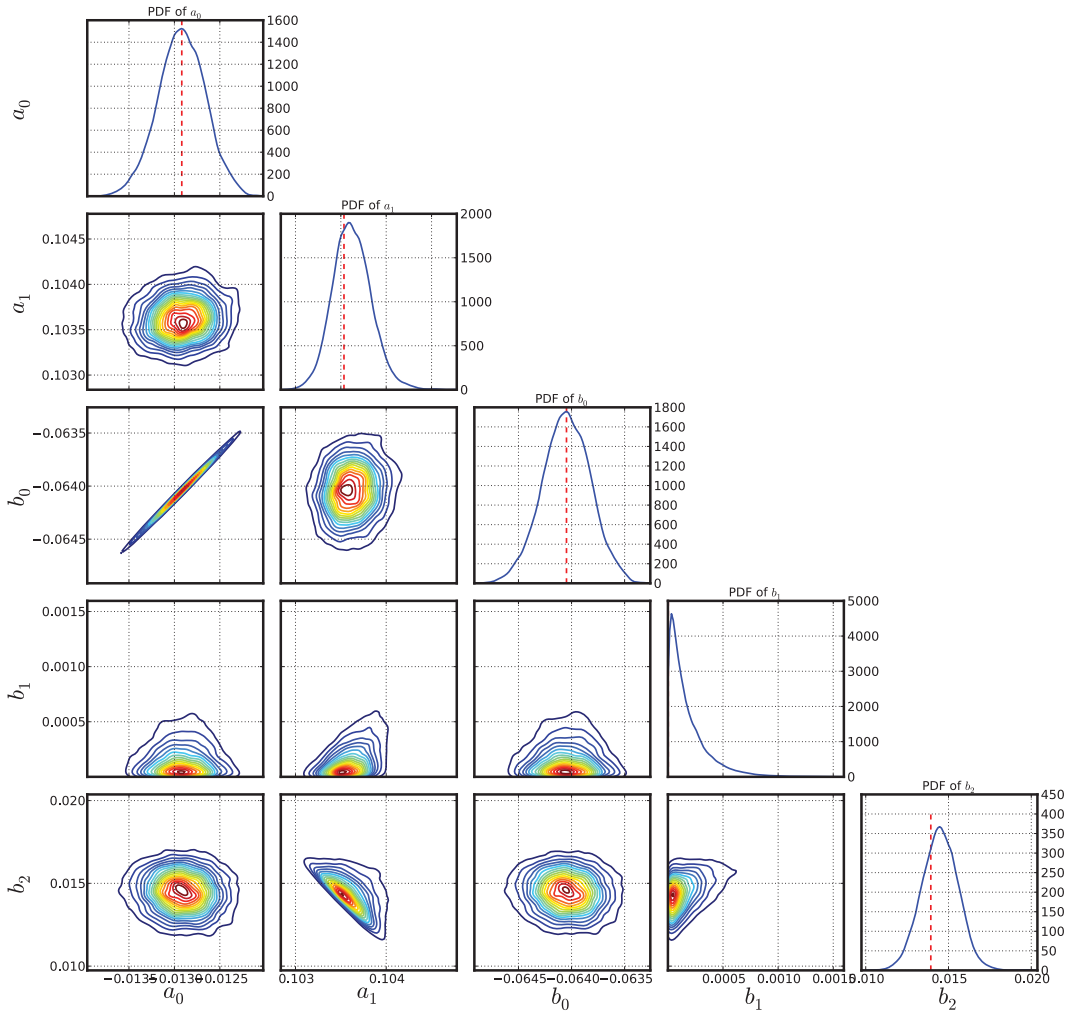

Figure 19 Marginal and joint posterior densities of the parameters $\alpha = (a_0, a_1, b_0, b_1, b_2)$. Densities are computed via KDE using 100, 000 MCMC samples. The parameters of the ABC likelihood are set to $\varepsilon = 0.001$ and $\gamma = 1$. The components of the MAP value α_{MAP} are highlighted by vertical red dashed lines on the 1D marginal plots. The MAP value for b_1 is very small, rendering the MAP predictive PDF of $(\ln A, E)$ an uncorrelated Gaussian in Fig. 18.

inclusion of equivalence ratio variation, exhibit the key features of the present model-to-model calibration procedure. The result of the calibration of the simple model in each case is a calibrated model whose uncertain predictions satisfy constraints of interest, being centered, in an averaged sense, on the detailed-model data, and having a range of uncertainty consistent with the range of discrepancy from the latter. Having more data, leads to a more precise consistency with these constraints.

By contrast, the conventional Bayesian calibration procedure, with a presumed additive statistical error model, $y = f(x, \boldsymbol{\lambda}) + \varepsilon$, leads to distinctly different results. Note that, as indicated earlier, there are two posterior predictions of interest. One is the PFP, being the result of pushing-forward the posterior density on model parameters through the fit-model $y = f(x, \boldsymbol{\lambda})$. The other, the PP, is the PFP through the data model $y = f(x, \boldsymbol{\lambda}) + \varepsilon$. In the above density-estimation ABC context, with the embedded ε, the two are the same by definition. Here, however, in the additive-noise context, they are clearly quite different. Whereas the PP exhibits predictions with a range of uncertainty consistent with the spread of the data, the PFP does not. This is as it should be, when the error term ε is due to measurement noise. In the present context, we submit that, in practice, the calibrated model $f(x, \boldsymbol{\lambda})$ would be used on its own for prediction, without the inclusion of the additive statistical model ε, in which case it is clearly the PFP that is relevant. The PFP resulting from the conventional calibration procedure, while being centered

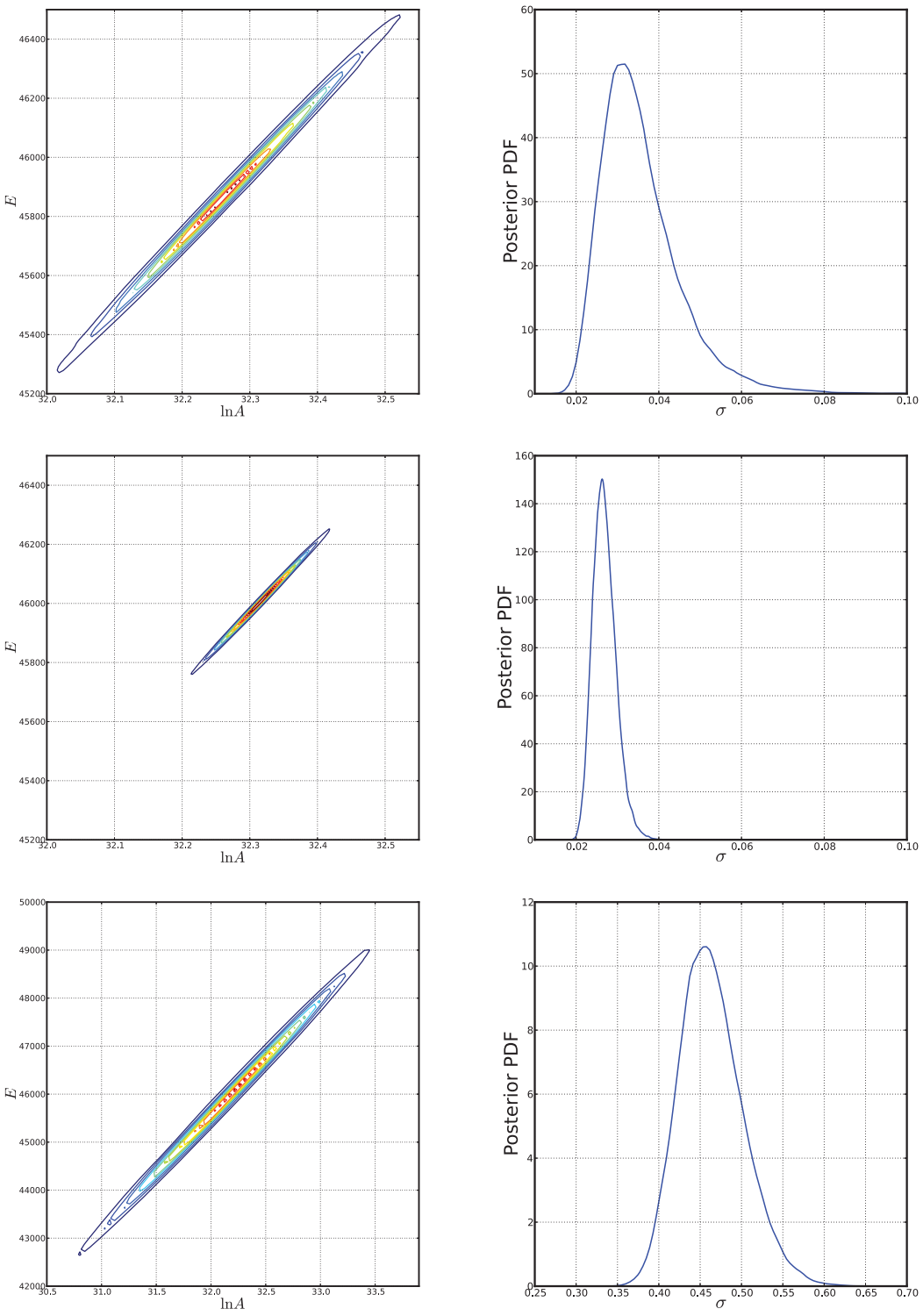

Figure 20 Marginal posterior distributions of the pair $\boldsymbol{\lambda} = (\ln A, E)$ (left column) and for σ (right column) for calibration with additive Gaussian noise assumption. The first row corresponds to the first case study with $N = 11$, the middle row illustrates the same case study only with $N = 51$, whereas the bottom row illustrates the densities for the second case study in which the data are collected across both initial temperature and equivalence ratio. The plots of the first case study (the top two rows) are shown on the same axes scales to facilitate visual comparison.

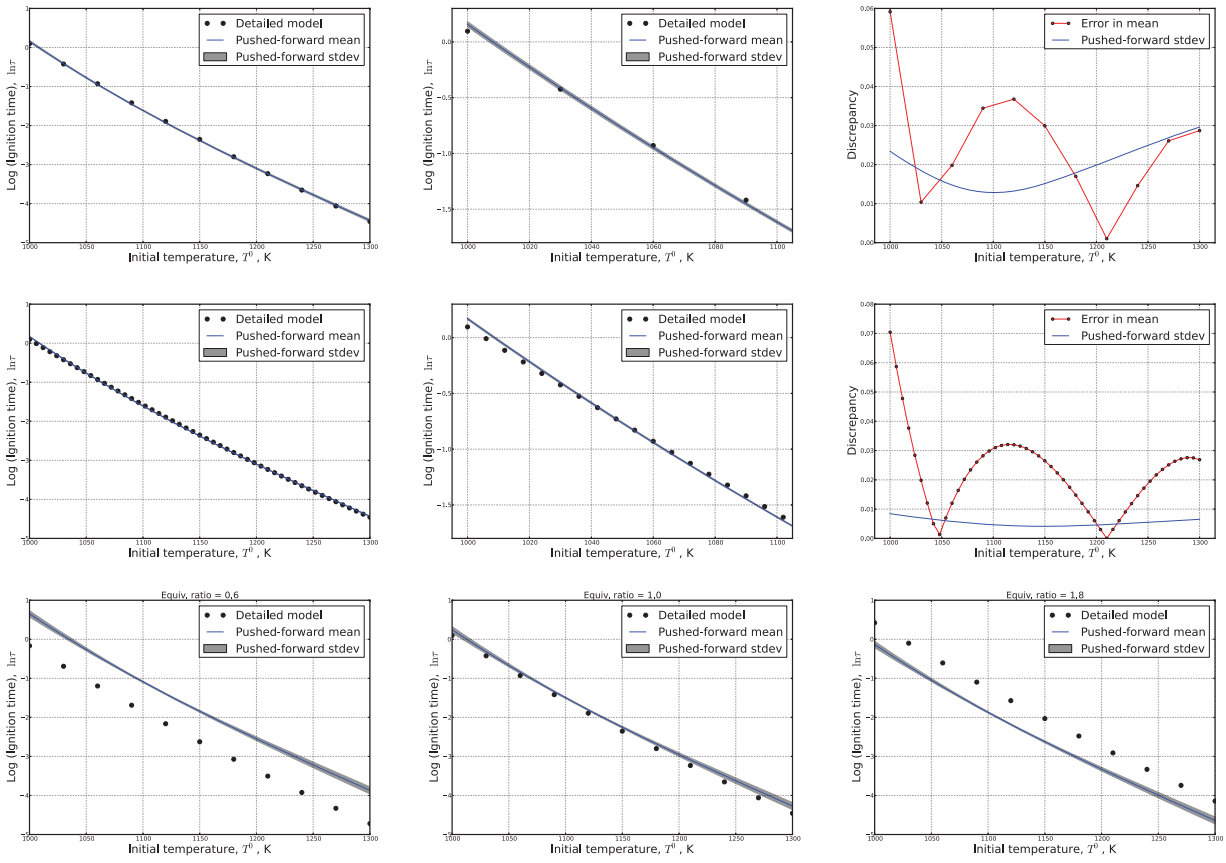

Figure 21 Results of calibration with additive Gaussian noise assumption. The top row shows the first case study with $N = 11$, the middle row is the same case study for $N = 51$, whereas the bottom row shows the second case study. For the first case study (first two rows), the middle column is simply a zoomed-in version of the left column plots, whereas the right columns illustrates the predictive standard deviations and the resulting mean error. For the second case study (bottom row), three "slices" are shown for three instances of Φ.

on the data (given no prior bias), has a range of uncertainty in predictions that, in general, does not reflect the discrepancy from the detailed model prediction and gets progressively smaller with additional data. We stress that, from a scientific model-to-model calibration perspective, the density-estimation context, where the resulting equivalent PFP/PP has the desired properties enforced with ABC, is of significantly higher relevance.

Considering specifically the two chemical-ignition cases presented above, whereas the structure of the predictive uncertain ignition time picture is similar in both cases, as per the imposed ABC-procedure constraints, it is of interest to examine the differences in the resulting uncertainty picture for $\boldsymbol{\lambda}$. Let us consider the small-ε cases in particular, where the MAP predictive and the PP densities $p(\boldsymbol{\lambda}|\mathcal{D})$ are essentially equivalent. We note the significant differences between the PP $\boldsymbol{\lambda}$ densities between Case 1 (only T^0 variation) and Case 2 (with (T^0, Φ) variation). In Case 1, the $\boldsymbol{\lambda}$-PP exhibits a high degree of correlation, being very nearly a narrow ridge with a positively sloped spine (Fig. 12). On the other hand, in Case 2, the $\boldsymbol{\lambda}$-PP is essentially uncorrelated (Fig. 18). The two densities are peaked roughly at the same MAP $(\ln A, E)$ value, but with clearly very different structure, leading to a predictive $\ln k(T)$ and ignition-time uncertainty that is significantly larger in Case 2, as compared to Case 1, consistent with the lack of ability of the simple model to accurately capture the result of variation in both T^0 and Φ on ignition time. Aside from the difference in the amplitude of predictive uncertainty, however, which is expected given the discrepancy with ignition time data in each case, it is not intuitively clear why the large difference in the $\boldsymbol{\lambda}$-PP correlation structure between the two cases. Analysis results, not shown, suggest that the Case 2 posterior density $p_{\mathrm{ABC}}(\alpha|\mathcal{D})$ is rather shallow in the vicinity of α_{MAP}, as it is evident that some densities $p(\boldsymbol{\lambda})$ that

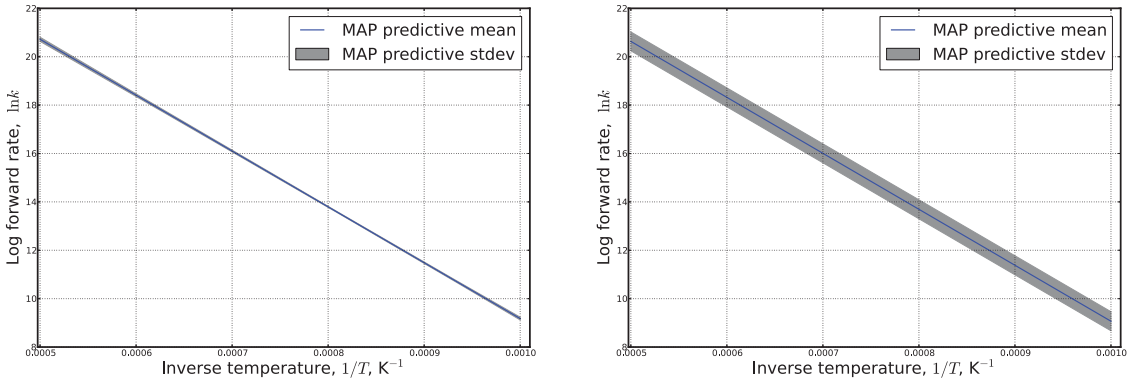

Figure 22 The predictions of the logarithm of the forward rate and the MAP predictive standard deviations. The left plot corresponds to the first case study, whereas the right plot illustrates the result for the second case study.

have strong correlation similar to that in Case 1 give a largely similar uncertainty in predicted ignition time. Nonetheless, it is evident that the uncorrelated density does provide better consistency with the imposed constraints, in the limit of small ε, leading to the findings in Fig. 18. As is well understood in general, while the MLE or MAP estimate (presuming priors of sufficiently large support) can be regarded to some extent objectively in the limit of large data and low noise, such that the question of having found the "correct" MLE/MAP value is well-posed, the same cannot be said regarding the uncertainty in the fitted parameters. It is a simple exercise to show that fitting a straight line to synthetic noisy straight-line data, with uninformative priors, can result in a positively, negatively, or uncorrelated posterior structure on the slope and intercept, depending on the range of the data, which is a subjective choice of the experimentalist. The same is not true for the MAP estimate, in the limit of large data and low noise. Thus, the drastic difference in the posterior structure resulting from the Case 1 and Case 2 data, while not intuitively clear, is nonetheless not entirely surprising. The key point, being guaranteed by the structure of the algorithm, is that in both cases the predictive uncertainty in ignition time satisfies constraints of interest in a way that a conventional, additive-noise likelihood model simply does not.

CONCLUSIONS

It is well-recognized [3,7,14,17,20,21,24,27–30] that the conventional mechanisms of statistical calibration of physical models often consider the model as "truth," neglecting model errors, and thus do not generally lead to uncertain predictions that are consistent with the data. In this regard, we developed a statistical calibration approach that embeds the model-to-truth discrepancy within model components. For the simplicity of presentation, we considered a specific detailed model as truth, therefore essentially performing model-to-model calibration. Our approach formulates the calibration problem as a problem of density estimation for model parameters of interest. With appropriate parameterization—multivariate normal or PC—of the joint density of model inputs, the problem is again reformulated as a typical calibration/parameter-estimation problem. While the Bayesian machinery is the tool of choice, its exact implementation reveals the ill-posedness of the reformulated model-to-model calibration. We therefore implemented ABC to form Bayesian likelihoods that are associated with the constraints driven by the goal of having predictive uncertainties that are consistent with the "truth" data spread. While the key aspects of the method are demonstrated on synthetic cases, we have also applied it to a simple chemical kinetic model that is calibrated with respect to ignition data from the (detailed) GRImech3.0 model. We illustrated two scenarios with data collected in different ways, across a range of initial temperatures only and across ranges of both initial temperature and equivalence ratio. In one case, the calibrated model matched very well the detailed model data, and, consequently, produced reasonably small predictive error bars. For the second case, the simple model was unable to capture the full range of data, and our method produced predictive errors that are consistent with this discrepancy between the calibrated model and the detailed "truth" model, unlike the conventional statistical calibration that would produce misleadingly small error bars in this case. Overall, we have demonstrated that our novel calibration technique properly accounts for the inherent model-to-model discrepancy therefore producing uncertain predictions that are consistent

Table AI A Summary of Predictive Quantities

Name	Formula	Mean	Variance
Posterior predictive, $Z^{\mathrm{PP}}(x)$	$\mathrm{E}_\alpha[Z(x;\alpha)]$	$\mathrm{E}_\alpha[\mu(\alpha;x)]$	$\mathrm{E}_\alpha[\sigma(\alpha;x)^2]+\mathrm{V}_\alpha[\mu(\alpha;x)]$
MAP predictive, $Z^{\mathrm{MP}}(x)$	$Z(x;\alpha_{\mathrm{MAP}})$	$\mu(\alpha_{\mathrm{MAP}};x)$	$\sigma(\alpha_{\mathrm{MAP}};x)^2$
Mean predictive, $Z^{\mathrm{MN}}(x)$	$Z(x;\alpha_{\mathrm{mean}})$	$\mu(\alpha_{\mathrm{mean}};x)$	$\sigma(\alpha_{\mathrm{mean}};x)^2$

with the available information. While our main focus, i.e., the model-to-model calibration, can be associated with model-to-data calibration with no data noise, we note that when both model errors and non-negligible data noise are present, the method needs conceptual reformulation, which is a matter of current work.

Finally, we note that the utilization of this approach for calibration of more general multistep reversible elementary kinetic mechanisms is a natural extension of the present work. Furthermore, in this context, it is feasible to imagine proposed reaction steps as hypothetical modifications to the model, where model error terms can be potentially embedded to explore their utility in explaining model error in specific predictions.

APPENDIX

The PP mean can be derived from the PP distribution (31) as follows:

$$\begin{aligned} Z^{\mathrm{PP}}_{\mathrm{mean}}(x) &= \int_z z\,\pi_{Z(x)}(z)\,dz \\ &= \int_\alpha \underbrace{\int_z z p(z|\alpha)\,dz}_{\mu(\alpha;x)}\, p(\alpha|\mathcal{D})\,d\alpha = \mathrm{E}_\alpha[\mu(\alpha;x)], \end{aligned} \tag{A1}$$

where $\mu(\alpha;x)=\mathrm{E}_\xi[f(x;\boldsymbol{\lambda}(\boldsymbol{\xi};\alpha))]$ is the forward-propagated mean of the function $f(x;\boldsymbol{\lambda}(\boldsymbol{\xi};\alpha))$ at location x for given α.

Similarly, the PP variance can be derived as

$$\begin{aligned} Z^{\mathrm{PP}}_{\mathrm{var}}(x) &= \int_z z^2\,\pi_{Z(x)}(z)\,dz - Z^{\mathrm{PP}}_{\mathrm{mean}}(x)^2 \\ &= \int_\alpha \underbrace{\int_z z^2 p(z|\alpha)\,dz}_{\sigma(\alpha;x)^2+\mu(\alpha;x)^2}\, p(\alpha|\mathcal{D})d\alpha - Z^{\mathrm{PP}}_{\mathrm{mean}}(x)^2 \\ &= \mathrm{E}_\alpha[\sigma(\alpha;x)^2]+\mathrm{V}_\alpha[\mu(\alpha;x)] \end{aligned} \tag{A2}$$

where $\sigma(\alpha;x)^2=\mathrm{V}_\xi[f(x;\boldsymbol{\lambda}(\boldsymbol{\xi};\alpha))]$ is the forward-propagated variance of the function $f(x;\boldsymbol{\lambda}(\boldsymbol{\xi};\alpha))$ at location x for given α, and V_α denotes variance with respect to the posterior distribution of α. Note that in practice the posterior distribution is described via samples of an MCMC chain; therefore, the expectation (E_α) and the variance (V_α) are computed via Monte-Carlo integration using MCMC samples.

Albeit not used in this paper, yet another predictive quantity can be obtained, by focusing on the mean posterior value $\alpha_{\mathrm{mean}}=\mathrm{E}_\alpha[\alpha]$, leading to $Z^{\mathrm{MN}}(x)=Z(x;\alpha_{\mathrm{mean}})$ with mean $Z^{\mathrm{MN}}_{\mathrm{mean}}(x)=\mu(\alpha_{\mathrm{mean}};x)$ and variance $Z^{\mathrm{MN}}_{\mathrm{var}}(x)=\sigma(\alpha_{\mathrm{mean}};x)^2$.

Note that, in general, the means of the three posterior predictions $Z^{\mathrm{PP}}_{\mathrm{mean}}(x)$, $Z^{\mathrm{MP}}_{\mathrm{mean}}(x)$, and $Z^{\mathrm{MN}}_{\mathrm{mean}}(x)$ are different from each other. Indeed, since generally, $\alpha_{\mathrm{MAP}}\neq\alpha_{\mathrm{mean}}$; therefore, $Z^{\mathrm{MP}}_{\mathrm{MAP}}(x)=\mu(\alpha_{\mathrm{MAP}};x)\neq\mu(\alpha_{\mathrm{mean}};x)=Z^{\mathrm{MP}}_{\mathrm{mean}}(x)$. Furthermore, since the function f is not linear with respect to $\boldsymbol{\lambda}$ in general, $Z^{\mathrm{PP}}_{\mathrm{mean}}(x)=\mathrm{E}_\alpha[\mu(\alpha;x)]=\mathrm{E}_\alpha\mathrm{E}_\xi[f(x;\boldsymbol{\lambda}(\boldsymbol{\xi};\alpha))]\neq\mathrm{E}_\xi[f(x;\boldsymbol{\lambda}(\boldsymbol{\xi};\alpha_{\mathrm{mean}}))]=\mu(\alpha_{\mathrm{mean}};x)=Z^{\mathrm{MN}}_{\mathrm{mean}}(x)$.

Table AI summarizes the three predictive quantities described in this section.

KS and RG acknowledge support for this work through the Scientific Discovery through Advanced Computing (SciDAC) program funded by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research. HNN acknowledges support for this work through the US Department of Energy (DOE), Office of Basic Energy Sciences (BES) Division of Chemical Sciences, Geosciences, and Biosciences. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94-AL85000.

BIBLIOGRAPHY

1. Masten, D.; Hanson, R.; Bowman, C. J Phys Chem 1990, 94, 7119–7128.
2. Miki, K.; Panesi, M.; Prudencio, E.; Prudhomme, S. Phys Plasmas 2012, 19(2), 023507.
3. Mosbach, S.; Hong, J. H.; Brownbridge, G. P.; Kraft, M.; Gudiyella, S.; Brezinsky, K. Int J Chem Kinet 2014, 46(7), 389–404.
4. Nagy, T.; Turányi, T. Int J Chem Kinet 2011, 43(7), 359–378.

5. Nagy, T.; Turányi, T. Reliab Eng Syst Safety 2012, 107, 29–34.
6. Varga, L.; Szabó, B.; Zsély, I. G.; Zempléni, A.; Turányi, T. J Math Chem 2011, 49(8), 1798–1809.
7. Donahue, N.; Clarke, J. Int J Kinet, Chem. 2004, 36(5), 259–272.
8. Olm, C.; Zsély, I. G.; Pálvölgyi, R.; Varga, T.; Nagy, T.; Curran, H. J.; Turányi, T. Combust Flame 2014, 161(9), 2219–2234.
9. Héberger, K.; Kemény, S.; Vidóczy, T. Int J Chem Kinet 1987, 19(3), 171–181.
10. Sivia, D. S.; Skilling, J. Data Analysis: A Bayesian Tutorial, 2nd ed.; Oxford University Press: New York, 2006.
11. Bernardo, J.; Smith, A. Bayesian Theory; Wiley Series in Probability and Statistics; Wiley: Chichester, UK, 2000.
12. Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. B. Bayesian Data Analysis, 2nd ed.; Chapman & Hall CRC Press: Boca Raton, FL, 2003.
13. Robert, C.; Casella, G. Monte Carlo Statistical Methods; Springer Texts in Statistics; Springer: Berlin, 2004.
14. Cailliez, F.; Pernot, P. J Chem Phys 2011, 134(5), 054124.
15. Mosbach, S.; Braumann, A.; Man, P. L.; Kastner, C. A.; Brownbridge, G. P.; Kraft, M. Combust Flame 2012, 159(3), 1303–1313.
16. Labahn, J. W.; Devaud, C. B.; Sipkens, T. A.; Daun, K. J. Combust Theory Model 2014, 18(3), 474–499.
17. Miki, K.; Cheung, S. H.; Prudencio, E. E.; Varghese, P. L. Int J Chem Kinet 2012, 44(9), 586–597.
18. Miki, K.; Panesi, M.; Prudencio, E. E.; Prudhomme, S. J Comput Phys 2012, 231(9), 3871–3886.
19. Miki, K.; Prudencio, E. E.; Cheung, S. H.; Terejanu, G. Combust Flame 2013, 160(5), 861–869.
20. Upadhyay, R.; Miki, K.; Ezekoye, O.; Marschall, J. Exp Therm Fluid Sci 2011, 35(8), 1588–1599.
21. Braman, K.; Oliver, T. A.; Raman, V. Combust Theory Model 2013, 17(5), 858–887.
22. Prager, J.; Najm, H.; Sargsyan, K.; Safta, C.; Pitz, W. Combust Flame 2013, 160, 1583–1593.
23. Prager, J.; Najm, H.; Zádor, J. Proc Combust Inst 2013, 34, 583–590.
24. Duque-Bernal, M.; Quintero-Arias, J.; Osorio-Viana, W.; Dobrosz-Gómez, I.; Fontalvo, J.; Gómez-García, M. Int J Chem Kinet 2013, 45(1), 10–18.
25. Kastner, C. A.; Braumann, A.; Man, P. L.; Mosbach, S.; Brownbridge, G. P.; Akroyd, J.; Kraft, M.; Himawan, C. Chem Eng Sci 2013, 89, 244–257.
26. Frenklach, M.; Wang, H.; Rabinowitz, M. J. Prog Energy Combus Sci 1992, 18(1), 47–73.
27. Feeley, R.; Frenklach, M.; Onsum, M.; Russi, T.; Arkin, A.; Packard, A. J Phys Chem A 2006, 110(21), 6803–6813.
28. Frenklach, M. Proc Combust Inst 2007, 31(1), 125–140.
29. Sheen, D.; You, X.; Wang, H.; Løvås, T. Proc Combust Inst 2009, 32(1), 535–542.
30. Xin, Y.; Sheen, D. A.; Wang, H.; Law, C. K. Combust Flame 2014, 161(12), 3031–3039.
31. Kass, R. E.; Raftery, A. J Am Stat Assoc 1995, 90, 773–795.
32. Beck, J.; Yuen, K. J Eng Mech 2004, 130(2), 192–203.
33. Janković, B. Int J Chem Kinet 2009, 41(1), 27–44.
34. Schwarz, G. et al. Ann Stat 1978, 6(2), 461–464.
35. Akaike, H. IEEE Trans Autom Control 1974, 19(6), 716–723.
36. Cavanaugh, J. E. Stat Prob Lett 1997, 33(2), 201–208.
37. Russi, T.; Packard, A.; Feeley, R.; Frenklach, M. J Phys Chem A 2008, 112(12), 2579–2588.
38. Russi, T.; Packard, A.; Frenklach, M. Chem Phys Lett 2010, 499(1), 1–8.
39. Kennedy, M. C.; O'Hagan, A. J R Stat Soc, Ser B 2001, 63(3), 425–464.
40. Franzelli, B.; Riber, E.; Sanjosé, M.; Poinsot, T. Combust Flame 2010, 157, 1364–1373.
41. Franzelli, B.; Riber, E.; Gicquel, L. Y.; Poinsot, T. Combust Flame 2012, 159, 621–637.
42. Valorani, M.; Creta, F.; Goussis, D.; Lee, J.; Najm, H. Combust Flame 2006, 146, 29–51.
43. Valorani, M.; Creta, F.; Donato, F.; Najm, H.; Goussis, D. Proc Combust Inst 2007, 31, 483–490.
44. Valorani, M.; Creta, F.; Li Brizzi, A.; Najm, H.; Goussis, D. In AIAA 46th AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, 2008.
45. Prager, J.; Najm, H.; Valorani, M.; Goussis, D. Proc Combust Inst 2009, 32(1), 509–517.
46. Brynjarsdottír, J.; O'Hagan, A. Inverse Problems. 2014, 30, 114007.
47. Berliner, L.; Jezek, K.; Cressie, N.; Kim, Y.; Lam, C.; van der Veen, C. J Glaciol 2008, 54(187), 705–714.
48. Strong, M.; Oakley, J.; Chilcott, J. J R Stat Soc, Ser C: Appl Stat 2012, 61(1), 25–45.
49. Strong, M.; Oakley, J. SIAM/ASA J Uncertain Quantif 2014, 2, 106–125.
50. Oliver, T.A.; Moser, R. D. J Phys: Conf Ser 2011, 318, 042032.
51. Cheung, S. H.; Oliver, T. A.; Prudencio, E. E.; Prudhomme, S.; Moser, R. D. Reliab Eng Syst Saf 2011, 96, 1137–1149.
52. Smith, R. Uncertainty Quantification: Theory, Implementation and Applications; SIAM Computational Science and Engineering, SIAM-Society for Industrial and Applied Mathematics: Philadelphia, PA, 2013.
53. Dwight, R. P.; Khodaparast, H.; Mottershead, J. E. In Proceedings of International Conference on Uncertainty in Structural Dynamics (ISMA/USD), Leuven, Belgium, 2012.
54. Ghanem, R.; Spanos, P. Stochastic Finite Elements: A Spectral Approach; Springer-Verlag: New York, 1991.
55. Debusschere, B.; Najm, H.; Pébay, P.; Knio, O.; Ghanem, R.; Le Maître, O. SIAM J Sci Comput 2004, 26(2), 698–719.
56. Phenix, B.; Dinaro, J.; Tatang, M.; Tester, J.; Howard, J.; McRae, G. Combust Flame 1998, 112, 132–146.

57. Reagan, M.; Najm, H.; Ghanem, R.; Knio, O. Combust Flame 2003, 132, 545–555.
58. Reagan, M.; Najm, H.; Debusschere, B.; Le Maître, O.; Knio, O.; Ghanem, R. Combust Theory Model 2004, 8, 607–632.
59. Reagan, M.; Najm, H.; Pébay, P.; Knio, O.; Ghanem, R. Int J Chem Kinet 2005, 37(6), 368–382.
60. Le Maître, O.; Najm, H.; Pébay, P.; Ghanem, R.; Knio, O. SIAM J Sci Comput 2007, 29(2), 864–889.
61. Navalho, J. E.; Pereira, J. M.; Ervilha, A. R.; Pereira, J. C. Combus Theory Model 2013, 17(6), 1067–1095.
62. Mendes, M.; Pereira, J.; Pereira, J. Combust Flame 2011, 158(3), 466–476.
63. Najm, H. Ann Rev Fluid Mech 2009, 41(1), 35–52.
64. Tomlin, A. Proc Combust Inst vol. 34, no. 1, pp. 159–176, 2013.
65. Goldsmith, C.; Tomlin, A.; Klippenstein, S. Proc Combust Inst 2013, 34(1), 177–185.
66. Esposito, G.; Sarnacki, B.; Chelliah, H. Combust Theory Model 2012, 16(6), 1029–1052.
67. Hébrard, É.; Tomlin, A. S.; Bounaceur, R.; Battin-Leclerc, F. Proc Combust Inst 2015, 35(1), 607–616.
68. Tomlin, A. S.; Agbro, E.; Nevrlỳ V.; Dlabka, J.; Vašinek, M. Int J Chem Kinet 2014, 46(11), 662–682.
69. Najm, H.; Berry, R.; Safta, C. Sargsyan, K.; Debusschere, B. Int J Uncertain Quantif 2014, 4(2), 111–132.
70. Najm, H.; Valorani, M. J Comput Phys 2014, 270, 544–569.
71. Khalil, M.; Lacaze, G.; Oefelein, J. C.; Najm, H. N. Proc Combust Inst 2015, 35(2), 1147–1156.
72. Sheen, D.; Wang, H. Combust Flame 2011, 158(12), 2358–2374.
73. Sheen, D. A.; Wang, H. Combust Flame 2011, 158(4), 645–656.
74. Sheen, D. A. Ph.D. thesis, University of Southern California, Los Angeles, CA, 2011.
75. Sheen, D.; Rosado-Reyes, C.; Tsang, W. Proc Combust Inst 2013, 34(1), 527–536.
76. Arnst, M.; Ghanem, R.; Soize, C. J Comput Phys 2010, 229(9), 3134–3154.
77. Gelman, A.; Meng, X.-L.; Stern, H. Stat Sinica 1996, 6, 733–807.
78. Haario, H.; Saksman, E.; Tamminen, J. Bernoulli 2001, 7, 223–242.
79. Beaumont, M.; Zhang, W.; Balding, D. J. Genetics 2002, 162(4), 2025–2035.
80. Marjoram, P.; Molitor, J.; Plagnol, V.; Tavaré, S. Proc Natl Acad Sci USA 2003, 100(26), 15324–15328.
81. Sisson, S.A.; Fan, Y. In Handbook of Markov Chain Monte Carlo; Chapman and Hall/CRC Press: BOca Raton, FL, 2011; pp. 313–338.
82. Ghanem, R. ASME J Appl Mech 1998, 65, 1004–1009.
83. Ghiocel, D.; Ghanem, R. J Eng Mech 2002, 128(1), 66–77.
84. Le Maître, O.; Reagan, M.; Najm, H.; Ghanem, R.; Knio, O. J Comput Phys 2002, 181, 9–44.
85. Smith, G.; Golden, D.; Frenklach, M.; Moriarty, N.; Eiteneer, B.; Goldenberg, M.; Bowman, C.; Hanson, R.; Song, S.; Gardiner Jr. W.; Lissianski, V.; Zhiwei, Q. GRI mechanism for methane/air, version 3.0, 7/30/99; available at http://combustion.berkeley.edu/gri-mech/; Accessed January, 2015.