

# Embedded Model Error Representation for Bayesian Model Calibration

Khachik Sargsyan\*, Xun Huan†, and Habib N. Najm†

January 20, 2018

## Abstract

Model error estimation remains one of the key challenges in uncertainty quantification and predictive science. For computational models of complex physical systems, model error, also known as structural error or model inadequacy, is often the largest contributor to the overall predictive uncertainty. This work builds on a recently developed framework of embedded, internal model correction, in order to represent and quantify structural errors, together with model parameters, within a Bayesian inference context. We focus specifically on a Polynomial Chaos representation with additive modification of existing model parameters, enabling a non-intrusive procedure for efficient approximate likelihood construction, model error estimation, and disambiguation of model and data errors’ contributions to predictive uncertainty. The framework is demonstrated on several synthetic examples, as well as on a chemical ignition problem.

*Keywords:* Model error, Structural error, Uncertainty quantification, Bayesian inference, Polynomial chaos, Approximate Bayesian computation

## 1 Introduction

Both computational capabilities and observational data availability have seen rapid advances in the past decade. While these improvements catalyzed and enabled the development of new, sophisticated methods for analyzing and assimilating data with complex computational models, considerable challenges remain. One such challenge involves the quantification of uncertainty in model predictions, and its accurate attribution to different uncertainty sources, all of which is important for targeted uncertainty reduction and improved predictability. While methods for quantifying data noise/error and associated parametric uncertainty have grown relatively mature, the quantification of *model error* (also known as structural error and model inadequacy) and associated predictive uncertainty is less developed. All models involve assumptions, and none is perfect. Models of complex physical systems are also often in error due to lack of proper understanding of underlying phenomena. Accurate attribution and quantification of model error in the process of model calibration is crucial for ensuring reliable predictions with meaningful estimates of predictive uncertainty, but yet this is rarely done. This critical gap is widely acknowledged in the literature [12, 51, 9, 16, 30].

Conventional model calibration seeks best-fit estimates of model parameters according to a penalty on the mismatch between model predictions and observational data. Statistical inference techniques for model calibration and parameter estimation typically ignore errors associated with the model itself—i.e., the model is assumed to be correct. However, every model has assumptions and therefore model error. If unaccounted for, model error can introduce bias in calibration and handicap predictive

---

\*Corresponding author: [ksargsy@sandia.gov](mailto:ksargsy@sandia.gov), Sandia National Laboratories, Livermore, CA 94550, USA.

†Sandia National Laboratories, Livermore, CA 94550, USA.

utility of the model [43, 66, 64]. Conventional statistical approaches for capturing model error typically estimate error correlations from observational data employing *ad hoc* covariance structures on quantities of interest (QoIs) [43, 35, 10]. This subsequently burdens the model with additive statistical mismatch terms that are QoI-specific, and lead to predictions that can violate the underlying physical constraints imposed through the models [66, 58].

A class of methods for model error quantification, inspired by the seminal work of Kennedy and O’Hagan [43], augments model output quantities of interest (QoIs) with *external* statistical terms for bias correction. For example, Wang *et. al.* [70] developed a two-step variation of the external discrepancy approach with sequential inference of model bias and the true output, while Joseph and Melkote [41] proposed a compromise between engineering and statistical approaches for explicit correction by adjusting the model at hand incrementally and sequentially using empirical Bayes methods retaining both the fitting power of statistical methods and the predictive power of engineering models. Higdon *et al.* [36] further developed the model correction framework and demonstrated it for physical systems. However, it is well-recognized that corrections at the QoI level produce predictions that could potentially violate physical laws in the model, unless ameliorated with special prior constructions [10]. More crucially, external, QoI-specific corrections do not provide any model error augmentation for extrapolative scenarios or for prediction of other QoIs [5, 53, 54]. Indeed, the resulting calibration information cannot be easily carried to making predictions of other model outputs. Besides, the model errors evaluated as explicit bias corrections are often entangled with data/measurement errors.

As a response to these limitations, *internal* model correction approaches have been gaining popularity, where select internal model components are augmented with statistical bias correction terms. He and Xiu [34] described a general framework for both internal embedding and external correction under a constrained optimization setting without quantifying the full Bayesian uncertainty. Strong *et al.* [67, 66] applied model internal correction in health-economic studies, and discussed information measures for the model improvement. Internal model corrections were also demonstrated in Reynolds-averaged Navier Stokes (RANS) applications [52, 22], large eddy simulation (LES) computations [38], molecular simulations [11], and chemical ignition modeling [33, 58].

In our framework, developed in [58], we embed model error in model components such as constitutive laws or phenomenological parameterizations rather than as additive corrections to model outputs. Specifically, we focus on an approach where corrections are added to select model input parameters, therefore allowing the model to be treated as a black-box. The framework is developed in a general Bayesian context, where existing model parameters are inferred simultaneously along with parameters that characterize model error. Additionally, we employ Polynomial Chaos (PC) [28, 72, 49, 44] to represent the augmented inputs, permitting effortless extraction and propagation of uncertainties, eventually leading to efficient Bayesian computation. The approach leads to a model error representation that is consistent with the underlying model physics, and can be employed as a diagnostic device to enable attribution of model errors to specific submodels, thus enabling targeted model improvement and optimal experimental design. The strength of the developed framework is highlighted in model-to-model calibration studies where the classical independent Gaussian discrepancy models are not defensible.

In this present work, we extend and enhance the probabilistic framework of [58] to enable calibration with respect to noisy observational data, and to provide a principled way of attributing predictive uncertainties to components due to data noise/error, model error, as well as any additional errors associated with the lack-of-information. This generalization allows the utilization of the construction with both computational and/or experimental data, while retaining the non-intrusive PC construction and the facile implementation of the method with black-box models. We detail a range of options for construction of likelihoods for model error estimation, and explore their utility in different scenarios. Overall, this work provides solid foundations for the earlier developments, and extends them to more general practical contexts.

The paper is organized as follows. Section 2 provides a general overview of model error quantification and its associated challenges. The embedded treatment is introduced in Section 3, along with the Bayesian machinery, PC representation, and predictive uncertainty attribution. The methodology is then demonstrated on synthetic problems in Section 4, and a chemical ignition application in Section 5. The paper ends with discussions and future work suggestions in Section 6.

## 2 The model error challenge

Consider a “truth” model  $g(x_i)$  that generates datum  $y_i$  at operating condition  $x_i$  (*e.g.* spatial and temporal coordinates) through the relationship  $y_i = g(x_i) + \epsilon_i$  where  $\epsilon_i$  is the measurement noise, for a total of  $i = 1, \dots, N$  data measurements. While the data can be observed, the truth model is unknown to us. Instead, we have access to a different model  $f(x_i; \boldsymbol{\lambda})$ , where  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$  is a  $d$ -dimensional parameter vector that can be estimated from experimental data. The model discrepancy, here defined to be the difference between the truth model and our model of interest, is  $\delta(x) := g(x) - f(x; \boldsymbol{\lambda})$ . The fit model  $f(x; \boldsymbol{\lambda})$  thus relates to the observations  $y_i$  via

$$y_i = \underbrace{f(x_i; \boldsymbol{\lambda}) + \delta(x_i)}_{\text{truth model } g(x_i)} + \epsilon_i. \quad (1)$$

Conventional parameter estimation (*i.e.* calibrating for  $\boldsymbol{\lambda}$ ) often ignores model errors, effectively assuming that  $f(x; \boldsymbol{\lambda})$  is able to exactly replicate  $g(x)$  for some  $\boldsymbol{\lambda}$ . However, this assumption typically does not hold for complex physical systems, where simplified, under-resolved, and poorly conceived models are often unavoidable. Consequently, the estimated parameter values would be biased, due to neglected model deficiencies. Bayarri *et al.* [6] present a clear illustration of this challenge. Moreover, ignoring model error, and thereby treating discrepancies between model predictions and available measurements as exclusively the result of uncorrelated data errors/noise results in a calibrated model whose predictive uncertainty can exhibit excessive overconfidence. This issue becomes quite evident as the number of data points  $N$  increases, as we illustrate through the following example.

Consider an exponential-growth model  $f(x; \boldsymbol{\lambda}) = \lambda_2 e^{\lambda_1 x} - 2$ , while the truth model includes ‘saturation’ and has the form  $g(x) = \tanh 3(x - 0.3)$ . Each synthetic datum is the truth model corrupted with a measurement noise in the form of independent additive Gaussian with zero mean and standard deviation  $\sigma = 0.1$ . As seen in Figure 1, as the number of data points increases, Bayesian calibration leads to shrinking uncertainty bands in the model prediction that do not adequately capture the mismatch between model prediction and data. In other words, the calibrated results without model error become overconfident around values that are, as illustrated elsewhere [6], potentially biased. Model error needs to be taken into account in order to address these challenges.

Moreover, and as already stated, the conventional Bayesian construction for calibration with model error explicitly represents the model error term  $\delta(x)$  in (1), for example, as a stochastic process (*e.g.* a Gaussian process) with a covariance structure that is optimized within a given class [43]. However, this approach has potential pitfalls when calibrating physical models for purposes of prediction [5, 10, 53, 54]. For example, there is no mechanism for utilizing the QoI-specific additive term  $\delta(x)$  for predictions of other QoIs from the same model  $f(x; \boldsymbol{\lambda})$ . A general statistical additive term without additional treatment from physical knowledge can also violate requisite physical constraints imposed by  $g(x)$ , and, presumably satisfied by  $f(x; \boldsymbol{\lambda})$  by construction. Lastly, the form (1) suffers identifiability difficulties since it is the sum of  $\epsilon_i$  and  $\delta(x_i)$  that is relevant in the measurement magnitude  $y_i$ . While progress has been made to incorporate physics-backed regularization for  $\delta(x)$  through informative priors [10], constructing them is often *ad hoc* and not feasible in general.

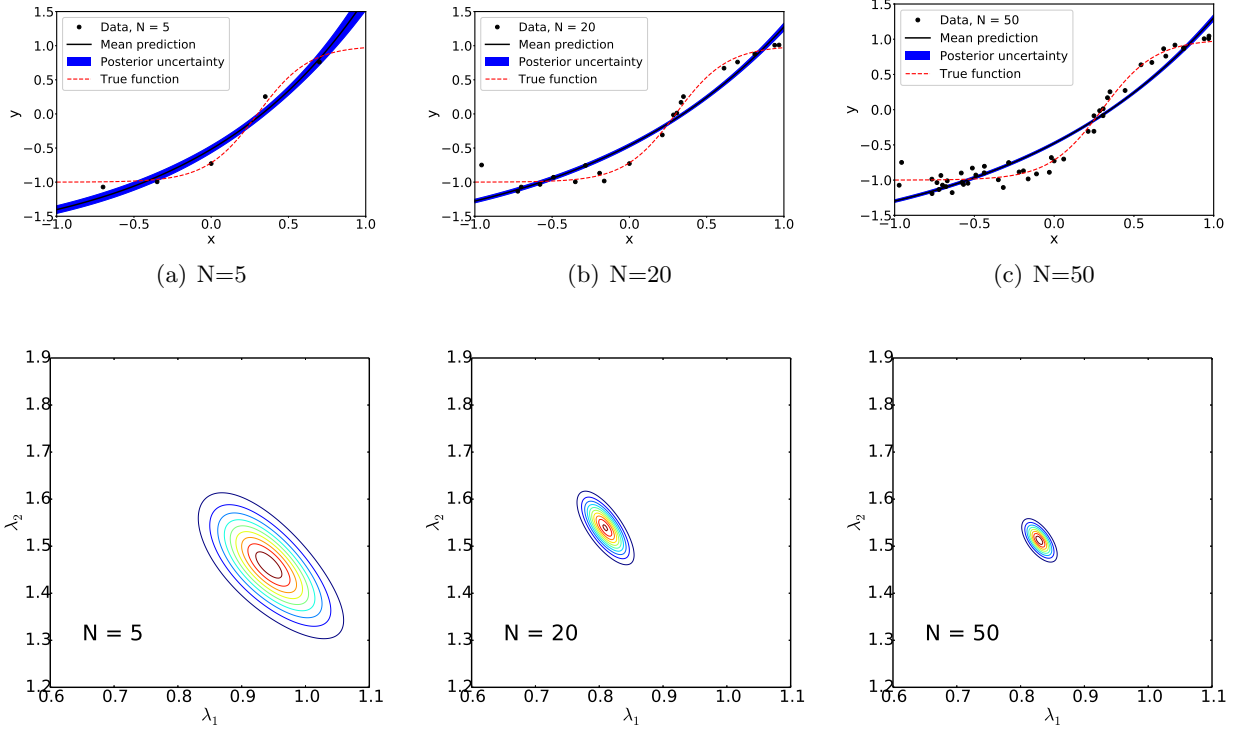


Figure 1: Demonstration of a classical Bayesian calibration – essentially a probabilistic least-squares fit – with increasing amount of data. Top row shows how pushed-forward (PF) errorbars, obtained by pushing posterior probability density functions (PDFs) through the model, shrink and do not capture the true discrepancy between data and the model. Correspondingly, the bottom row shows how calibrated parameter distributions shrink around potentially wrong values of  $\lambda_1$  and  $\lambda_2$ .

### 3 Embedded model error representation

We develop a calibration approach that accounts for model error via probabilistic embedding *within* the model, in elements of  $f(x; \boldsymbol{\lambda})$ . For example, the embedding can be carried out in an intrusive manner, via physics-backed phenomenological parameterizations that are not yet part of the model, effectively enhancing the model as  $\tilde{f}(x; \boldsymbol{\lambda}, \boldsymbol{\theta})$  to include new parameters  $\boldsymbol{\theta}$ . In this work, we demonstrate a simpler, non-intrusive approach in which a subset of parameters in  $\boldsymbol{\lambda}$  are selected and augmented by an additive discrepancy adjustment  $\boldsymbol{\delta}$ , and aim to allow the embedded model  $f(x; \boldsymbol{\lambda} + \boldsymbol{\delta})$  to produce predictive uncertainties that are consistent with the data<sup>1</sup>. This idea in its simplest form has already been demonstrated in [58] for synthetic examples and a simple chemical kinetics model, in which model parameters  $\boldsymbol{\lambda}$  were cast as probabilistic quantities with a requirement that their randomness, while respecting the physical constraints *by construction*, be consistent with the data in some sense, *i.e.*  $f(x_i; \boldsymbol{\lambda}) \approx y_i$ . The embedded model error representation is generally parameterized by  $\boldsymbol{\alpha}$ , and obeys a probability density function  $\pi_{\boldsymbol{\delta}}(\cdot; \boldsymbol{\alpha})$ , or is explicitly written as  $\boldsymbol{\delta}(\boldsymbol{\alpha}, \boldsymbol{\xi})$  to highlight the stochastic dimension in  $\boldsymbol{\xi}$ . The task of estimating parameters  $\boldsymbol{\lambda}$  is thus reformulated as an estimation of the augmented set of parameters  $\tilde{\boldsymbol{\alpha}} = (\boldsymbol{\lambda}, \boldsymbol{\alpha})$ . We note that the resulting uncertainty inflation, a term coined in [54], does not necessarily improve the validity of the inadequate model, but rather allows for meaningful calibrated predictions endowed with a degree of uncertainty that captures model predictive

<sup>1</sup>With some abuse of notation, we use vector notation  $\boldsymbol{\delta}$  for the internal correction in contrast to the explicit model discrepancy  $\delta(x)$ .

inadequacy, while remaining consistent with the physical constraints required by the physics.

The specific selection of model components for embedding is a problem-dependent task. In principle, one can select existing model components where simplifications, approximations, or phenomenological modeling have been employed. In [38], the authors employed global sensitivity analysis (GSA) in order to isolate one or two most impactful parameters for model error embedding. However, GSA only uses model information and is not data-driven. A more rigorous and comprehensive approach would involve Bayesian model selection based on evidence computation to determine the best probabilistic embedding. Note that it is also possible to envision more complicated scenarios, *e.g.* an  $x$ -dependent random process discrepancy term  $\delta(x, \alpha, \xi)$ .

The overall data model with additive model error embedding can be rewritten as

$$y_i \approx H_i(\tilde{\alpha}) \equiv h_i(\xi, \epsilon; \tilde{\alpha}) = f(x_i; \lambda + \delta(\alpha, \xi)) + \epsilon_i \text{ for } i = 1, \dots, N, \quad (2)$$

or, in vector notation,

$$\mathbf{y} \approx \mathbf{H}(\tilde{\alpha}) \equiv \mathbf{h}(\xi, \epsilon; \tilde{\alpha}) = \mathbf{f}(\lambda + \delta(\alpha, \xi)) + \epsilon, \quad (3)$$

where  $\mathbf{h}(\xi, \epsilon; \tilde{\alpha}) = (h_1(\xi, \epsilon; \tilde{\alpha}), \dots, h_N(\xi, \epsilon; \tilde{\alpha}))$  is a random vector, parameterized by  $\tilde{\alpha}$ , induced by  $\xi$  and  $\epsilon$ , as a result of the model evaluations pushed-forward through  $f(x_i; \cdot)$ , with the addition of data noise  $\epsilon$ . We define corresponding upper case quantities  $H_i(\tilde{\alpha})$  and  $\mathbf{H}(\tilde{\alpha})$  to simplify the notation, with an understanding that these are random quantities. Without loss of generality, and for ease of illustration, the components of the measurement error vector  $\epsilon$  will be modeled as independent, identically distributed normal random variables with zero mean and fixed standard deviation  $\sigma$ .

The problem of calibrating model parameters  $\lambda$  thus becomes a *density estimation* problem for the augmented stochastic input  $\lambda + \delta(\alpha, \xi)$ , or a parameter estimation problem for  $\tilde{\alpha} = (\lambda, \alpha)$ . Uncertainty for epistemic parameters can be reduced with increased data size, which translates to a corresponding reduction of prediction uncertainty. While this is sensible when the discrepancy between the model prediction  $f(x_i; \lambda)$  and the data  $y_i$  is due to measurement noise only, it is not desirable when this discrepancy includes model error (one would not expect model error to disappear even with infinite data). The uncertainty in the calibrated model prediction  $f(x_i; \lambda)$  thus needs to reflect this residual uncertainty due to model error, and the embedded approach allows such an implementation via additional stochastic dimensions encoded in  $\xi$ ; this will be described with more details in Section 3.2.

### 3.1 Bayesian inference of model error representation

The reformulated parameter estimation problem for  $\tilde{\alpha} = (\lambda, \alpha)$  can be thought of as a density estimation problem for the embedded random input  $\mathbf{\Lambda}(\tilde{\alpha}) = \lambda + \delta(\alpha; \xi)$ , and is tackled as Bayesian parameter estimation of  $\tilde{\alpha}$ . Bayesian methods are well-suited for dealing with uncertainties from different sources, from intrinsic noise in the system to parametric uncertainty and experimental errors. Although it is usually more computationally involved compared to, for example, regularized optimization techniques, the Bayesian framework provides a rigorous platform for capturing the *state-of-knowledge* about quantities of interest before and after assimilating the data. Furthermore, Bayesian techniques are very convenient for dealing with *nuisance* parameters, *i.e.* parameters that are generally unknown and are not of intrinsic interest, say, for prediction purposes. The key relationship for Bayesian inference is Bayes formula [8, 62, 13], which in this context reads as

$$p(\tilde{\alpha}|\mathbf{y}) = \frac{p(\mathbf{y}|\tilde{\alpha})p(\tilde{\alpha})}{p(\mathbf{y})}. \quad (4)$$

The *prior* probability density  $p(\tilde{\alpha})$  and the *posterior* probability density  $p(\tilde{\alpha}|\mathbf{y})$  represent degrees of belief about model parameters  $\tilde{\alpha}$  before and after the data  $\mathbf{y}$  is incorporated, respectively. The *evidence*  $p(\mathbf{y})$  often plays the role of a normalizing constant. However, it becomes a critical measure

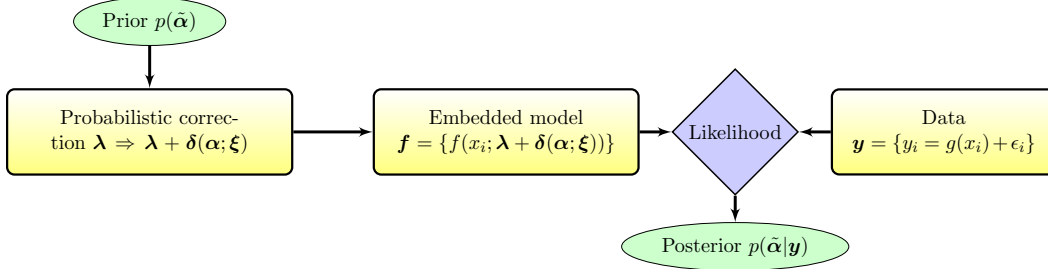


Figure 2: Schematic of the Bayesian inference of  $\tilde{\alpha} = (\lambda, \alpha)$ , *i.e.* estimation of model parameters ( $\lambda$ ) and stochastic model error correction parameters ( $\alpha$ ).

when multiple models are compared against each other given the same set of available data. The key factor in Eq. (4) is the *likelihood*  $\mathcal{L}_{\mathbf{y}}(\tilde{\alpha}) = p(\mathbf{y}|\tilde{\alpha})$  that relates the data to the model parameters. Figure 2 demonstrates the schematic of the embedded model error estimation framework within the Bayesian paradigm.

The posterior distribution is difficult to evaluate analytically for general likelihoods and priors, and is also challenging to estimate numerically due to a usually high dimensional  $\tilde{\alpha}$ . In such cases, Markov chain Monte Carlo (MCMC) methods are typically used, in which one generates a Markov chain whose stationary distribution is the posterior distribution [25, 29]. The posterior distribution  $p(\tilde{\alpha}|\mathbf{y})$  provides the current state-of-knowledge about the parameter of interest  $\tilde{\alpha}$ , while the maximum a posteriori (MAP) value

$$\tilde{\alpha}_{\text{MAP}} = \arg \max_{\tilde{\alpha}} p(\tilde{\alpha}|\mathbf{y}) \quad (5)$$

is also of interest as the vector of most probable value  $\tilde{\alpha}$  conditioned on measurements  $\mathbf{y}$ . A common point estimate for the density of the embedded random input  $\mathbf{\Lambda}(\tilde{\alpha}) = \lambda + \delta(\alpha; \xi)$  is characterized by the MAP value of  $\tilde{\alpha}$ , which can either be extracted from the full MCMC procedure, or computed via optimization techniques. Whether optimization or MCMC is used, they all require multiple evaluations of the likelihood and the prior in Eq. (4). Construction of appropriate likelihoods and, to a smaller extent, prior distributions are therefore crucial components of the presented approach.

### 3.1.1 Likelihood construction

The construction of a justifiable likelihood is perhaps the most critical step for obtaining the posterior probability distribution. From the data model (2) one can write the *true* likelihood as

$$\mathcal{L}_{\mathbf{y}}(\tilde{\alpha}) = \pi_{\mathbf{H}(\tilde{\alpha})}(\mathbf{y}), \quad (6)$$

where  $\pi_{\mathbf{H}(\tilde{\alpha})}(\mathbf{y})$  is the PDF of the data prediction vector  $\mathbf{H}(\tilde{\alpha}) = \mathbf{h}(\xi, \epsilon; \tilde{\alpha})$  defined in Eq. (3). Therefore, the true likelihood evaluation (6) requires estimation of an  $N$ -dimensional PDF as  $\mathbf{y} \in \mathbb{R}^N$ . This can be performed by sampling of both  $\xi$  and  $\epsilon$  for a given  $\tilde{\alpha}$ , followed by a Kernel Density Estimation (KDE) to obtain a smooth estimate of  $\pi_{\mathbf{H}(\tilde{\alpha})}(\mathbf{y})$  [59, 60]. For example, with  $R$  samples of  $\mathbf{H}(\tilde{\alpha})$  the KDE-estimated likelihood is written as

$$\mathcal{L}_{\mathbf{y}}(\alpha) = \pi_{\mathbf{H}(\tilde{\alpha})}(\mathbf{y}) \approx \frac{1}{R} \sum_{r=1}^R K(\mathbf{h}^{(r)} - \mathbf{y}). \quad (7)$$

Typically, Gaussian kernels  $K(\mathbf{z}) = (2\pi)^{-d/2} |\mathbf{W}|^{-1/2} e^{-\mathbf{z}^T \mathbf{W}^{-1} \mathbf{z}/2}$  are used. The bandwidth matrix  $\mathbf{W}$ , required to be symmetric and positive definite, is often selected as a diagonal matrix with entries



proportional to the marginal standard deviations of the samples. Note that this procedure has to be repeated at each MCMC sample  $\tilde{\alpha}$  and is computationally infeasible in most practical cases.

The true likelihood is degenerate for model-to-model calibration studies when there is no measurement error, as recognized in [58] and, in a slightly different context, in [3]. The degeneracy is a direct consequence of the fact that, if no data error is expected, there is *zero* probability that the model  $f(x; \lambda)$  is able to replicate the ‘truth’  $g(x)$  across all  $x$  no matter how well one tunes parameter  $\lambda$ , unless  $g(\cdot)$  itself is a special case of  $f(\cdot; \lambda)$  for some fixed value of  $\lambda$ . With data noise, while the likelihood is not degenerate anymore, it may still lead to posteriors that are hard to sample from. Besides, as already indicated, the true likelihood computation requires a high-dimensional KDE at each likelihood evaluation. Below, several approximating options for the true likelihood, that trade-off computational feasibility and accuracy, are listed.

- **Independent-component approximation:** The likelihood can be approximated as a product of its marginal components as

$$\mathcal{L}_{\mathbf{y}}^{\text{IC}}(\tilde{\alpha}) = \prod_{i=1}^N \pi_{h_i}(y_i), \quad (8)$$

where each marginal PDF  $\pi_{h_i}(y_i)$  is estimated via sampling and KDE. That is, having constructed  $R$  samples of  $\mathbf{h}$ , one approximates the marginal PDFs as

$$\pi_{h_i}(y_i) \approx \frac{1}{R} \sum_{r=1}^R K(h_i^{(r)} - y_i) \text{ for all } i = 1, \dots, N, \quad (9)$$

where, again, Gaussian kernels  $K(z) = (2\pi)^{-1/2} w^{-1} e^{-z^2/(2w^2)}$  are typically employed. Note that KDE estimates in this case are one-dimensional and computationally much more feasible compared to Eq. (7). Nevertheless, the dependence on the KDE bandwidth  $w$ , sample size  $R$ , and the general expense of sampling at each likelihood computation step, can make this approach computationally challenging, albeit to a lesser degree than the computation of the true likelihood via Eq. (7).

- **Multivariate normal approximation:** In this case, one can extract the mean and the covariance of the random vector  $\mathbf{h}$  and, instead of estimating the true joint PDF  $\pi_{\mathbf{h}}(\mathbf{y})$ , evaluate a multivariate normal (MVN) corresponding to the mean and the covariance. If random sampling is employed for extracting the first two moments of  $\mathbf{h}$ , one has

$$\boldsymbol{\mu}^h \approx \frac{1}{R} \sum_{r=1}^R \mathbf{h}^{(r)} \quad \text{and} \quad \mathbf{C}^h \approx \frac{1}{R-1} \sum_{r=1}^R \left( \mathbf{h}^{(r)} - \boldsymbol{\mu}^h \right) \left( \mathbf{h}^{(r)} - \boldsymbol{\mu}^h \right)^T, \quad (10)$$

then one arrives at the likelihood

$$\mathcal{L}_{\mathbf{y}}^{\text{MVN}}(\tilde{\alpha}) = (2\pi)^{-\frac{N}{2}} |\mathbf{C}^h|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}^h)^T (\mathbf{C}^h)^{-1} (\mathbf{y} - \boldsymbol{\mu}^h)}. \quad (11)$$

- **Independent normal approximation:** This option is a combination of the independent-component and the multivariate normal approximations. With random sampling, the mean and the variance of the components of  $\mathbf{h}$  are computed as

$$\mu_i^h \approx \frac{1}{R} \sum_{r=1}^R h_i^{(r)} \quad \text{and} \quad (\sigma_i^h)^2 = \frac{1}{R-1} \sum_{r=1}^R (h_i^{(r)} - \mu_i^h)^2, \quad (12)$$

followed by the likelihood estimate

$$\mathcal{L}_{\mathbf{y}}^{\text{IN}}(\tilde{\alpha}) = (2\pi)^{-\frac{N}{2}} \prod_{i=1}^N (\sigma_i^h)^{-1} e^{-\frac{(y_i - \mu_i^h)^2}{2(\sigma_i^h)^2}}. \quad (13)$$

Likelihood type	Formula	Eq.
True likelihood	$\mathcal{L}_{\mathbf{y}}(\tilde{\boldsymbol{\alpha}}) = \pi_{\mathbf{h}}(\mathbf{y})$	(6)
Independent-component	$\mathcal{L}_{\mathbf{y}}^{\text{IC}}(\tilde{\boldsymbol{\alpha}}) = \prod_{i=1}^N \pi_{h_i}(y_i)$	(8)
Multivariate normal	$\mathcal{L}_{\mathbf{y}}^{\text{MVN}}(\tilde{\boldsymbol{\alpha}}) = (2\pi)^{-\frac{N}{2}}  \mathbf{C}^h ^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}^h)^T \mathbf{C}^{h-1} (\mathbf{y} - \boldsymbol{\mu}^h)}$	(11)
Independent normal	$\mathcal{L}_{\mathbf{y}}^{\text{IN}}(\tilde{\boldsymbol{\alpha}}) = (2\pi)^{-\frac{N}{2}} \prod_{i=1}^N (\sigma_i^h)^{-1} \exp\left(-\frac{(y_i - \mu_i^h)^2}{2(\sigma_i^h)^2}\right)$	(13)
ABC; mean and stdev	$\mathcal{L}_{\mathbf{y}}^{\text{ABC}}(\tilde{\boldsymbol{\alpha}}) = \frac{1}{\epsilon\sqrt{2\pi}} \prod_{i=1}^N \exp\left(-\frac{(\mu_i^h - y_i)^2 + (\sigma_i^h - \gamma \mu_i^h - y_i )^2}{2\epsilon^2}\right)$	(14)

Table 1: A summary of potential options for likelihood construction, including the corresponding equation number. Note that all PDFs and moments of  $\mathbf{h}$  and their components depend on  $\tilde{\boldsymbol{\alpha}}$ , but are dropped for simplicity.

- **Moment-matching approximation:** To avoid computing expensive likelihoods, and in order to build constraints driven by the purposes of the modeler, one can construct other forms of approximate likelihoods, inspired by Approximate Bayesian Computation (ABC), otherwise called likelihood-free methods [7, 47, 61]. This approach measures the discrepancy between a chosen set of statistics of model outputs and the corresponding estimates from the data. In this regard, component-wise means and variances are common choices. The ABC formulation relies on a kernel function, typically a Gaussian  $K(z) = e^{-z^2/2}/\sqrt{2\pi}$ , and a chosen distance measure between statistics of interest on data prediction ( $S_{\mathbf{h}}$ ) and observed data ( $S_{\mathbf{y}}$ ), as well as a pre-selected ‘tolerance’ parameter  $\epsilon$  to arrive at

$$\mathcal{L}_{\mathbf{y}}^{\text{ABC}}(\tilde{\boldsymbol{\alpha}}) = \epsilon^{-1} K(\epsilon^{-1} \rho(S_{\mathbf{h}}, S_{\mathbf{y}})) = \frac{1}{\epsilon\sqrt{2\pi}} e^{-\frac{\rho(S_{\mathbf{h}}, S_{\mathbf{y}})^2}{2\epsilon^2}}. \quad (14)$$

It is of interest for predictive purposes to require agreement, on average, between (a) the mean prediction and the data, and (b) the predictive standard deviation and the discrepancy between the mean prediction and the data. That is, besides matching the means  $\boldsymbol{\mu}^h \approx \mathbf{y}$ , one should additionally constrain the standard deviation as  $\boldsymbol{\sigma}^h \approx \gamma|\boldsymbol{\mu}^h - \mathbf{y}|$ , where the absolute value is understood to be taken element-wise, and  $\gamma > 0$  is an additional parameter defined by modeler. In this case, the ABC likelihood takes the form

$$\mathcal{L}_{\mathbf{y}}^{\text{ABC}}(\tilde{\boldsymbol{\alpha}}) = \frac{1}{\epsilon\sqrt{2\pi}} \prod_{i=1}^N \exp\left(-\frac{(\mu_i^h - y_i)^2 + (\sigma_i^h - \gamma|\mu_i^h - y_i|)^2}{2\epsilon^2}\right). \quad (15)$$

Note that all PDFs and moments of  $\mathbf{h}$  or their components depend on  $\tilde{\boldsymbol{\alpha}}$ , albeit dropped for simplicity in Eqs. (7)-(15).

A proof-of-concept demonstration of such an ABC approach has been performed in [58] in the context of calibrating a low-fidelity chemical kinetic model with respect to (noise free) simulation data from a higher-fidelity model. As a direct consequence of the ABC likelihood, the resulting calibrated uncertain model outputs were shown to be centered on the data, at the same time exhibiting a degree of uncertainty consistent with the average discrepancy from the data.

Table 1 summarizes the likelihood options in a concise way. Note that, in this section, the PDF of  $\mathbf{h}$ , as well as, when necessary, computation of its moments, are estimated via random sampling, requiring a potentially large number  $R$  of model evaluations. With the non-intrusive polynomial chaos (PC) approach described further in Section 3.2.2, one can (a) reduce the sampling cost by replacing model evaluations with PC evaluations, if KDE estimates are necessary, and (b) evaluate moments exactly from PC representations, without additional cost.



### 3.1.2 Prior construction

Another major ingredient of Bayes formula (4) is the prior PDF  $p(\tilde{\alpha})$  on parameters  $\tilde{\alpha} = (\lambda, \alpha)$ . The prior, in the absence of information to the contrary, is often separated for convenience, *i.e.*  $p(\tilde{\alpha}) = p(\lambda)p(\alpha)$ . Prior selection is a known conceptual challenge for any Bayesian method. In this context, while one selects  $p(\lambda)$  according to some physical considerations, *e.g.* expert knowledge, or a result of a previous calibration, the additional difficulty arises from the fact that  $\alpha$  are not physically-meaningful model parameters, but rather parameters that define the discrepancy term  $\delta(\alpha, \xi)$ . At the same time, prior selection offers several opportunities for imposing specific constraints or regularization. It is important to select appropriate priors that capture the initial knowledge on  $\alpha$  before any observational or simulation data is available. For bounded embeddings this would typically entail priors with specific irregular support, while for unbounded embeddings, such as a zero-mean multivariate normal (MVN)  $\delta(\alpha, \xi)$  the constraints can be derived from the requisite constraints on the covariance structure. The latter is a reasonably mature area in Bayesian hierarchical modeling [46, 1, 17], with several options available, such as both Wishart [15] and inverse-Wishart priors [42], Cholesky-factor priors [63], reference priors [73], shrinkage priors [69], separation-based priors that separate standard deviation and correlation coefficients [40, 26, 4], and matrix logarithm [37]. The prior selection will not be made the focus of the current paper; unless otherwise specified, we will restrict ourselves to uniform priors, potentially with physics-based support constraints.

### 3.1.3 Posterior sampling and MCMC challenges

Sampling of the posterior distribution via MCMC is a challenging task due to the high-dimensionality of  $\tilde{\alpha}$  and the near-degeneracy discussed in Section 3.1.1. One approach of mitigating the high-dimensionality is to embed the model error representation in a few parameters at a time, instead of the full input  $\lambda$ . This enables attribution of model errors to specific parameters and provides grounds for attributing model errors to certain submodels. Another way of mitigating the high-dimensionality is to resort to simple posterior maximization, via standard optimization methods, instead of the full MCMC sampling. This is a key advantage of the embedded representation – even if one resorts to the best value  $\tilde{\alpha}_{\text{MAP}}$ , it still leads to prediction with uncertainties that capture model errors, albeit without a component relevant to the posterior distribution.

Nevertheless, sampling of the posterior distribution via MCMC or even simply maximizing the posterior distribution may be a difficult task in general, due to the structure of the posterior distribution itself. It may be multimodal as well as peaked along lower-dimensional manifolds, making MCMC sampling extremely challenging. Structural improvements to the posterior distribution may be achieved by appropriate regularization via prior selection.

## 3.2 Prediction with model error

The key advantage of the developed method is highlighted when one uses the calibrated model for prediction purposes. Explicit additive representation of the model discrepancy function  $\delta(x)$  in [43] makes the predictions challenging since it may violate physical constraints and is specific to the observable quantities used for the calibration. On the contrary, in our approach, model discrepancy is captured *within* the model, which makes definitions of predictive quantities relatively straightforward and meaningful. Each fixed value of the vector  $\tilde{\alpha} = (\lambda, \alpha)$  induces a random input vector  $\Lambda(\tilde{\alpha}) = \lambda + \delta(\alpha, \xi)$  which in turn leads to a stochastic process  $F(x; \tilde{\alpha})$  that is the result of pushing forward the stochastic germ  $\xi$  through the model  $f(x; \Lambda(\tilde{\alpha}))$ . The mean and covariance of this process are

denoted by  $\mu^f(x; \tilde{\alpha})$  and  $C^f(x, x'; \tilde{\alpha})$ , respectively, where

$$\mu^f(x; \tilde{\alpha}) = \mathbb{E}_{\xi}[F(x; \tilde{\alpha})] \quad (16)$$

$$C^f(x, x'; \tilde{\alpha}) = \mathbb{C}_{\xi}[F(x, x'; \tilde{\alpha})]. \quad (17)$$

The MAP value  $\tilde{\alpha}_{\text{MAP}}$  can be used to arrive at the best PDF for  $\mathbf{\Lambda}$  that encompasses the augmented model error representation. This input can subsequently be pushed forward through the model  $f(x; \mathbf{\Lambda})$  to obtain the *MAP pushed-forward* process  $f_{\text{MPF}}(x) = F(x; \tilde{\alpha}_{\text{MAP}})$ , with moments denoted as  $\mu_{\text{MPF}}(x)$  and  $C_{\text{MPF}}(x, x')$  which can be computed via sampling similar to (10) for general  $x$ , or via non-intrusive spectral projection with polynomial chaos described further in Section 3.2.2.

Furthermore, accounting for the full posterior distribution of  $\tilde{\alpha}$ , one can compute the moments of *posterior pushed-forward* process  $F(x; \tilde{\alpha})$  for any value of  $x$  as a combination of two stochastic dimensions (model-error  $\xi$  and posterior range sampled by MCMC samples  $\tilde{\alpha}$ ):

$$\mu_{\text{PF}}(x) = \mathbb{E}_{\xi} \mathbb{E}_{\tilde{\alpha}}[F(x; \tilde{\alpha})] = \mathbb{E}_{\tilde{\alpha}}[\mu^f(x; \tilde{\alpha})], \quad (18)$$

and, due to variance decomposition formula,

$$C_{\text{PF}}(x, x') = \mathbb{E}_{\tilde{\alpha}}[C^f(x, x'; \tilde{\alpha})] + \mathbb{C}_{\tilde{\alpha}}[\mu^f(x; \tilde{\alpha})], \quad (19)$$

where  $\mathbb{E}_{\tilde{\alpha}}$  and  $\mathbb{C}_{\tilde{\alpha}}$  denote the mean and covariance with respect to the posterior distribution  $p(\tilde{\alpha}|\mathbf{y})$ , respectively, and can be computed with MCMC samples and standard estimators. Pushed-forward variance for each fixed  $x$  can be computed simply by  $\sigma_{\text{PF}}^2(x) = C_{\text{PF}}(x, x)$ , and is decomposed as

$$\sigma_{\text{PF}}^2(x) = \underbrace{\mathbb{E}_{\tilde{\alpha}}[\sigma^f(x; \tilde{\alpha})^2]}_{\text{model error}} + \underbrace{\mathbb{V}_{\tilde{\alpha}}[\mu^f(x; \tilde{\alpha})]}_{\text{posterior uncertainty}} \quad (20)$$

which consists of contributions due to the embedded model error as well as due to posterior uncertainty stemming from the quality and amount of data. The key feature of the developed approach – and one to be illustrated in numerical examples below – is that, with increasing amount of data, the model error component  $\mathbb{E}_{\tilde{\alpha}}[\sigma^f(x; \tilde{\alpha})^2]$  does not decrease – rather it saturates to a fixed value. At the same time, the data error encoded into the posterior variance term  $\mathbb{V}_{\tilde{\alpha}}[\mu^f(x; \tilde{\alpha})]$  typically reduces with increasing amount of data.

While the pushed-forward process can be useful for prediction purposes for general values of  $x$ , in order to compare with data, it is also important to consider the *posterior predictive* distribution [27], which is by definition the combination of the pushed-forward process at the design conditions  $x_i$  and the data noise model  $\epsilon_i$ . The  $i$ -th component of the posterior predictive random vector is then  $F(x_i; \tilde{\alpha}) + \epsilon_i$  with mean and covariance

$$\mu_{\text{PP}}(x_i) = \mathbb{E}_{\tilde{\alpha}}[\mu^f(x_i; \tilde{\alpha})] = \mu_{\text{PF}}(x_i), \quad (21)$$

$$C_{\text{PP}}(x_i, x_j) = \mathbb{E}_{\tilde{\alpha}}[C^f(x_i, x_j; \tilde{\alpha}) + \sigma^2 \delta_{ij}] + \mathbb{C}_{\tilde{\alpha}}[\mu^f(x_i; \tilde{\alpha})] = C_{\text{PF}}(x_i, x_j) + \sigma^2 \delta_{ij}. \quad (22)$$

The posterior predictive covariance (22) should be contrasted with an imposed parameterized covariance – typically a square exponential – when modeling the model discrepancy term explicitly in an additive fashion [43]. The major conceptual difference is that the covariance obtained via model error embedding is directly informed by the model structure, rather than parameterized in a model-independent way.

Note that often  $\sigma$  can be viewed as a hyperparameter to be inferred together with  $\tilde{\alpha}$ , in which case the last term  $\sigma^2$  in Eq. (22) is replaced with its posterior average  $\mathbb{E}_{\tilde{\alpha}}[\sigma^2]$ . This is demonstrated in the chemical kinetics example in Section 5.1.

Both likelihood computation via any of the options from Table 1, and predictive/pushed-forward quantity computations, require sampling of model input  $\mathbf{\Lambda}(\tilde{\boldsymbol{\alpha}}) = \boldsymbol{\lambda} + \boldsymbol{\delta}(\boldsymbol{\alpha}, \boldsymbol{\xi})$  and subsequent evaluations of the model  $f(x, \mathbf{\Lambda}(\tilde{\boldsymbol{\alpha}}))$  with given  $\tilde{\boldsymbol{\alpha}}$ , or ability to compute the pushed-forward moments  $\mu^f(x; \tilde{\boldsymbol{\alpha}})$  and  $C^f(x, x'; \tilde{\boldsymbol{\alpha}})$ . Polynomial Chaos (PC) machinery provides a flexible mechanism of representing random variables allowing computationally inexpensive sampling, moment evaluation, as well as, generally, convenient tools for forward uncertainty propagation [28, 72, 44].

### 3.2.1 Polynomial chaos representation of augmented inputs

While the present framework remains valid with any parameterization of the stochastic model discrepancy term  $\boldsymbol{\delta}(\boldsymbol{\alpha}, \boldsymbol{\xi})$ , the focus of this paper is specifically on Polynomial Chaos (PC) representations, since they allow flexible parameterization of a wide class of random variables, with efficient sampling, moment evaluation and forward propagation machinery. In order to avoid crowding the notation, and for demonstrating proof-of-concept, assume that model error representation is embedded in *all* components of  $\boldsymbol{\delta}$ . In such case, the components of the model input are written as an expansion

$$\Lambda_j(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\xi}) = \lambda_j + \delta_j(\boldsymbol{\alpha}, \boldsymbol{\xi}) = \lambda_j + \sum_{k=1}^{K_j-1} \alpha_{kj} \Psi_k(\boldsymbol{\xi}), \text{ for } j = 1, \dots, d \quad (23)$$

with respect to standard orthogonal polynomials  $\Psi_k(\boldsymbol{\xi})$  of an independent-component standard random vector  $\boldsymbol{\xi}$ . The two most commonly used PC expansions are a) Gauss-Hermite, *i.e.* the *germ*  $\boldsymbol{\xi}$  is a vector of *i.i.d.* Gaussian random variables, and, correspondingly, the  $\Psi_k(\boldsymbol{\xi})$  are Hermite polynomials that are orthogonal with respect to the PDF of  $\boldsymbol{\xi}$ ,  $\pi_{\boldsymbol{\xi}}(\boldsymbol{\xi}) = \frac{e^{-\boldsymbol{\xi}^2/2}}{\sqrt{2\pi}}$ , and b) Legendre-Uniform, in which  $\boldsymbol{\xi}$  is a vector of *i.i.d.* standard uniform random variables on  $[-1, 1]$ , and the  $\Psi_k(\boldsymbol{\xi})$  are Legendre orthogonal polynomials. The latter is a more meaningful parameterization for model inputs that are bounded. The orthogonality of polynomial bases is written as

$$\int_{\boldsymbol{\xi}} \Psi_k(\boldsymbol{\xi}) \Psi_j(\boldsymbol{\xi}) \pi_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} = \begin{cases} \|\Psi_k\|^2, & \text{if } k = j \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

where the norm  $\|\Psi_k\| = \left( \int_{\boldsymbol{\xi}} \Psi_k^2(\boldsymbol{\xi}) \pi_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \right)^{1/2}$  is usually precomputed and stored a priori. Typically, with an understanding that each probabilistically cast  $\delta_j$  corresponds to a stochastic dimension, one chooses  $\boldsymbol{\xi}$  to have the same size as  $\boldsymbol{\delta}$ . The order of each expansion in Eq. (23) is selected a priori, denoted by  $p_j$ , leading to  $K_j = (d + p_j)! / (d! p_j!)$  terms. Note that formulating model input adjustments as PC expansions (23), the model-error estimation is essentially reformulated as a parameter estimation for deterministic quantities  $\alpha_{kj}$  that comprise the set  $\tilde{\boldsymbol{\alpha}} = \{\alpha_{kj}\}_{j=1, \dots, d}^{k=0, \dots, K_j-1}$ , in which, out of convenience, we denoted  $\lambda_j = \alpha_{0,j}$  for all  $j = 1, \dots, d$ .

Below we discuss two commonly used embeddings that allow efficient Bayesian calibration.

- *Multivariate normal input:* A multivariate normal assumption for  $\boldsymbol{\delta}(\boldsymbol{\alpha}, \boldsymbol{\xi})$  is a convenient way of parameterizing the internal model-error correction, particularly in cases when there is no restriction on the range of input parameters. It, as does the uniform embedding discussed below, also has the practical advantage of being first-order in  $\boldsymbol{\xi}$ , thus eliminating difficulties associated with inferring high-order PC coefficients (23). Indeed, the first-order PC expansion for  $\boldsymbol{\delta}(\boldsymbol{\alpha}, \boldsymbol{\xi})$ , corresponding to an MVN, which leads to

$$\Lambda_j(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\xi}) = \lambda_j + \delta_j(\boldsymbol{\alpha}, \boldsymbol{\xi}) = \lambda_j + \sum_{k=1}^d \alpha_{kj} \xi_k, \text{ for } j = 1, \dots, d \quad (25)$$

is a special case of Eq. (23) for  $p_j = 1$  for all  $j = 1, \dots, d$ . It highlights one key challenge of such PC representations – many coefficient sets  $\boldsymbol{\alpha} = \{\alpha_{kj}\}_{j=1, \dots, d}^{k=0, \dots, d}$  may correspond to the same joint PDF for  $\mathbf{\Lambda}$  and therefore will lead to the same likelihood value and, consequently, to a multimodal posterior distribution or a posterior distribution that is constant along certain low-dimensional manifolds. Barring judicious choice of priors for  $\boldsymbol{\alpha}$ , this poses practical difficulties for optimization or MCMC methods. Nevertheless, for the linear case (25), a simple re-parameterization resolves this ambiguity. Specifically, one can propose a ‘triangular’ linear PC form

$$\delta_j(\xi_1, \dots, \xi_j) = \sum_{k=1}^j \alpha_{kj} \xi_k, \text{ for } j = 1, \dots, d \quad (26)$$

with an additional constraint  $\alpha_{kd} > 0$  for  $k = 1, \dots, d$  enforced via Bayesian priors. In this case, there is a one-to-one correspondence between any MVN PDF of  $\mathbf{\Lambda}$  and the set of coefficients  $\tilde{\boldsymbol{\alpha}} = \{\alpha_{kj}\}_{j=0, \dots, d}^{k=0, \dots, j}$ , effectively ruling out a large set of potentially challenging, degenerate posteriors.

- *Uniform i.i.d. input:* In physical models, one often has bounds for input parameters that cannot be exceeded under any circumstances. In such cases, simple Gaussian additive adjustments to the model inputs may not be the best option. We employ Legendre-Uniform PC expansions for such cases that are bounded. However, in general this may require prior constraints on  $\tilde{\boldsymbol{\alpha}}$  that are technically challenging to implement. To simplify, and for the sake of demonstration, a linear Legendre-Uniform PC with *i.i.d.* embedding can be used,

$$\Lambda_j(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\xi}) = \lambda_j + \delta_j(\boldsymbol{\alpha}, \boldsymbol{\xi}) = \lambda_j + \alpha_{1j} \xi_j, \text{ for } j = 1, \dots, d \quad (27)$$

corresponding to model-error correction  $\delta_j(\xi_j) = \alpha_{1j} \xi_j$ , for  $\xi_j \in \mathcal{U}[-1, 1]$ . The *i.i.d.* embedding allows simple prior constraints

$$\begin{cases} \lambda_j + \alpha_{1j} \leq a_j \\ \lambda_j - \alpha_{1j} \geq b_j. \end{cases} \quad (28)$$

on  $\tilde{\boldsymbol{\alpha}} = (\boldsymbol{\lambda}, \boldsymbol{\alpha})$  to enforce parameters within given ranges  $\Lambda_j = \lambda_j + \alpha_{1j} \xi_j \in [a_j, b_j]$ . The prior (28), together with  $\alpha_{1j} \geq 0$  (added due to invariance, and to avoid bimodal posteriors) corresponds to a triangular region in the  $(\lambda_j, \alpha_{1j})$  space, a subset of  $\mathbb{R}^2$ .

### 3.2.2 Uncertainty propagation and predictive moment estimation

Representing inputs of the function  $f(x; \mathbf{\Lambda}(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\xi}))$  with PC expansions (23), or its simpler forms (25) and (27), allows efficient propagation of the uncertainties through  $f(x; \cdot)$  via non-intrusive spectral projection (NISP), as follows. One constructs a PC expansion for the output for each  $x$  as

$$f(x; \mathbf{\Lambda}(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\xi})) \simeq \sum_{k=0}^{K-1} f_k(x; \tilde{\boldsymbol{\alpha}}) \Psi_k(\boldsymbol{\xi}), \quad (29)$$

where the coefficients  $f_k(x; \tilde{\boldsymbol{\alpha}})$  are found via orthogonal projection that is numerically computed with quadrature integration, using a precomputed quadrature point-weight pairs  $(\boldsymbol{\xi}^{(q)}, w_q)$  for  $q = 1, \dots, Q$ ,

$$\begin{aligned} f_k(x; \tilde{\boldsymbol{\alpha}}) &= \frac{1}{\|\Psi_k\|^2} \int_{\boldsymbol{\xi}} f(x; \mathbf{\Lambda}(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\xi})) \Psi_k(\boldsymbol{\xi}) \pi_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \approx \\ &\approx \frac{1}{\|\Psi_k\|^2} \sum_{q=1}^Q w_q f(x; \mathbf{\Lambda}(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\xi}^{(q)})) \Psi_k(\boldsymbol{\xi}^{(q)}). \end{aligned} \quad (30)$$

The PC coefficients depend on  $\tilde{\alpha}$  since the relationship  $\Lambda(\tilde{\alpha}, \xi)$  encapsulated in the input PC expansion (23) is parameterized by  $\tilde{\alpha}$ . It is also important to recognize that this PC propagation approach essentially relies on *sampling*, too, since the model evaluations are driven by the underlying quadrature sampling of  $\xi$ . Such sampling is, however, much more efficient and accurate for a large class of models, compared to Monte-Carlo based uncertainty propagation approaches [21, 57].

Note that the classical calibration strategy that ignores the model error and assumes  $f(x; \lambda)$  replicates the ‘truth’  $g(x)$ , is merely a special case of the developed model-error embedding framework for 0-th order PC expansion in (23), *i.e.*  $\Lambda_j(\tilde{\alpha}, \xi) = \lambda_j$ , with no explicit model-error dimension ( $\xi$ ) dependence, and  $\tilde{\alpha} \equiv \lambda$ .

Next, we return to the predictive process  $F(x; \tilde{\alpha}) = f(x; \Lambda(\tilde{\alpha}, \xi))$  described in Section 3.2 in terms of PC representations. Both generic predictions for arbitrary  $x$  and likelihood computations for  $x_i$ ’s require either KDE estimation or moment computation. This entails extensive sampling of the function outputs and may become prohibitively expensive. For sampling, the PC representation (29) can serve as a ‘surrogate’ approximation of the function and be sampled instead of it. However, the crucial advantage of the PC representation is highlighted in moment-based likelihood cases, Eqs. (11), (13) and (14). In these cases, only moments are required to compute the likelihoods, and the PC expansion (29) offers analytical, closed forms for them, circumventing sampling.

The predictive stochastic process  $F(x; \tilde{\alpha})$  is now written as a PC expansion

$$F(x; \tilde{\alpha}) \simeq \sum_{k=0}^{K-1} f_k(x; \tilde{\alpha}) \Psi_k(\xi) \quad (31)$$

from which moments can be easily extracted as

$$\mu^f(x; \tilde{\alpha}) \simeq f_0(x; \tilde{\alpha}) \quad \text{and} \quad C^f(x, x'; \tilde{\alpha}) \simeq \sum_{k=1}^{K-1} f_k(x; \tilde{\alpha}) f_k(x'; \tilde{\alpha}) \|\Psi_k\|^2. \quad (32)$$

The overall pushed-forward moments are written as

$$\mu_{\text{PF}}(x) = \mathbb{E}_{\tilde{\alpha}}[\mu^f(x; \tilde{\alpha})] \simeq \mathbb{E}_{\tilde{\alpha}}[f_0(x; \tilde{\alpha})] \quad (33)$$

$$\begin{aligned} C_{\text{PF}}(x, x') &= \mathbb{E}_{\tilde{\alpha}}[C^f(x, x'; \tilde{\alpha})] + \mathbb{C}_{\tilde{\alpha}}[\mu^f(x; \tilde{\alpha})] \simeq \\ &\simeq \underbrace{\sum_{k=1}^{K-1} \mathbb{E}_{\tilde{\alpha}}[f_k(x; \tilde{\alpha}) f_k(x'; \tilde{\alpha})] \|\Psi_k\|^2}_{\text{Model error}} + \underbrace{\mathbb{C}_{\tilde{\alpha}}[f_0(x; \tilde{\alpha})]}_{\text{Posterior uncertainty}}. \end{aligned} \quad (34)$$

Table 2 summarizes the predictive process  $F(x; \tilde{\alpha})$ , both for fixed  $\tilde{\alpha}$  and posterior average with respect to  $\tilde{\alpha}$ . It also lists its moments, and provides formulae specific for the PC representation.

Correspondingly, the data model (2) now takes the form, as a combination of pushed-forward process evaluated at  $x_i$ ’s and measurement errors,

$$y_i \approx \sum_{k=0}^{K-1} f_k(x_i; \tilde{\alpha}) \Psi_k(\xi) + \sigma \xi_{d+i}, \quad \text{for } i = 1, \dots, N, \quad (35)$$

from which one can extract moments as in (32). The right-hand-side of Eq. (35) includes all components of predictive uncertainty (model error  $\xi$  and data error  $\xi_{i+d}$ ) in a PC form for given  $\tilde{\alpha}$ . One can further build a PC expansion given posterior samples of  $\tilde{\alpha}$  arriving at an overall PC representation that also includes posterior uncertainty. Details of this procedure are given in Appendix A. Further in this text, we will keep  $\tilde{\alpha}$  as the uncertain dimension due to the posterior PDF, and the associated variance contributions will be accounted for via estimators that are based on MCMC samples.

-	Process	Mean	Covariance
Fixed $\tilde{\alpha}$	$F(x; \tilde{\alpha})$	$\mu^f(x; \tilde{\alpha})$	$C^f(x, x'; \tilde{\alpha})$
PC appr.	$\sum_{k=0}^{K-1} f_k(x; \tilde{\alpha}) \Psi_k(\xi)$	$f_0(x; \tilde{\alpha})$	$\sum_{k=1}^{K-1} f_k(x; \tilde{\alpha}) f_k(x'; \tilde{\alpha}) \ \Psi_k\ ^2$
PF, $\mathbb{E}_{\tilde{\alpha}}$	$\mathbb{E}_{\tilde{\alpha}}[F(x; \tilde{\alpha})]$	$\mathbb{E}_{\tilde{\alpha}}[\mu^f(x; \tilde{\alpha})]$	$\mathbb{E}_{\tilde{\alpha}}[C^f(x, x'; \tilde{\alpha})] + \mathbb{C}_{\tilde{\alpha}}[\mu^f(x; \tilde{\alpha})]$

Table 2: A summary of the predictive process, its moments, PC form and posterior average.

When the predictive process  $F(x; \tilde{\alpha})$  is evaluated at specific locations  $x_i$  at which the data is taken, it allows direct comparison with the data points  $y_i$ , thus facilitating likelihood computations. The corresponding moments can be written as

$$\mu_i(\tilde{\alpha}) = \mathbb{E}_{\xi} f(x_i; \mathbf{\Lambda}(\tilde{\alpha}, \xi)) \simeq f_0(x_i; \tilde{\alpha}), \quad (36)$$

$$\begin{aligned} C_{ij}(\tilde{\alpha}) &= \mathbb{E}_{\xi} [(f(x_i; \mathbf{\Lambda}(\tilde{\alpha}, \xi)) - \mu_i(\tilde{\alpha})) (f(x_j; \mathbf{\Lambda}(\tilde{\alpha}, \xi)) - \mu_j(\tilde{\alpha}))] \simeq \\ &\simeq \sum_{k=1}^{K-1} f_k(x_i; \tilde{\alpha}) f_k(x_j; \tilde{\alpha}) \|\Psi_k\|^2 + \delta_{ij} \sigma^2. \end{aligned} \quad (37)$$

$$\sigma_i^2(\tilde{\alpha}) = \mathbb{E}_{\xi} [f(x_i; \mathbf{\Lambda}(\tilde{\alpha}, \xi)) - \mu_i(\tilde{\alpha})]^2 \simeq \sum_{k=1}^{K-1} f_k^2(x_i; \tilde{\alpha}) \|\Psi_k\|^2 + \sigma^2, \quad (38)$$

Note that in the equations above, as well as before in this text, the symbol  $\simeq$  highlights the fact that the result is based on a PC expansion which approximates equality and extra care is needed, *e.g.*, in terms of appropriate truncation of the expansion. Quadrature point sampling and consequent function evaluations are still present during the construction of the PC expansion, but for reasonably smooth forward functions they are much more efficient than Monte-Carlo based approaches for moment computation. Besides, as shown in the next section, one can pre-construct and employ a polynomial surrogate *before* the inference, allowing efficient MCMC without expensive model evaluations for each  $\tilde{\alpha}$ , *i.e.* at each MCMC step.

### 3.2.3 Surrogate construction and overall algorithm

For most physical application problems, the model  $f(x; \mathbf{\Lambda})$  is quite expensive, or difficult to efficiently evaluate during MCMC. Here we describe a procedure to pre-construct a surrogate  $f_s(x; \mathbf{\Lambda}) \approx f(x; \mathbf{\Lambda})$ , to alleviate the cost. The surrogate  $f_s(x; \mathbf{\Lambda})$  can be constructed as an approximation of the function  $f(x; \mathbf{\Lambda})$  over the joint space  $(x, \mathbf{\Lambda})$ , or over  $\mathbf{\Lambda}$  only, for a select set of  $x_i$  that have been utilized during calibration and prediction. Here, we will describe and employ the latter approach for simplicity. For each  $x_i$ , consider a surrogate form

$$f(x_i; \mathbf{\Lambda}) \approx f_s(x_i; \mathbf{\Lambda}) = \sum_{k=0}^{K-1} s_{ik} L_k(\mathbf{\Lambda}), \quad (39)$$

where  $L_k(\mathbf{\Lambda})$ 's are scaled multivariate Legendre polynomials, with inputs scaled to their respective ranges  $\lambda_j \in [a_j, b_j]$  for  $j = 1, \dots, d$ . Similar to the general PC construction described earlier, the truncation, *i.e.* the number of polynomial bases  $K$ , is driven by a predefined rule, *e.g.* according to a total degree of retained polynomials.

The polynomial fit (39) is constructed via least-squares regression, using an arbitrary set of  $R$  model evaluations  $\{f(x_i, \mathbf{\lambda}_r)\}_{r=1}^R$ . In other words, the coefficients  $c_{ik}$  are the solution of the minimization problem for each  $i$

$$\arg \min_{c_{ik}} \sum_{r=1}^R \left( f(x_i, \mathbf{\lambda}_r) - \sum_{k=0}^{K-1} c_{ik} L_k(\mathbf{\lambda}_r) \right)^2. \quad (40)$$



Note that the simple least-squares solution is analytically available and does not require an optimization engine. Namely, the solution vector  $\mathbf{c}_i = (c_{i0}, \dots, c_{iK})^T$  of size  $K \times 1$  is written as

$$\mathbf{c}_i = (P^T P)^{-1} P^T \mathbf{f}_i, \quad (41)$$

where  $\mathbf{f}_i = (f(x_i, \boldsymbol{\lambda}_1), \dots, f(x_i, \boldsymbol{\lambda}_R))^T$  is the model evaluations' vector of size  $R \times 1$ , and  $P$  is the basis evaluation matrix of size  $R \times K$  with entries  $P_{rk} = L_k(\boldsymbol{\lambda}_r)$ . Note that the least-squares solution (41) is employed instead of NISP-like orthogonal projection, since it enables an surrogate error metric, leave-one-out error which can in principle be added to the predictive uncertainty budget. Besides, the regression-based surrogate is more general and does not require basis orthogonality.

Now, when we construct the PC propagation (29) via NISP, the surrogate  $f_s(x; \boldsymbol{\lambda})$  replaces the actual forward model  $f(x; \boldsymbol{\lambda})$  in the quadrature integration (30). Moreover, if the surrogate is built as a polynomial expansion (39) with the same order as the NISP in Eq. (29), then no extra approximation error is induced due to NISP. The only additional error is due to the surrogate approximation itself which can be estimated, *e.g.* via leave-one-out cross-validation as demonstrated in [38]. In this work, for the numerical demonstrations in Section 4, no surrogate is employed as the functions are synthetic and easy-to-evaluate, while for the chemical kinetics application in Section 5 the surrogate errors are negligible and are skipped for the clarity of presentation. The overall mechanism of surrogate-enhanced model error inference and prediction is listed in Algorithm 1.

## 4 Numerical demonstrations

In this section, we illustrate the strengths of the embedded model-error strategy with a few numerical examples. Note that we skip the surrogate construction step as listed in Algorithm 1, since the synthetic model evaluations are not costly.

**Demo 1 (consistent predictive errorbars)** Let us revisit the motivational example from Section 2. That is, we calibrate the model  $f(x; \boldsymbol{\lambda}) = \lambda_2 e^{\lambda_1 x} - 2$ , with noisy data measured from an underlying 'truth' model  $g(x) = \tanh 3(x - 0.3)$ . As Figure 1 highlighted, the conventional calibration approach results in posterior PF uncertainty which is not representative of the correct discrepancy between the calibrated model and the underlying 'truth' model. Now, we employ the embedded model error calibration as follows. Consider a 'triangular' MVN additive term for the parameter  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$  as in Eq. (26), and employ the ABC likelihood (14) with  $\gamma = 1$  and  $\epsilon = 0.0001$ . The choice of  $\gamma$  specifies the requirement that, on average, the  $1\sigma$  range of the predictive uncertainty is equal to  $\gamma \times$  the discrepancy between the data and the mean prediction. The choice of  $\epsilon \ll 1$  specifies the stringency with which the ABC kernel density enforces the desired constraints on the predictive mean and standard deviation. Figure 3(a) clearly demonstrates that the resulting pushed-forward prediction errorbars are representative of the true discrepancy, unlike the classical case, shown in the  $N = 50$  subplot of Figure 1. Moreover, Figures 3(b) and (c) demonstrate posterior predictive PDFs of  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$  for embedded (in which case  $\boldsymbol{\lambda}$  is effectively cast as a random variable  $\boldsymbol{\Lambda}$  and plotted is PDF induced by posterior *and* variability of  $\boldsymbol{\Lambda}$  itself) and classical cases, respectively, as the number of data points increases. It can be seen that model-error embedding guards against posterior shrinkage.

**Demo 2 (targeted model improvement)** For another illustration, consider a 'truth' model  $g(x) = e^{-0.5x} + e^{-2x}$  mimicking a physical system that involves two decay scales. Assuming no measurement noise, the data at 10 equidistant locations in  $x \in [0, 5]$  come from this model exactly. Now, consider a single-decay model  $f(x; \boldsymbol{\lambda}) = e^{-(\lambda_1 + \lambda_2 x)}$ , as if the analyst is unaware of the second decay source. A classical calibration for  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$  leads to wrong values for the decay rate, and fails to capture the double-exponential dynamics as shown in the left plot of Figure 4 – clearly the uncertain prediction

---

**Algorithm 1:** The full workflow of the surrogate-enhanced model error inference algorithm.

---

**Surrogate construction:**

- Note: this step is skipped for the demo problems, and only applied to the chemistry problem in Section 5.
- Choose  $R$  input samples  $\lambda_r$ , for  $r = 1, \dots, R$ :
  - we chose uniform random set of points distributed in given parameter ranges.  
 $\lambda_{ir} \in [a_i, b_i]$ .
- Evaluate the forward model  $R$  times and extract  $f(x_i; \lambda_r)$  at each design condition  $x_i$ , for  $i = 1, \dots, N$  and  $r = 1, \dots, R$ .
- Build polynomial surrogate (39):
  - we employed least-squares solution (41).

**Model error inference:**

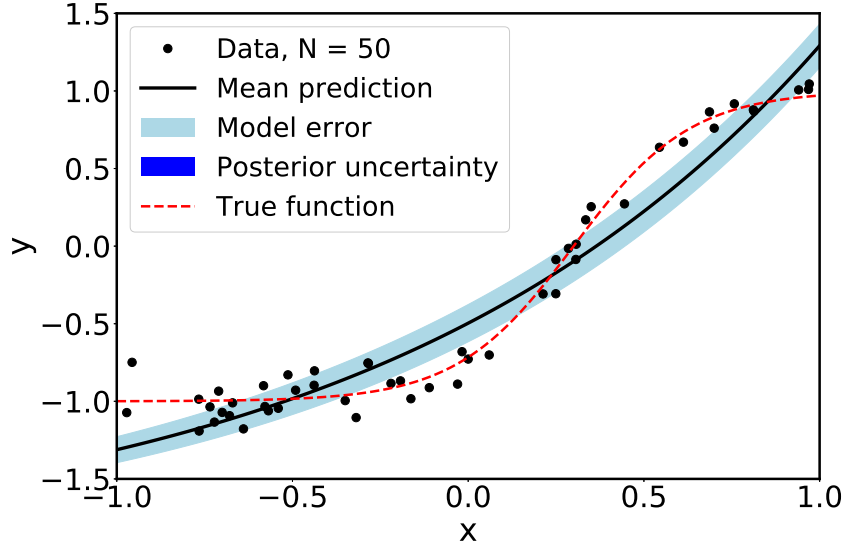
- Select embedding type (see Section 3.2.1):
  - we chose MVN input (25) for demo problems with no parameter range constraints, and the uniform i.i.d. inputs (27) otherwise.
- Select a prior (see Sections 3.1.2 and 3.2.1).
- Select a feasible likelihood form from Table 1:
  - for most studies here, we chose independent-normal, or ABC.
- Run MCMC inference:
  - we used adaptive MCMC (AMCMC) algorithm, introduced in [31].
  - at each MCMC step, in order to compute the likelihood, PC propagation (35) and moment estimation (36)-(38) is completed via integration by quadrature (30).
  - for the chemistry problem in Section 5, the expensive model evaluations required by (30) are replaced by the surrogate evaluations (39).

**Prediction:**

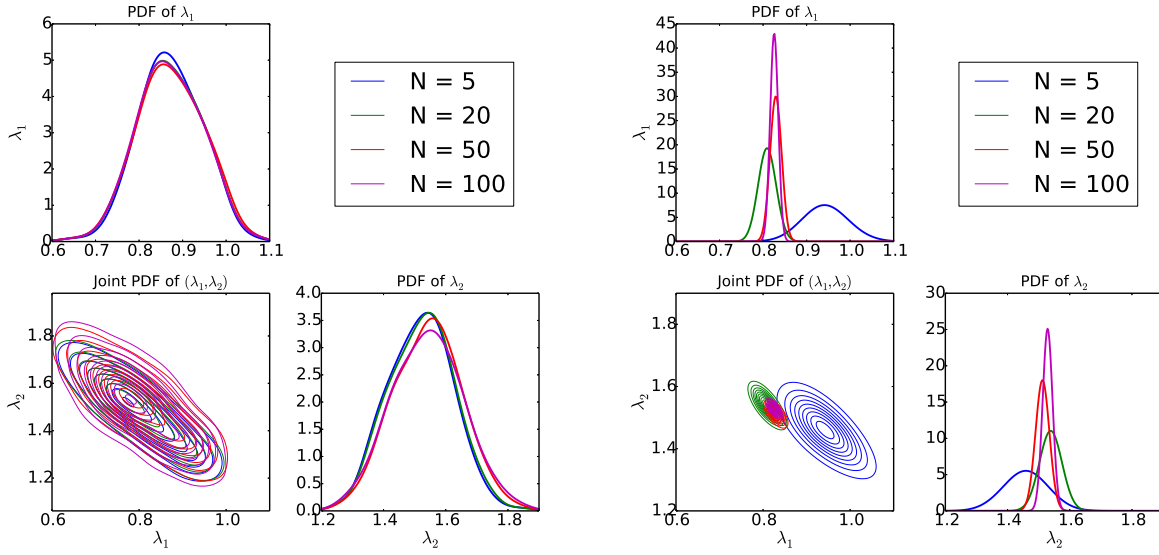
- Note: the predictions can be made for general design conditions  $x$ , even if data is collected at specific values  $x_i$ , for  $i = 1, \dots, N$ .
  - Predictive moments of  $f(x; \lambda)$  are computed as given in Eqs. (33) and (34), and are summarized in Table 2.
- 

is not consistent with the range of discrepancy from the truth. On the other hand, augmenting with an MVN discrepancy term to  $\lambda$  and calibrating for the best values of the MVN parameters  $\alpha$ , via the ABC likelihood (14), leads to an uncertain model prediction that captures the discrepancy from the data better (middle plot). One can envision a next step, *e.g.* embed more complicated ‘physics’, albeit still not the true double-exponential. For example, one can keep the single-exponential description and propose a quadratic-exponential model,  $f_2(x; \lambda) = e^{-(\lambda_1 + \lambda_2 x + \lambda_3 x^2)}$  with an additional parameter  $\lambda_3$ . As the right plot suggests, this allows better predictions with smaller model errors. The present approach allows such embedding of model error terms in targeted model components, and the evaluation of the resulting uncertain predictions for model comparison/selection studies.

**Demo 3 (convergence with increasing  $N$ )** Figure 5 demonstrates in more detail how the data error reduces with increasing  $N$  while the model error term does not. This illustration uses a fractional-power polynomial ‘truth’ function  $g(x) = 6 + x^2 - (x + 1)^{3.5}$  defined on  $x \in [-1, 1]$  and observations that are corrupted by a Gaussian, *i.i.d.* noise term of standard deviation  $\sigma = 0.5$ . Four simple



(a) PF Prediction with ABC



(b) ABC Posterior

(c) Classical Posterior

Figure 3: Illustration of the pushed-forward predictions with ABC likelihood, as well as comparison of the posterior PDFs as  $N$  grows for classical case and the case with model error augmentation with ABC likelihood.

models are calibrated with respect to their parameter vector  $\boldsymbol{\lambda}$ , (a) linear,  $f(x; \boldsymbol{\lambda}) = \lambda_0 + \lambda_1 x$ , (b) quadratic,  $f(x; \boldsymbol{\lambda}) = \lambda_0 + \lambda_1 x + \lambda_2 x^2$ , (c) cubic,  $f(x; \boldsymbol{\lambda}) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3$ , and (d) true model,  $f(x; \boldsymbol{\lambda}) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 (x + 1)^{3.5}$ . Clearly, while the data error component (as given in the decomposition (20)) reduces with increasing amount of data, the model error ‘saturates’ at a limiting value driven by the quality of the model, for all the approximate models. Figure 6 illustrates this effect further for the quadratic model case, with four different values of  $N$ , demonstrating the fact that the model error component of the variance remains relatively constant with increasing  $N$ .

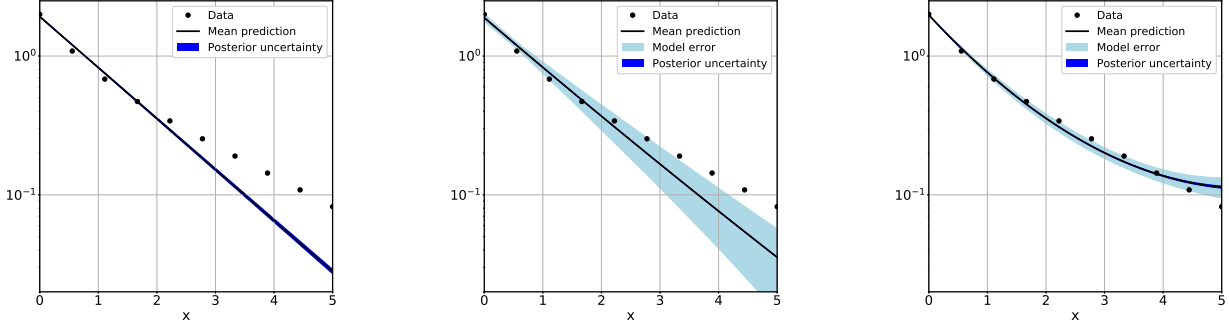


Figure 4: Calibration of a single-exponent model  $f(x; \lambda) = e^{-(\lambda_1 + \lambda_2 x)}$  with additive Gaussian noise (left) fails to predict the truth model  $g(x) = e^{-0.5x} + e^{-2x}$ , while the calibration with embedded model error leads to uncertain predictions that capture the truth model (middle). A quadratic-exponential model  $f_2(x; \lambda) = e^{-(\lambda_1 + \lambda_2 x + \lambda_3 x^2)}$  captures the truth better with smaller uncertainties (right). The uncertainty bands correspond to one PF standard deviation from the PF means, as defined in Eqs. (33) and (34).

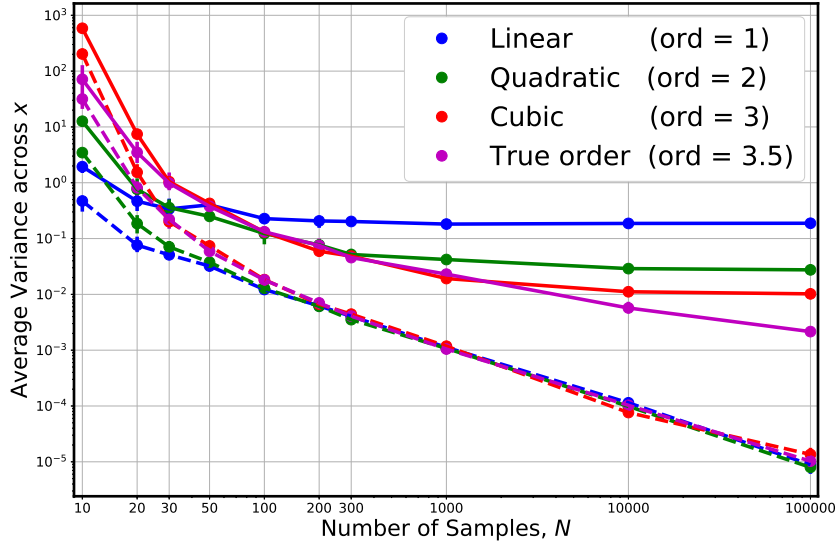


Figure 5: Demonstration of the decay of predictive variance components as the number of data points increases. The solid lines correspond to model error, while the dashed lines correspond to data error, according to the pushed-forward variance decomposition in (20). Besides the true fractional-power model, three simple models with varying degree of accuracy are used, clearly demonstrating the saturation of the model error component to a limiting value, while the data error due to  $N$  keeps reducing. The points in the plot indicate averages across all values of  $x_i$ , and, in order to obtain smoother convergence curve, 100 replica simulations have been performed and only median results with 25%-75% quantiles are reported.

## 5 Ignition model

This section illustrates the method on an example problem taken from chemical kinetics. Consider the chemical model with the classical two-step mechanism described by [71], which accounts for the

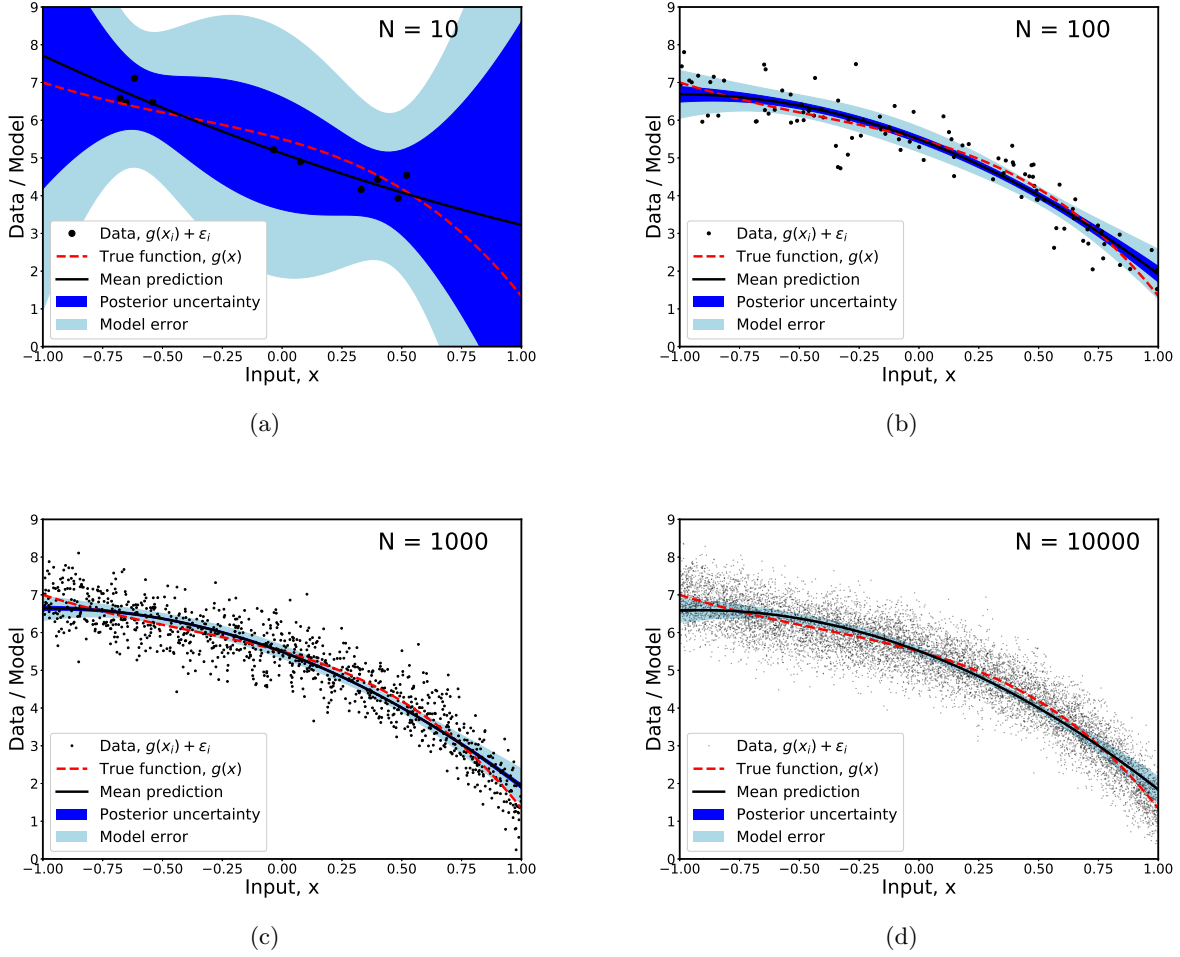
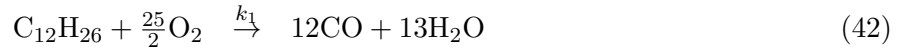


Figure 6: Demonstration of the pushed-forward predictions given noisy observations of a ‘truth’ function  $g(x) = 6 + x^2 - (x + 1)^{3.5}$  being calibrated by a quadratic model  $f(x; \lambda) = \lambda_0 + \lambda_1 x + \lambda_2 x^2$  via MVN model error embedding in all parameters  $\lambda$ . The four frames (a, b, c, d) correspond to the four  $N$  values ( $10, 10^2, 10^3, 10^4$ ), as indicated, respectively. The model error component (light blue region) of the pushed-forward variance decomposition in (20) remains relatively unchanged after enough data is collected, while posterior uncertainty keeps reducing. The average fit variance corresponds to the convergence curve for the quadratic case in Figure 5.

incomplete oxidation of n-dodecane



The reaction rates are modeled using Arrhenius laws, with the parameters taken from [71, 20]

$$k_1 = Ae^{(-\frac{E}{RT})}[\text{C}_{12}\text{H}_{26}]^{0.25}[\text{O}_2]^{1.25} \quad (44)$$

$$k_{2f} = 3.98 \times 10^{14} e^{(-\frac{40000}{RT})}[\text{CO}][\text{H}_2\text{O}]^{0.5}[\text{O}_2]^{0.25} \quad (45)$$

$$k_{2b} = 5 \times 10^8 e^{(-\frac{40000}{RT})}[\text{CO}_2]. \quad (46)$$

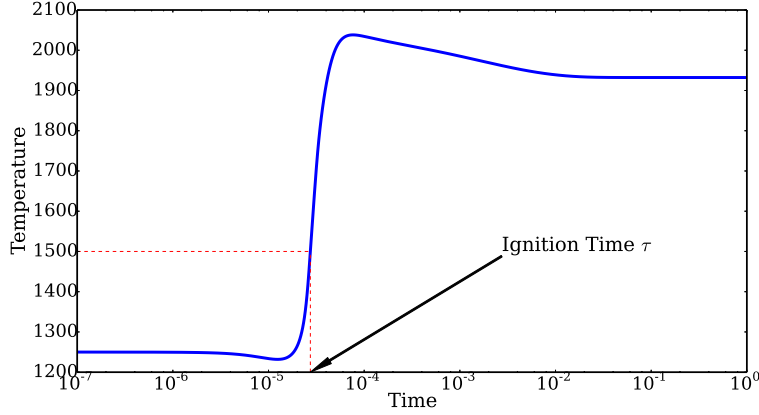


Figure 7: An example of temperature profile and ignition time definition.

where  $R$  is the ideal gas constant,  $T$  is the temperature in Kelvin, and  $[X]$  the molar concentration of species  $X$  in units of  $\text{mol}/\text{cm}^3$ . The units of time and energy are sec and cal, respectively. Rate  $k_1$  is the rate of progress of fuel oxidation. The pre-exponential factor  $A$  is parameterized as

$$\ln A = a_1 + a_2 e^{a_3 \phi} + a_4 \tanh((a_5 + a_6 \phi)T_0 + a_7), \quad (47)$$

in which it is modeled as a function of initial temperature,  $T_0$ , and equivalence ratio,  $\phi$ . Such a special form allows accounting for the desired behavior of the autoignition delay time in  $(T_0, \phi)$  space; e.g., in the Negative Temperature Coefficient (NTC) region [45]. Varying the Arrhenius parameters with mixture state is a well known technique that has been applied elsewhere with good success [23, 24, 48]. Together with the activation energy  $E$ , the parameter  $a_1$  is of interest, *i.e.*  $\lambda = (E, a_1)$ , while for the purposes of the demonstrations in this work, we fix the rest of the parameters at their nominal values  $(a_2, a_3, a_4, a_5, a_6, a_7) = (-2.13, -2.05, 1.89, -0.01, 2.87 \cdot 10^{-4}, 8.43)$ , found in [32]. The operating conditions for a given ignition simulation are the pressure (in atm), the equivalence ratio, and the scaled inverse of the initial temperature. In terms of the notation in this work, the operating conditions are  $x = (P, \phi, 1000/T_0)$ , and the model output of interest  $f(x; \lambda)$  is the logarithm of the ignition time,  $\ln(\tau)$ . The ignition time is defined as the time when the temperature reaches 1500 K, as demonstrated in Figure 7. The simulations of the 2-step mechanism are performed using TChem [55], a software toolkit for the analysis of complex kinetic models.

We demonstrate calibration of this 2-step model in two case studies: (a) experimental shock-tube data from [68], and (b) simulation data generated from the skeletal mechanism by [50] with 255 species and 2289 reactions.

## 5.1 Calibration with experimental data

Here we employ experimental shock-tube ignition data from [68] collected in a heated, high-pressure shock tube for two fixed values of equivalence ratio ( $\phi = 0.5$  and  $\phi = 1.0$ ), over a range of variation of the two other operating conditions, pressure ( $P$ ) and initial temperature ( $T_0$ ). As such, the logarithm of ignition time,  $\ln(\tau)$ , collected under varying conditions  $x = (P, \phi, 1000/T_0)$ , is the data generation model  $g(x)$ . Overall, there are 42 data samples, or quartets,  $\{(P, \phi, 1000/T_0)_i, \ln(\tau_i)\}_{i=1}^{42}$ , 16 of which correspond to  $\phi = 0.5$ , while the other 26 are collected under equivalence ratio  $\phi = 1.0$ . We note that the two other operating conditions,  $P$  and  $T_0$ , are measured rather than strictly enforced, and the authors in [68] somewhat heuristically evaluate the overall noise in ignition time measurement as  $\pm 10\%$ . Besides, the definition of ignition time in the experiments differs somewhat from the definition employed in the model-to-be-calibrated. However, the difference is of little consequence. Given the



sharp temperature increase and overall shape of the temperature evolution as in Figure 7, the minor discrepancy in the definition of ignition time can essentially be regarded as data noise. Overall, given  $\ln(1.1\tau) \approx \ln(\tau) + 0.1$  and  $\ln(0.9\tau) \approx \ln(\tau) - 0.1$ , we note that the nominal standard deviation is  $\sigma = 0.1$ . For the inference, we assume an *i.i.d.* additive data noise model on  $\ln(\tau)$ , distributed as a normal random variable with vanishing mean and constant standard deviation to be inferred.

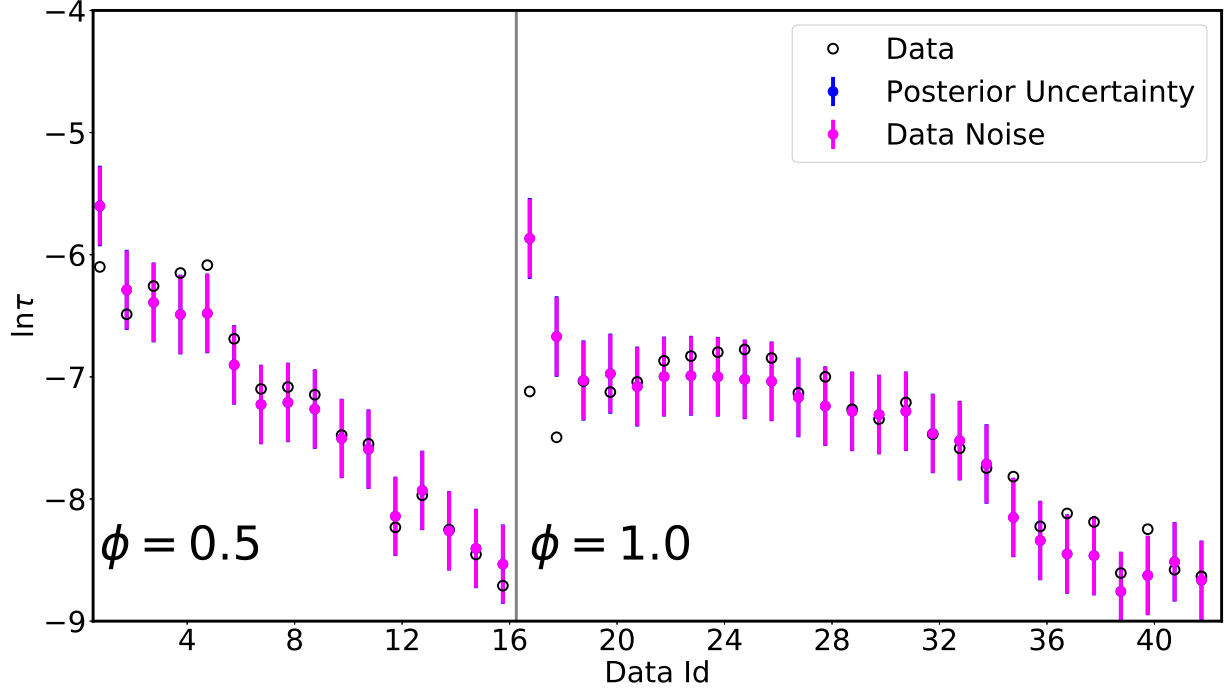
Following Algorithm 1, we build a polynomial surrogate for the output  $\ln(\tau)$  of the two-step mechanism (42)-(43) at all operating conditions  $x_i$ , for  $i = 1, \dots, N$  and with  $N = 42$ . The polynomial surrogate is chosen to be 3-rd order, with the coefficients found by regression described in Section 3.2.3, using  $R = 100$  input parameters  $\lambda = (E, a_1)$ , sampled uniformly over ranges  $E \in [10000, 40000]$  and  $a_1 \in [15, 37]$ . The polynomial regression allows analytical extraction of leave-one-out (LOO) error as a measure of its accuracy [14], and can subsequently augment predictive uncertainty budget, as demonstrated in [39]. For the examples shown, the LOO errors are on the order of  $10^{-2}$ , while the observable  $\ln \tau$  varies in the range between  $-11$  and  $-6$ , therefore we ignore surrogate error in the present discussion. Moreover, as we employ a third-order polynomial surrogate model, the PC-based NISP propagation with third-order truncation provides an exact extraction of mean and variance of the PF predictions. The data noise is assumed constant, independent across operating conditions  $i = 1, \dots, 42$ , and is inferred together with the model parameters. The goal of the demonstration is to highlight the ability of the model error approach to differentiate model error from data noise appropriately. In other words, we demonstrate that including embedded model error term allows estimating data noise standard deviation much more accurately, while the classical inference without accounting for model error overestimates the data standard deviation, compensating for unaccounted model error. Figure 8 demonstrates the posterior predictive uncertainty that can be derived similar to Eq. (22) as

$$\sigma_{\text{PP}}^2(x_i) = \underbrace{\mathbb{E}_{\tilde{\alpha}}[\sigma^f(x_i; \tilde{\alpha})^2]}_{\text{model error}} + \underbrace{\mathbb{V}_{\tilde{\alpha}}[\mu^f(x_i; \tilde{\alpha})]}_{\text{posterior uncertainty}} + \underbrace{\mathbb{E}_{\sigma}[\sigma^2]}_{\text{data noise}}. \quad (48)$$

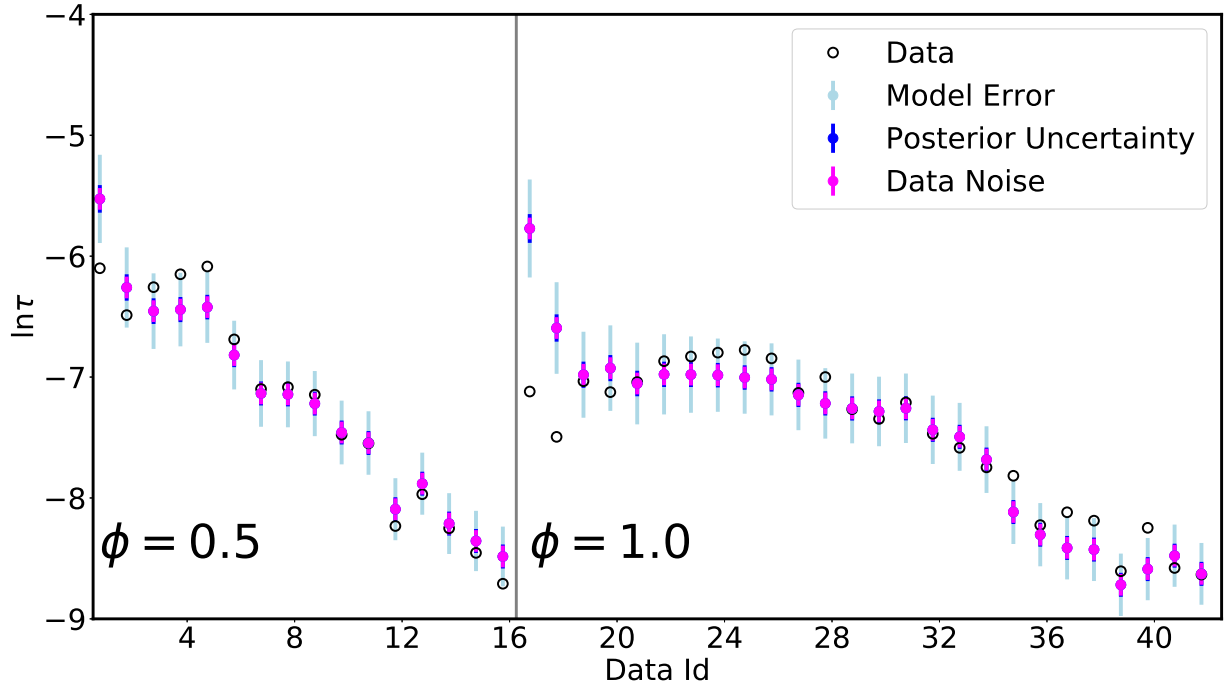
This equation is similar to the pushed forward uncertainty (20), with the additional data noise term. The classical calibration, as Figure 8(a) shows, captures the experimental data well *on average*, with posterior uncertainties, due to low amount of the data, being consistent with the data. However, it includes no model error by construction, and attributes most of the posterior predictive uncertainty to data noise, thus overestimating the latter. When including model error, as shown in Figure 8(b), with MVN embedding (11) in the two main parameters  $\lambda = (E, a_1)$  of the 2-step model, and independent-normal likelihood (13), one achieves similar posterior predictive variance, but with an estimate of data noise size that is closer to the nominal value of  $\sigma = 0.1$ . This more accurate estimation of  $\sigma$  is highlighted in Figure 9, in which the MCMC samples from both cases are shown, together with their respective PDFs and the nominal value of  $\sigma$ .

## 5.2 Calibration with complex-model data

Consider a more complex mechanism (albeit still a “skeletal” mechanism for this fuel) with 255 species and 2289 reactions defined in [50]. We simulate ignition with this mechanism, using TChem [55], at arbitrary operating conditions. We select a uniform  $11 \times 11 \times 11$  grid in the  $x = (P, \phi, 1000/T_0)$  space varying these conditions in the ranges  $x \in [11, 61]$  atm,  $\phi \in [0.5, 2.5]$ ,  $1000/T_0 \in [0.8, 1.3]$  K $^{-1}$ . The logarithm of the ignition time for this mechanism is the data generation model  $g(x)$  in the notation of this work. The ignition time is defined as the time when temperature reaches 1500 K. As in the previous case, the model-to-be-calibrated is the two-step mechanism from (42)-(43), for which we prebuild 3-rd order polynomial surrogates for all  $N = 1331$  operating conditions, using  $R = 100$  random parameter samples of  $\lambda = (E, a_1)$ , sampled from ranges  $E \in [10000, 40000]$  and  $a_1 \in [15, 37]$ . Once again, the PC-based NISP propagation provides exact values for the PF moments. Unlike the



(a) No model error



(b) Model error

Figure 8: Demonstration of uncertainty decomposition of the predictions, according to Eq. (48). The classical Bayesian calibration, with inferred constant data noise  $\sigma$  is shown in the first plot, while the results with the embedded model error approach are shown in the second plot. Note that in both cases, the posterior uncertainty contribution is relatively small.

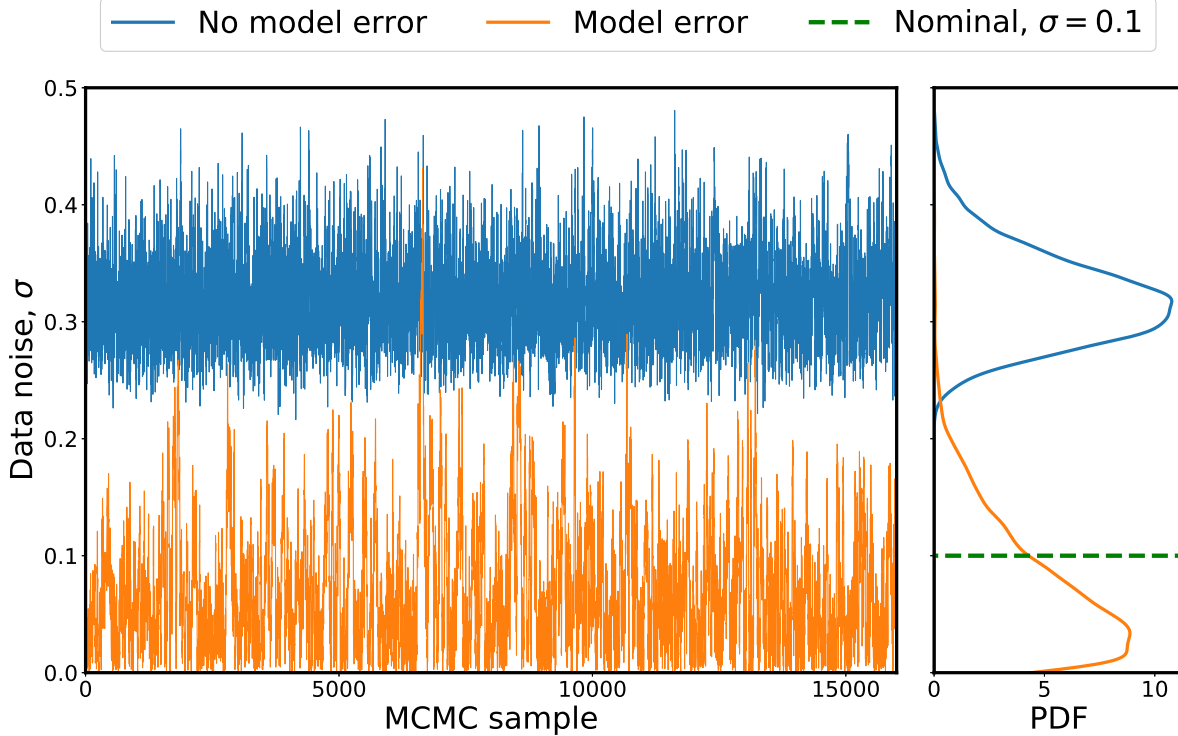


Figure 9: Demonstration of posterior samples of data noise standard deviation with or without including model error.

experimental data from the previous case, the complex-model simulation produces data with no noise, and the discrepancy between the two-step mechanism and the complex model is purely due to model error. As such, the main assumption underlying the classical calibration, i.e. the *i.i.d.* nature of data noise is unfounded. Still, we proceed with the classical calibration for demonstration, assuming constant data noise  $\sigma$  that is inferred together with the model parameters  $\lambda = (E, a_1)$ . What we have found is that the posterior variance  $\mathbb{V}_\lambda[f(x; \lambda)]$ , being the only contributor to the PF uncertainty, is negligibly small due to abundance of data for calibration.

Indeed, we have found that while the calibrated 2-step mechanism captures the detailed mechanism reasonably well *on average*, the PF uncertainty is by no means representative of the true discrepancy. At the same time, as Figure 10 shows, the embedded model error calibration (again, we employed MVN embedding (11) and independent-normal likelihood (13)) leads to a predictive uncertainty component that fairly describes the true discrepancy across all 1331 data points. This figure shows the true discrepancy from complex-model data to the mean prediction of the two-step mechanism, with a cloud of errorbars behind it corresponding to the PF variance components according to the decomposition (20). Further, Figure 11 illustrates the posterior joint PDFs for  $(E, a_1)$ . While the classical calibration (on the left) leads to an over-confident PDF, the model-error embedded approach leads to posterior PDFs that are representative of actual model discrepancies, as its PF predictions show in Figures 12 and 13. These results demonstrate the model-to-model calibration for a small subset of the 1331 operating conditions. Namely, they illustrate the ignition time dependence on initial temperature for 4 ‘corner’ cases of pressure and equivalence ratio, (11, 0.5), (11, 2.5), (61, 0.5) and (61, 2.5). Again, it is clear that the classical calibration (Figure 12) underestimates the predictive uncertainties, while the model-error embedding (Figure 13) allows predictive uncertainties to capture the discrepancy across a wide range of conditions.

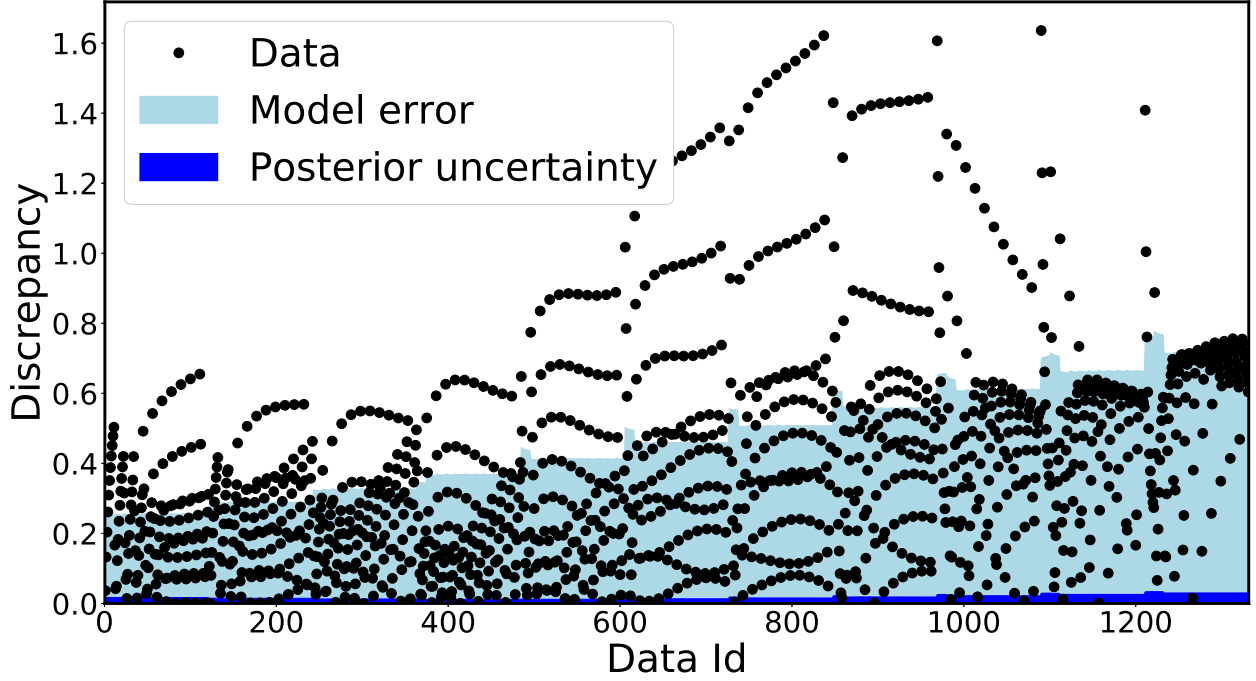


Figure 10: Demonstration of the 2-step mechanism  $f(x, \lambda)$  fit to simulation data from the detailed mechanism  $g(x)$ , across 1331 operating conditions for varying  $x = (P, \phi, 1000/T_0)$  on a uniform  $11^3$  grid. The data points correspond to the actual discrepancy between complex-model data and two-step mechanism mean predictions, while the colored cloud demonstrates the PF standard deviations according to the variance decomposition (20). The MVN-embedded model-error calibration leads to PF uncertainties that are on average representative of the actual model discrepancies, unlike the classical calibration (results not shown), which includes negligibly small PF uncertainty.

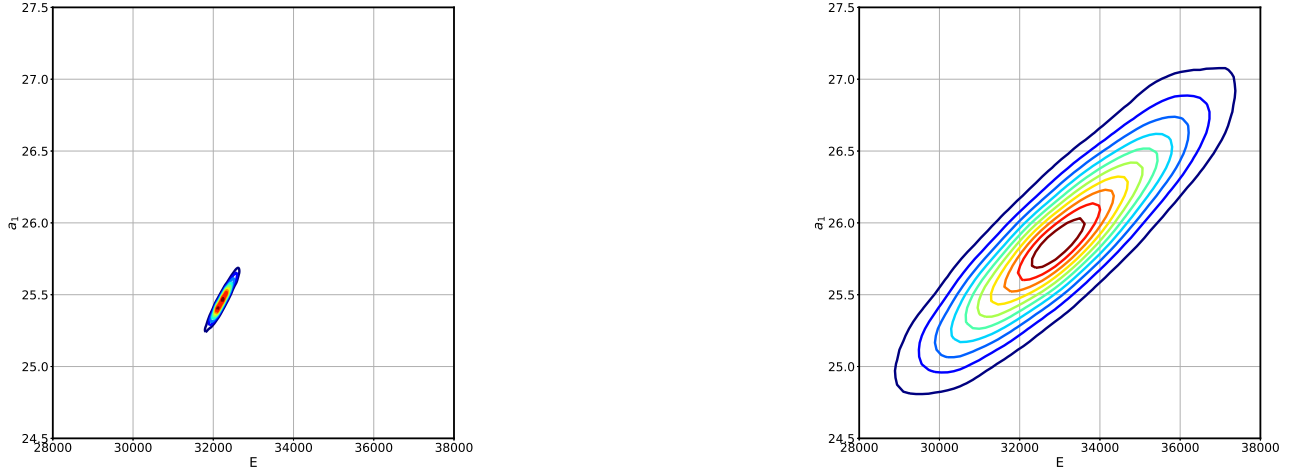


Figure 11: Joint posterior PDFs for  $(E, a_1)$  for (left) classical calibration, and (right) model-error embedded calibration. The classical approach leads to overconfident posterior PDFs as predictive errorbars have shown in Figure 10.

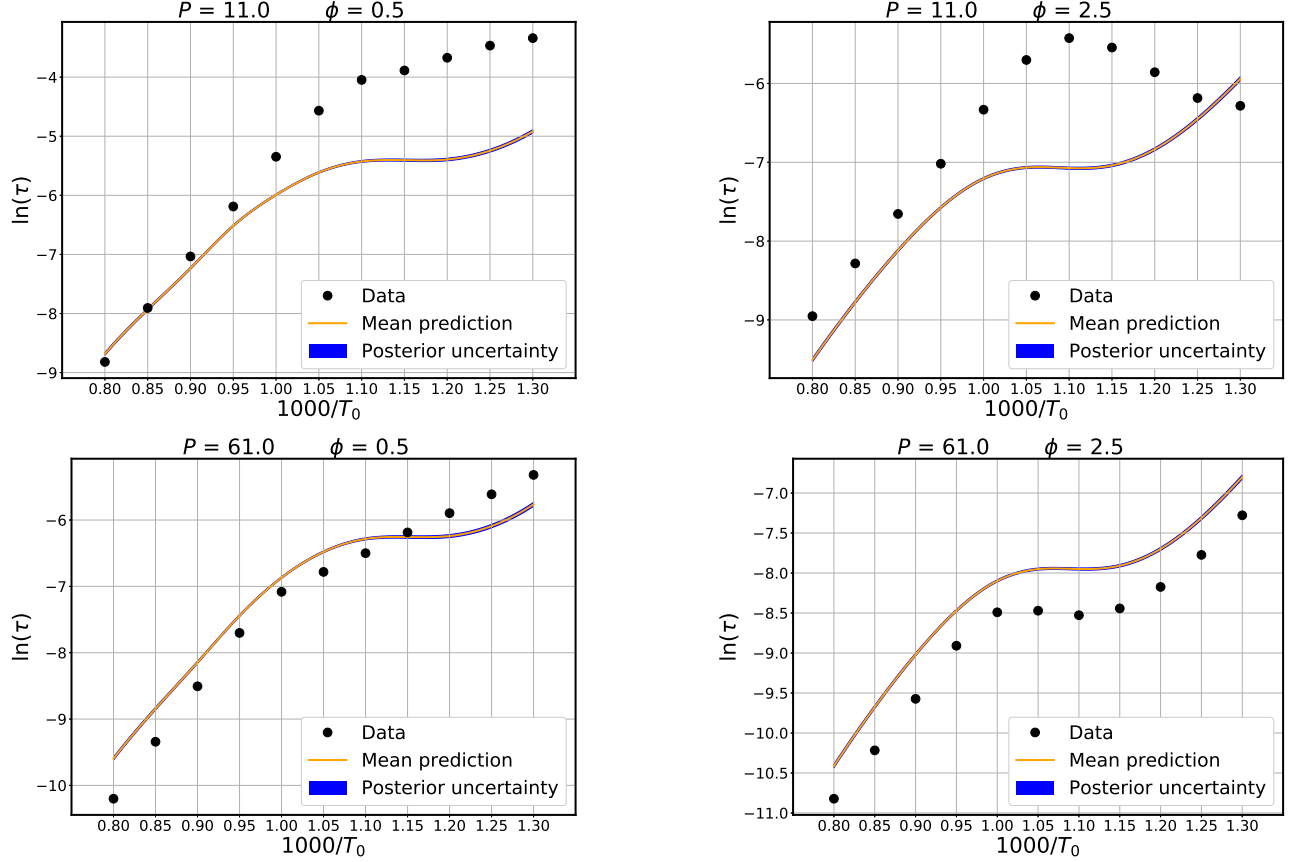


Figure 12: Demonstration of PF uncertainties of classical calibration. The output of interest,  $\ln(\tau)$  is shown as a function of the inverse initial temperature, for 4 ‘corner’ cases of the two other operating conditions, the equivalence ratio  $\phi$  and the pressure  $P$ .

## 6 Discussion

The framework presented in this work takes essential steps towards representation and quantification of model errors. Specifically, we present a principled approach for representing model structural error by embedding it in the model itself. The Bayesian machinery is employed for estimation of embedded model error parameterization together with original model parameters. At the same time, polynomial chaos expansions with pre-built surrogates allow efficient likelihood estimation and uncertainty propagation. The approach shows promising results in various contexts/scenarios that are well-recognized to highlight fundamental challenges associated with model error assessment. For example, model-to-model calibration is a special case of the developed methodology. In such cases, there is no observational data, *i.e.*  $\epsilon_i = 0$  in Eq. (1), one strives to calibrate a lower-fidelity model  $f(x; \lambda)$  with respect to *simulation* data from a higher-fidelity model that is effectively assumed to be the ‘truth’  $g(x)$ . This scenario particularly highlights the deficiency of the conventional calibration approaches that assume *i.i.d* error between the high- and low-fidelity models, while the embedded model error calibration allows a principled way to construct model-to-model discrepancy terms that are directly propagated through the physical model and satisfy physical constraints by definition. This work extends this no-noise approach, developed in [58], to incorporate experimental/observational data leading to full attribution of predictive uncertainties to various components, including uncertainties associated with both structural errors and data amount/quality.

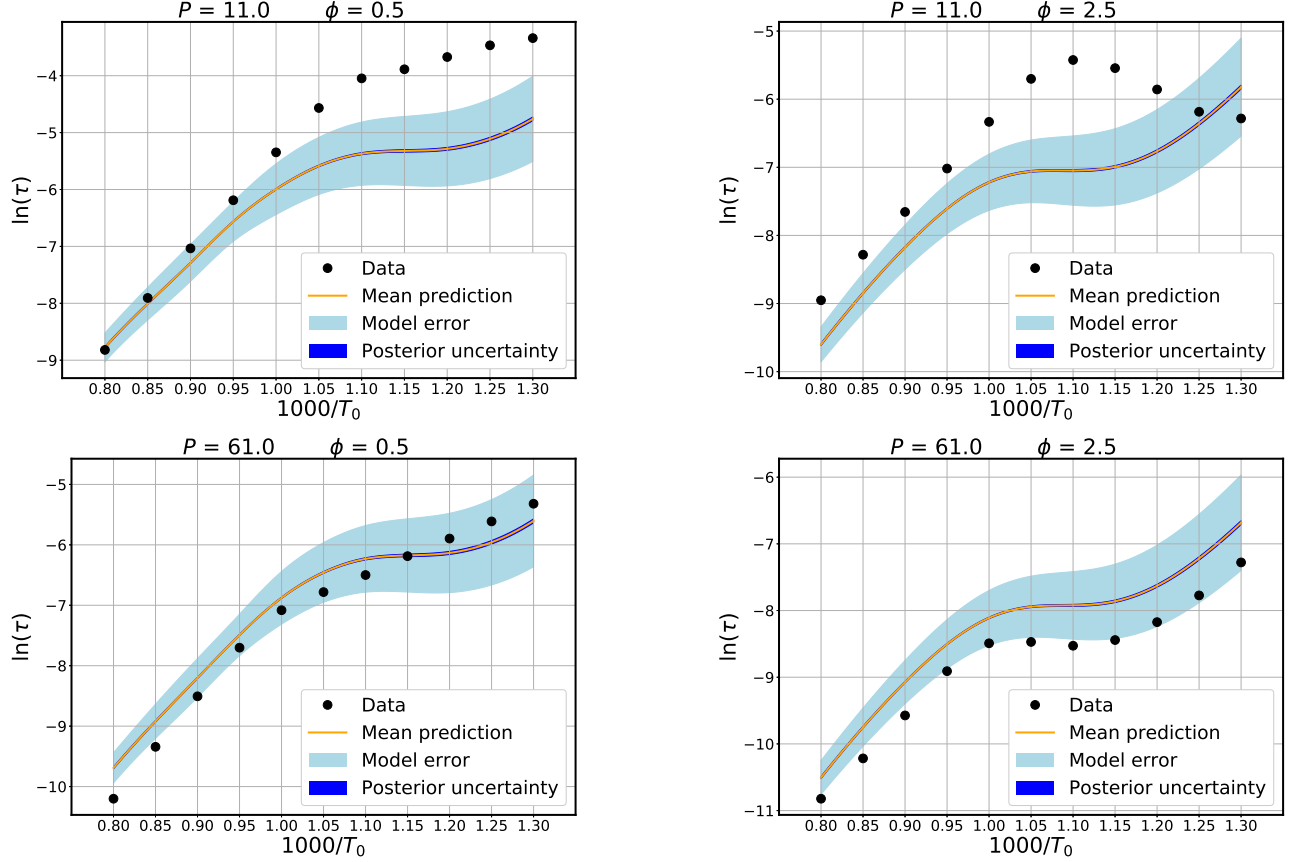


Figure 13: Demonstration of PF uncertainties of calibration with model error embedding. The output of interest,  $\ln(\tau)$  is shown as a function of the inverse initial temperature, for 4 ‘corner’ cases of the two other operating conditions, the equivalence ratio  $\phi$  and the pressure  $P$ .

As a future extension, one can tackle hierarchical models, as hierarchical structure is quite common in physical models. There are also potential opportunities for exploring connections between this embedded model error approach and multi-fidelity model analysis. Another advantage of the present approach is highlighted when the predictive quantities are computed for *another* QoI that relies on  $\lambda$  or a subset thereof. Two different extrapolation modes for prediction are implied, (a) extrapolating  $f(x; \lambda)$  to design conditions  $x$  that are outside the range of  $x_i$ ’s used for calibration, and (b) extrapolating to a different observable or model, *e.g.*  $h(x'; \lambda)$ , potentially at qualitatively different design conditions  $x'$ . The latter in particular highlights the strength of the proposed approach compared to conventional statistical approaches of explicit modeling of model discrepancy as an additive correction on a specific observable [43]. Such additive corrections do not have any predictive meaning when applied to other observable QoIs, while the present approach of embedding model error representations inside the models allows predictions, with quantified uncertainties, for categorically different models or model scenarios. It is worth noting that extrapolation in *any* sense is a dangerous task – and this work is not guaranteeing accurate extrapolative predictions for general models of interest. Rather, by “correcting” the model in physically consistent ways, the approach retains the predictive strength of the physical model intact, whether for prediction of other QoIs, or, to the extent that the physical model is “valid” over a broad set of operating conditions, for extrapolation. The approach lays a foundation for making such predictions that are at least meaningful and informed by the inferred model error magnitude.



While, generally throughout the paper, it is assumed that both the model  $f(x; \boldsymbol{\lambda})$  and the ‘truth’ or high-fidelity model  $g(x)$  are deterministic, the developed machinery in principle generalizes to stochastic models that exhibit internal variability. Such internal variability will simply add an extra uncertainty component to the pushed-forward process  $F(x; \tilde{\boldsymbol{\alpha}})$  (if the model-to-be-calibrated  $f(x; \boldsymbol{\lambda})$  is stochastic) in Eq. (20), or to the posterior predictive random variable (if the high-fidelity model  $g(x)$  is stochastic), effectively playing a role of observational error. In a PC context, one can envision extra elements in the augmented germ (52) to capture internal stochasticity of  $f(x; \boldsymbol{\lambda})$  or  $g(x)$ .

Finally, it is also important to highlight some major challenges for the embedded model error construction that are only partially addressed in this paper or relegated to future work. For example, the dimensionality increase in the associated calibration problem renders the MCMC sampling quite challenging, suggesting the use of advanced MCMC sampling schemes that are well-tuned to sample in data-informed lower-dimensional manifolds or nearly-degenerate posteriors. Also, the Bayesian problem can often be prior-dominated, making appropriate prior selection a crucial part of the methodology. Further, in general, identifiability is an issue for inferring both model structural error parameterization and physical parameters at the same time, particularly in the limit of low amount of high-fidelity or observational data [2].

The calibration with model error embedding, including a variety of likelihood, prior, and embedding form options, is implemented as a part of UQTk v3.0 [19, 18], a lightweight C++/Python library for a range of basic UQ tasks developed at Sandia National Laboratories in Livermore, CA.

## Acknowledgments

The authors would like to thank Jason Bender, Chi Feng, Youssef Marzouk and Cosmin Safta for comments and suggestions throughout this work. The authors acknowledge the support by the Defense Advanced Research Projects Agency (DARPA) program on Enabling Quantification of Uncertainty in Physical Systems (EQUIPS). Also, KS and HNN acknowledge support for this work through the Scientific Discovery through Advanced Computing (SciDAC) program funded by the U.S. Department of Energy (DOE), Office of Science, Advanced Scientific Computing Research (ASCR). HNN also acknowledges support for this work through the DOE, Office of Basic Energy Sciences (BES), Division of Chemical Sciences, Geosciences, and Biosciences. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA-0003525.

## Appendices

### A Attribution of predictive uncertainties

Eq. (35) can be viewed as a  $(d+N)$ -dimensional PC expansion with respect to the augmented PC germ  $\hat{\xi} = (\xi_1, \dots, \xi_d, \xi_{d+1}, \dots, \xi_{d+N})$  in which the first  $d$  stochastic dimensions correspond to the embedded model error, and the last  $N$  correspond to the error associated with data collection.

One can proceed further with the PC paradigm and represent the posterior random vector  $\tilde{\boldsymbol{\alpha}}$  as a PC expansion of dimension  $d_{\tilde{\boldsymbol{\alpha}}}$ , the number of entries in  $\tilde{\boldsymbol{\alpha}}$ ,

$$\tilde{\alpha}_j \simeq \sum_{k=0}^{K_{\tilde{\boldsymbol{\alpha}}}-1} a_{jk} \Psi_k(\xi_{d+N+1}, \dots, \xi_{d+N+j}), \quad \text{for } j = 1, \dots, d_{\tilde{\boldsymbol{\alpha}}}. \quad (49)$$

This can be accomplished via the Rosenblatt transformation, using posterior samples obtained from MCMC. As before, one will need to truncate at an appropriate PC order  $p_{\tilde{\alpha}}$ , leading to a number of terms  $K_{\tilde{\alpha}} = (d_{\tilde{\alpha}} + p_{\tilde{\alpha}})! / (d_{\tilde{\alpha}}! p_{\tilde{\alpha}}!)$ . Then, similar to the PC propagation of  $\mathbf{\Lambda}$  through the model  $f(x; \mathbf{\Lambda})$ , one can propagate the PCE of  $\tilde{\alpha}$  in Eq. (49), through each of the coefficients  $f_k(x_i; \tilde{\alpha})$  to obtain

$$f_k(x_i; \tilde{\alpha}) \simeq \sum_{m=0}^{\tilde{K}-1} f_{kim} \Psi_m(\xi_{d+N+1}, \dots, \xi_{d+N+d_{\alpha}}) \quad (50)$$

for some truncation number  $\tilde{K}$ . Plugging this into the data model (35), one obtains the fully expanded PC representation

$$y_i \approx \sum_{k=0}^{K-1} \sum_{m=0}^{\tilde{K}-1} f_{kim} \Psi_m(\xi_{d+N+1}, \dots, \xi_{d+N+d_{\alpha}}) \Psi_k(\boldsymbol{\xi}) + \sigma \xi_{d+i} \quad \text{for } i = 1, \dots, N. \quad (51)$$

This is a generic PC representation with respect to the fully augmented PC germ

$$\hat{\boldsymbol{\xi}} = \underbrace{(\xi_1, \dots, \xi_d)}_{\text{Model error}}, \underbrace{(\xi_{d+1}, \dots, \xi_{d+N})}_{\text{Measurement error}}, \underbrace{(\xi_{d+N+1}, \dots, \xi_{d+N+d_{\alpha}})}_{\text{Posterior uncertainty}}. \quad (52)$$

Even if the resulting PC expansion (51) has a very large stochastic dimensionality of  $d_{full} = d + N + d_{\alpha}$ , one can employ generic PC tools for variance attribution – without additional cost – in order to attribute the overall uncertainty of the data model into specific components/dimensionalities, such as model error, measurement error and posterior uncertainty. Namely, the right-hand side of Eq. (51) can be written in a general PC form

$$y_i \approx \sum_{\mathbf{p} \in \mathcal{S}} f_{\mathbf{p}} \Psi_{\mathbf{p}}(\hat{\boldsymbol{\xi}}), \quad (53)$$

where  $\mathcal{S}$  denotes the set of multiindices, *i.e.* dimension-wise orders,  $\mathbf{p} = (p_1, \dots, p_{d_{full}})$ , induced by the specific form of the expansion (51). Now the total variance of this expansion  $V_{total} = \sum_{\mathbf{p} \in \mathcal{S}} f_{\mathbf{p}}^2 \|\Psi_{\mathbf{p}}\|^2$  can be attributed to fractional contributions corresponding to specific subsets of the germ, via Sobol sensitivity indices [65, 56]. For example, the main Sobol sensitivity index corresponding to a subset of dimensions  $\hat{\boldsymbol{\xi}}_- \subset \hat{\boldsymbol{\xi}}$  is

$$S_{\hat{\boldsymbol{\xi}}_-} = \frac{1}{V_{total}} \sum_{\mathbf{p} \in \mathcal{S}_-} f_{\mathbf{p}}^2 \|\Psi_{\mathbf{p}}\|^2, \quad (54)$$

where  $\mathcal{S}_- \subset \mathcal{S}$  is the subset of multiindices that involve *only* dimensions in  $\hat{\boldsymbol{\xi}}_-$ . When specific subsets of the full germ, highlighted in Eq. (52), are taken, one recovers a decomposition that is essentially the PC-based counterpart of the general variance decomposition (20). Such uncertainty attribution decomposes the overall uncertainty and allows informed decision-making in selecting the submodels/parameterizations that contribute most to it.

## References

- [1] I. Alvarez, J. Niemi, and M. Simpson. Modeling structural uncertainties in reynolds-averaged computations of shock/boundary layer interactions. In *Annual Conference on Applied Statistics and Agriculture*, pages 71–82, 2014.
- [2] P. D. Arendt, D. W. Apley, and W. Chen. Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *Journal of Mechanical Design*, 134(10):100908, 2012.

- [3] M. Arnst, R. Ghanem, and C. Soize. Identification of Bayesian posteriors for coefficients of chaos expansions. *Journal of Computational Physics*, 229(9):3134–3154, 2010.
- [4] J. Barnard, R. McCulloch, and X.-L. Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, pages 1281–1311, 2000.
- [5] M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C. H. Lin, and J. Tu. A framework for validation of computer models. *Technometrics*, 49(2):138–154, 2007.
- [6] M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu. A framework for validation of computer models. *Technometrics*, 49(2):138–154, 2007.
- [7] M. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian Computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.
- [8] J. Bernardo and A. Smith. *Bayesian Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd, Chichester, England, 2000.
- [9] L. Bojke, K. Claxton, M. Sculpher, and S. Palmer. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value in Health*, 12(5):739–749, 2009.
- [10] J. Brynjarsdottir and A. O’Hagan. Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30, 2014.
- [11] F. Cailliez and P. Pernot. Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation. *The Journal of chemical physics*, 134(5):054124, 2011.
- [12] K. Campbell. Statistical calibration of computer simulations. *Reliability Engineering & System Safety*, 91(10):1358–1363, 2006.
- [13] B. P. Carlin and T. A. Louis. *Bayesian Methods for Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL, 2011.
- [14] R. Christensen. *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer-Verlag New York, New York, NY, 3rd edition, 2002.
- [15] Y. Chung, A. Gelman, S. Rabe-Hesketh, J. Liu, and V. Dorie. Weakly informative prior for point estimation of covariance matrices in hierarchical models. *Journal of Educational and Behavioral Statistics*, 40(2):136–157, 2015.
- [16] J. A. Curry and P. J. Webster. Climate science and the uncertainty monster. *Bulletin of the American Meteorological Society*, 92(12):1667–1682, 2011.
- [17] M. Daniels and R. Kass. Nonconjugate Bayesian Estimation of Covariance Matrices and Its Use in Hierarchical Models. *J. Am. Stat. Assoc.*, 94(448):1254–1263, 1999.
- [18] B. Debusschere, K. Sargsyan, C. Safta, and K. Chowdhary. The uncertainty quantification toolkit (uqtk). In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*. Springer, 2016.
- [19] B. Debusschere, K. Sargsyan, C. Safta, and K. Chowdhary. UQ Toolkit. <http://www.sandia.gov/UQToolkit>, 2017.

- [20] F. L. Dryer and I. Glassman. High-temperature oxidation of CO and CH<sub>4</sub>. In *Proceedings of the Combustion Institute*, volume 14, pages 987–1003, 1973.
- [21] M. S. Eldred. Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design. *AIAA Paper*, 2274:2009, 2009.
- [22] M. Emory, R. Pecnik, and G. Iaccarino. Modeling structural uncertainties in reynolds-averaged computations of shock/boundary layer interactions. In *49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, page 479, 2011.
- [23] E. Fernandez-Tarrazo, A. L. Sánchez, A. Linan, and F. A. Williams. A simple one-step chemistry model for partially premixed hydrocarbon combustion. *Combustion and Flame*, 147(1):32–38, 2006.
- [24] B. Franzelli, E. Riber, M. Sanjosé, and T. Poinso. A two-step chemical scheme for kerosene-air premixed flames. *Combustion and Flame*, 157(7):1364–1373, July 2010.
- [25] D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC, Boca Raton, FL, 2006.
- [26] A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–534, 2006.
- [27] A. Gelman, X.-L. Meng, and H. Stern. Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. *Statistica Sinica*, 6:733–807, 1996.
- [28] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer Verlag, New York, 1991.
- [29] W. R. Gilks. *Markov chain Monte Carlo*. Wiley Online Library, 2005.
- [30] H. Gupta, M. Clark, J. Vrugt, G. Abramowitz, and M. Ye. Towards a comprehensive assessment of model structural adequacy. *Water Resour. Res.*, 48:W08301, 2012.
- [31] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242, 2001.
- [32] L. Hakim, G. Lacaze, M. Khalil, H. Najm, and J. Oefelein. Modeling auto-ignition transients in reacting Diesel jets. *ASME J. Eng. Gas Turb. Power*, 138(11):112806–112806–8, 2016. Paper #: GTP-16-1054.
- [33] L. Hakim, G. Lacaze, M. Khalil, K. Sargsyan, H. Najm, and J. Oefelein. Probabilistic parameter estimation in a 2-step chemical kinetics model for n-dodecane jet autoignition. *Combustion Theory and Modeling*, 2017. in review.
- [34] Y. He and D. Xiu. Numerical strategy for model correction using physical constraints. *Journal of Computational Physics*, 313:617–634, 2016.
- [35] D. Higdon, J. Gattiker, B. Williams, and M. Rightley. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482), 2008.
- [36] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafo, and R. D. Ryne. Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466, 2004.

- [37] C.-W. Hsu, M. S. Sinay, and J. S. Hsu. Bayesian estimation of a covariance matrix with flexible prior specification. *Annals of the Institute of Statistical Mathematics*, 64(2):319–342, 2012.
- [38] X. Huan, C. Safta, K. Sargsyan, G. Geraci, M. S. Eldred, Z. P. Vane, G. Lacaze, J. C. Oefelein, and H. N. Najm. Global Sensitivity Analysis and Quantification of Model Error for Large Eddy Simulation in Scramjet Design. In *19th AIAA Non-Deterministic Approaches Conference*, number 2017-1089, Grapevine, TX, 2017.
- [39] X. Huan, C. Safta, K. Sargsyan, G. Geraci, M. S. Eldred, Z. P. Vane, G. Lacaze, J. C. Oefelein, and H. N. Najm. Global Sensitivity Analysis and Quantification of Model Error in Scramjet Computations. *AIAA Journal*, 2017. in press.
- [40] A. Huang, M. P. Wand, et al. Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452, 2013.
- [41] V. R. Joseph and S. N. Melkote. Statistical adjustments to engineering models. *Journal of Quality Technology*, 41(4):362, 2009.
- [42] R. E. Kass, R. Natarajan, et al. A default conjugate prior for variance components in generalized linear mixed models (comment on article by browne and draper). *Bayesian Analysis*, 1(3):535–542, 2006.
- [43] M. C. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B*, 63(3):425–464, 2001.
- [44] O. Le Maître and O. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, New York, NY, 2010.
- [45] I. A. Leenson and G. B. Sergeev. Negative temperature coefficient in chemical reactions. *Russian Chemical Reviews*, 53(5):417, 1984.
- [46] T. Leonard and J. S. Hsu. Bayesian inference for a covariance matrix. *The Annals of Statistics*, pages 1669–1696, 1992.
- [47] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci USA*, 100(26):15324–15328, 2003.
- [48] A. Misdariis, O. Vermorel, and T. Poinsot. A methodology based on reduced schemes to compute autoignition and propagation in internal combustion engines. *Proceedings of the Combustion Institute*, 35(3):3001–3008, 2015.
- [49] H. Najm. Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, 41(1):35–52, 2009.
- [50] K. Narayanaswamy, P. Pepiot, and H. Pitsch. A chemical mechanism for low to high temperature oxidation of n-dodecane as a component of transportation fuel surrogates. *Comb. Flame*, 161:867–884, 2014.
- [51] A. O’Hagan. Bayesian inference with misspecified models: Inference about what? *Journal of Statistical Planning and Inference*, 143(10):1643 – 1648, 2013.
- [52] T. A. Oliver and R. D. Moser. Bayesian uncertainty quantification applied to RANS turbulence models. *J. Phys.: Conf. Ser.*, 318, 2011.

- [53] T. A. Oliver, G. Terejanu, C. S. Simmons, and R. D. Moser. Validating predictions of unobserved quantities. *Computer Methods in Applied Mechanics and Engineering*, 283:1310–1335, 2015.
- [54] P. Pernot. The parameters uncertainty inflation fallacy. *arXiv preprint arXiv:1611.04295*, 2016.
- [55] C. Safta, H. Najm, and O. Knio. TChem - a software toolkit for the analysis of complex kinetic models. *Sandia Report*, SAND2011-3282, 2011. <http://www.sandia.gov/tchem>.
- [56] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, 2004.
- [57] K. Sargsyan. Surrogate models for uncertainty propagation and sensitivity analysis. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*. Springer, 2016.
- [58] K. Sargsyan, H. N. Najm, and R. Ghanem. On the Statistical Calibration of Physical Models. *International Journal for Chemical Kinetics*, 47(4):246–276, 2015.
- [59] D. Scott. *Multivariate Density Estimation. Theory, Practice and Visualization*. Wiley, New York, 1992.
- [60] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.
- [61] S. A. Sisson and Y. Fan. *Handbook of Markov Chain Monte Carlo*, chapter Likelihood-free Markov chain Monte Carlo, pages 313–338. Chapman and Hall/CRC Press, 2011.
- [62] D. S. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial, Second Edition*. Oxford University Press, 2006.
- [63] M. Smith and R. Kohn. Parsimonious covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association*, 97(460):1141–1153, 2002.
- [64] R. Smith. *Uncertainty Quantification: Theory, Implementation and Applications*. SIAM Computational Science and Engineering, Philadelphia, PA, 2013.
- [65] I. M. Sobol. Theorems and examples on high dimensional model representation. *Reliability Engineering and System Safety*, 79:187–193, 2003.
- [66] M. Strong and J. Oakley. When is a model good enough? Deriving the expected value of model improvement via specifying internal model discrepancies. *SIAM/ASA J Uncertainty Quantification*, 2:106–125, 2014.
- [67] M. Strong, J. Oakley, and J. Chilcott. Managing structural uncertainty in health economic decision models: a discrepancy approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(1):25–45, 2012.
- [68] S. S. Vasu, D. F. Davidson, Z. Hong, V. Vasudevan, and R. K. Hanson. n-Dodecane oxidation at high-pressures: Measurements of ignition delay times and OH concentration time-histories. *Proceedings of the Combustion Institute*, 32(1):173–180, 2009.
- [69] H. Wang and N. S. Pillai. On a class of shrinkage priors for covariance matrix estimation. *Journal of Computational and Graphical Statistics*, 22(3):689–707, 2013.



- [70] S. Wang, W. Chen, and K.-L. Tsui. Bayesian validation of computer models. *TECHNOMETRICS*, 51(4):439–451, 2009.
- [71] C. Westbrook and F. Dryer. Simplified Reaction Mechanisms for the Oxidation of Hydrocarbon Fuels in Flames. *Combustion Science and Technology*, 27:31–43, 1981.
- [72] D. Xiu and G. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.
- [73] R. Yang and J. Berger. Estimation of a Covariance Matrix Using the Reference Prior. *The Annals of Statistics*, 22(3):1195–1211, 1994.