14th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP14
Dublin, Ireland, July 9-13, 2023

Polynomial Chaos Surrogate Construction for Stochastic Models with Parametric Uncertainty

Joy N. Mueller
Postdoctoral Appointee, Sandia National Laboratories, Livermore, USA

Khachik Sargsyan
Principal Member of Technical Staff, Sandia National Laboratories, Livermore, USA

Habib N. Najm
Senior Scientist, Sandia National Laboratories, Livermore, USA

ABSTRACT: Engineering and applied science relies on computational experiments to rigorously study physical systems. The mathematical models used to probe these systems are highly complex, and sampling-intensive studies often require prohibitively many simulations for acceptable accuracy. Surrogate models provide a means of circumventing the high computational expense of sampling such complex models. In particular, Polynomial chaos expansions (PCEs) have been successful for uncertainty quantification studies of deterministic models where the dominant source of uncertainty is parametric. We discuss an extension to conventional PCE surrogate modeling to enable surrogate construction for stochastic computational models that have intrinsic noise in addition to parametric uncertainty. We develop a PCE surrogate on a joint space of intrinsic and parametric uncertainty, enabled by Rosenblatt transformations. We then take advantage of closed-form solutions for computing PCE Sobol indices to perform a global sensitivity analysis of the model which explicitly quantifies the intrinsic noise contribution to the overall model output variance. Additionally, the resulting joint PCE is generative in the sense that it allows generating random realizations at any input parameter setting that are statistically approximately equivalent to realizations from the underlying stochastic model. The methodology is demonstrated on a chemical catalysis example model.

1. INTRODUCTION

Uncertainty quantification and model calibration studies of complex computational models are often prohibitively expensive. In order to alleviate the expense, one often constructs a surrogate model or a response surface approximation of the input-output map of the underlying model. The surrogates are constructed using a feasible number of training simulations spanning a range of input conditions. Polynomial chaos expansions (PCEs) specifically have been successful for surrogate modeling and uncertainty quantification for a range of engineering applications [5, 9, 11]. Typically, PCEs are used in the context of deterministic models with uncertain inputs. They serve as a convenient technique for uncertainty propagation, moment estimation, and variance decomposition due to the orthogonality property of polynomial bases [19].

While local/wavelet-based PCE constructions can be used to represent non-smooth functions, in most practical applications the effectiveness of (global) PCEs relies on the smoothness of the input-output map comprising the model. As such, they are not well suited for noisy/non-smooth input-output maps. As a remedy, PCEs constructed for noisy systems are built for integrated output quantities of interest (QoIs), e.g. expectations of model output. In this work we strive to delay the averag-

ing and consequent loss of information until after the surrogate is constructed. That is, we would like to develop a *generative* surrogate approximation to the model at the sample realization level. Generative models, such as variational autoencoders [8], generative adversarial networks [6] or normalizing flows [12], have gained popularity in machine learning literature. In engineering literature, a similar idea for stochastic PCE construction based on maximum likelihood estimation of the underlying distributions has been presented albeit with a focus on intrinsic noise only [20]. Another approach effectively builds a parameter-space surrogate for each random realization of the random seed, arriving at probabilistic description of global sensitivity indices [7]. In this work, we remain in the PCE framework by employing the Rosenblatt transformation [13] to incorporate the generative, stochastic dimensions while keeping the surrogate approximation valid in the parametric dimensions. Consequently, the PCE framework allows immediate analytical expressions for global sensitivity indices and consequent variance decomposition into parametric uncertainty and intrinsic noise. Our joint intrinsic-parametric PCE construction explicitly decouples the intrinsic noise from the parametric uncertainty, allowing model exploration from both a conventional PCE surrogate perspective (averaging over intrinsic noise) and from the viewpoint of probability density function (PDF) comparison for fixed parameter configurations.

The paper is organized as follows. In Section 2, we describe the basic building blocks of the methodology, PCEs and Rosenblatt maps. Section 3 describes the joint PCE construction. Next, Sections 4 and 5 describe the application problem of interest and results of applying our method to it, respectively. We close with concluding remarks in Section 6.

2. METHODOLOGY

Consider a stochastic model $\boldsymbol{y} = f(\boldsymbol{\lambda}, \omega)$ where the forward function f depends on a set of input parameters $\boldsymbol{\lambda} \in \mathbb{R}^L$ and outputs a random vector $\boldsymbol{y} \in \mathbb{R}^d$. With ω we denote an element of the sample space explicitly indicating the stochasticity of the model. The underlying assumption is that f is expensive to evaluate directly, and therefore it is useful to identify a simpler model whose input-output behavior closely mimics the underlying forward function. The simple model acts as a surrogate representation for f and can be used in its place to perform sample-intensive studies of the system of interest such as uncertainty propagation, sensitivity analysis, or parameter estimation.

2.1. Polynomial Chaos Expansions

Polynomial chaos expansions are functional representations of random variables and provide convenient machinery for uncertainty quantification in computational models. Typically, uncertain inputs are written as a polynomial expansion

$$\boldsymbol{\lambda} \simeq \sum_{p=0}^{P-1} \boldsymbol{a}_p \Psi_p(\boldsymbol{\xi}) \tag{1}$$

in terms of orthogonal polynomials Ψ_p of a vector of *i.i.d.* standard random variables $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_L)$. Multivariate polynomials in turn are written as products of univariate polynomials $\Psi_p(\boldsymbol{\xi}) = \prod_{\ell=1}^{L} \psi_{p_\ell}(\xi_\ell)$, where each scalar index $p = p(p_1, \ldots, p_\ell)$ accounts for the corresponding *multi-index* or multivariate degree $(p_1, \ldots, p_\ell)$. We will employ Legendre-Uniform PCEs (Legendre polynomials as functions of standard uniform random variables), as is typical for input parameters with compact support. Moreover, we will assume a first-order, independent-component PCE for each parameter $\lambda_i = \frac{b_i+a_i}{2} + \frac{b_i-a_i}{2}\xi_i$ given a physically meaningful interval of uncertainty $[a_i, b_i]$, for $i = 1, \ldots, L$. This simplifying assumption is made without loss of generality: we largely rely on a correspondence $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\xi})$ between the physical parameter $\boldsymbol{\lambda}$ and the parametric *germ* $\boldsymbol{\xi} \in \text{Unif}[-1, 1]^L$.

Conventionally, for a *deterministic* model $g(\boldsymbol{\lambda})$, one writes a PCE representation

$$g(\boldsymbol{\lambda}) \simeq \sum_{p=0}^{P-1} \boldsymbol{c}_p \Psi_p(\boldsymbol{\xi}) \tag{2}$$

with the same germ $\boldsymbol{\xi}$, and finds coefficients $\boldsymbol{c}_p$ by regression using a set of N model evaluations $g(\boldsymbol{\lambda}^{(n)})$ at select training inputs $\boldsymbol{\lambda}^{(n)} = \boldsymbol{\lambda}(\boldsymbol{\xi}^{(n)})$, for $n = 1, \ldots, N$. Together with the map $\boldsymbol{\lambda}(\boldsymbol{\xi})$, the PCE

in Eq. (2) acts as a surrogate representation for the deterministic model $\boldsymbol{y} = g(\boldsymbol{\lambda})$ [15].

Representing random variables with PCEs offers a number of advantages. Importantly for this work, the first two moments can be extracted from a PCE via weighted sum-of-squares of the expansion coefficients and magnitudes of the basis functions, namely:

$$\mathbb{E}[\boldsymbol{y}] = c_0, \qquad \mathbb{V}[\boldsymbol{y}] = \sum_{p=1}^{P-1} c_p^2 ||\Psi_p||^2. \tag{3}$$

Furthermore, PCE-based surrogate models enable variance-based decomposition and subsequent extraction of Sobol indices. The Sobol indices decompose the output variance into fractional variance contributions of each parameter or group of parameters [14]. Here we examine the main-effect (first-order) sensitivities, which measure the influence of each parameter in isolation on the output variance. These indices are defined as

$$S_i := \frac{\mathbb{V}_{\lambda_i}[\mathbb{E}_{\lambda_{-i}}[g(\boldsymbol{\lambda} \mid \lambda_i)]]}{\mathbb{V}[\boldsymbol{y}]}, \tag{4}$$

where $\mathbb{V}_{\lambda_i}$ denotes the variance with respect to the ith parameter and $\mathbb{E}_{\lambda_{-i}}$ the expectation with respect to all remaining parameters. Similar to the first two moments in Eq. (3), due to orthogonality of PCE bases, one can extract exact formulae for the sensitivity indices of a PCE surrogate model as weighted sums of squares of PCE coefficients [3, 15, 19].

2.2. The Rosenblatt Map and Stochastic PCEs

Underpinning the PCE surrogate construction for the deterministic forward model described above is the explicit mapping between uncertain parameters $\boldsymbol{\lambda}$ – and consequently, model outputs $g(\boldsymbol{\lambda})$ – and the PCE germ $\boldsymbol{\xi}$. In this work, we refer to these PCE surrogates as 'parametric PCEs.'

Now, suppose that our model has intrinsic noise, i.e. for each fixed value of $\boldsymbol{\lambda}^\star$ its output is a random quantity $f(\boldsymbol{\lambda}^\star)$. For ease of presentation, although less rigorous, we write the forward function as $f(\boldsymbol{\lambda}, \omega)$ to explicitly highlight a sample space element ω. Surrogate construction is not as straightforward as described in Section 2.1, since there is no explicit control on ω. Having said that, one can construct a PCE representation of a random variable $\boldsymbol{y} = f(\boldsymbol{\lambda}^\star, \omega) \in \mathbb{R}^d$ given M samples $\{\boldsymbol{y}^{(m)} = f(\boldsymbol{\lambda}^\star, \omega^{(m)})\}_{m=1}^M$, relying on the Rosenblatt transformation [13] and kernel density estimation (KDE) [18]. The Rosenblatt transformation is the multivariate generalization of the cumulative distribution function (CDF) theorem, mapping any random variable to a set of *i.i.d.* uniform random variables $(\zeta_1, \ldots, \zeta_d)$. In fact, the inverse of the Rosenblatt transformation $R^{-1}(\boldsymbol{\zeta}) = \boldsymbol{y}$ essentially provides a mapping from ω to the germ $\boldsymbol{\zeta}$ (see [16] for details). In other words, R^{-1} acts as a deterministic forward map allowing construction of the PCE similar to Section 2.1. Consequently, the PCE is written as

$$\boldsymbol{y} = R^{-1}(\boldsymbol{\zeta}) \simeq \sum_{s=0}^{S-1} \boldsymbol{z}_s \Psi_s(\boldsymbol{\zeta}). \tag{5}$$

This expansion is built purely on samples of the intrinsic noise in the model, and consequently we refer to it as a 'stochastic PCE'.

3. Joint Stochastic-Parametric PCE

With the parametric and stochastic PCE constructions in mind, we now describe our algorithm to arrive at a PCE surrogate on a joint space $(\boldsymbol{\lambda}, \omega)$ of parametric and stochastic dimensions.

Suppose we have an ensemble of parameter configurations $\boldsymbol{\lambda}^{(n)}$ $(n = 1, \ldots, N)$ and, for each fixed $\boldsymbol{\lambda}^{(n)}$, we obtain M samples of the stochastic model $f(\boldsymbol{\lambda}^{(n)}, \omega^{(n,m)})$. For each parametric sample, we build a multi-output stochastic PCE according to Eq. (5) which takes the form

$$f(\boldsymbol{\lambda}^{(n)}, \omega) \simeq \sum_{s=0}^{S-1} \boldsymbol{z}_s^{(n)} \Psi_s(\boldsymbol{\zeta}). \tag{6}$$

The next step in our joint PCE construction is to build a parametric PCE for the coefficients $\boldsymbol{z}_s(\boldsymbol{\lambda})$, $s = 0, \ldots, S-1$, where the samples $\{\boldsymbol{z}_s^{(n)}\}_{n=1}^N$ will serve as training data for the parametric PCE regression. The PCE will then take the form

$$\boldsymbol{z}_s(\boldsymbol{\lambda}) \approx \boldsymbol{z}_s^{\mathrm{PCE}}(\boldsymbol{\lambda}) \simeq \sum_{p=0}^{P-1} \boldsymbol{a}_{sp} \Psi_p(\boldsymbol{\xi}) \tag{7}$$

Replacing the coefficients in Eq. (6) with this coefficient surrogate Eq. (7), we obtain a joint parametric-stochastic PCE representation

$$f(\boldsymbol{\lambda},\omega) \simeq \sum_{s=0}^{S-1} \left(\sum_{p=0}^{P-1} \boldsymbol{a}_{sp} \Psi_p(\boldsymbol{\xi}) \right) \Psi_s(\boldsymbol{\zeta}). \tag{8}$$

In other words, we arrive at a joint PCE by building regression-based stochastic PCEs from inverse-Rosenblatt maps for each parameter $\boldsymbol{\lambda}$ followed by another regression with respect to the parameter $\boldsymbol{\lambda}$ itself. Note that one can rewrite the nested expansion Eq. (8) as a single joint expansion with a higher-dimensional multivariate basis

$$f(\boldsymbol{\lambda},\omega) \simeq \sum_{j=0}^{J-1} \boldsymbol{c}_j \Psi_j(\boldsymbol{\xi},\boldsymbol{\zeta}) \tag{9}$$

where the PCE germs $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ correspond to their 'physical' counterparts $\boldsymbol{\lambda}$ and ω, respectively.

4. CO Oxidation on RuO_2

To illustrate the methodology, we consider a chemical catalysis model featuring a set of uncertain rate coefficients as well as intrinsic noise due to the randomness of chemical reactions. Specifically, consider CO oxidation on a RuO_2 surface, represented at atomic scales as a lattice with alternating coordinatively unsaturated (`cus`) sites and bridge (`br`) sites where unimolecular adsorption (desorption) of CO and dissociative adsorption (desorption) of O_2 may occur. The adsorbed species can make diffusive hops to neighboring vacant sites and nearest-neighbor O and CO can react, producing gaseous CO_2. The system is modeled using 22 permissible processes, where a 'process' on the lattice is specified by event type, chemical species, and lattice site(s) involved – for example, diffusion of CO from a `cus` site to a `br` site (see [1, 4] for a more thorough discussion of the mechanism).

Associated with each process is a reaction rate constant which is treated as an uncertain parameter. The CO_2 formation processes and species diffusion between identical lattice sites are modeled as irreversible 'forward' processes with associated forward rate constants k_f, while diffusion between non-identical lattice sites and adsorption/desorption are treated as reversible processes. The reversible processes have forward and reverse rate constants k_f and k_r that are thermodynamically balanced.

The uncertain input parameter vector $\boldsymbol{\lambda}$ consists of logarithms of 15 reaction rates, specified as uniform random variables around some nominal values with a scale factor vector $\boldsymbol{r}$:

$$\boldsymbol{\lambda} := \{\boldsymbol{k} \,|\, \ln \boldsymbol{k} \sim \mathrm{Unif}(\ln \boldsymbol{k}_{\mathrm{nom}} - \ln \boldsymbol{r}, \ln \boldsymbol{k}_{\mathrm{nom}} + \ln \boldsymbol{r})\}.$$

The QoI considered here is the collection of mean coverage fractions for each species at a fixed time point, $\boldsymbol{y}_{\mathrm{cvg}}(t_f) = (y^{\mathrm{O}}_{\mathrm{cvg}}(t_f), y^{\mathrm{CO}}_{\mathrm{cvg}}(t_f))$. For notational convenience, we drop explicit dependence on the fixed time t_f. We also extract any necessary results for the evolution of vacant sites via a conservation law $y^{\mathrm{vac}}_{\mathrm{cvg}}(t_f) = 1 - y^{\mathrm{O}}_{\mathrm{cvg}}(t_f) - y^{\mathrm{CO}}_{\mathrm{cvg}}(t_f)$.

Figure 1 demonstrates the evolution of QoI $y^{\mathrm{CO}}_{\mathrm{cvg}}(t)$ for varying parameter values. Clearly, the coverage fraction of the chemical system changes significantly with time. The effect of the parameters $\boldsymbol{\lambda}$ is illustrated by the divergence of the sample solution means to distinct trajectories over time, while the magnitude of intrinsic noise ω is indicated by the colored band surrounding each sample mean. Looking ahead to global sensitivity analysis, the fractional variance contributions of intrinsic noise and parametric uncertainty are expected to be dominated by the latter at the final time, since the colored bands seem well-separated.

5. Results

We illustrate the construction of the joint PCE surrogate for the test case described in the previous section modeling the mean species-coverages of CO oxidation over time on a 40×40 RuO_2 lattice. The uncertain parameter vector is a set of 15 forward (log) reaction rate constants $\boldsymbol{\lambda} = (\ln k_1, \ldots, \ln k_{15})$. Here we generate $N = 150$ parameter configurations sampled around nominal values given in [4] with an uncertainty scale factor of $r = 3$ for each parameter. For each fixed parameter configuration, we take $M = 500$ samples of the intrinsically stochastic model. We first examine the quality of the stochastic and parametric PCEs separately to demonstrate that each step in the surrogate construction is sufficiently accurate.

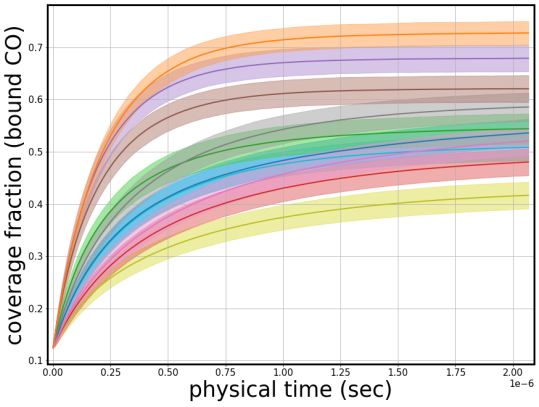

Figure 1: Coverage fraction time evolution for CO. Each color corresponds to a fixed parameter configuration $\boldsymbol{\lambda}^{\star}$. The ensemble means are plotted as thick lines and the standard deviation represented as a band around the mean.

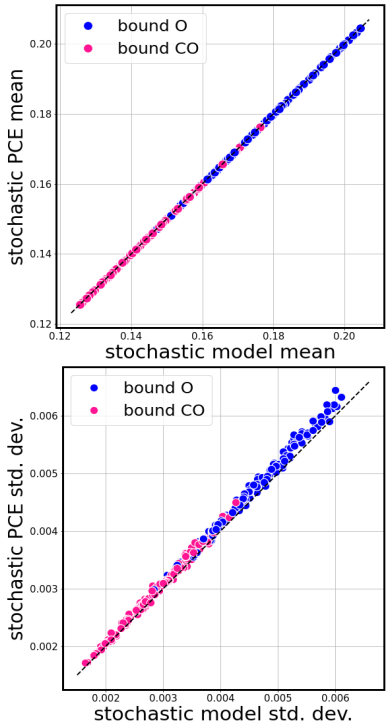

Figure 2: Parity plots of the mean ($E_{rRMSE} = 2.5 \times 10^{-4}$) and standard deviation ($E_{rRMSE} = 3.9 \times 10^{-2}$) of species-coverage fractions predicted by the stochastic PCE for all $\boldsymbol{\lambda}$. The statistics are colored by species in the RuO_2 synthetic model: bound O (blue) and bound CO (magenta).

It follows that we must consider multiple measures of error related to the PCE goodness-of-fit, as the stochastic and parametric PCEs define fundamentally different approximations of the model outputs and cannot be evaluated in the same way. In either case, we quantify how well the PCE approximation agrees with the model results via the relative root mean-squared error (rRMSE) defined as

$$E_{\text{rRMSE}} := \sqrt{\frac{\sum_{n=1}^{N} (v_n - v_n^{\text{PCE}})^2}{\sum_{n=1}^{N} v_n^2}} \tag{10}$$

where v_n and v_n^{PCE} represent the true and PCE-predicted values (either point estimates or moments), respectively.

5.1. Stochastic PCE

We limited the stochastic PCEs to first-order expansions of the germ data, as higher-order expansions appeared to offer little gain in accuracy while rapidly increasing the number of parametric surrogates needed to construct the joint PCE.

Due to lack of control over $\boldsymbol{\omega}$ and $\boldsymbol{\zeta}$, the training samples of $\boldsymbol{y}_{\text{cvg}}$ and the stochastic PCE predictions are only guaranteed to agree in distribution. Therefore, we assess the overall performance of the stochastic PCEs by comparing statistics of the training model simulations and the PCE-predicted species coverage fractions. For each $\boldsymbol{\lambda}$, we take advantage of Eq. (3) to find the analytical means and standard deviations for the PCE predictions. We then plot each component of the predicted means (standard deviations) against the true sample means (standard deviations) in a parity plot. Figure 2 demonstrates this comparison, indicating that the stochastic PCEs predict the sample means with high accuracy. A similar comparison of the predicted standard deviations reveals that the stochastic PCEs generally over-estimate the variance of each distribution (see Figure 2). The choice of bandwidth used to construct the Rosenblatt maps is a likely culprit, and we believe that the bandwidth based on Silverman's rule-of-thumb [18] is not optimal for moment approximation and leads to slight but systematic over-estimation of predicted variances in the stochastic samples.

5.2. Parametric PCE

Since the elements of $\boldsymbol{\lambda}$ are distributed uniformly on $[\ln \boldsymbol{k}_{\text{nom}} - \ln \boldsymbol{r}, \ln \boldsymbol{k}_{\text{nom}} + \ln \boldsymbol{r}]$ and the parametric

samples are independent, we use Legendre polynomials as the expansion basis and obtain the simple rescaling relationship as a first-order Legendre-Uniform PCE,

$$\lambda_i = \ln k_{\text{nom},i} + \xi_i \ln r_i, \text{ for } i = 1,\ldots,L. \qquad (11)$$

As described in Section 2, we then construct a parametric PCE for the stochastic PCE coefficients $\boldsymbol{z}_s$ as functions of the rescaled input, or germ, $\boldsymbol{\xi}$.

The parametric PCEs are chosen to have at most third-order expansions of the coefficient data $\boldsymbol{z}_s$, and the optimal orders for each s are selected via model evidence which is analytically available for this regression step [15]. We reduce the number of expansion terms in the parametric PCE by estimating the coefficients with Bayesian Compressive Sensing (BCS), which adaptively finds a sparse, optimal basis set for the PCE [2, 10, 17]. We examined the accuracy of the parametric PCEs using randomized 80%:20% training/testing splits of the data with the rRMSE metric on coefficient approximation, followed by a final training of the parametric surrogate using all samples.

Unlike the stochastic PCEs, the parametric PCE preserves pointwise mappings between the PCE germ $\boldsymbol{\xi}$, parametric inputs $\boldsymbol{\lambda}$, and $\boldsymbol{z}_s(\boldsymbol{\lambda}) = (z_s^{\text{O}}(\boldsymbol{\lambda}), z_s^{\text{CO}}(\boldsymbol{\lambda}))$ coefficient values, hence we do not need to resort to moment comparisons and can examine pointwise parity plots of the true and PCE-predicted outputs.

In Figure 3, we group the parametric PCE coefficients according to species and examine the approximation quality of each coefficient. The close fit of the models to the training data for each coefficient is supported by the computed rRMSE, where we find $E_{\text{rRMSE}} = 3.0 \times 10^{-3}$ and $E_{\text{rRMSE}} = 3.1 \times 10^{-3}$ for $\boldsymbol{z}_s^{\text{O}}$ and $\boldsymbol{z}_s^{\text{CO}}$, respectively.

5.3. Joint PCE and Global Sensitivity Analysis

As discussed in Sections 1 and 2, the joint surrogate is generative in nature, and its quality must be assessed in distribution, requiring sampling of the joint PCE (9). We analyze the joint PCE surrogate predictions by comparing the means and standard deviations of the training data and the PCE-predicted samples. As seen in Figure 4, the tendency of the stochastic PCE to overestimate the predicted variance that was observed above appears to propagate through the joint PCE surrogate, resulting in predicted variance that is also slightly skewed toward larger values.

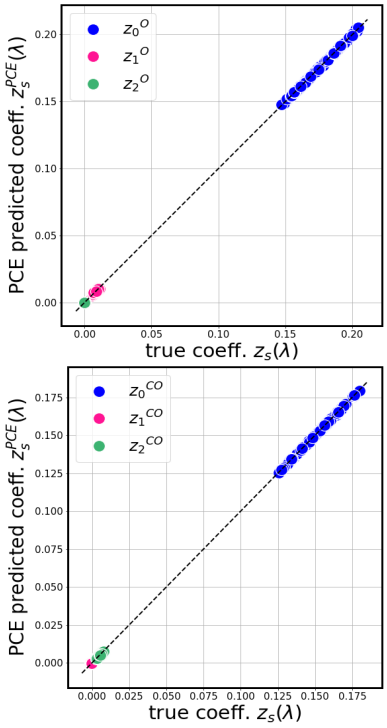

Figure 3: Parity plots of parametric PCE predicted coefficients $\boldsymbol{z}_s^{PCE}(\boldsymbol{\lambda})$ constructed in Eq. (7) vs true coefficients $\boldsymbol{z}_s(\boldsymbol{\lambda})$ found in Eq. (6) for all ensemble members. The coefficients are grouped by species.

Finally, Figure 5 illustrates the PDFs of the full set of training data, with both parametric variation and intrinsic stochasticity, compared to samples obtained from the joint PCE surrogate. It is evident that the joint PCE model serves as an accurate generative approximation to the model.

As a bi-product of the joint PCE construction (9), one can extract global sensitivity indices and distinguish the contribution of intrinsic noise from parametric uncertainty. As illustrated in Figure 6, we find the adsorption rates to be the largest contributors to the output variance. The Sobol index from the intrinsic model noise is essentially a group sensitivity that is found by summing the effects of all $\boldsymbol{\zeta}$. The parametric uncertainties overall seem to dominate the intrinsic noise as one would expect intuitively based on Figure 1.

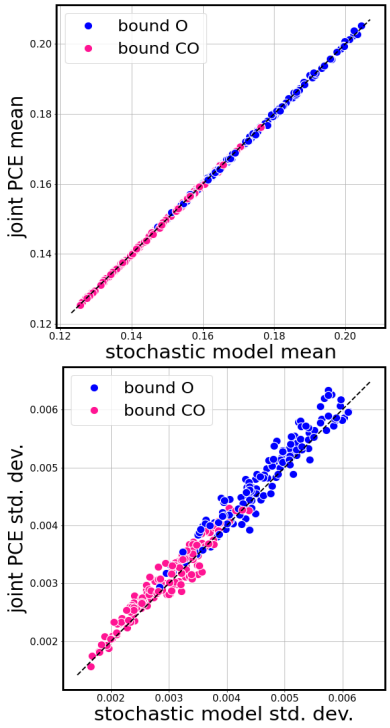

Figure 4: Parity plots of the mean ($E_{rRMSE} = 1.9 \times 10^{-3}$) and standard deviation ($E_{rRMSE} = 5.8 \times 10^{-2}$) of the species coverage fractions predicted by the joint PCE surrogate for all $\boldsymbol{\lambda}$. The statistics are colored by species in the RuO_2 synthetic model: bound O (blue) and bound CO (magenta).

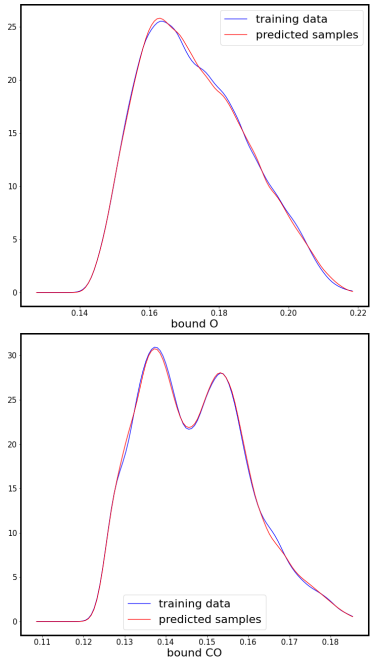

Figure 5: Comparison of probability density functions for each output species. Samples of the training data (blue) are overlayed with samples taken from the joint PCE surrogate (red).

6. Conclusions

We developed a joint PCE model for representing both parametric uncertainty and intrinsic noise of a stochastic model in a compact surrogate form. The construction combines the conventional parametric PCE with a stochastic PCE component constructed with the help of the inverse Rosenblatt transformation. The joint PCE surrogate is generative as the underlying stochastic germ is represented in a way that can be resampled to generate new realizations of the underlying stochastic model.

The joint PCE construction enables analytical extraction of global sensitivity indices via variance decomposition. This allows separating the two sources of uncertainty: intrinsic noise and parametric uncertainty. Decomposing the output variance in this way allows us to assess the impact that intrinsic noise has on the model predictions and can be used to inform the size and quality of the data set used to construct the surrogate. The parametric PCE component of the algorithm, while relatively routine, also includes model selection and basis selection strategies to address the factorial growth of the basis set with the model order and input dimension. We demonstrated the results on a heterogeneous catalysis model example. Extending the joint PCE construction to spatio-temporal random field realizations is the subject of ongoing work.

Acknowledgements: This work is supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Division of Chemical Sciences, Geosciences and Biosciences, as part of the Computational Chemistry Sciences Program. Sandia National Laboratories is a multi-mission 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-NA0003525.

References

[1] M. Andersen, C. Panosetti, and K. Reuter. A practical guide to surface kinetic monte carlo simulations. *Frontiers in Chemistry*, 7(202), 2019.

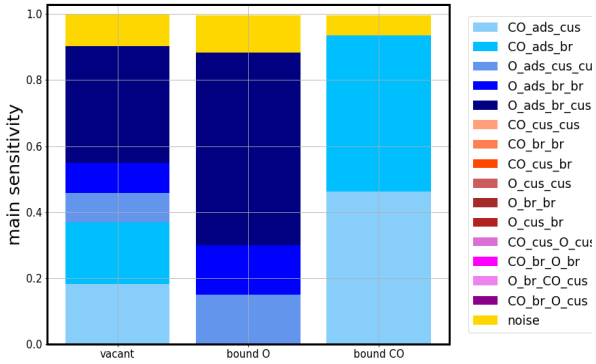

Figure 6: Main sensitivity indices for uncertain parameters and noise in the RuO_2 problem.

[2] S. Babacan, R. Molina, and A. Katsaggelos. Bayesian compressive sensing using Laplace priors. *IEEE Transactions on Image Processing*, 19(1):53–63, 2010.

[3] T. Crestaux, O. Le Maître, and J. Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161–1172, 2009.

[4] S. Döpking and S. Matera. Error progagation in first-principles kinetic monte carlo simulation. *Chemical Physical Letters*, 674:28–32, 2017.

[5] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer Verlag, New York, 1991.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[7] J. L. Hart, A. Alexanderian, and P. A. Gremaud. Efficient computation of sobol' indices for stochastic models. *SIAM Journal on Scientific Computing*, 39(4):A1514–A1530, 2017.

[8] D. P. Kingma and M. Welling. Auto-encoding variational Bayes, 2013.

[9] O. Le Maître and O. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, New York, NY, 2010.

[10] N. Lüthen, S. Marelli, and B. Sudret. Sparse polynomial chaos expansions: Literature survey and benchmark. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):593–649, 2021.

[11] H. Najm. Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, 41(1):35–52, 2009.

[12] D. J. Rezende and S. Mohamed. Variational inference with normalizing flows, 2015.

[13] M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3):470 – 472, 1952.

[14] A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo, E. Commission, and J. R. C. Ispra. Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering & System Safety*, 91(10-11):1109–1125, 2006.

[15] K. Sargsyan. Surrogate models for uncertainty propagation and sensitivity analysis. In . H. O. R. Ghanem, D. Higdon, editor, *Handbook of Uncertainty Quantification*, pages 673–698. Springer, 2017.

[16] K. Sargsyan, B. Debusschere, H. Najm, and O. Le Maître. Spectral representation and reduced order modeling of the dynamics of stochastic reaction networks via adaptive data partitioning. *SIAM Journal on Scientific Computing*, 31(6):4395–4421, 2010.

[17] K. Sargsyan, C. Safta, H. N. Najm, B. J. Debusschere, D. Ricciuto, and P. Thornton. Dimensionality reduction for complex models via Bayesian compressive sensing. *International Journal for Uncertainty Quantification*, 4(1):63–93, 2014.

[18] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

[19] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008.

[20] X. Zhu and B. Sudret. Stochastic polynomial chaos expansions to emulate stochastic simulators. *International Journal for Uncertainty Quantification*, 2023.