The Journal of Open Source Software

UncertainSCI: A Python Package for Noninvasive Parametric Uncertainty Quantification of Simulation Pipelines

Jess Tate[1], Zexin Liu[1,2], Jake A Bergquist[1, 3, 4], Sumientra Rampersad[5, 6], Dan White[1], Chantel Charlebois[1, 3], Lindsay Rupp[1, 3, 4], Dana H Brooks[6], Rob S MacLeod[1, 3], and Akil Narayan[1, 2]

1 Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT, USA **2** Mathematics Department, University of Utah, Salt Lake City, UT, USA **3** Biomedical Engineering Department , University of Utah, Salt Lake City, UT, USA **4** Nora Eccles Cardiovascular Research and Training Institute, University of Utah, Salt Lake City, UT, USA **5** Physics Department, University of Massachusetts, Boston, MA, USA **6** Electrical and Computer Engineering Department, Northeastern University, Boston, MA, USA

DOI: 10.21105/joss.04094

Software

- Review
- Repository
- Archive

Editor: Pending Editor

Submitted: 21 January 2022
Published: 23 January 2022

Summary

We have developed UncertainSCI (*UncertainSCI*, 2020) as an open-source tool designed to make modern uncertainty quantification (UQ) techniques more accessible in biomedical simulation applications. UncertainSCI is implemented in Python with a noninvasive interface to meet our software design goals of 1) numerical accuracy, 2) simple application programming interface (API), 3) adaptability to many applications and methods, and 4) interfacing with diverse simulation software. Using a Python implementation in UncertainSCI allowed us to utilize the popularity and low barrier-to-entry of Python and its common packages and to leverage the built-in integration and support for Python in common simulation software packages and languages. Additionally, we used noninvasive UQ techniques and created a similarly noninvasive interface to external modeling software that can be called in diverse ways, depending on the complexity and level of Python integration in the external simulation pipeline. We have developed and included examples applying UncertainSCI to relatively simple 1D simulations implemented in Python, and to bioelectric field simulations implemented in external software packages, which demonstrate the use of UncertainSCI and the effectiveness of the architecture and implementation in achieving our design goals. Figure Figure 1 illustrates the use of UncertainSCI in computing UQ with modeling pipelines for bioelectricity simulations.

The Journal of Open Source Software

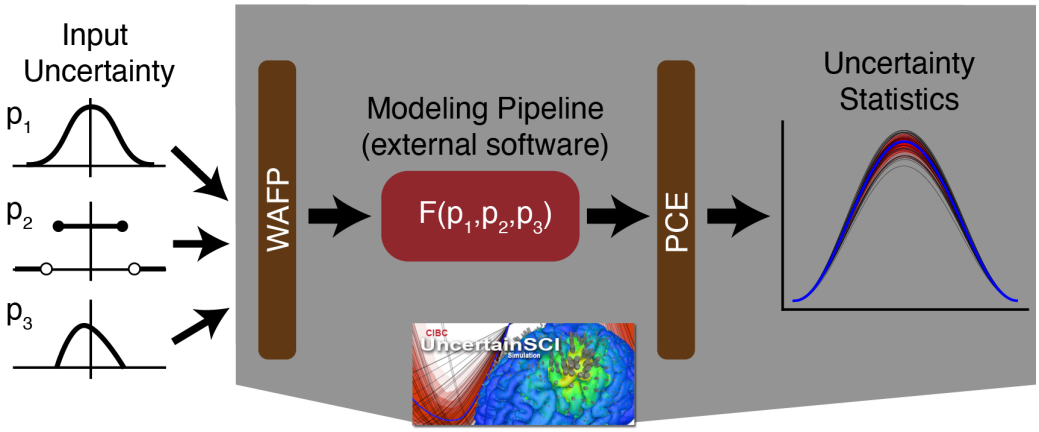

Figure 1: User pipeline for UncertainSCI. After the user inputs parameter distributions, UncertainSCI will compute an efficient sampling scheme. The parameter samples are run through the targeted modeling pipeline, which can be implemented in external software tools. The computed solutions are collected and compiled into relevant statistics with UncertainSCI.

Statement of need

Biomedical computer models include many input parameters that do not have precisely defined values, for example because their value defines physiological processes that are not uniform across patients. As such, any simulation output necessarily has some uncertainty associated with the uncertain value of the input parameter. Exploration and quantification of this model output uncertainty is challenging when more than a single parameter is present; biomedical computer models often have 5-20 such parameters. Quantification of this uncertainty through UQ techniques provides statistics and sensitivity information, a critical component when evaluating the relative impact of parameter variation on the solution accuracy. While the need and importance of UQ in clinical modeling is generally accepted, automated tools for implementing UQ techniques remain evasive for many researchers. UncertainSCI has been use to quantify uncertainty in multiple modeling pipelines, including: cardiac tissue modeling (Bergquist et al., 2021; Rupp et al., 2020, 2021), electrocardiographic (ECG) simulation (Geneser et al., 2005; Swenson et al., 2011), ECG imaging (ECGI) (Tate et al., 2019), transcranial current stimulation (tCS) modeling (Rampersad et al., 2021; Tate et al., 2021), and electrocorticography (ECoG) stimulation (Rampersad et al., 2021; Tate et al., 2021).

Mathematics

In UncertainSCI, we quantify forward parametric uncertainty in cardiac simulations using polynomial chaos expansions (PCE) (Xiu, 2010). PCE approximates the dependence of a quantity of interest (QoI) that is the model output from the forward simulation on a finite number of random parameters via a multivariate polynomial function of those parameters. With u the QoI (scalar-valued for simplicity), and $p \in \mathbb{R}^d$ the vector of uncertain parameters, the PCE approach builds the function u_N, given as,

$$u(p) \approx u_N(p) = \sum_{j=1}^{N} c_j \phi_j(p), \tag{1}$$

where the $\{\phi_j\}_{j=1}^N$ functions are multivariate polynomials, and the coefficients $\{c_j\}_{j=1}^N$ are learned through an ensemble of collected data,

$$\{p_j\}_{j=1}^M \xrightarrow[\text{Forward model } u]{} \{u(p_j)\}_{j=1}^M \xrightarrow[\text{Emulator training}]{} \{c_j\}_{j=1}^N . \tag{2}$$

Typically one seeks convergence of u_N to u in an L^2-type norm weighted by the probability density of the parameters p. The polynomial function u_N constitutes an emulator for QoI u, from which statistics of the QoI, including the mean, variance, and parameter sensitivities, are efficiently computed from the polynomial. UncertainSCI uses a particular type of pseudo-random sampling for the parameter design $\{p_j\}_{j=1}^M$, along with a particular weighted least-squares approach for emulator training, both of which are recently developed advances in high-dimensional approximation. The entire procedure (2) is non-intrusive, since the forward model need only be queried at a particular set of samples.

The efficiency of PCE for analysis of UQ in a forward simulation depends on efficient selection of parameter samples $\{p_j\}_{j=1}^M$. The goal is to use as few samples M as possible while still ensuring that u_N is an accurate emulator for u. UncertainSCI uses a two-step approach to strategically sample the parameter space: 1. A discrete candidate set is generated via random sampling with respect to a ''biased'' probability measure μ, that is distinct from (but related to) the probability distribution of the parameter p. 2. A weighted D-optimal design is sought by subsampling this discrete candidate set. UncertainSCI uses a greedy approach to produce an approximation to such a design.

The probability measure μ that must be randomly sampled is a distribution that exploits a concentration of measurable phenomena to provably increase the quality of the candidate set (Cohen & Migliorati, 2017). Sampling from this distribution when the components of the parameter vector p are independent is computationally efficient, having complexity that is linear in the number of parameters d (Narayan, 2018). The relatively large candidate set generated from this random sampling is pruned via subsampling using a weighted D-optimal design optimization. UncertainSCI's algorithm for this approach approximately computes a weighted D-optimal design via the weighted approximate Fekete points (WAFP) procedure (Burk et al., 2020; Guo et al., 2018), which greedily maximizes a weighted matrix determinant. The result is a geometrically unstructured parameter design of M samples for use in the pipeline (2).

Once the experimental design is created through the WAFP procedure, an ensemble of forward simulations $\{u(p_j)\}_{j=1}^M$ is collected from the simulation software, and UncertainSCI produces a PCE emulator u_N through a (weighted) least-squares procedure. From this emulator, UncertainSCI can compute statistics, sensitivities, residuals, and cross-validation metrics, and can adaptively tune the complexity of the PCE emulator based on a user-prescribed tolerance and/or computational budget.

Acknowledgements

This project was supported by grants from the National Institute of Biomedical Imaging and Bioengineering (U24EB029012) from the National Institutes of Health.

References

Bergquist, J. A., Zenger, B., Rupp, L. C., Narayan, A., Tate, J. D., & MacLeod, R. S. (2021). Uncertainty quantification in simulations of myocardial ischemia. *Computing in Cardiology*, *48*.

Burk, K. M., Narayan, A., & Orr, J. A. (2020). Efficient sampling for polynomial chaos-based uncertainty quantification and sensitivity analysis using weighted approximate fekete

points. *International Journal for Numerical Methods in Biomedical Engineering*, *36*(11), e3395. https://doi.org/https://doi.org/10.1002/cnm.3395

Cohen, A., & Migliorati, G. (2017). Optimal weighted least-squares methods. *SMAI Journal of Computational Mathematics*, *3*, 181–203. https://doi.org/10.5802/smai-jcm.24

Geneser, S. E., Choe, S., Kirby, R. M., & Macleod, R. S. (2005). The influence of stochastic organ conductivity in 2D ECG forward modeling: A stochastic finite element study. *Proceedings of the IEEE Engineering in Medicine and Biology Society 27th Annual International Conference*, 5528–5531.

Guo, L., Narayan, A., Yan, L., & Zhou, T. (2018). Weighted Approximate Fekete Points: Sampling for Least-Squares Polynomial Approximation. *SIAM Journal on Scientific Computing*, *40*(1), A366–A387. https://doi.org/10.1137/17M1140960

Narayan, A. (2018). Computation of induced orthogonal polynomial distributions. *Electronic Transactions on Numerical Analysis*, *50*, 71–97. https://doi.org/10.1553/etna_vol50s71

Rampersad, S., Charlebois, C., Tate, J. D., MacLeod, R. S., Brooks, D. H., & Narayan, A. (2021, May). Quantification of uncertainty due to tissue conductivity variability in simulations of brain stimulation. *10th International IEEE EMBS Conference on Neural Engineering*.

Rupp, L. C., Bergquist, J. A., Zenger, B., Gillette, K., Narayan, A., Tate, J. D., Plank, G., & MacLeod, R. S. (2021). The role of myocardial fiber direction in epicardial activation patterns via uncertainty quantification. *Computing in Cardiology*, *48*.

Rupp, L. C., Liu, Z., Bergquist, J. A., Rampersad, S., White, D., Tate, J. D., Brooks, D. H., Narayan, A., & MacLeod, R. S. (2020). Using UncertainSCI to quantify uncertainty in cardiac simulations. *2020 Computing in Cardiology*, 1–4. https://doi.org/10.22489/CinC.2020.275

Swenson, D. J., Geneser, S. E., Stinstra, J. G., Kirby, R. M., & MacLeod, R. S. (2011). Cardiac position sensitivity study in the electrocardiographic forward problem using stochastic collocation and BEM. *Annals of Biomedical Engineering*, *30*(12), 2900–2910.

Tate, J. D., Good, W. W., Zemzemi, N., Boonstra, M., Dam, P. van, Brooks, D. H., Narayan, A., & MacLeod, R. S. (2019). Uncertainty quantification of the effects of segmentation variability in ECGI. In *Functional imaging and modeling of the heart* (pp. 515–522). Springer-Cham. https://doi.org/https://doi.org/10.1007/978-3-030-78710-3_49

Tate, J. D., Rampersad, S., Charlebois, C., Liu, Z., Bergquist, J. A., White, D., Rupp, L. C., Brooks, D. H., Narayan, A., & MacLeod, R. S. (2021). Uncertainty quantification in brain stimulation using UncertainSCI. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *14*(6), 1659–1660. https://doi.org/10.1016/j.brs.2021.10.226

UncertainSCI. (2020). https://www.sci.utah.edu/cibc-software/uncertainsci.html

Xiu, D. (2010). *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press. ISBN: 0-691-14212-2