Koi-utils: a Framework for Uncertainty Quantification in Cardiac Simulation

Eli Child[1,2], Asher Merrill[1,3], Matthias AF Gsell[4,5], Anton J Prassl[5], Gernot Plank[4,5], Akil Narayan[1,2], Karli Gillette[1,2]

[1] Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, USA,
[2] Department of Biomedical Engineering, University of Utah, Salt Lake City, USA,
[3] Department of Mathematics, University of Utah, Salt Lake City, USA,
[4] BioTechMed-Graz, Graz, Austria,
[5] Gottfried Schatz Research Center, Medical University of Graz, Graz, Austria

Abstract

Cardiac digital twins (CDTs) for electrophysiology (EP) show promise for predicting cardiac function and response to therapy. However, their predictions depend on uncertain model inputs and assumptions. Uncertainty quantification (UQ) is therefore needed to determine how these uncertainties affect predicted cardiac behavior. Existing UQ tools are not integrated with cardiac-specific experiment preparation and simulation, leaving investigators to assemble disjointed, study-specific workflows.

We present koi-utils*, an open-source workflow for forward parametric UQ in cardiac EP simulation. It reproducibly integrates parameter sampling, plausibility screening, ForCEPSS experiment preparation, openCARP simulation, biomarker extraction, surrogate modeling and evaluation, and global sensitivity analysis (GSA).*

We demonstrate the complete workflow in a fixed-S2 arrhythmia-inducibility study using a standardized ventricular-tissue sheet model. From prescribed parameter distributions, koi-utils *automatically completed the analysis, and its emulators reproduced simulator-derived biomarkers in an independent validation set.* Koi-utils *provides a practical, reproducible basis for incorporating UQ into cardiac modeling studies.*

1. Introduction

Cardiac digital twins (CDTs) for electrophysiology (EP) show promise for patient-specific prognostics and arrhythmia therapy planning [1]. Their predictions, however, depend on model inputs that are incompletely observed, estimated from heterogeneous data, or deliberately simplified. Uncertainty quantification (UQ) is therefore needed to determine how these uncertainties affect predicted EP behavior. Beyond statistical methods, UQ of cardiac EP requires cardiac-specific experiment preparation, simulation, and biomarker calculation. General-purpose UQ libraries do not integrate these tasks with established cardiac simulation software, leaving disjointed, study-specific workflows that limit reproducibility and hinder UQ adoption.

We therefore present *koi-utils*, an open-source workflow for forward parametric UQ in cardiac EP simulation. It integrates parameter sampling and plausibility screening, ForCEPSS experiment preparation [2], openCARP or CARP-EP (NumeriCor GmbH, Graz, Austria) simulation, biomarker extraction, Gaussian process (GP) emulation and independent evaluation, and global sensitivity analysis (GSA) (Figure 1). Consistent stage inputs and outputs support reproducibility and study-specific modification of individual methods. We demonstrate the workflow in a geometrically simplified, fixed-S2 arrhythmia-inducibility study inspired by Virtual Arrhythmia Risk Prediction (VARP) [2, 3].

2. Methods

Koi-utils conducts a cardiac EP study from prescribed parameter distributions to uncertainty and sensitivity estimates (Figure 1). It coordinates sampling and plausibility screening, ForCEPSS experiment preparation and prepacing, openCARP or CARP-EP simulation, biomarker extraction, surrogate modeling, independent evaluation, and sensitivity analysis. A study requires a cardiac mesh, a JSON experiment plan containing model settings and parameter distributions, and a YAML workflow file. The plan is structured according to a schema introduced by ForCEPSS and maintained in the open-source `carputils` framework [2].

We demonstrated the workflow in a small-scale yet physiologically complex arrhythmia-inducibility study using the geometrically simplified ForCEPSS sheet model and inspired by VARP [2, 3]. The study varied cellular conductances, tissue conduction velocities (CVs), pacing

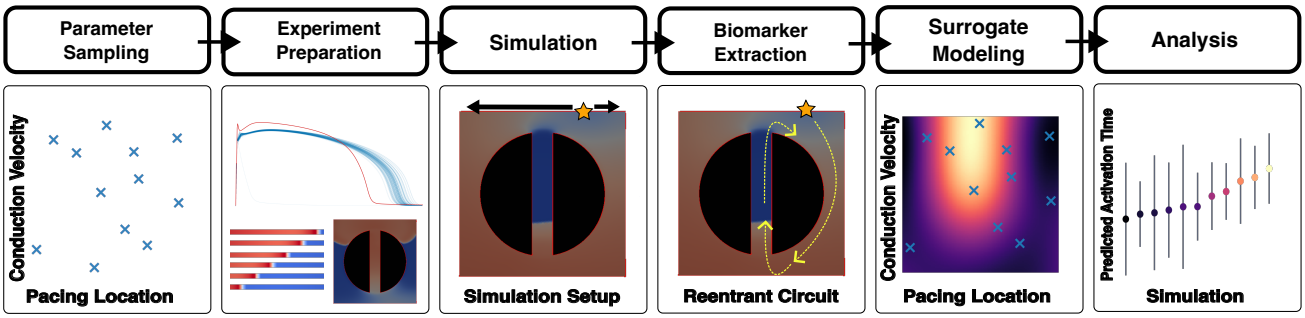

Figure 1. The *koi-utils* workflow, from parameter sampling and experiment preparation through simulation, surrogate modeling, and sensitivity analysis.

location, and S1–S2 coupling interval. GP emulators were trained and used for GSA. *Koi-utils* and the example study will be available at `github.com/Gillette-Lab/koi-utils/`.

2.1. Example Study and Inputs

Our example study represents a standard cardiac EP setting in which small changes in tissue, cellular, and pacing properties can alter arrhythmia inducibility. Using Mesher, we generated a 40×40 mm planar sheet with 12,009 nodes and 23,616 triangular elements at nominal 400 μm resolution and fibers aligned along the x-axis. Normal myocardium surrounded a passive scar traversed by a viable isthmus and a compromised border zone (Figure 2). S1 and S2 stimuli shared a variable location along the top edge. Normal myocardium used the default ten Tusscher–Panfilov settings. In the border zone and isthmus, ionic-conductance scaling factors varied by approximately $\pm 20\%$ around the VARP infarct-remodeling values [3].

2.2. Parameter Sampling

Koi-utils samples all parameter distributions specified in the JSON experiment plan. The workflow file specifies the method, sample count, and random seed; sampled values are written to `sampling.csv`. Latin hypercube sampling (LHS), Monte Carlo, and Saltelli designs are supported.

For the example study, we generated 5,000 LHS samples of nine independent uniform parameters; 100 Monte Carlo samples provided independent validation. Healthy-tissue longitudinal CV ranged from 0.7 to 0.9 m/s. Border-zone and isthmus longitudinal CV ranged from 0.15 to 0.30 m/s, with a shared off-axis CV of 0.075–0.15 m/s. In these regions, maximal conductances were scaled relative to model defaults by 0.30–0.46 for G_{Na}, 0.25–0.37 for G_{CaL}, 0.24–0.36 for G_{Kr}, and 0.16–0.24 for G_{Ks}. Pacing position spanned the top edge, and the S1–S2 coupling interval ranged from 300 to 360 ms.

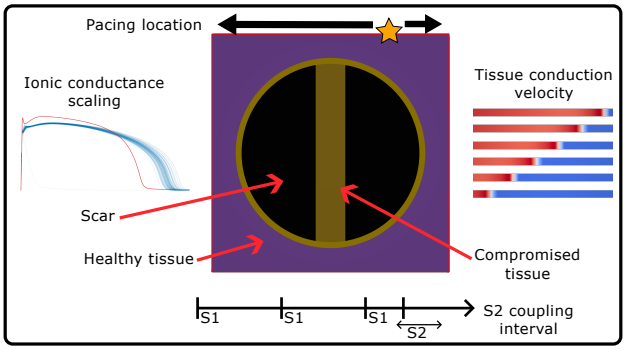

Figure 2. VARP-inspired example study. Normal myocardium surrounds a passive scar traversed by a viable isthmus and compromised border zone. Co-located S1 and S2 stimuli are applied at a sampled top-edge position; annotations indicate the varied conduction velocities, ionic conductances, and coupling interval.

2.3. Plausibility Screening

Plausibility screening prevents common input and model-setup errors from propagating through many costly cardiac simulations. *Koi-utils* checks prescribed physiological and numerical limits, including CV magnitude and directional ordering, time-step and stimulus settings, and mesh resolution and quality. Severity levels help investigators select samples for exclusion or review while retaining the original sampled design.

Screening identified off-axis CVs in the example study below the prescribed physiological limit of 0.1 m/s in approximately 1,700 of 5,000 training samples and 31 of 100 validation samples.

2.4. Experiment Preparation

Before application of a study protocol, each parameter sample requires cardiac-specific experiment prepara-

tion. Using standardized ForCEPSS, *koi-utils* applies cellular EP initialization, tissue-conductivity calibration, and tissue-scale limit-cycle pacing to each sample [2].

For the example, preparation used a 600 ms basic cycle length, 100 cycles for cellular initialization and tissue prepacing, and conductivity calibration for the 400 μm mesh and numerical configuration.

2.5. Simulation

The simulation stage executes prepared cardiac runs with either openCARP [4] or CARP-EP and retains outputs for post-processing. The example used CARP-EP, restored each ForCEPSS-prepared state, recorded all -40 mV activation-threshold crossings, and terminated after 40 ms without activation or at the specified end time.

2.6. Biomarker Extraction

Simulation outputs generally require reduction to physiologically interpretable biomarkers before modeling and analysis. *Koi-utils* provides standard biomarker calculations and accepts study-specific biomarker calculations; extracted values are collected in `biomarkers.csv`.

For the example, standard calculations provided the S1 total activation time, defined as the activation-time range, and activation-time standard deviation. Sustained activity required threshold-crossing activity without a 40 ms quiescent interval for 1,000 ms after S2. S2 capture required activation of at least one node outside the directly stimulated region within 60 ms after S2; the S2-propagated tissue fraction quantified tissue subsequently activated by that wave. A study-specific third-activation recurrence biomarker identified whether any node crossed the activation threshold at least three times, indicating recurrent activation but not the topology of a reentrant circuit.

2.7. Surrogate Modeling

GP emulators approximate the relationship between the nine uncertain inputs and each biomarker, evaluating cardiac model responses with substantially fewer simulations while retaining predictive uncertainty. *Koi-utils* fits normalized GP regression for each continuous biomarker and latent GP classification for each binary outcome. Regressors return posterior summaries in the biomarker's original units; classifiers return predicted event probabilities. All input parameters and continuous biomarker values were standardized before fitting. For continuous biomarkers, the workflow file specified fixed values for the kernel length scale and noise. For binary biomarkers, *koi-utils* optimized the kernel length scale and amplitude within specified ranges and repeated the optimization from two additional starting points to reduce dependence on the initial values. The specified and fitted values were recorded with the models and included in the independent evaluation. The emulators were fitted to all 5,000 training samples across the rectangular parameter domain; independent validation assessed predictions at the 69 points retained within the feasible region after screening. Implementations use UncertainSCI and scikit-learn [5, 6].

2.8. Analysis

The final stage assesses emulator performance and parameter importance. Continuous predictions are evaluated by posterior-mean error and interval coverage; binary predictions by Brier score and area under the receiver operating characteristic curve (AUC).

For GSA, surrogate-based Sobol indices quantified the variation in each predicted biomarker attributable to each input alone (first-order) and including interactions with other inputs (total-order). The analysis used posterior means for continuous biomarkers and predicted event probabilities for binary outcomes across the rectangular parameter domain defined by uniform distributions in the experiment plan. Indices were calculated with SALib [7].

3. Results

Koi-utils automatically generated samples, prepaced the cardiac models with ForCEPSS, ran simulations, extracted biomarkers, fitted and independently evaluated the GP emulators, and calculated GSA indices in the example setup. Using ForCEPSS and CARP-EP demonstrated compatibility with established CARP-based studies, while consistent inputs and outputs enable reproducible modification or extension of individual stages. Independent evaluation showed that the emulators reproduced continuous activation measures and distinguished binary arrhythmia outcomes sufficiently for sensitivity analysis.

The GP emulators were fitted using all 5,000 training samples and independently evaluated on the 69 validation cases retained after screening, all of which had complete biomarker values.

For S2 capture, sustained activity, and third-activation recurrence, the classifiers achieved ROC AUCs of 0.933–0.997 and Brier scores of 0.033–0.076 in independent validation. For S1 total activation time, activation-time standard deviation, and S2-propagated tissue fraction, posterior means explained 81.9–93.5% of observed variance ($R^2 = 0.819$–0.935), and 95% posterior intervals contained 95.7–100% of validation measurements.

The S1–S2 coupling interval was the dominant input for S2 capture, sustained activity, third-activation recurrence, and S2-propagated tissue fraction, with total-order Sobol indices of 97.8%, 65.6%, 93.9%, and 94.7%, respectively. Border-zone/isthmus off-axis CV dominated S1 total ac-

tivation time and activation-time standard deviation, with total-order indices of 73.6% and 72.4%, respectively. For sustained activity, the difference between first- and total-order indices (36.3% versus 65.6%) indicated interactions with other inputs.

4. Discussion

Koi-utils makes UQ more practical for cardiac EP modelers by coordinating steps from parameter sampling through simulation and analysis. General-purpose UQ packages provide sampling, surrogate modeling, and sensitivity analysis, but not cardiac-specific experiment preparation and simulation. *Koi-utils* incorporates ForCEPSS experiment preparation [2], openCARP or CARP-EP simulation, biomarker extraction, GP modeling and evaluation, and GSA. By reducing study-specific scripting and consistently recording inputs and settings while retaining established model definitions and simulation tools, *koi-utils* can broaden the use and improve the reproducibility of cardiac EP UQ.

The example study produced the expected qualitative results. The S1–S2 coupling interval dominated S2 capture, S2-propagated tissue fraction, and third-activation recurrence, consistent with the dependence of premature-stimulus responses on recovery from the preceding activation. Border-zone/isthmus off-axis CV dominated S1 total activation time and activation-time standard deviation, consistent with the top-edge pacing location and overall slower CV in the compromised tissue. For sustained activity, the substantially larger total-order than first-order index for coupling interval indicated contributions both directly and through interactions with other inputs. Independent validation showed that the emulators captured these continuous and binary responses sufficiently well for the demonstrated sensitivity analysis.

Current limitations concern workflow flexibility and the scope of the example study. Adaptive sampling is not yet supported, and preparation results are not reused when samples share the relevant cellular, tissue, and pacing inputs, increasing computational cost. Continuous-emulator settings and the kernel form and optimization ranges for binary biomarkers were selected for this example and may require adjustment in other studies. The fixed-S2 example differs from the full VARP protocol and clinical pacing workflows. Third-activation recurrence does not establish reentrant-circuit topology. The Sobol indices are conditional on the selected independent parameter distributions and fitted emulators, do not propagate emulator uncertainty, and are not clinical risk estimates.

Future development will prioritize model-guided adaptive sampling, reuse of equivalent prepacing states, and high-performance computing integration. Support for correlated input distributions and model-discrepancy analysis would broaden the range of uncertainty analyses addressed by the workflow. These developments would further reduce the effort and computational cost of cardiac EP UQ and support risk-aware analysis as CDTs advance toward clinical use.

Acknowledgments

This research was funded by the National Science Foundation (NSF) [grant 24-559], awarded jointly to Dr. Karli Gillette and Dr. Akil Narayan. This research was also funded in whole or in part by the Austrian Science Fund (FWF) [grant 10.55776/ESP592] awarded to Dr. Karli Gillette.

References

[1] Bhagirath P, Strocchi M, Bishop MJ, Boyle PM, Plank G. From bits to bedside: entering the age of digital twins in cardiac electrophysiology. Europace 2024;26(12):euae295.

[2] Gsell MA, Neic A, Bishop MJ, Gillette K, Prassl AJ, Augustin CM, Vigmond EJ, Plank G. Forcepss—a framework for cardiac electrophysiology simulations standardization. Computer Methods and Programs in Biomedicine 2024; 251:108189.

[3] Arevalo HJ, Vadakkumpadan F, Guallar E, Jebb A, Malamas P, Wu KC, Trayanova NA. Arrhythmia risk stratification of patients after myocardial infarction using personalized heart models. Nature communications 2016;7(1):11437.

[4] Plank G, Loewe A, Neic A, Augustin C, Huang YL, Gsell MA, Karabelas E, Nothstein M, Prassl AJ, Sánchez J, Seemann G, Vigmond E. The opencarp simulation environment for cardiac electrophysiology. Computer Methods and Programs in Biomedicine 2021;208:106223.

[5] Narayan A, Liu Z, Bergquist JA, Charlebois C, Rampersad S, Rupp L, Brooks D, White D, Tate J, MacLeod RS. Uncertainsci: Uncertainty quantification for computational models in biomedicine and bioengineering. Computers in Biology and Medicine 2023;152:106407.

[6] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 2011;12:2825–2830.

[7] Iwanaga T, Usher W, Herman J. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. Socio Environmental Systems Modelling May 2022;4:18155. URL `https://sesmo.org/article/view/18155`.

Address for correspondence:

Karli Gillette
72 S Central Campus Dr, Salt Lake City, UT, USA 84112
karli.gillette@utah.edu