SimCenter Community Roundtable
"Surrogate modeling of site, building, and bridge system performance in regional earthquake simulation"

October 10, 2024

This SimCenter Community Roundtable meeting was organized by the **Working Group on Regional Simulation of Earthquakes**.

By reducing the computational burden that would otherwise be required using detailed analyses, data-driven surrogate models offer unprecedented opportunities to incorporate high-resolution behavior of local site and structural system response in regional simulations. This community roundtable will focus on the question, "How to create useful and usable surrogate models for earthquake engineering – status quo, limitations, and opportunities." The roundtable will feature three speakers to prompt discussion. Topics to be addressed include (1) illustrative use cases to define key features and applications of surrogate models, (2) a framework for incorporating high-dimension surrogate models in regional simulations, (3) best practices for developing and sharing surrogate models, and (4) assessing the additional uncertainty introduced into regional simulations by surrogate models.

Host: Gregory Deierlein, Stanford University

Presentations and Key Ideas

1. **"Mechanics-informed machine learning for geospatial modeling of soil liquefaction: global surrogate models for simulation and near-real-time response"**
   Presenter: Brett Maurer, University of Washington.
   This presentation described mechanics-informed decision-tree surrogate models to predict liquefaction probability based on geospatial ground features. Prof. Maurer described high-dimensional inputs, combining local and regional characteristics, and combining physics-based and machine-learning models.

2. **"Surrogate Modeling for Regional Seismic Risk-Based Assessment of Building Inventories"**
   Presenter: Henry Burton, UCLA.
   This presentation applied surrogate models (XG-Boost) for efficient regional seismic risk assessment to predict high-resolution building performance with mid-resolution computation cost. Prof. Burton discussed different output choices (i.e. structural behavior or consequence) and validation measures in different spatial scales.

3. **"Surrogate Modeling for Efficient Regional Seismic Fragility Assessment of Bridge Infrastructure through Active Learning and Deep Learning"**
   Presenter: Tim Xie, McGill University.
   This presentation applied (Gaussian Process) surrogate modeling of individual bridge-specific fragility models to reduce computation effort for regional fragility assessment of transportation systems. Prof. Xie discussed the importance of capturing output uncertainty and applied active learning approaches to improve training speed.

Discussion Highlights

- Difference between ML/surrogate: ML is a tool that can be applied to develop surrogate models, but the term "surrogate" primarily emphasizes replacing detailed computational models with simpler computationally efficient models ranging from regression to data-driven ML models.

- What to consider when choosing a surrogate: (1) Training data size, e.g., Decision tree or GP for small data, XGBoost or generative models for extensive data, (2) Ability to capture randomness in the system (e.g., GP, XGBoost) (3) Purpose, e.g., trade-off between reliability (accuracy) and simplicity (general applicability).
- Discussion on the general applicability of surrogate: (1) Trade-off exists between resolution and broader applicability, (2) Applicability depends on the required quality of the output, e.g., resolution, fidelity, aggregation-scale, etc., (3) General applicability can be improved by adding a final step of combining global model with the local information, e.g., opportunities in transfer learning, and (4) Tools to help automate training of surrogate models would facilitate their adoption and use.
- For the regional structures portfolio, should we surrogate across archetypes or per archetype? One needs to consider both computational efficiency and accuracy. Understanding how different the behaviors are across archetypes and preventing extrapolations is key. A physics-informed implementation may be desired for surrogating across archetypes.
- Choosing input and outputs of surrogate model: Regional workflow allows different entry and exit points of surrogates. It is desirable to surrogate structural responses directly in principle, but there can be other choices.
- Data sharing: The sharing of accompanying model and data should be a "must" in an engineering publication
- Notes to SimCenter: (1) Provide tools to reduce the effort in collecting real-world inventory data, e.g., bridge class; (2) Better streamline surrogate workflow inside SimCenter tools, e.g., from the EE-UQ to R2D simulations; (3) Play a role in educating users and preventing abuse of ML/data science techniques in the community; (4) A key strength of the SimCenter workflow is its commitment to preserving the model "resolution" while avoiding over-generalization.

More Information

Additional SimCenter Community Roundtable meetings can be found at https://simcenter.designsafe-ci.org/collaborate/scr/.

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2131111.