UCLA Samueli School of Engineering

Machine Learning in Infrastructure Risk and Resilience Assessment: A Constructive Critique

Henry V. Burton, PhD, S.E.,

Associate Professor and Presidential Chair in Structural Engineering

University of California, Los Angeles

NHERI SimCenter Machine Learning Training in Natural Hazards Engineering

September 26, 2022

Overview

- Machine Learning Applications in Natural Hazards Engineering (ML-NHE): A Timeline
- High Level Basics of Machine Learning
- Supervised Machine Learning: Theory, Intuition and Model Development Workflow
- Types of Supervised Machine Learning Algorithms
- When to Consider ML Applications in NHE

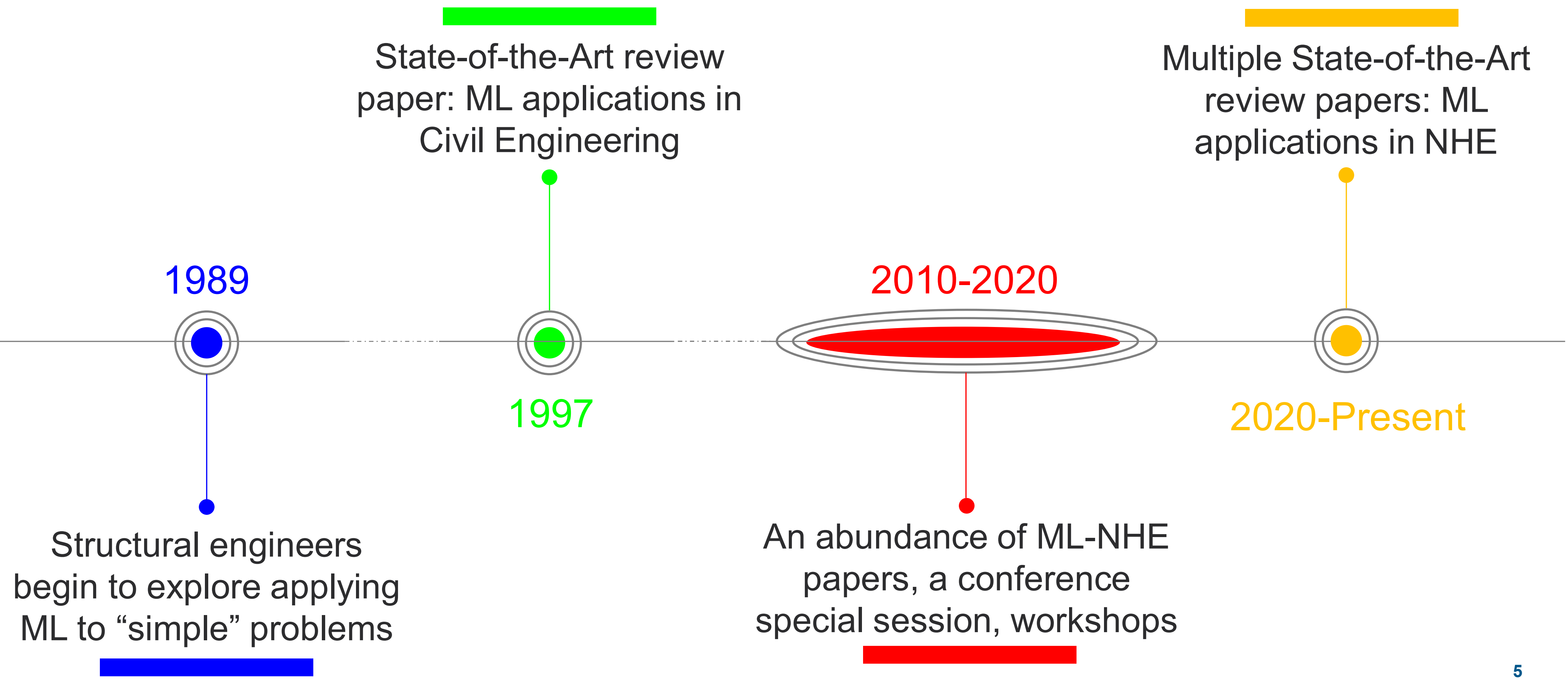
Overview

- Opportunities and Challenges: The Next Frontier in ML-NHE
- Final Thoughts

Overview

- Machine Learning Applications in Natural Hazards Engineering (ML-NHE): A Timeline
- High Level Basics of Machine Learning
- Supervised Machine Learning: Theory, Intuition and Model Development Workflow
- Types of Supervised Machine Learning Algorithms
- When to Consider ML Applications in NHE

ML Applications in Natural Hazards Engineering: A Timeline



State-of-the-Art Review Papers

- Reich, Y. (1997). Machine learning techniques for civil engineering problems. *Computer-Aided Civil and Infrastructure Engineering*, 12(4), 295-310.

- Xie, Y., Ebad Sichani, M., Padgett, J. E., & DesRoches, R. (2020). The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. *Earthquake Spectra*, 36(4), 1769-1801.

- Sun, W., Bocchini, P., & Davison, B. D. (2020). Applications of artificial intelligence for disaster management. *Natural Hazards*, 103(3), 2631-2689.

- Sun, H., Burton, H. V., & Huang, H. (2021). Machine learning applications for building structural design and performance assessment: State-of-the-art review. *Journal of Building Engineering*, 33, 101816.

State-of-the-Art Review Papers

- Mostafa, K., Zisis, I., & Moustafa, M. A. (2022). Machine Learning Techniques in Structural Wind Engineering: A State-of-the-Art Review. *Applied Sciences*, 12(10), 5232.

- Wang, X., Mazumder, R. K., Salarieh, B., Salman, A. M., Shafieezadeh, A., & Li, Y. (2022). Machine Learning for Risk and Resilience Assessment in Structural Engineering: Progress and Future Trends. *Journal of Structural Engineering*, 148(8), 03122003.

Overview

- Machine Learning Applications in Natural Hazards Engineering (ML-NHE): A Timeline
- High Level Basics of Machine Learning
- Supervised Machine Learning: Theory, Intuition and Model Development Workflow
- Types of Supervised Machine Learning Algorithms
- When to Consider ML Applications in NHE

Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL)

Artificial intelligence

The ability of a machine to perform cognitive functions associated with the human mind, including perceiving, reasoning, learning, and problem solving.

Machine learning

A branch of artificial intelligence that uses algorithms to find patterns in large datasets and make predictions about the future, enabling machines to learn without receiving explicit programming instruction. The three major types of ML include **supervised learning, unsupervised learning**, and reinforcement learning.

Deep learning

A branch of machine learning that uses artificial neural networks to ingest vast amounts of data (i.e., big data), often producing more accurate results than traditional ML approaches.

Categories of Machine Learning

MAJOR TYPES OF MACHINE LEARNING		
Supervised learning	**Unsupervised learning**	**Reinforcement learning**
Uses training data and feedback from humans to detect patterns and predict next values *Example*: Predicting the drift capacity of reinforced concrete walls	Explores unlabeled input data without being provided an explicit output variable to detect patterns *Example*: Creating subgroups a set of buildings based on structural similarities.	Learns to detect patterns or perform a task by trying to maximize the rewards it receives for its actions *Example*: Finding the sequence of pipe repairs in an earthquake-damaged water network that minimizes the cumulative loss of service.

Categories of Supervised Learning

TYPES OF SUPERVISED LEARNING	
Regression The outputs or response variables are continuous *Examples*: • Drift capacity of reinforced concrete walls • Drift demands in steel moment frames • Collapse capacity of woodframe buildings	**Classification** The outputs or response variables are categorical *Example*: • The damage state of a structure • The failure mode of a reinforced concrete wall .

Overview

- Machine Learning Applications in Natural Hazards Engineering (ML-NHE): A Timeline
- High Level Basics of Machine Learning
- Supervised Machine Learning: Theory, Intuition and Model Development Workflow
- Types of Supervised Machine Learning Algorithms
- When to Consider ML Applications in NHE

General Supervised Learning Formulation

n is the number of training data samples

$\phi(\cdot)$ is the loss measure between the true and predicted value of the response variable.

$f(\boldsymbol{x}_i;\theta)$: predicted response from the ML model based on the feature $\boldsymbol{x}_i$

$$\underbrace{\underset{\theta,\lambda}{argmin}}\frac{1}{n}\sum_{i=1}^{n}\phi\Big(\underbrace{y_i},\overbrace{f(\boldsymbol{x}_i;\theta)}\Big)+\underbrace{\lambda\Omega(\theta)}$$

$\lambda\Omega(\theta)$: regularization term that penalizes the model based on its complexity through some regularizing function $\Omega(\theta)$

θ is the set of model parameters

Optimize θ and λ to minimize the total loss

y_i: response variable for observation i

Ordinary Least Squares (OLS) as a Supervised Learning Model

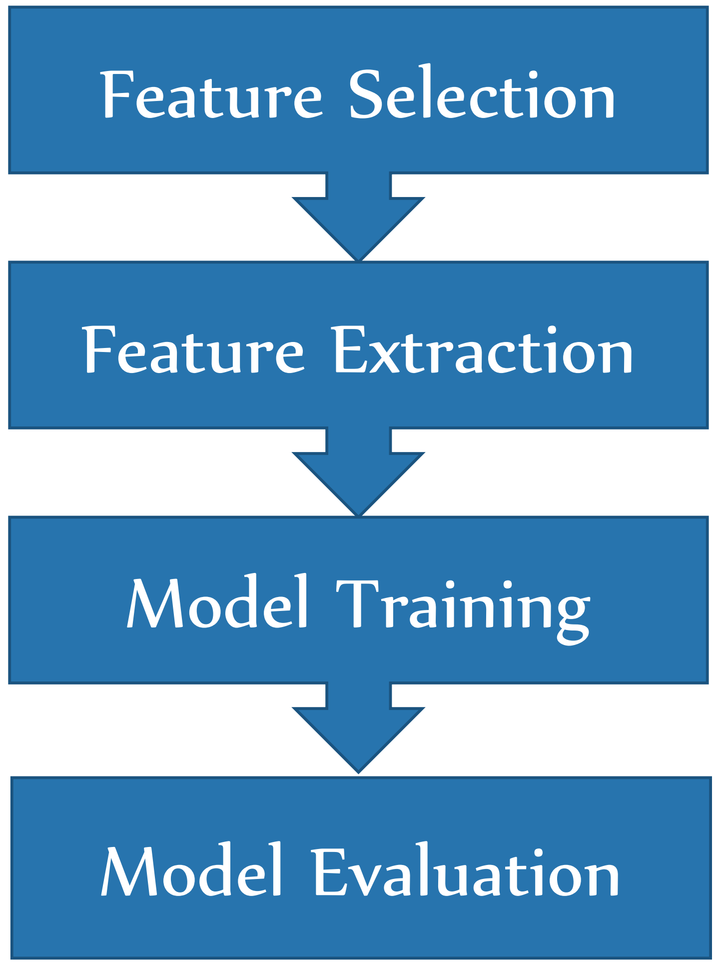
In ordinary least squares (OLS) regression, which is arguably the simplest form of supervised learning, the residual sum of squares (RSS) serves as the loss function. model)

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{\epsilon} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \qquad \underbrace{\boldsymbol{RSS} = (\boldsymbol{y} - \boldsymbol{X\beta})^{\boldsymbol{T}}\,(\boldsymbol{y} - \boldsymbol{X\beta})}_{\text{Loss function } (\phi(\cdot)) \text{ in OLS}}$$

$$\underbrace{\widehat{\boldsymbol{\beta}}_{OLS} = \underset{\beta}{argmin}(RSS) = \left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}}_{\text{Parameters to be optimized } (\theta) \text{ in OLS}}$$

- $\widehat{\boldsymbol{\beta}}_{OLS}$: the predictor coefficients
- $\boldsymbol{y}$ is the observed response variables and $\boldsymbol{X}$ is the feature matrix
- $\boldsymbol{\epsilon}$ is a vector of residuals

Supervised Machine Learning Workflow



Overview

- Machine Learning Applications in Natural Hazards Engineering (ML-NHE): A Timeline
- High Level Basics of Machine Learning
- Supervised Machine Learning: Theory, Intuition and Model Development Workflow
- Types of Supervised Machine Learning Algorithms
- When to Consider ML Applications in NHE

Types of Supervised ML Algorithms: Regularized Least Squares

- Least Absolute Shrinkage and Selection (LASSO)
- Ridge Regression
- Elastic Net
- Kernel formulation (applicable to all of the above)

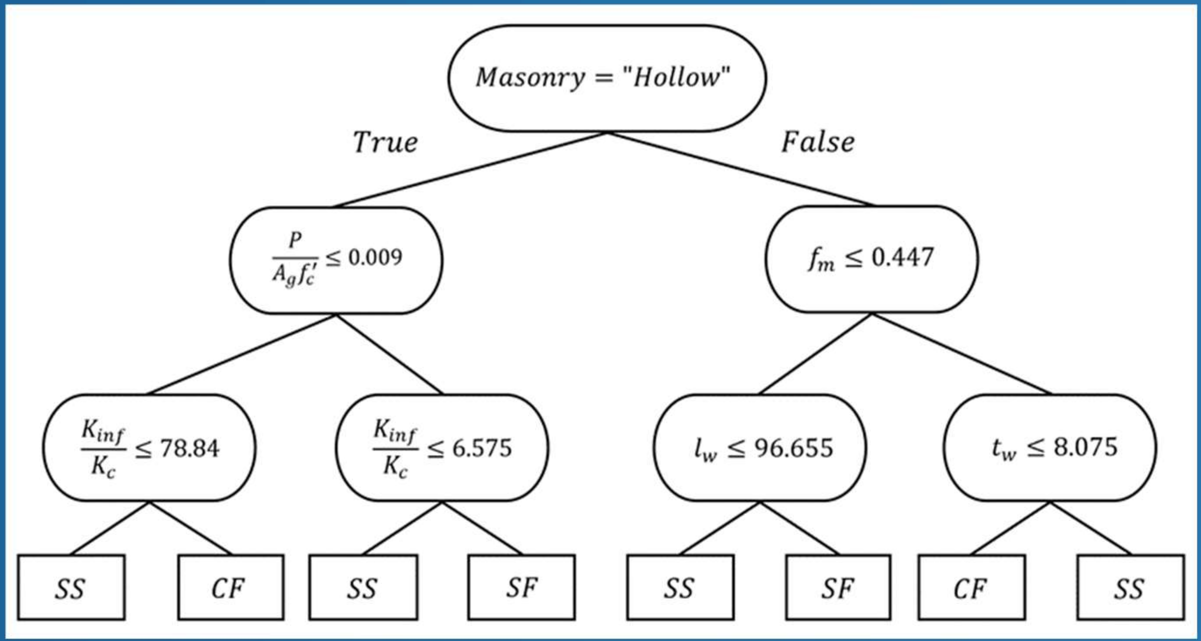
$$\underset{\theta,\lambda}{argmin}\frac{1}{n}\sum_{i=1}^{n}\phi\big(y_i, f(\boldsymbol{x}_i;\theta)\big)+\lambda\Omega(\theta)$$

$$\underset{\beta}{argmin}\frac{1}{n}\sum_{i=1}^{n}(y_i-\boldsymbol{\beta}\boldsymbol{x}_i)^2$$

$$\underset{\beta,\lambda}{argmin}\frac{1}{n}\sum_{i=1}^{n}(y_i-\boldsymbol{\beta}\boldsymbol{x}_i)^2+\lambda\Omega(\beta)$$

Types of Supervised ML Algorithms: Decision Tree-Based Algorithms

- Decision Trees
- Random Forests
- Extreme Gradient Boosting (XGBoost)
- Adaptive Gradient Bosting

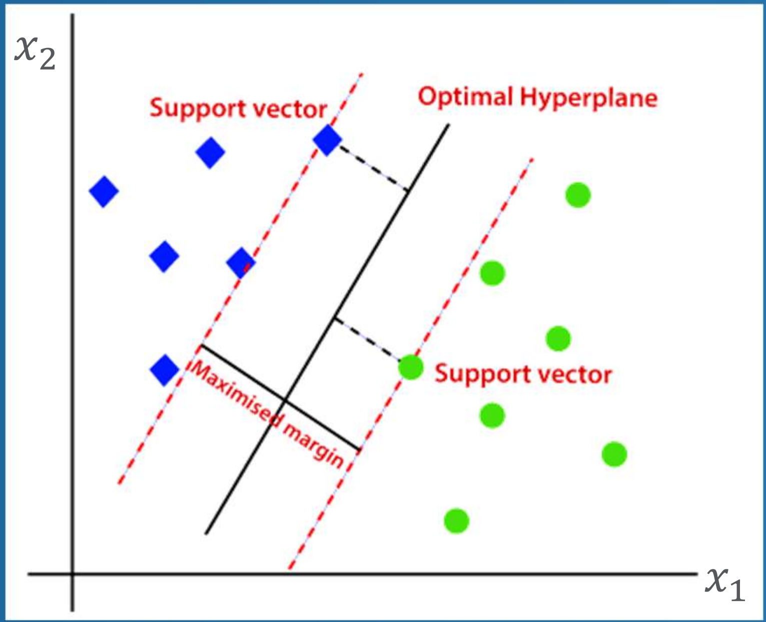


Decision Tree for Classifying Infill Failure Mechanisms

Types of Supervised ML Algorithms: Support Vector Machines

- Support Vector Machines (SVMs) separates hyperplanes

- The output is produced in the form of an optimal hyperplane that categorizes new observations.

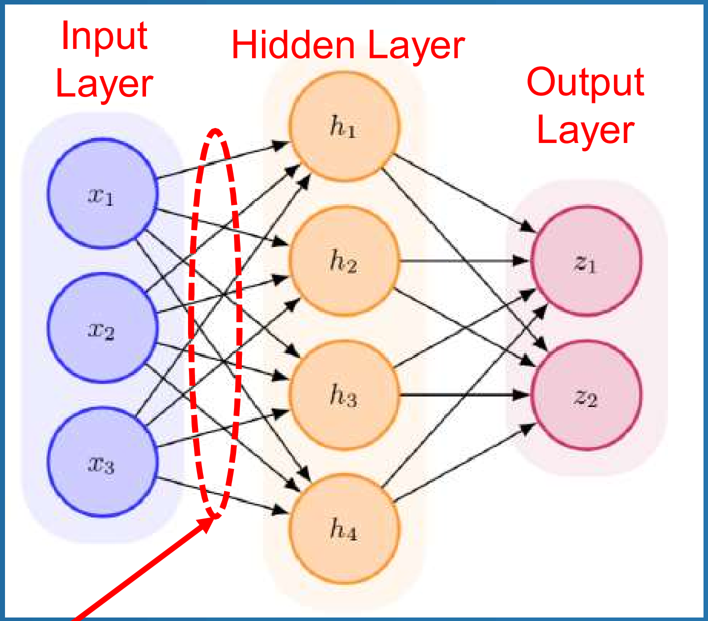


$$\underset{\beta,b}{argmin}\frac{1}{2}\boldsymbol{\beta}\boldsymbol{\beta}^{\boldsymbol{T}} \text{ subject to } |y_i - (\boldsymbol{\beta}\boldsymbol{x}_i - b)| \leq \varepsilon$$

Types of Supervised ML Algorithms: Artificial Neural Networks

- Artificial neural networks (ANNs) operate by recursively (layer-by-layer) applying a series weights ($\boldsymbol{W}_i$) and biases ($\boldsymbol{b}_i$) to the feature matrix ($\boldsymbol{x}_i$)

- The output layer is fed into a loss function which is optimized for prediction.



$\boldsymbol{W}_1$ is the weight matrix

First Layer: $h = f(W_1 x + b_1)$

Second (output) Layer: $z = W_2 h + b_2$

Overview

- Machine Learning Applications in Natural Hazards Engineering (ML-NHE): A Timeline
- High Level Basics of Machine Learning
- Supervised Machine Learning: Theory, Intuition and Model Development Workflow
- Types of Supervised Machine Learning Algorithms
- When to Consider ML Applications in NHE

ML-NHE Applications: When it Makes Sense

- Any problem where you were going to use an empirical model anyway.
- Anything involving extracting information from images and/or text (e.g., retrieval of information from “harsh” conditions using robots).
- Optimizing the sequencing of activities such as post-event inspections or scheduled maintenance (reinforced learning).
- Any complex and computationally expensive multi-step workflow that has stochasticity

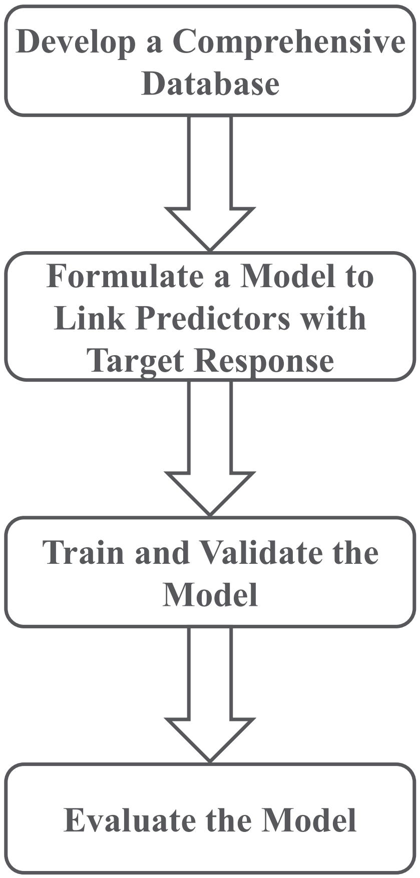
ML-NHE Applications: When it Makes Sense

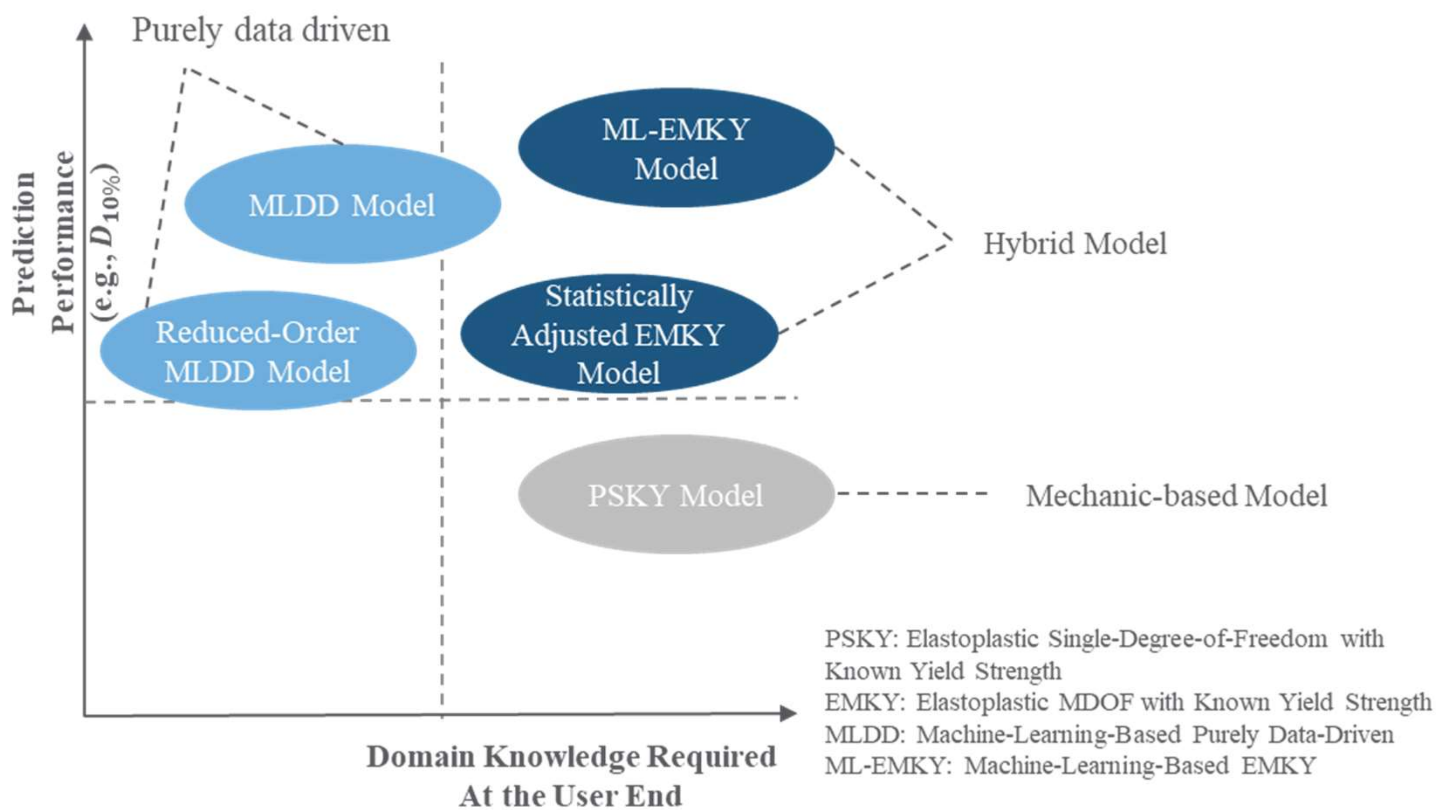
- If it allows you to acquire new knowledge about the problem at hand.

Overview

- Opportunities and Challenges: The Next Frontier in ML-NHE
- Final Thoughts

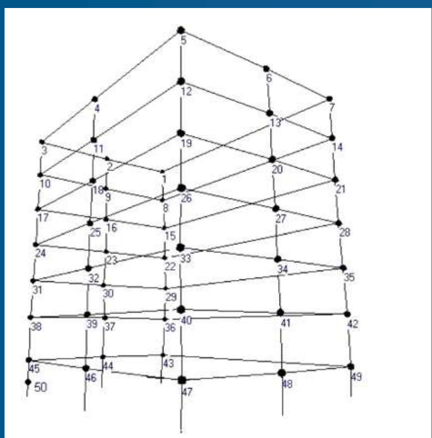
Surrogate Modeling for Computational Efficiency





Guan, X., Burton, H., Shokrabadi, M., & Yi, Z. (2021). Seismic drift demand estimation for steel moment frame buildings: From mechanics-based to data-driven models. Journal of Structural Engineering, 147(6), 04021058.

Machine Learning for Efficiency (Surrogate Models)



Structural Model

β_1

Component-Level Testing

β_2

System-Level Testing

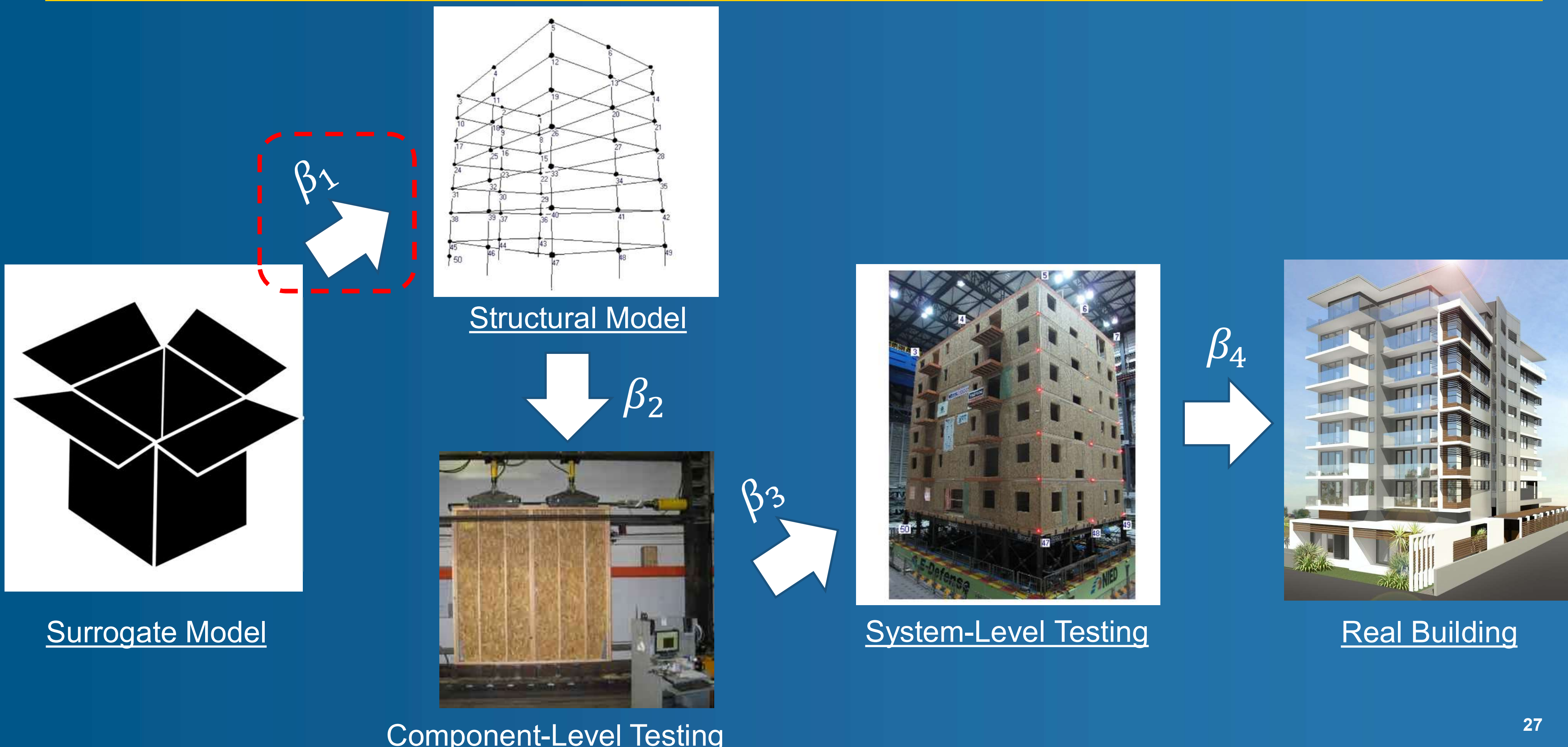
β_3

Real Building

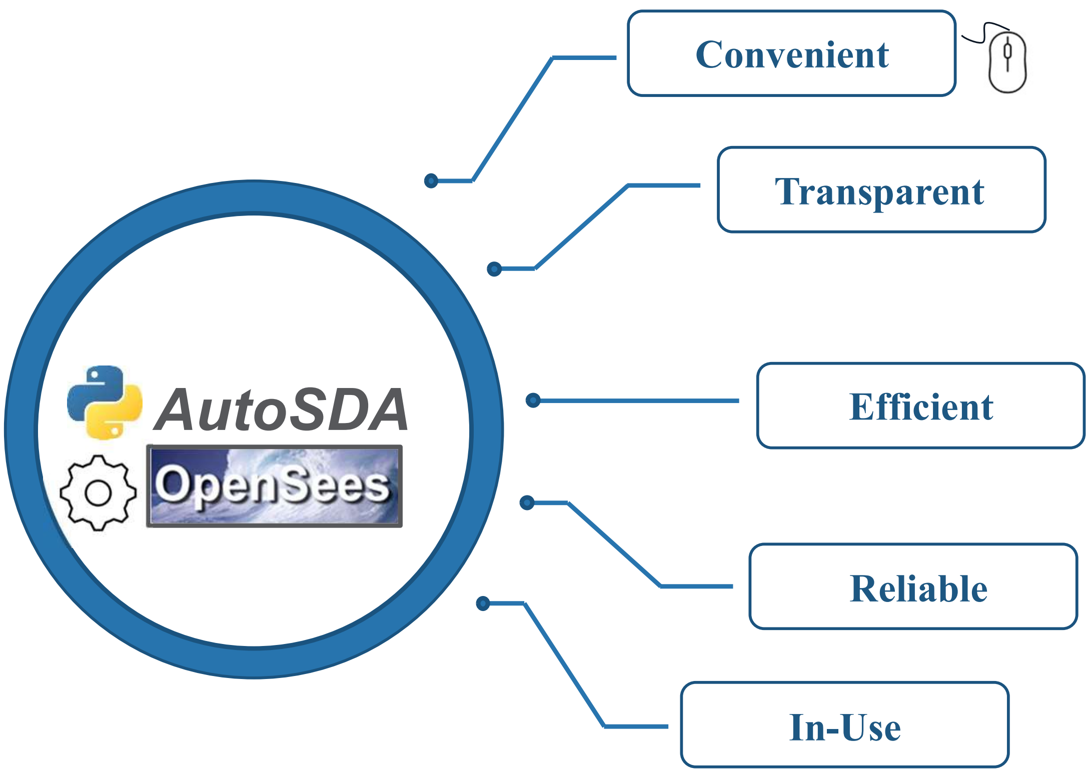
β_i: Error or uncertainty introduced at step i

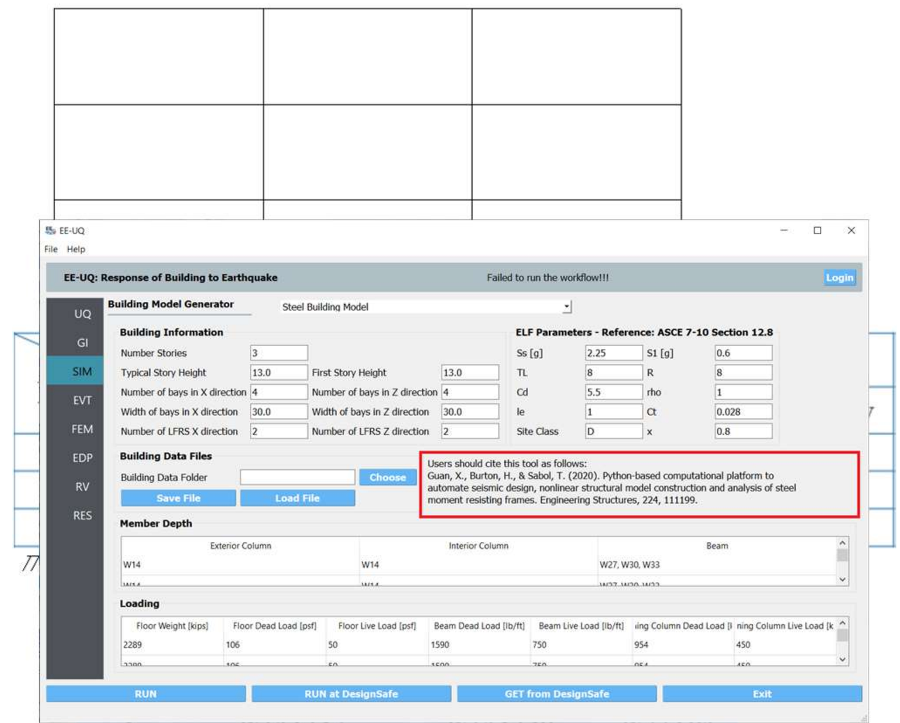
β_{TOT}: Total error or uncertainty

Machine Learning for Efficiency (Surrogate Models)



What About Automation for Efficiency?





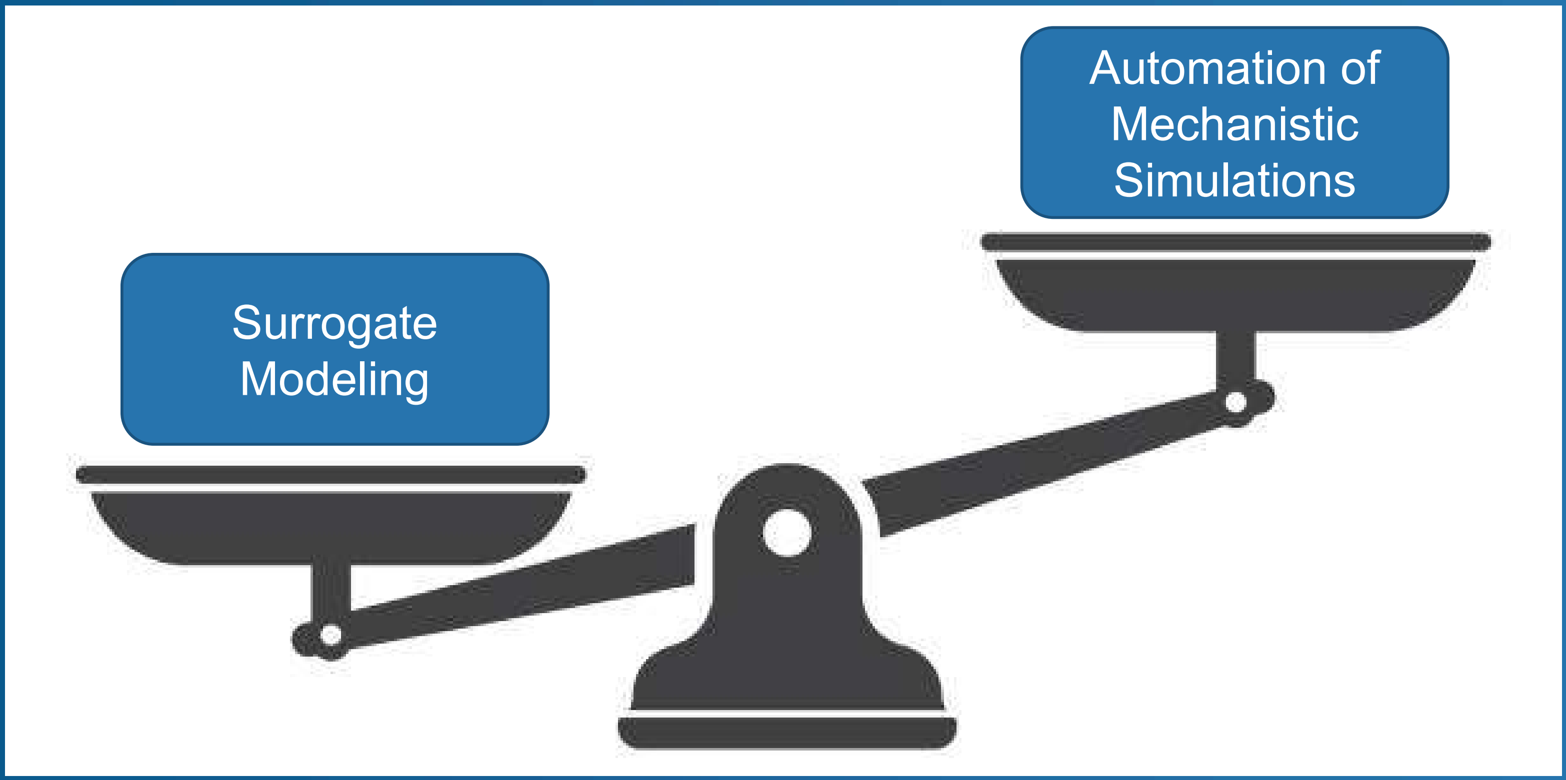
NSF NHERI EE-UQ framework

NHERI: Natural Hazards Engineering Research Infrastructure

EE-UQ: Earthquake Engineering with Uncertainty Quantification

Guan, X., Burton, H., & Sabol, T. (2020). Python-based computational platform to automate seismic design, nonlinear structural model construction and analysis of SMRFs. Engineering Structures, 224, 111199

Balancing the use of Automation and Surrogate Modeling for Achieving Scale in Regional Risk Assessments



Predicting the Drift Capacity of Reinforced Concrete Walls with Special Boundary Elements (SBE) using ML

Abdullah and Wallace (2019) developed an empirical equation to predict the drift capacity of SBE walls based on a dataset that comprises 164 test. The equation is specified in ACI 318-19.

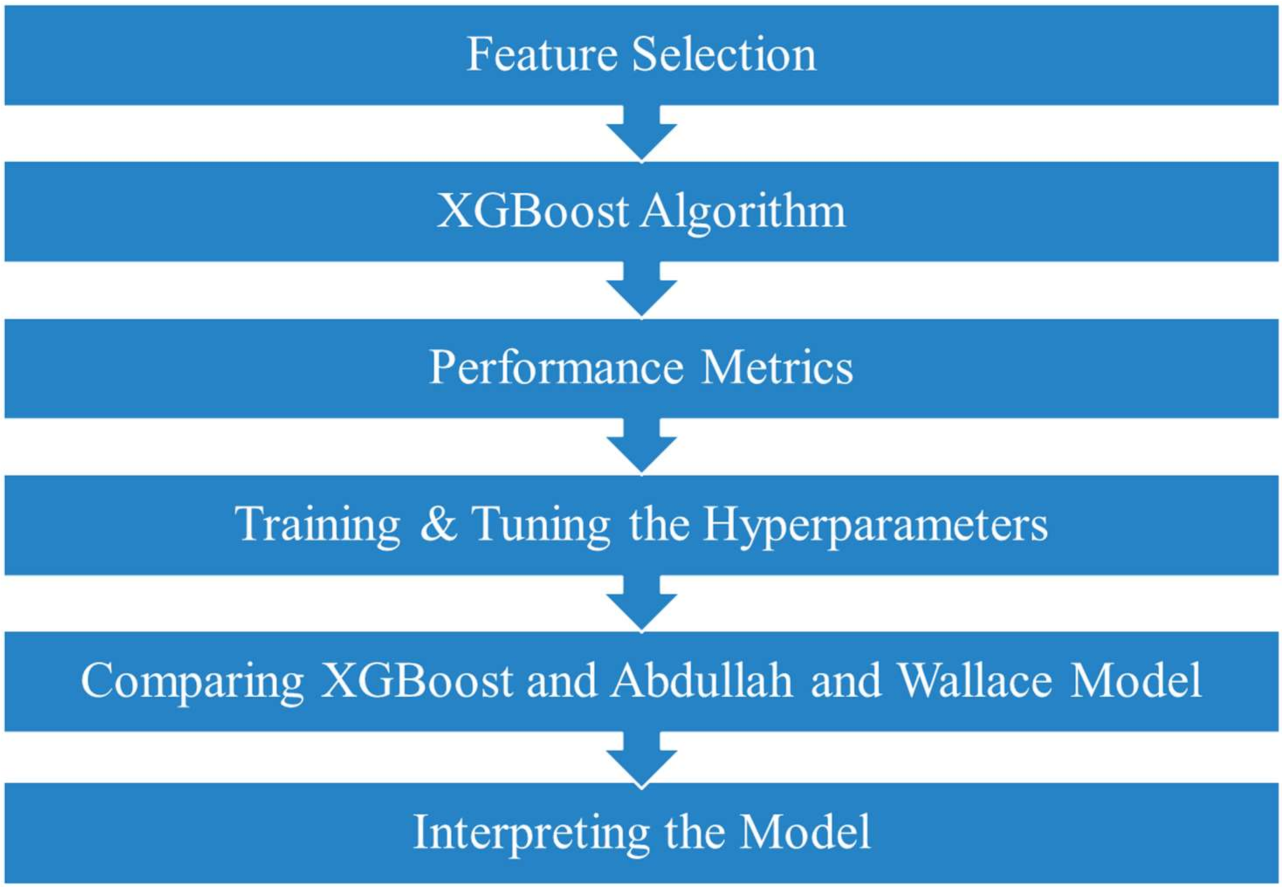
$$\frac{\delta_c}{h_w}(\%) = 3.85 - \frac{\lambda_b}{\alpha} - \frac{v_{max}}{10\sqrt{f_c'(psi)}}$$

$\lambda_b = \frac{l_w b}{c^2}$ is a slenderness ratio

$\alpha = 60$ for overlapping hoops and 45 for single hoops with crossties

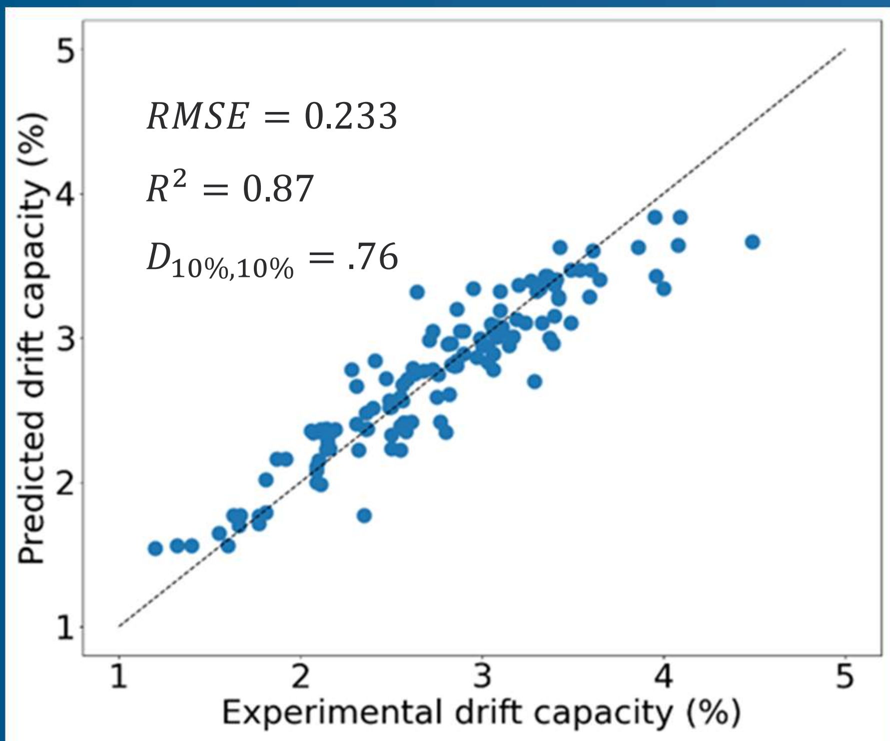
v_{max}: maximum shear stress and f_c': concrete compressive strength

Overview of ML Model Development Workflow

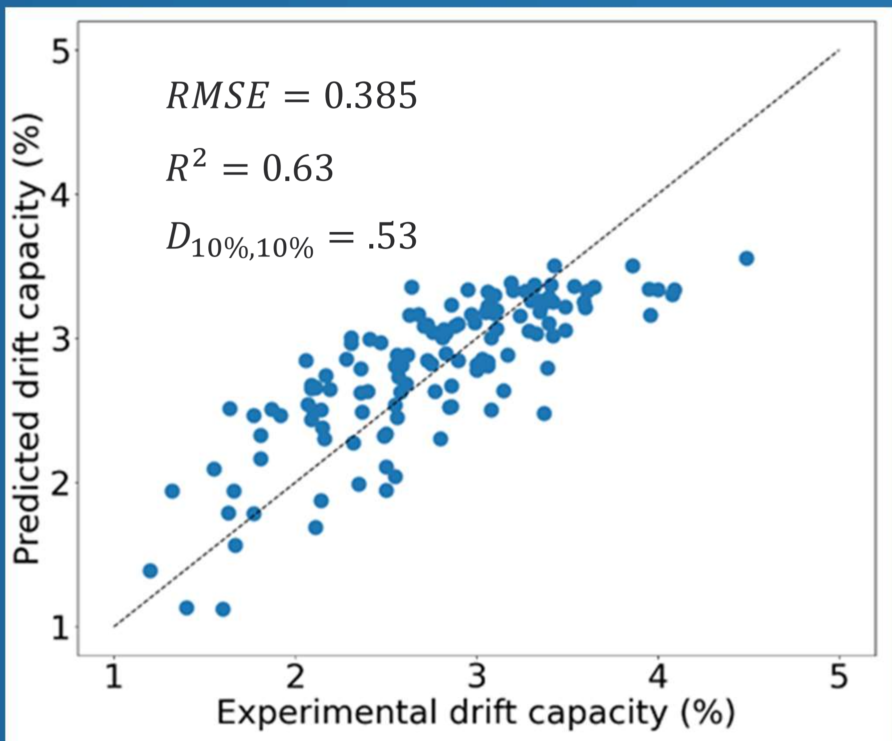


Aladsani, M. A., Burton, H., Abdullah, S. A., & Wallace, J. W. (2022). Explainable Machine Learning Model for Predicting Drift Capacity of Reinforced Concrete Walls. ACI Structural Journal, 119(3).

Comparing Performance of XGBoost and Abdullah and Wallace Models



XGBoost



Abdullah and Wallace

Aladsani, M. A., Burton, H., Abdullah, S. A., & Wallace, J. W. (2022). Explainable Machine Learning Model for Predicting Drift Capacity of Reinforced Concrete Walls. ACI Structural Journal, 119(3).

What is the Value of Increased Predictive Accuracy?

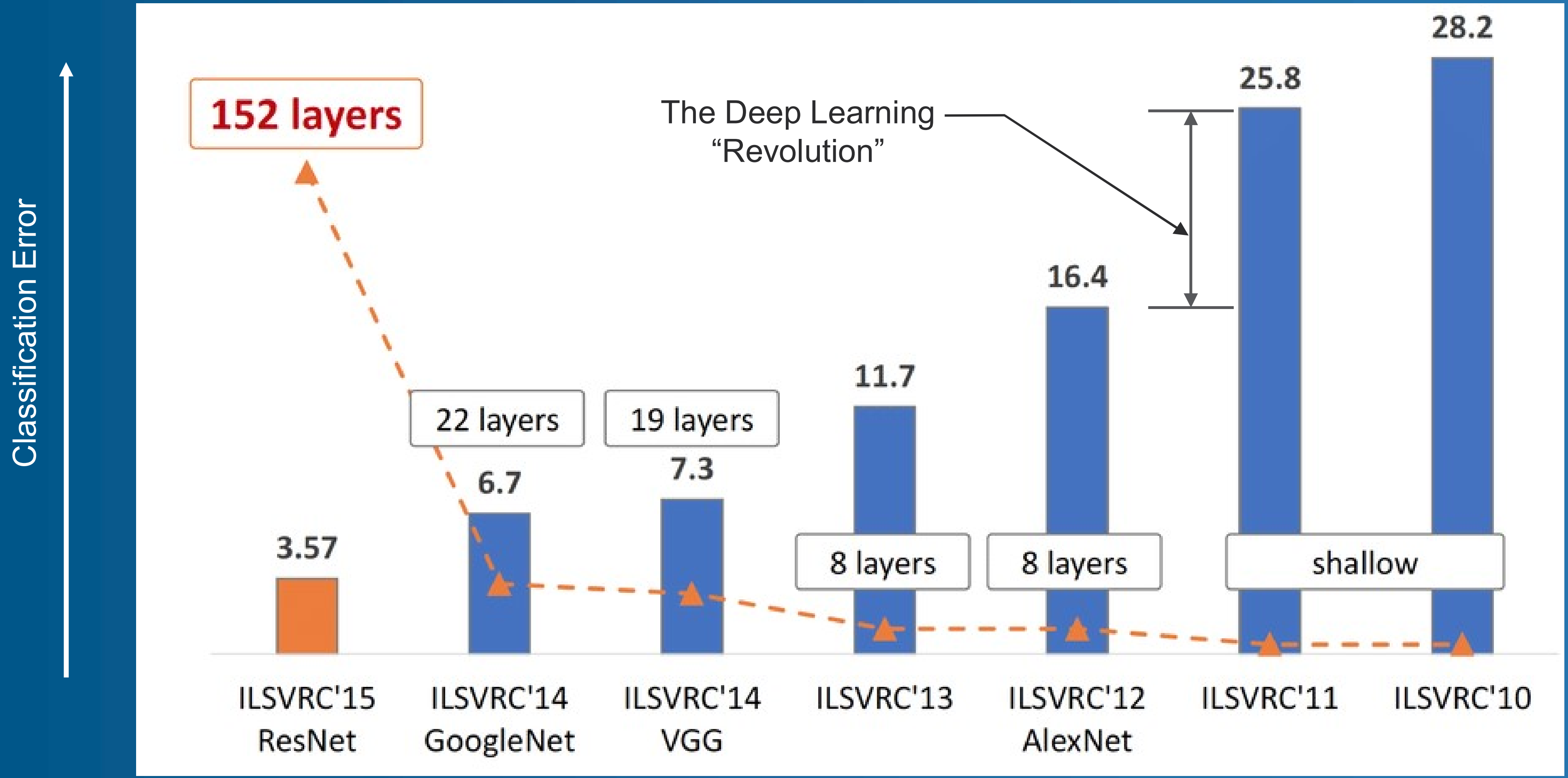
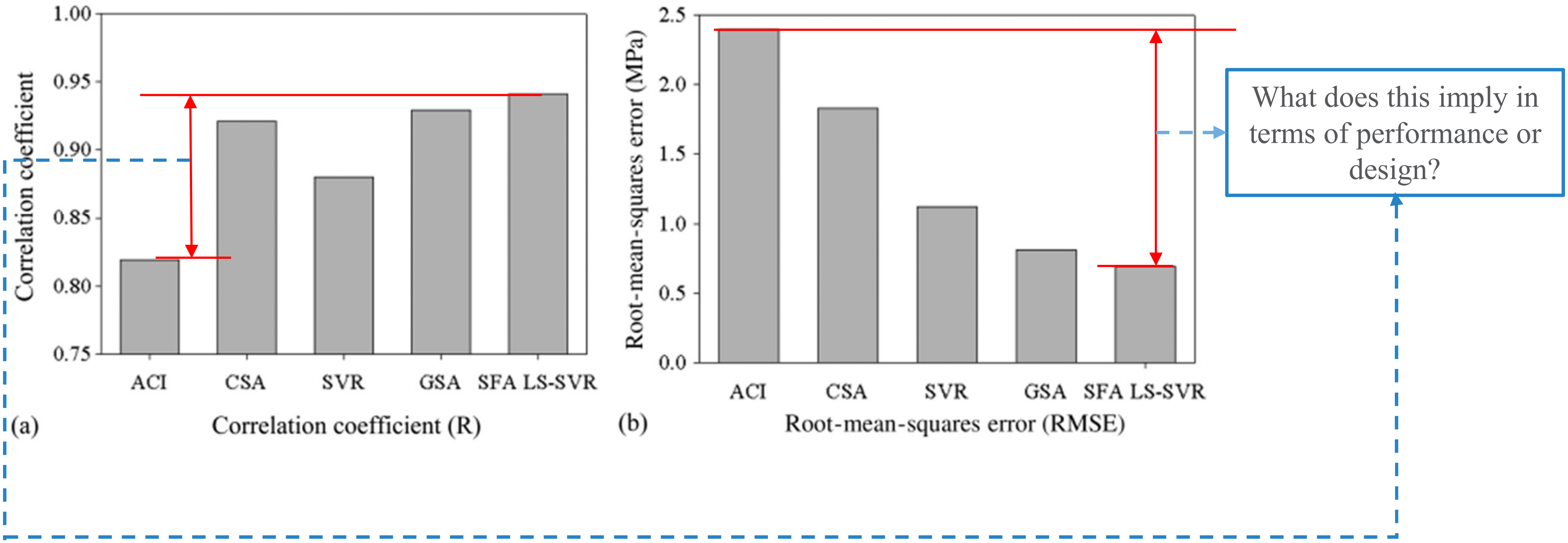


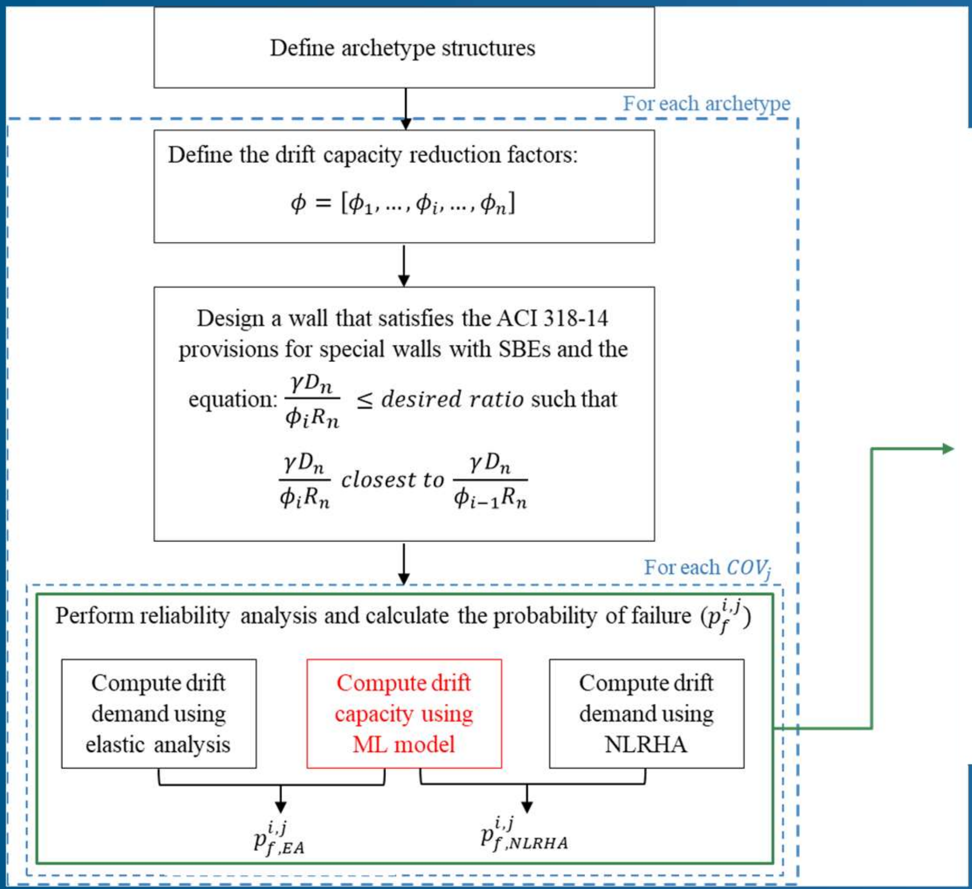
Image courtesy Nguyen et al. 2017

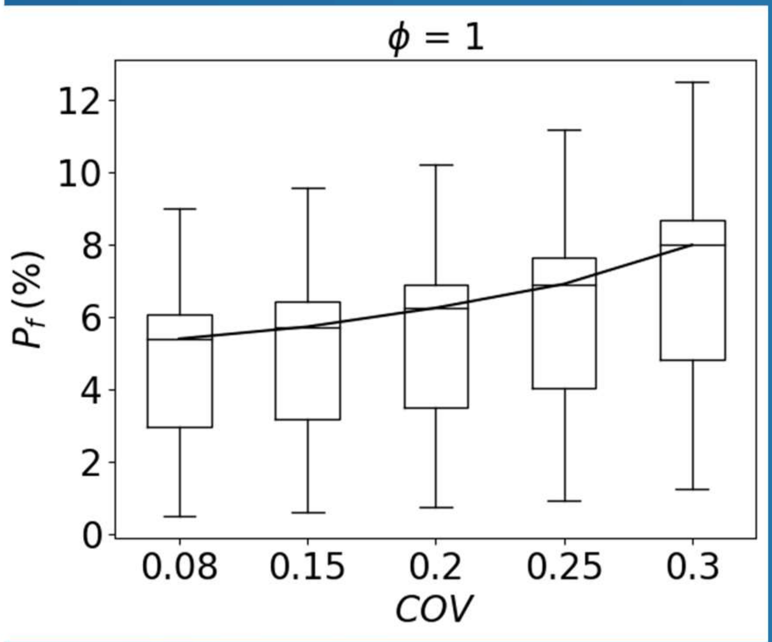
What is the Value of Increased Predictive Accuracy?



Chou, J., Ngo, N., & Pham, A. (2016). Shear Strength Prediction in Reinforced Concrete Deep Beams Using Nature-Inspired Metaheuristic Support Vector Regression. Journal of Computing in Civil Engineering, 30(1)

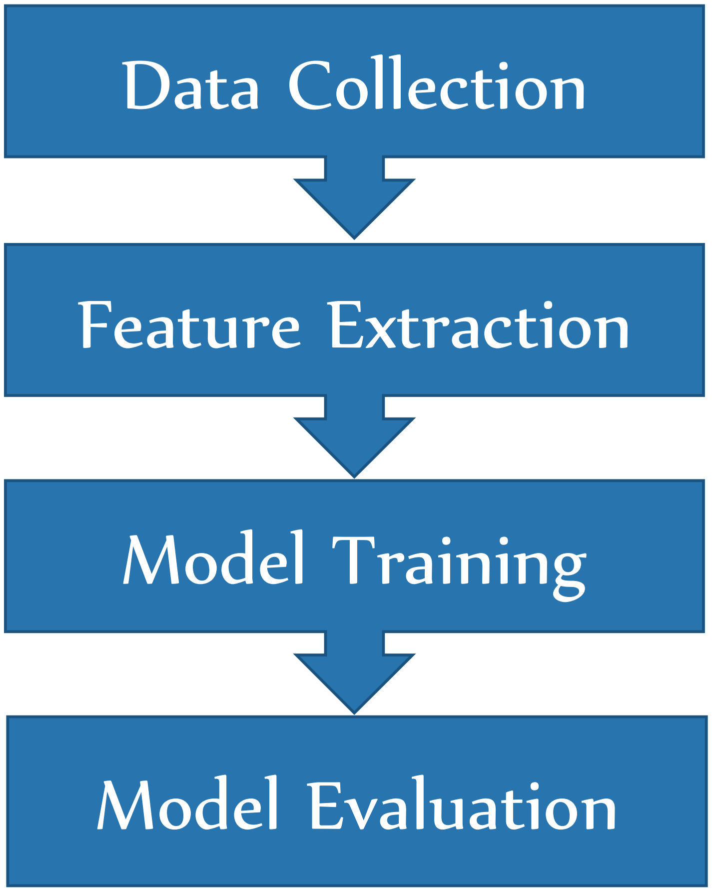
The Value of Increased Predictive Accuracy: A Reliability-Based Evaluation





Aladsani, M. A., and Burton, H. (2022). A reliability-based approach to quantify the benefits of machine learning predictive models in terms of the reduced uncertainty. Proceedings of the 12th National Conference on Earthquake Engineering, Salt Lake City, Utah.

Machine Learning for Regional Impact Assessment (An Alternative to HAZUS)



Predicted Class

Actual Class	G	Y	R	
G	141 (20.7%)	129 (18.9%)	0 (0%)	52%
Y	74 (10.9%)	302 (44.3%)	4 (0.6%)	79%
R	7 (1.0%)	21 (3.1%)	4 (0.6%)	13%
	64%	67%	50%	66%

Multiple trees are "in parallel"

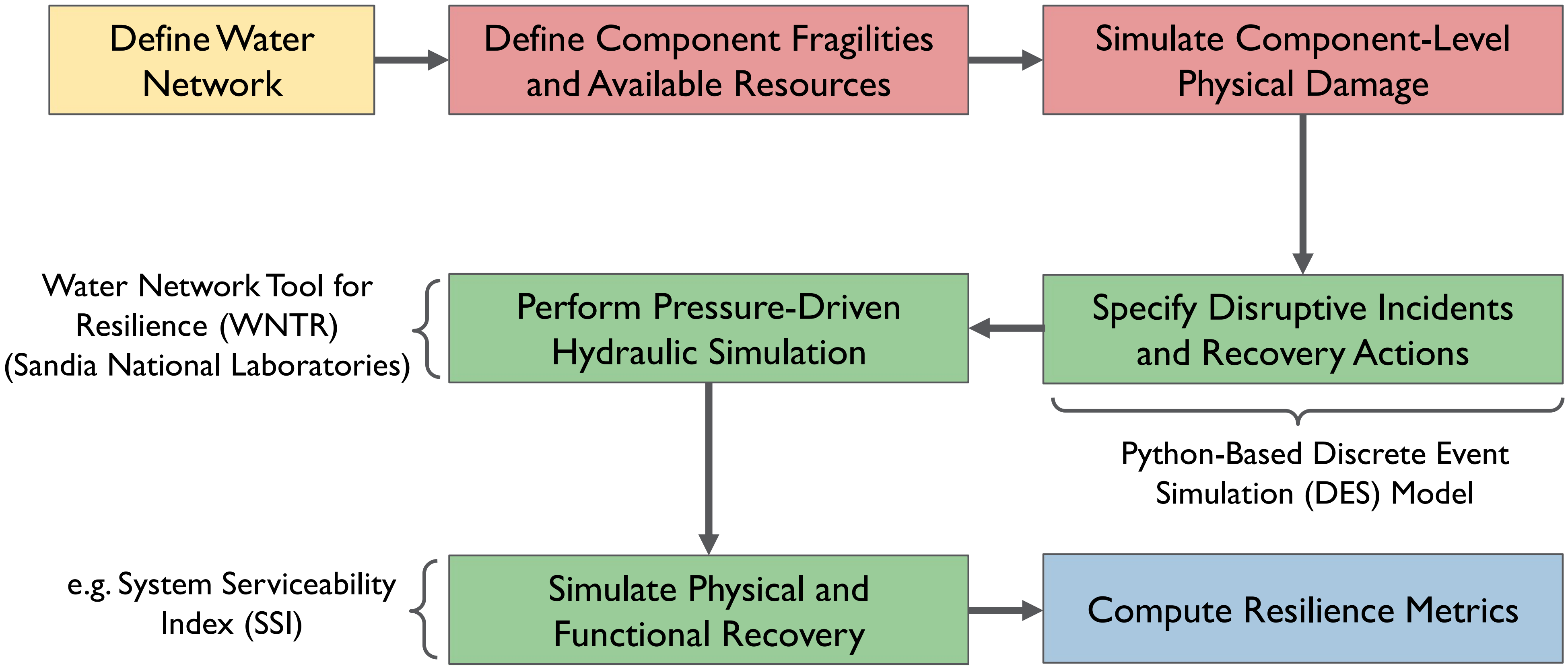
Testing Set

Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., & Burton, H. V. (2020). Classifying earthquake damage to buildings using machine learning. Earthquake Spectra, 36(1), 183-208.

Machine Learning for Predicting Regional Impacts: Some Critical Questions

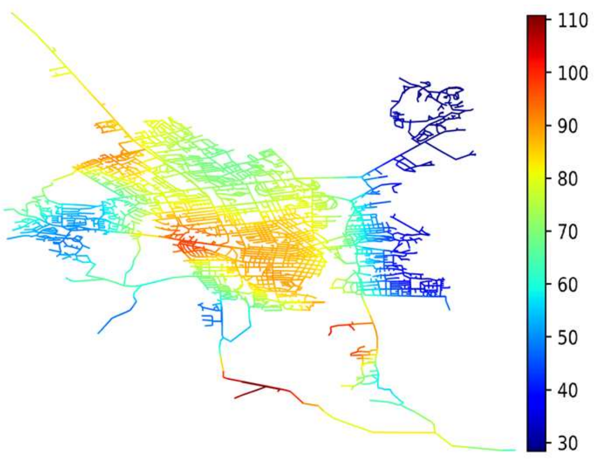
- How does the ML approach perform relative to HAZUS?
- Can we make the ML results as interpretable (or even more so) than HAZUS?
- What if I applied my model trained on data from the 2014 South Napa earthquake to the 1994 Northridge earthquake?
- How many events need to be considered in the training data to make the ML model “generalizable enough” ?
- Does a practitioner have to be knowledgeable about ML to apply the model?

Active Learning for Event Subset Selection in Distributed Infrastructure Risk and Resilience Assessment



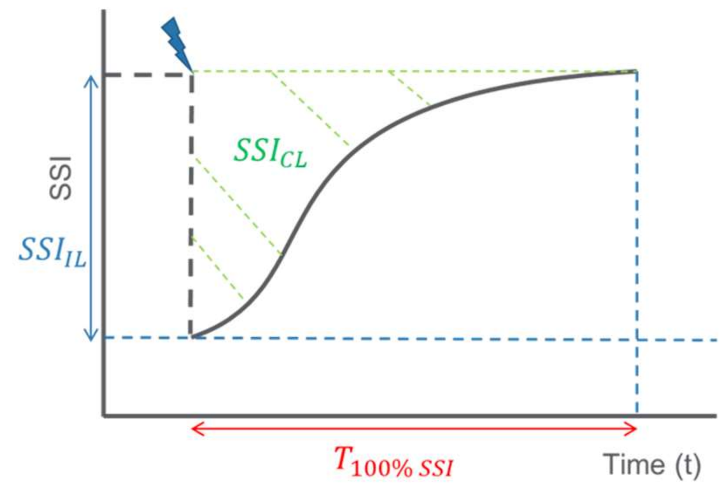
Tomar, A., Burton, H., Mosleh, A., Lee, J. Y., (2020) "Hindcasting the Functional Loss and Restoration of the Napa Water System Following the 2014 Earthquake Using Discrete-Event Simulation," *ASCE Journal of Infrastructure Systems*, 10.1061/(ASCE) IS.1943-555X.0000574

Active Learning for Event Subset Selection in Distributed Infrastructure Risk and Resilience Assessment



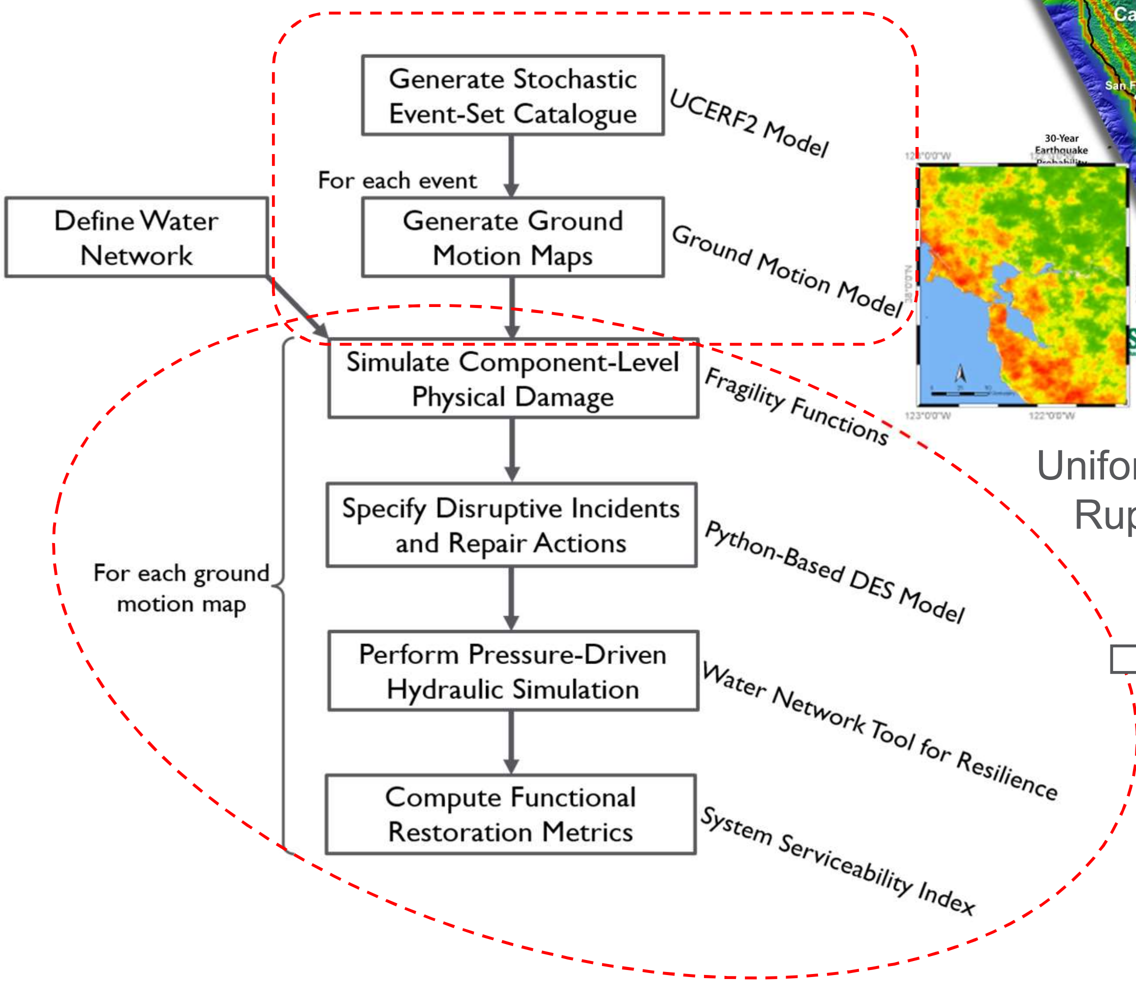
Single Scenario

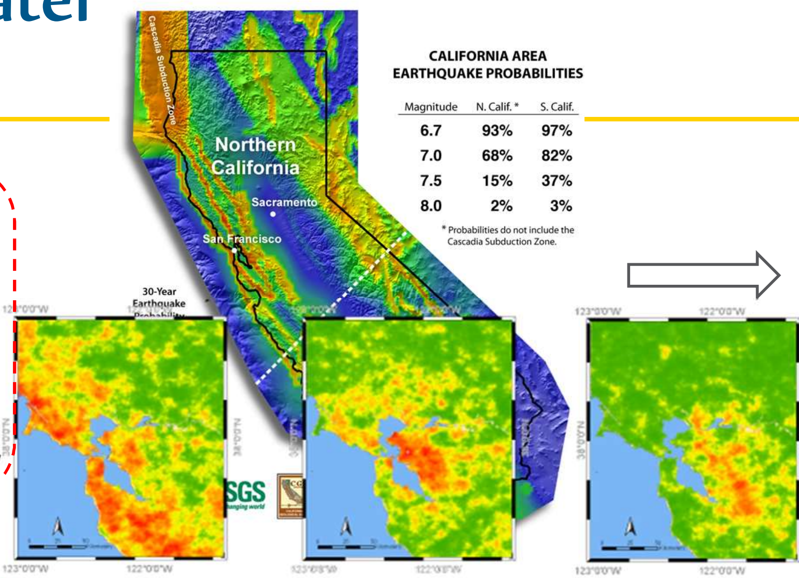
End-to-End
Simulation Model



SSI_{IL}	SSI Immediate Loss
$T_{100\% \, SSI}$	Time to reach 100% of SSI
SSI_{CL}	SSI Cumulative Loss

Risk-Based Assessment of Water System Resilience





Markhvida et al. (2018)

Uniform California Earthquake Rupture Forecast 2 Model

For Napa Region

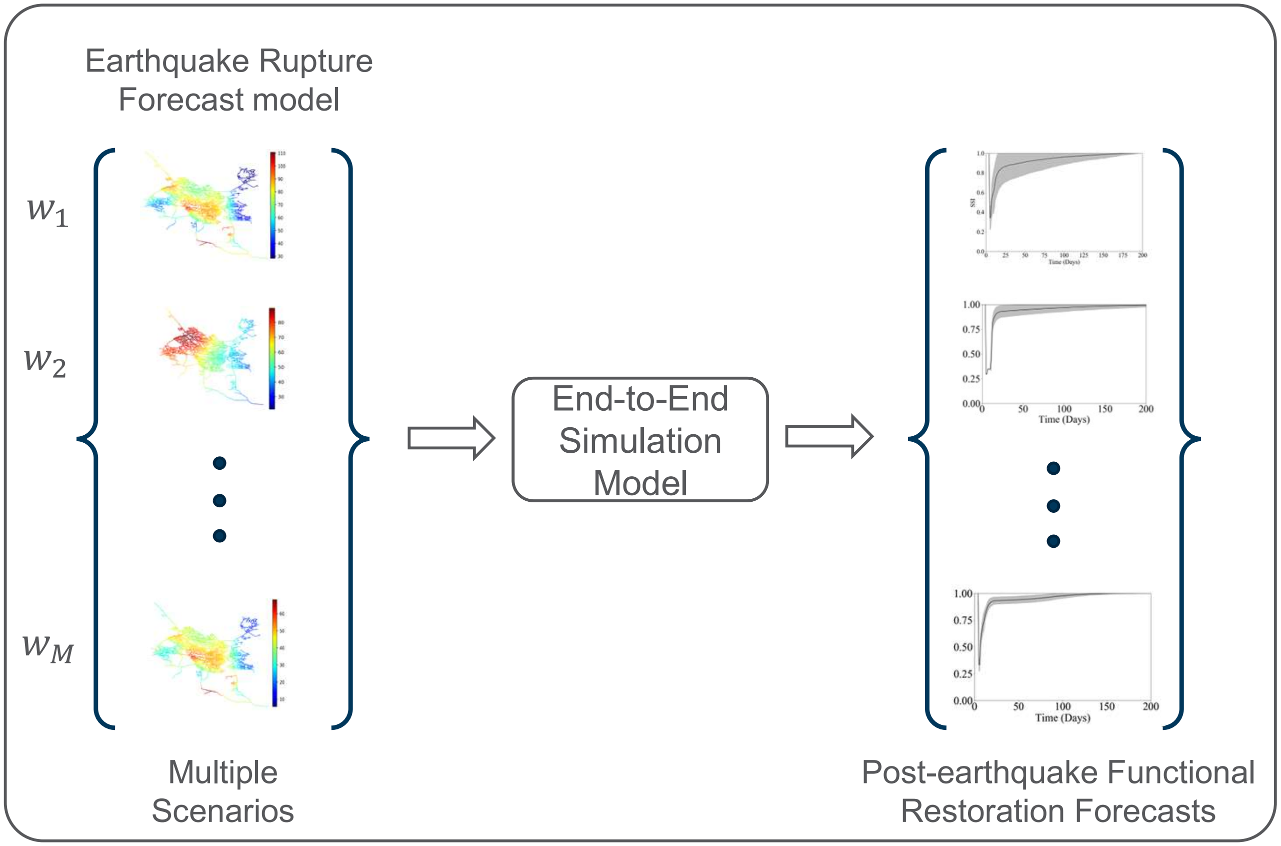
2343 Events

2343 X 50 Spatially Correlated Ground Motion Maps

End-to-End Simulation Model

2343 X 1000 Simulations

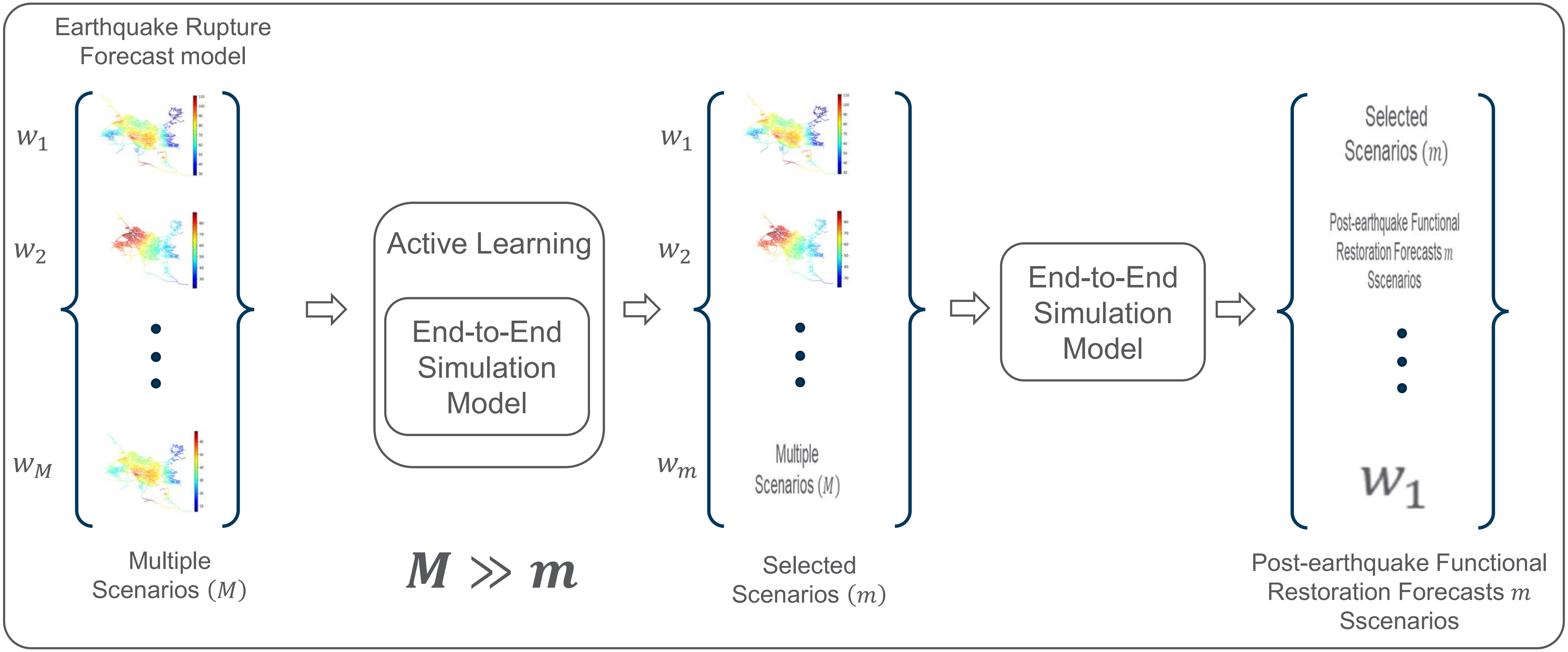
Computational Challenges of Stochastic Event Set-Based Assessments



- A typical simulation takes between 2 minutes to 6 hours
- We used parallel computing
- Set of all possible scenarios using rupture forecast models:
  - UCERF2 – 2343 scenarios
  - UCERF3 – 236905 scenarios
- Analyzing larger system
  - Napa - 612 km of pipeline
  - Los Angeles County ~ 12000 km of pipelines
- Some network performance metrics require heavy computational power
  - Centrality measures (betweenness)

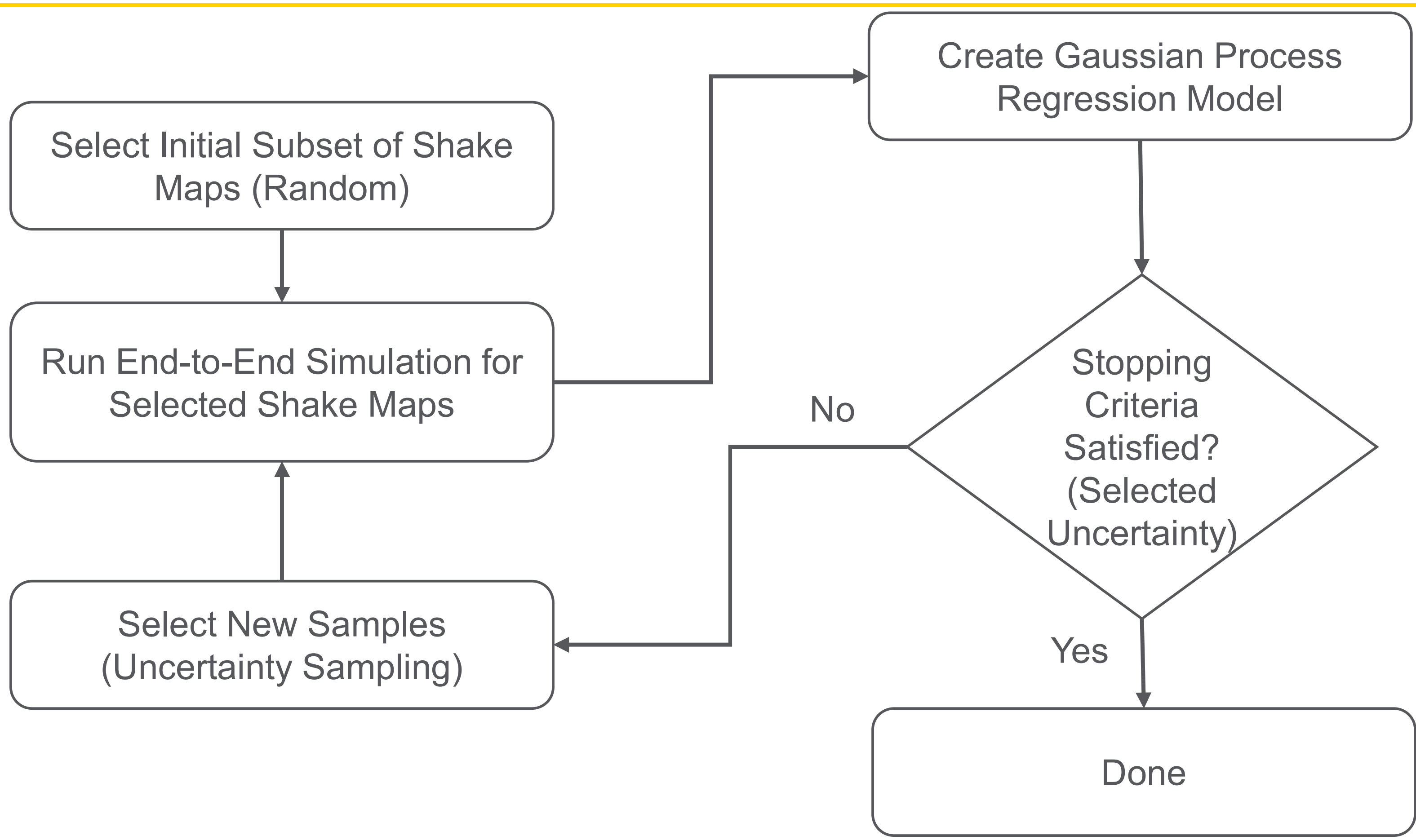
Tomar, A., & Burton, H. V. (2021). Active learning method for risk assessment of distributed infrastructure systems. Computer-Aided Civil and Infrastructure Engineering, 36(4), 438-452.

Shaking Intensity Map Subset Selection using Active Learning



Tomar, A., & Burton, H. V. (2021). Active learning method for risk assessment of distributed infrastructure systems. Computer-Aided Civil and Infrastructure Engineering, 36(4), 438-452.

Active Learning for Map Subset Selection



Tomar, A., & Burton, H. V. (2021). Active learning method for risk assessment of distributed infrastructure systems. Computer-Aided Civil and Infrastructure Engineering, 36(4), 438-452.

Selected Uncertainty Stopping Criterion

The variance of the top-*m* selected samples (i.e., *m* most uncertain cases) at each learning cycle will be a good signal of the confidence in the current model.

$$SC_{SU}(C) = \begin{cases} 1, & \dfrac{\sum_{x \in C} Var(x)}{m} < \delta_{SU} \\ 0, & otherwise \end{cases}$$

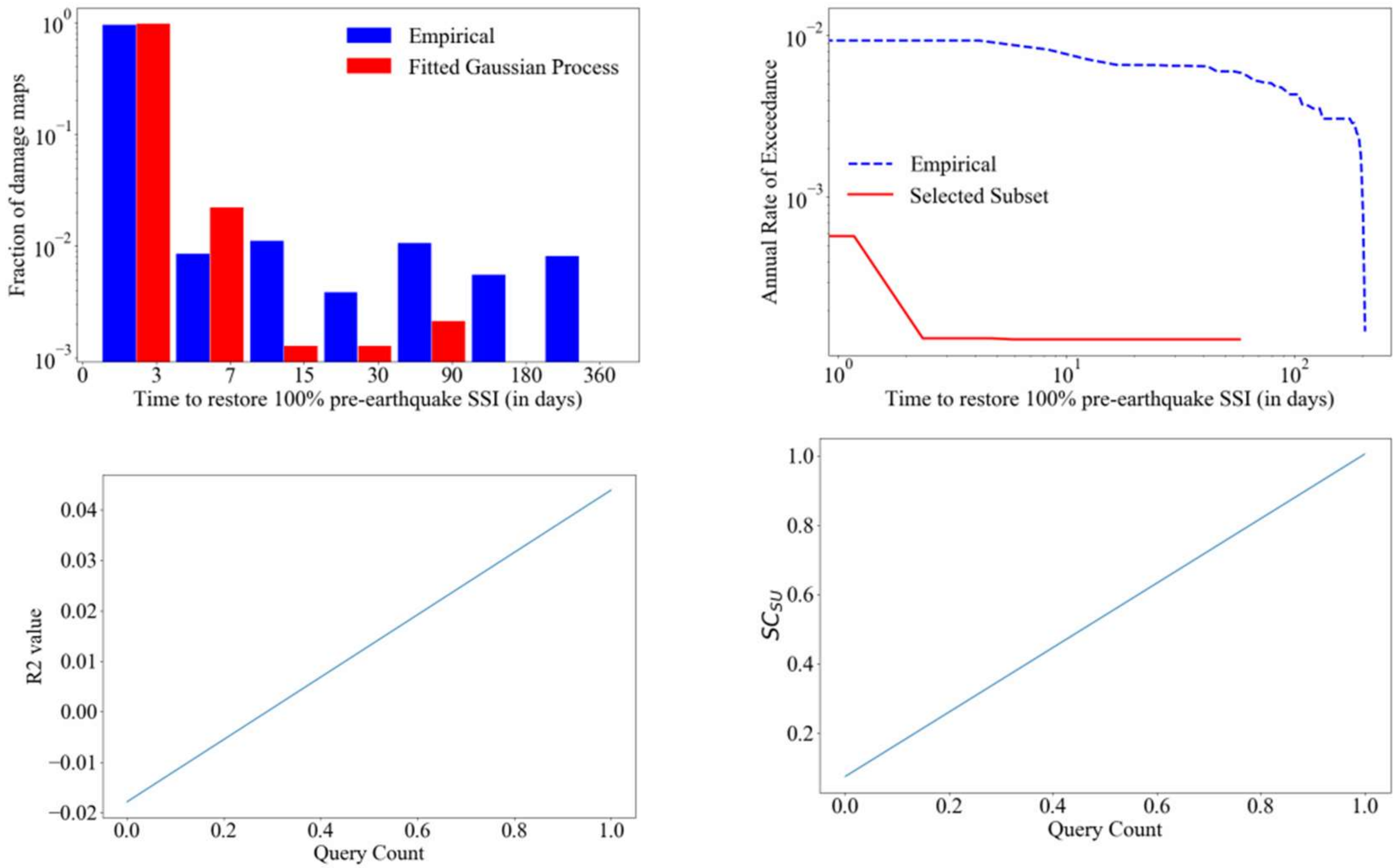
where δ_{SU} is the user-predefined variance threshold,

C is the set of top-*m* selected unlabeled samples,

$Var(x)$ evaluates the variance on set C based on the current model.

Tomar, A., & Burton, H. V. (2021). Active learning method for risk assessment of distributed infrastructure systems. Computer-Aided Civil and Infrastructure Engineering, 36(4), 438-452.

Active Learning Results for $T_{100\%SSI}$



Tomar, A., & Burton, H. V. (2021). Active learning method for risk assessment of distributed infrastructure systems. Computer-Aided Civil and Infrastructure Engineering, 36(4), 438-452.

Active Learning: Magic Bullet, or No?

- Active learning is a powerful because it allows to learn from incrementally generated mechanistic simulations.
- The earlier approach to “event subset selection” used an optimization framework where the mechanistic assessment for the complete event set was used.
- The complete event set mechanistic simulation is not required in active learning.
- However, careful thought and effort needs to be placed in establishing the stopping criteria.

Overview

- Opportunities and Challenges: The Next Frontier in ML-NHE
- Final Thoughts

Final Thoughts

- For us (natural hazard engineers), ML is “just a tool”, but it can be an extremely powerful tool in the right context (and if applied correctly).
- It’s not about the algorithms themselves, it’s about the problems that are able to solve by applying them.
- In our current paradigm, all machine learning models are based on associational (not causal) relationships (more on this later).
- As a community, we need to move ML-NHE applications beyond “exploration” to real “problem solving”.

The End


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS