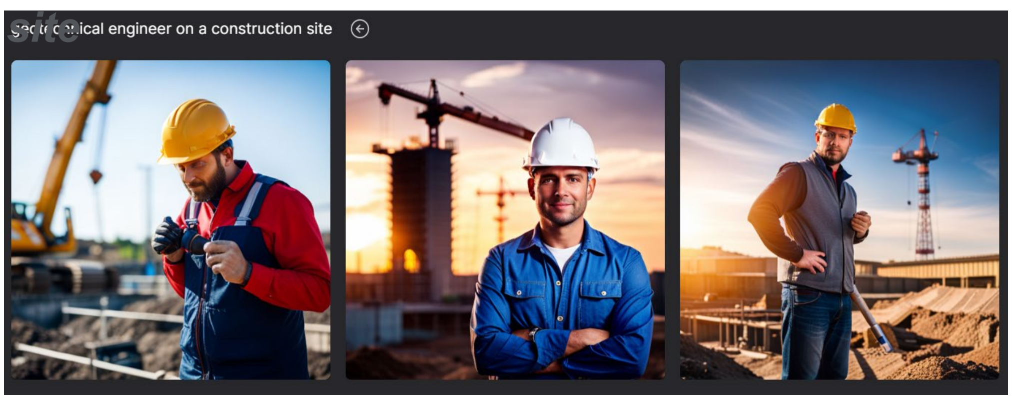
Towards reliable data science for data-driven landslide susceptibility modeling

06/12/2024

Te Pei, Ph.D.
Assistant Professor in Civil Engineering

Artificial intelligence (AI) & machine learning (ML) are changing our lives

Prompt: ***Geotechnical engineer on a construction site***



Images generated using Stable Diffusion

ChatGPT

Model: GPT-4

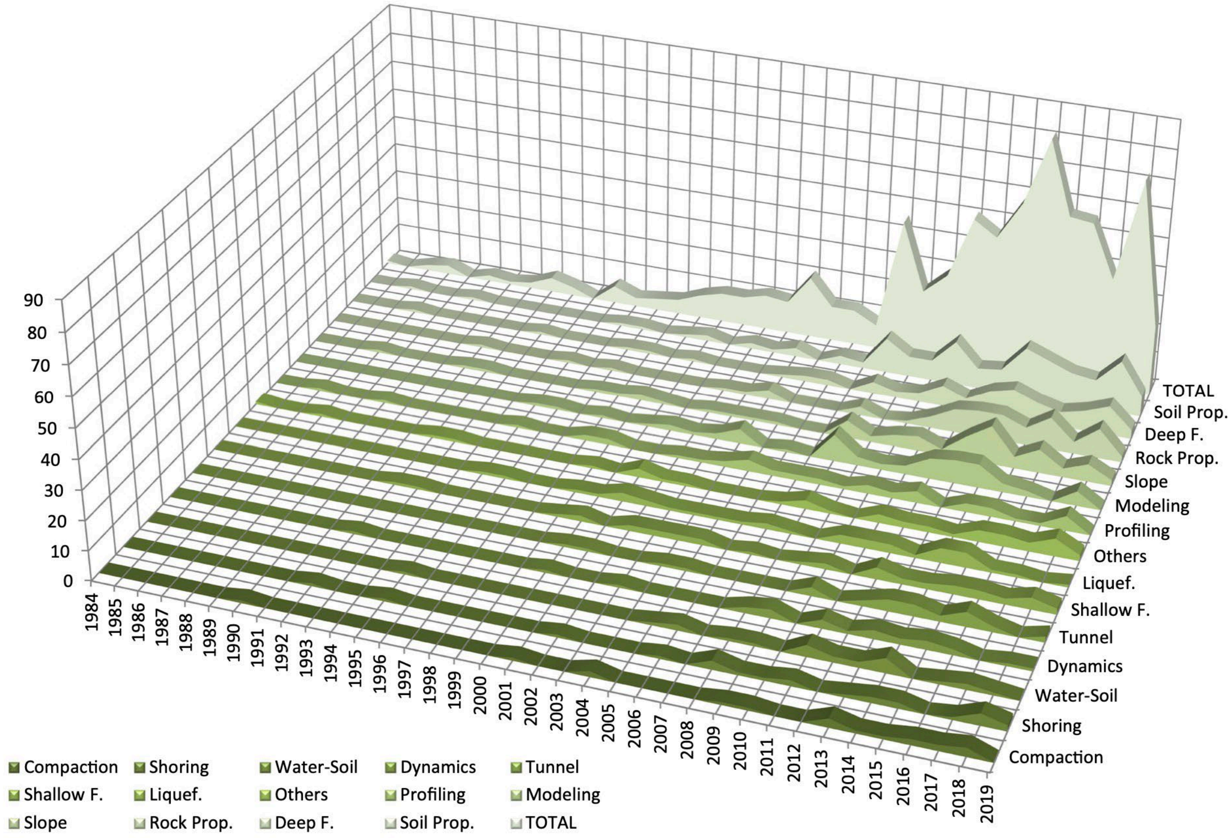
PE How can AI change the future of geotechnical engineering?

Artificial Intelligence (AI) can potentially revolutionize the field of geotechnical engineering in a number of ways, impacting everything from design and analysis to the construction and monitoring of earthworks and structures. Here are a few ways how AI might affect geotechnical engineering in the future:

1. **Risk Prediction and Mitigation**: Geotechnical engineers often need to assess and manage risks associated with natural hazards like earthquakes, landslides, or soil liquefaction. AI can help with this, using historical data and complex algorithms to predict where and when these hazards are most likely to occur, and what their impact could be. This can help engineers design and implement mitigation strategies more effectively.
2. **Design Optimization**: AI and Machine Learning (ML) algorithms can be used to optimize geotechnical design processes. For example, they can help engineers determine the most effective types of foundations for specific soil

AI&ML has been widely applied in Geotech

- Over 600 papers have been published on applications of various AI techniques to geotechnical engineering problems during the last three decades



Ebid (2021)

However pure data-driven ML models have limitations

- Pure-data driven ML models often act unexpectedly in parts of the input space not covered by the training and validation datasets

Example of a toy problem
(https://www.tensorflow.org/lattice)

Several challenges facing the Geotech community in adopting AI & ML

Data scarcity

- high-quality databases with sufficient samples are difficult to obtain

Generalization capability

- models only learn rules based on a particular dataset and have poor performance on new data

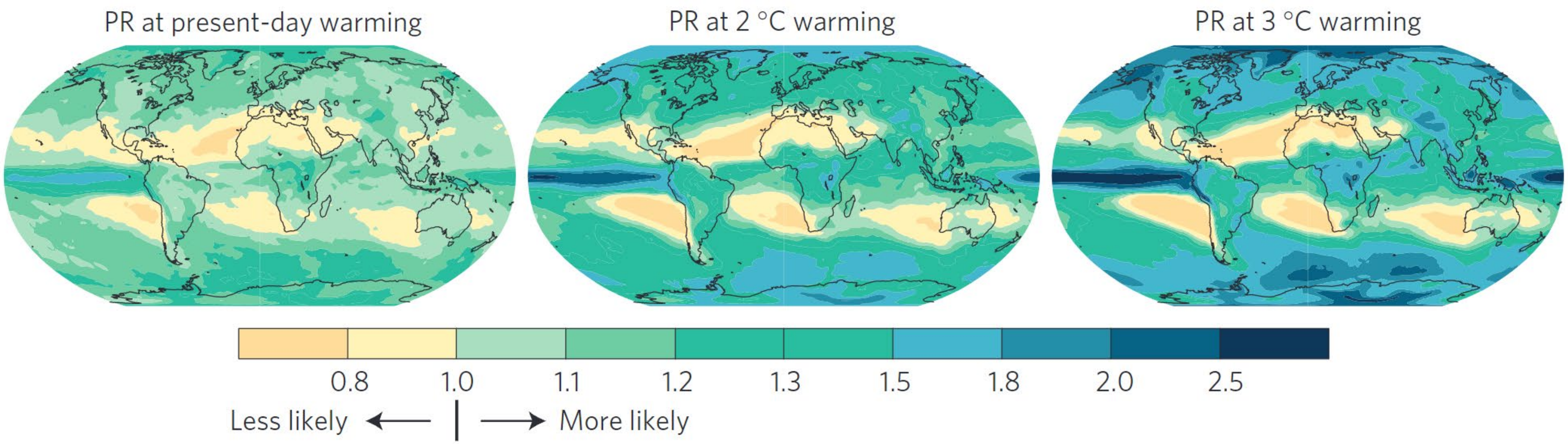
Explainability and physics consistency

- model predictions may violate common sense

Landslides are major natural disasters

Landslides are major natural disasters

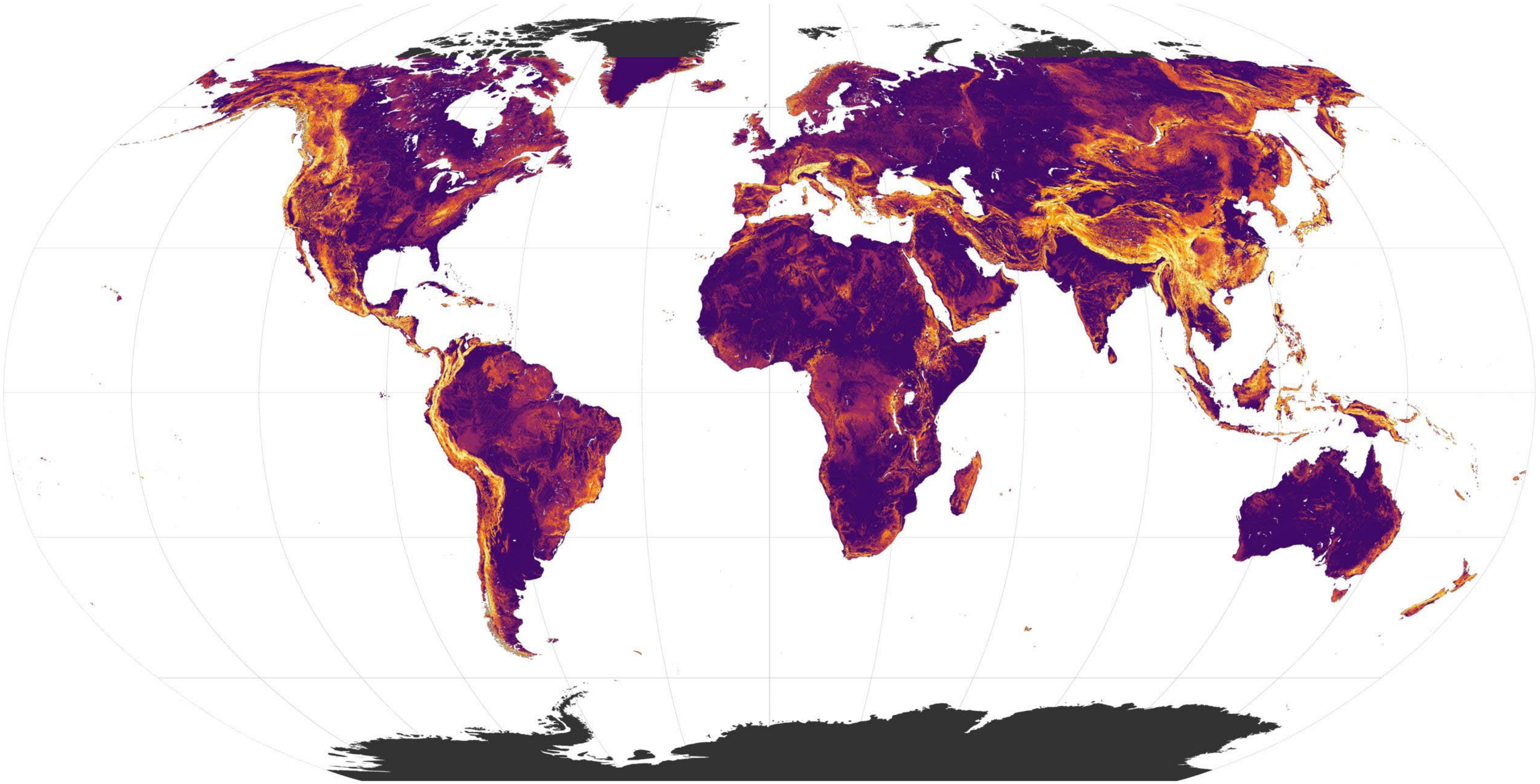
- Landslides can be triggered by earthquakes, volcanic eruptions, and **precipitation**
- Heavy precipitation including **rainfall** and **snowmelt** is the most common landslide trigger



Change in probability of heavy precipitation (Fischer and Knutti 2015)

More extreme precipitation is expected under current climate projection

Landslides are major natural disasters



NASA global landslide susceptibility estimate

Understand when and where landslide will occur can protect communities

- However, challenges exist, for example:

Complexity of triggering and failure mechanisms

Heterogeneity of hillslope environment

Incomplete or uncertain data

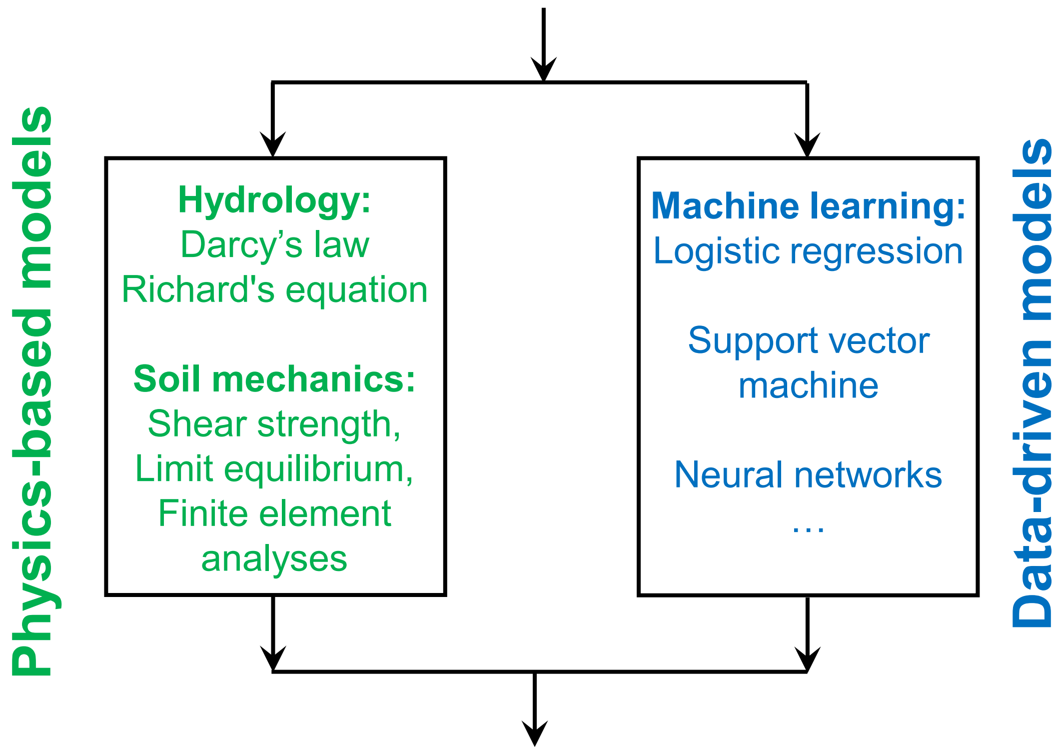
Limitations of predictive models

(GPT plotted this figure)

Both physics-based and data-driven methods can be used to study landslide risk

Input: *X* $(x_1, x_2 \ldots x_i)$

(precipitation, soil properties, groundwater, slope geometry, etc.)



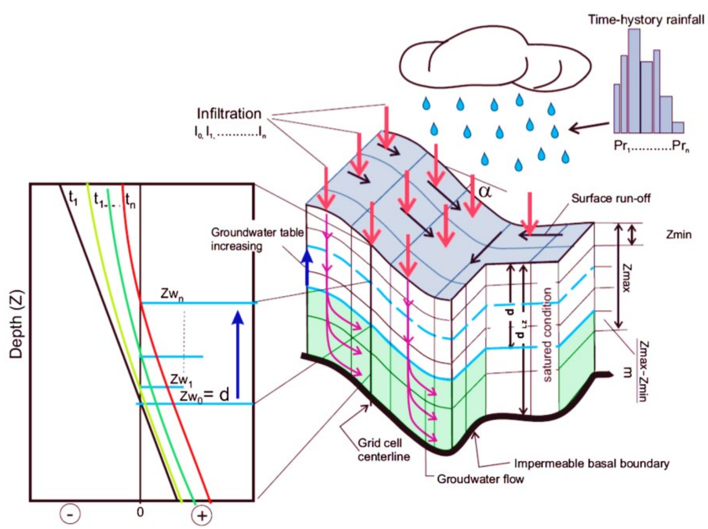
Output: *Y* $(y_1, y_2 \ldots y_i)$

(factor of safety, slope failure risk, etc.)

However, they both have inherent limitations

Physics-based models:

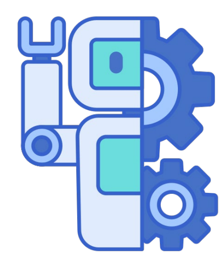
- Physically consistent results
- Performance can be significantly affected by quality of input data
- Applicable to site-specific analysis or small regions



From Grelle et al. (2013)

Machine learning models:

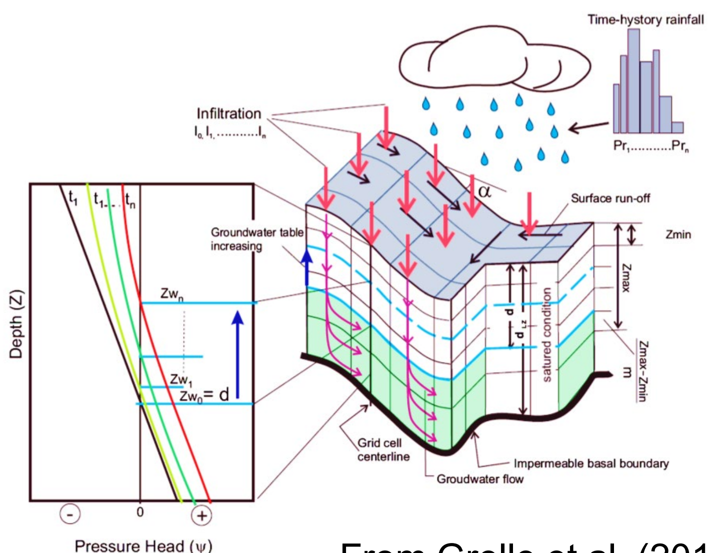
- Applicable to large regions
- Performance can be affected by data distribution
- Poor performance on out-of-domain samples
- Results may violate physics
- Poor interpretability



However, they both have inherent limitations

Physics-based models:

- Physically consistent results
- Performance can be significantly affected by quality of input data
- Applicable to site-specific analysis or small regions



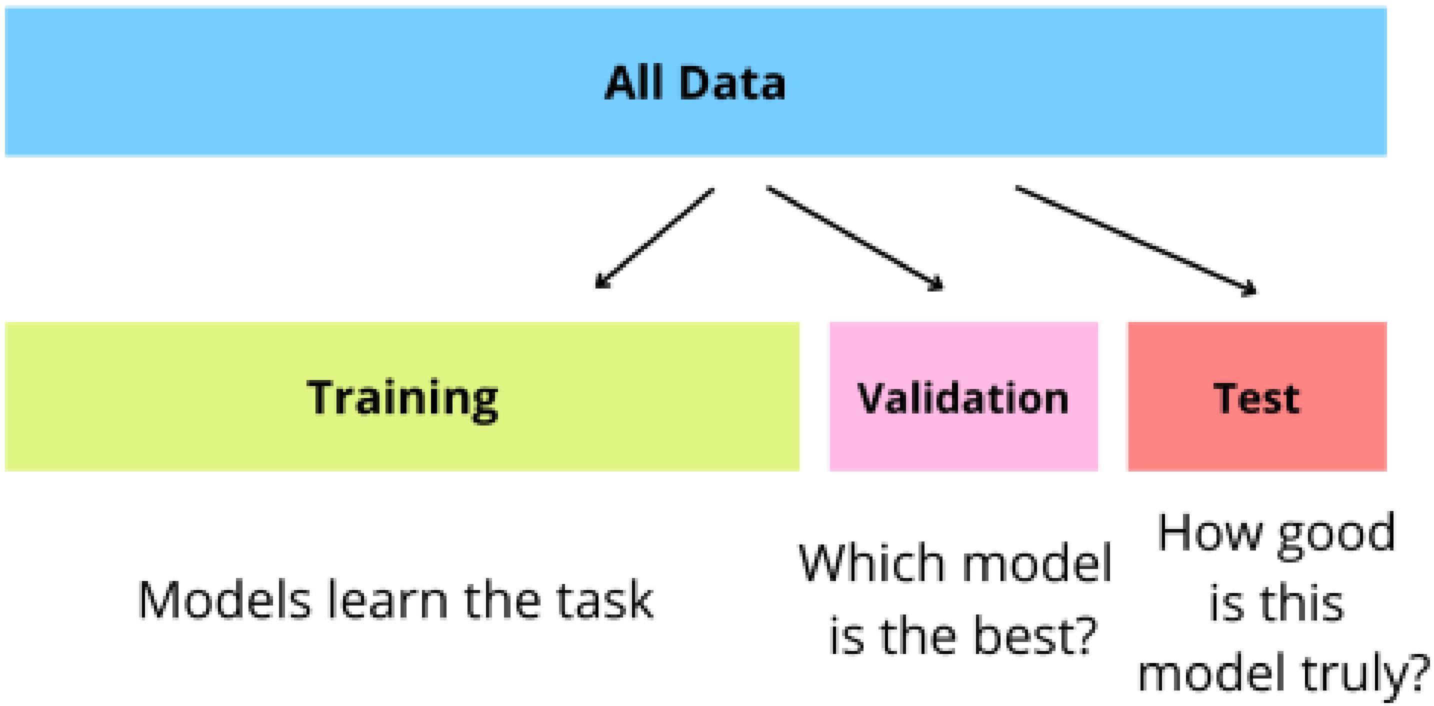
From Grelle et al. (2013)

Machine learning models:

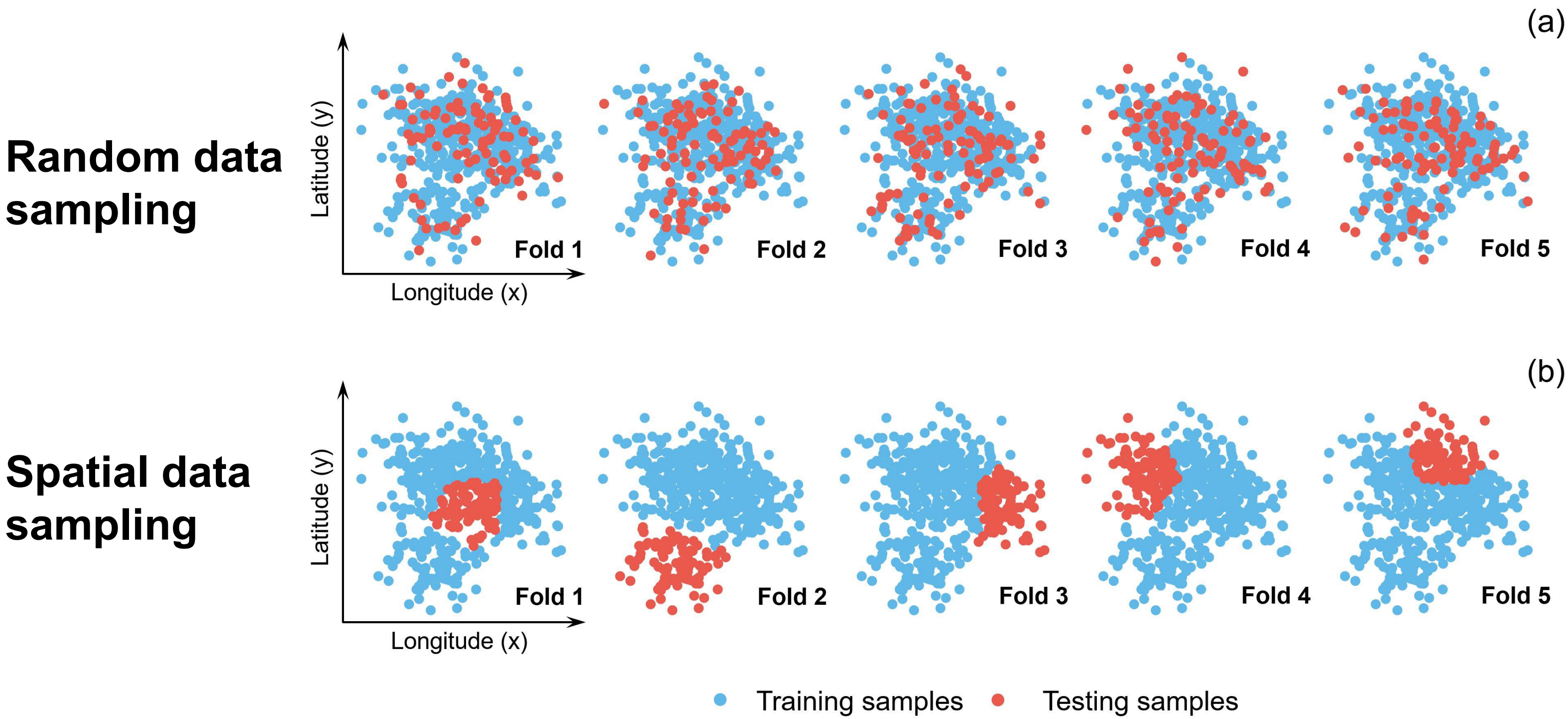
- Applicable to large regions
- Performance can be affected by data distribution
- Poor performance on out-of-domain samples
- Results may violate physics
- Poor interpretability

Currently most popular methods due to AI and remote sensing developments

Generic model validation approach can not deal with geospatial data



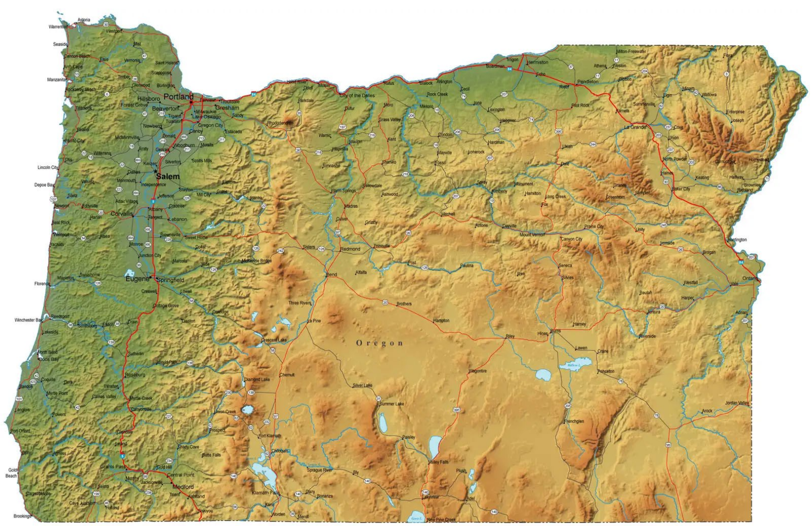
Generic model validation approach can not deal with geospatial data

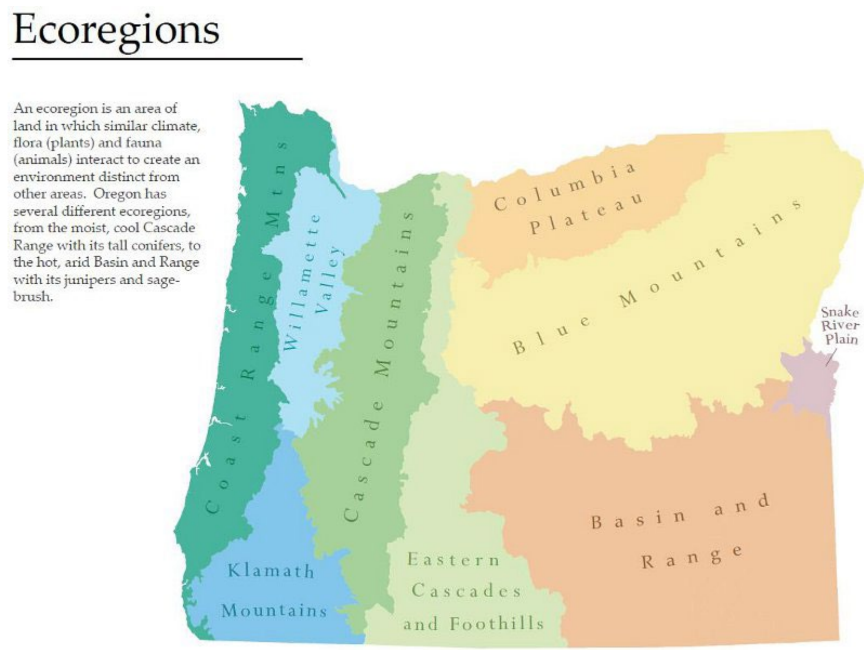


Spatial data autocorrelation matters!

What happens if you ignore data dependency

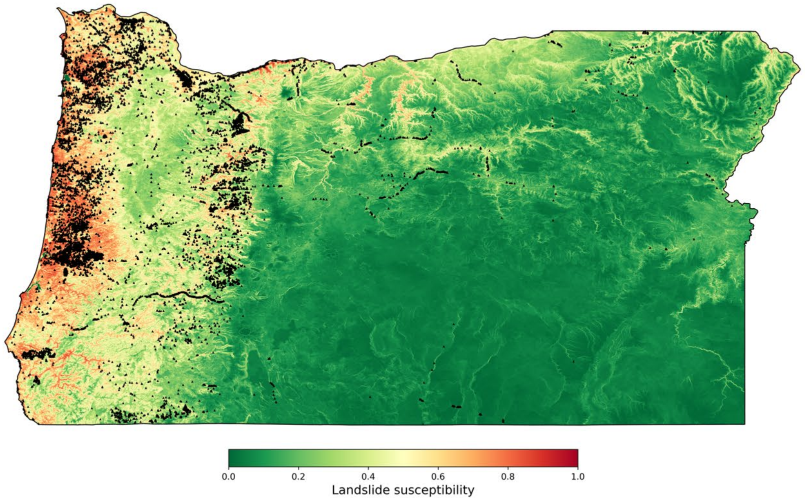
Oregon is heavily affected by landslide and has a diverse eco-environment

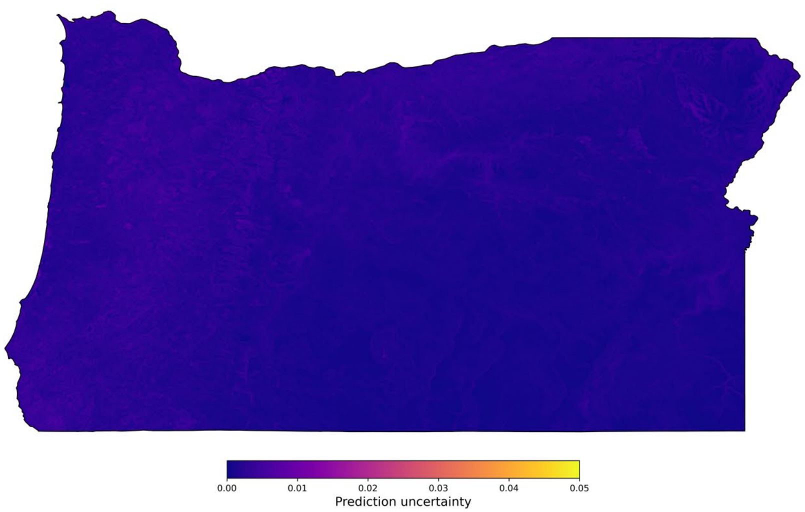




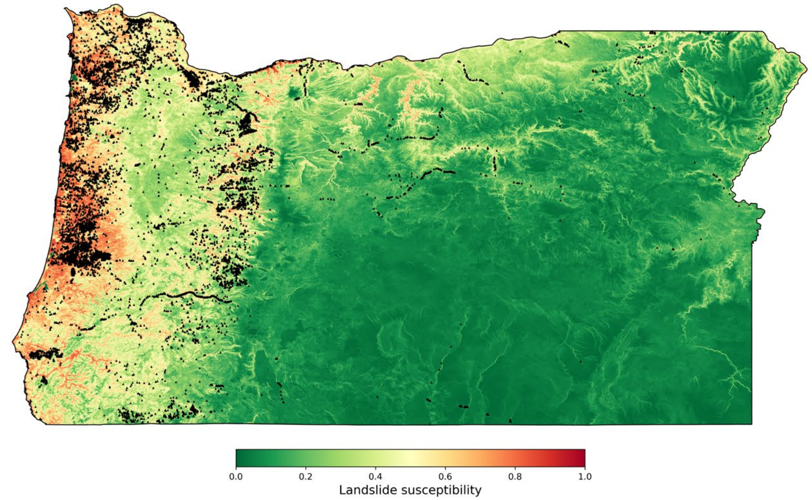
What happens if you ignore data dependency

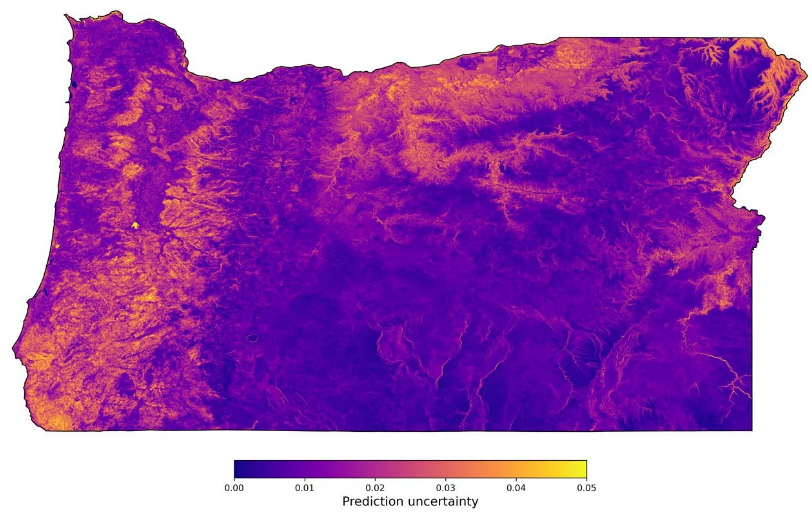
Model based on random sampling





Model based on spatial sampling





What do we want to do:

We want to develop best practices for responsible and reliable data science applications for geohazards modeling

What you will learn:

- **Learn how to use earth observation data and model them**
- **Learn how to code advanced AI/ML models and use them to understand natural hazards**
- **Understand the power and limitations of AI/ML and how to use them responsibly**

Interested?

Just drop me an email and we can discuss more: tpei@ccny.cuny.edu

Te Pei

Assistant Professor

Main Affiliation
Civil Engineering

Additional Departments/Affiliated Programs
- CUNY-CREST Institute
- Earth System Science and Environmental Engineering

Areas of Expertise/Research
- Geohazards and Georisks
- Geotechnical Engineering
- Machine Learning

Building
Steinman Hall

Office
103

Phone
(212) 650-5143

Email
tpei@ccny.cuny.edu