Introduction

Site characterization is a cornerstone of geotechnical and rock engineering. “Data-driven site characterization” refers to site characterization methodologies relying on data that can be derived from experimental/site observations, model predictions, and/or experts’ knowledge. For site-specific data, it includes data collected from the current project, data archived from previous stages of the project, and/or data populated from the neighboring site, among others. Despite the rapid proliferation and improvement of characterization techniques and the increasing trend of data availability, it is an open question that what data-driven site characterization (DDSC) can achieve and how useful are the outcomes for practice. Given the ongoing digital transformation in the geoengineering field, this “added value of data” question has been of major interest among research and practical communities, especially in the context of risk and uncertainty quantification.

This machine learning competition is organized as a student competition event at the Geo-risk 2023 Conference which is held in Arlington, Virginia, USA, July 23–26, 2023 (https://www.geo-risk.org/). The contest will last 3 months. Geotechnical engineers (students and practitioners) are invited to take part in the competition. Participants are encouraged to share their methodologies, experience, and relevant materials in the Discussion forums. Kaggle (https://www.kaggle.com) is used as the platform for hosting this competition. The results of the contest will be presented at the conference.

Problem statement

The task is to perform probabilistic three-dimensional geological modeling (i.e., predicting the soil stratification of a site at unobserved locations) and uncertainty quantification using sparse cone penetration testing (CPT) soundings and a limited number of boreholes at selected locations within the site of interest.

The horizontal layout (i.e., plan view) of these in-situ tests is shown in Figure 1. The complete dataset contains 26 CPT–sounding logs and 9 borehole logs. In the contest, all 26 CPT soundings and 5 boreholes out of 9 are provided to the participants for predicting the stratification of the reserved 4 boreholes: BH_1, BH_3, BH_5, and BH_9. The reserved 4 boreholes form a testing set and will be used for evaluating the performance of each participating group. Detailed geometry of the site and location of the in-situ tests are shown in Figure 1. The dataset (26 CPT soundings and 5 boreholes) can be accessed at (https://www.kaggle.com/competitions/georisk2023-3d-geological-modeling-using-cpt-data/data). Participants are asked to develop/apply machine learning, geo-statistical, or other data-driven techniques to:

1) Interpret the geological stratigraphic configuration (i.e., different soil layers and boundary locations along depth direction) of each CPT sounding;
2) Generate the inferred 2D (or 3D, which is preferred if possible) soil stratigraphic profiles at self-defined cross sections (or volume if the 3D model is created) that pass through (or include if the 3D model is created) the 4 reserved testing boreholes;
3) Perform uncertainty quantification on the outcome from 2), there is no specific requirement on the uncertainty quantification technique;
4) Report the final deliverable in an extended abstract in English with maximum 2 pages of text and a file of PowerPoint slides submitted to the Georisk 2023 student contest session and three groups will be chosen to present the results.

Let's hope some novel/improved data-driven techniques can help us to do a fantastic interpretation and prediction job on this site that can add great value for making site characterization decisions.

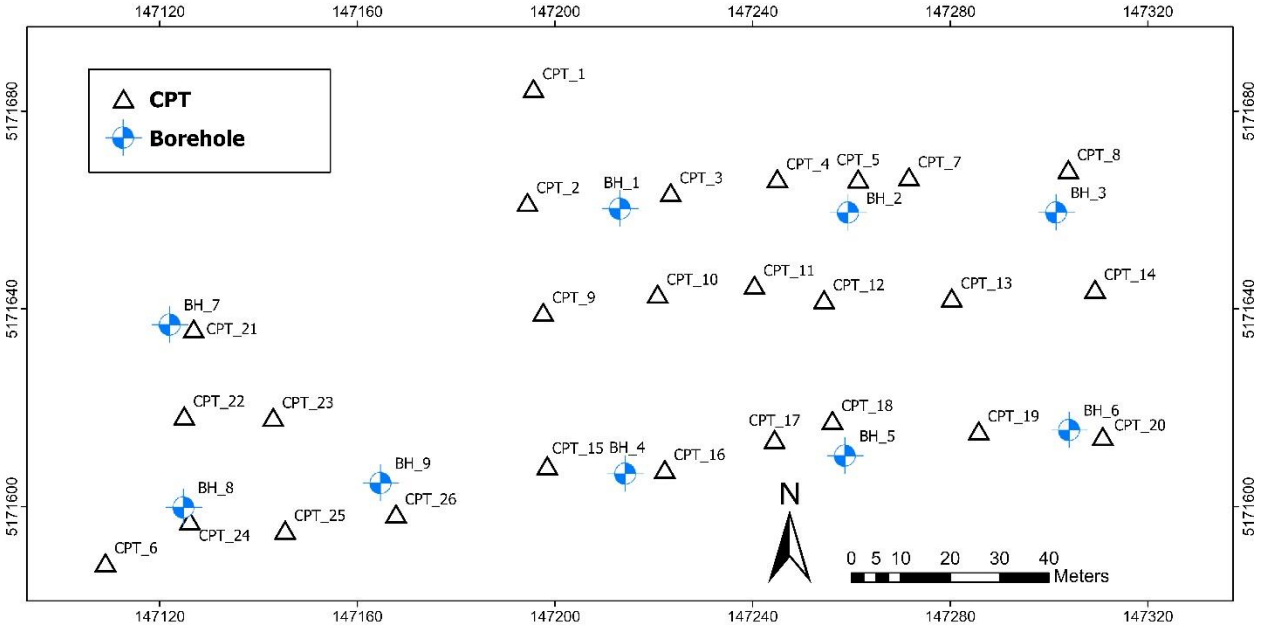

Figure 1. Spatial distribution of CPT soundings and boreholes.

Guidelines for participation

Target audience

Geotechnical engineers (students and practitioners) are invited to take part in the competition. The target teams or individuals should have some background in using machine learning, random field, and geostatistics for solving geotechnical site characterization problems. It is recommended that participants have experience working with 1D, 2D, or 3D geospatial data, Bayesian inference, and uncertainty quantification.

To attend the competition you have to sign in via an invitation-link. This is provided in various geotechnical websites (e.g. Georisk 2023, TC304, and TC309 websites).

Submitting contributions

The participants in the contest are required to submit the following 2 deliverables:

1) A Kaggle project deliverable. Contributions should be submitted via this Kaggle in-class competition. The evaluation criteria are described in detail in the "evaluation" tab.

2) An extended abstract in English with maximum 2 pages of text and a file of PowerPoint slides to hwang12@udayton.edu and Cc'd to yichuan.zhu@temple.edu.

Notice that it is also possible to only deliver in the Kaggle-competition, but your total result will then not be evaluated for presentation at the conference.

Summary and presentations of results

The chosen three best deliveries (based on Kaggle-score, extended abstract, and slides) will be invited to present their work at the conference Georisk 2023 in a special event dedicated to the contest.

A GI RAM-TC304-TC309 award committee will review the extended abstracts/slides and select the winner(s) of the Georisk 2023 event. An award certificate will be given to the winner during the conference. Depending on the number of participants, several encouragement awards may also be given.

The winning teams will be invited to summarize the results into a technical paper.

Communication and support platform

The discussion forum on the Kaggle platform will be used for questions, support, and hints for improvement during the event.

Scoring criteria

Participants are required to provide the probability of observing soil types along depth at testing boreholes. This will be used to evaluate two factors: (1) the accuracy of the estimation of stratification at borehole locations, and (2) the precision in the estimation of stratification. While multiple thin layers or only few layers can be reasonable, any over- or under-estimation will be evaluated by the competition judges. Additionally, the robustness of the methodology will also be assessed when the testing borehole has fewer CPT surroundings, as in the case of BH_9 compared to BH_5 in Figure 1.

Timeframe and important dates

Applicants fill out the participation form by March 31.
The contest event runs from 03/21/2023 to 06/21/2023, a total of 3 months.
06/21/2023: deadline of deliveries.
06/23/2023 – 06/30/2023: the competition judges will evaluate the received deliveries.
07/01/2023: the committee will announce the three best deliveries, and invite the winning teams to present their result at a special event on 07/25/2023.