TC corner

ISSMGE TC309/TC304/TC222 Third ML dialogue on "Data-Driven Site Characterization (DDSC)"

Introduction

ISSME TC309 (Machine Learning and Big Data) organised a dialogue on 28 July 2019 before the second International Workshop on Machine Learning and Big Data in Geoscience at Tongji University, Shanghai, China, 28 – 30 July 2019. The ISSMGE TC304/TC309/TC210 Second Machine Learning (ML) Dialogue for Geotechnics 2019 was held on 14 December 2019 following the 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019), 11-13 December, National Taiwan University, Taipei, Taiwan. The ISSMGE TC309/TC304/TC222 Third Machine Learning in Geotechnics Dialogue (3MLIGD) was hosted online by the Norwegian Geotechnical Institute on 3 December 2021, 9 am – 12 noon (Oslo).

The organising team for 3MLGID consists of Zhongqiang Liu (Chair), Zijun Cao (Secretary), Kok-Kwang Phoon, and Jianye Ching. It is jointly organised by TC309 chaired by Zhongqiang Liu (Machine Learning and Big Data), TC304 chaired by Jianye Ching (Engineering Practice of Risk Assessment and Management), and TC222 chaired by Magnus Rømoen (Geotechnical BIM and Digital Twin). A total of 15 academics and 21 practitioners took part in this online dialogue (a snapshot of some participants is shown in Figure 1).

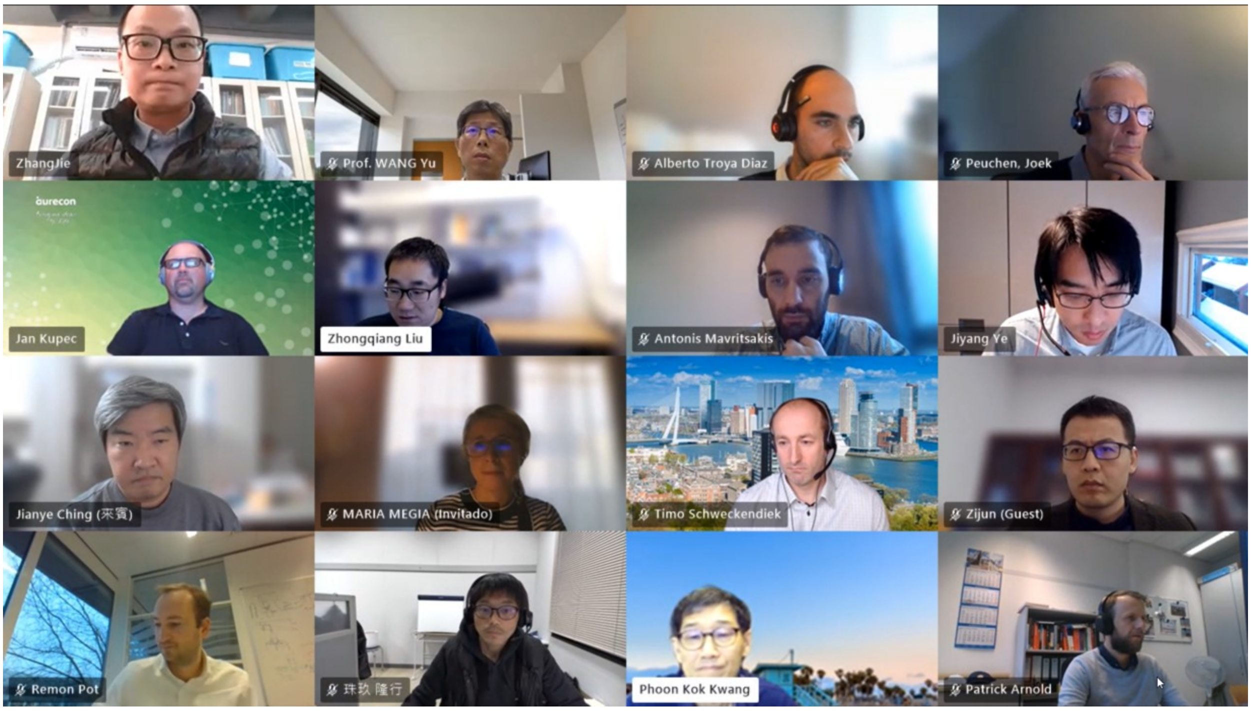

Figure 1. Snapshot of some participants online

The Third Machine Learning in Geotechnics Dialogue (3MLIGD) started off with two short presentations on data-driven site characterization and geotechnical BIM and digital twin to provide a brief overview of the emerging agenda in data-centric geotechnics, challenges, ideas for collaboration, and a glimpse of current and future developments. This is followed by 4 small group discussions and 1 general discussion. The objectives of the dialogue are for academia and industry to jointly identify: (1) topics of potential/actual value to industry, (2) ideas to foster collaboration between academia and industry, and (3) platforms to accelerate progress. A summary of the dialogue is provided below. A full report is scheduled to be published as: Phoon et al. (2022). Report for ISSMGE TC309/TC304/TC222 Third ML dialogue on "Data-Driven Site Characterization (DDSC)". Georisk: Assessment & Management of Risk for Engineered Systems & Geohazards.

TC corner

ISSMGE TC309/TC304/TC222 Third ML dialogue on "Data-Driven Site Characterization (DDSC)" (Con't)

Current Trends

Data-driven site characterization (Kok-Kwang Phoon, Appointed Board Member ISSMGE)
There is no doubt that geotechnical engineering, along with all other industries, will be transformed in the foreseeable future as the digital revolution gathers pace. The intent of data-centric geotechnics is to create a platform for geotechnical practice to leverage on emerging digital technologies such as BIM, digital twin, machine learning, AI, and blockchains to deliver services that meet the need of the end-users and the community rather than to build infrastructure as an end goal. By definition, data-centric geotechnics must adopt a "data first" agenda that is "fit for practice" and addresses real-world challenges.

Research in developing tools for data-centric geotechnics is primarily focused on site characterization currently. Data-driven site characterization (DDSC) is defined as any site characterization methodology that relies solely on measured data, both site-specific data collected for the current project and existing data of any type collected from past stages of the same project or past projects at the same site, neighboring sites, or beyond. Three challenges can be readily identified: (1) ugly data, (2) site recognition, and (3) stratification. The first two challenges are fundamental to data-centric geotechnics and data-driven site characterization. Data from one specific site follow several "ugly" attributes: Multivariate, Uncertain and Unique, Sparse, Incomplete, potentially Corrupted, and spatially variable or MUSIC-3X. Classical statistics is not applicable. A Big Indirect Data (BID) consisting of data from multiple sites is heterogeneous because of inter-site variability. The site recognition challenge involves combining site-specific data with BID in the presence of inter-site variability.

Geotechnical BIM and Digital Twin (Mats Kahlström, TC222 secretary)
Building Information Modelling (BIM) is a relatively new concept to the construction industry and geotechnical engineering. BIM has in recent years become an important topic within infrastructure, transportation, building and construction, earthworks, and buried structures. The degree of uncertainty inherent in geotechnical engineering and site characterization, associated with the need for effective collaboration throughout all project phases, makes BIM an attractive methodology. Further, Digital Shadows and Digital Twins have gained increased interest in recent years, promoting connection between the digital and physical worlds. Using Digital Shadows of Digital Twins in projects may be considered as a way to further develop the use of the observational method in geotechnical engineering.

Small Group Discussions

Following the two presentations, participants were divided into 4 small groups (see Table 1). The structure of the small group discussions took guidance in part from the recommendations made by 2MLIGD:

1. Focus on a project deemed "high value" by the industry.
2. Focus on specific ideas for collaboration between academia and industry.
3. Focus on grand challenges that can be discussed in the 4MLIGD.

Table 1. Organization of group discussions

Group	Name	Affiliation
Overall	Zhongqiang Liu	Norwegian Geotechnical Institute
	Kok-Kwang Phoon	Singapore University of Technology and Design
	Jianye Ching	National Taiwan University, Taiwan
1	Jinhui Li (facilitator)	Harbin Institute of Technology, Shenzhen
	Zijun Cao (reporter)	Wuhan University
	Jan Kupec	Aurecon
	Xu Li	Beijing Jiaotong University
	Remon Pot	Deltares
	Hao Shen	Tetra Tech Coffey, Australia
	Takayuki Shuku	Okayama University
	Limin Zhang	The Hong Kong University of Science and Technology

TC corner

ISSMGE TC309/TC304/TC222 Third ML dialogue on "Data-Driven Site Characterization (DDSC)" (Con't)

2	Timo Schweckendiek (facilitator)	Deltares /Delft University of Technology
	Antonis Mavritsakis (reporter)	Deltares
	Patrick Arnold	GuD
	Yu Wang	City University Hong Kong
	Joek Peuchen	Fugro
	Nezam Bozorgzadeh	Norwegian Geotechnical Institute
	Mingliang Zhou	Tongji University
	Cornel du Toit	Institute of Mine Seismology
3	Jie Zhang (facilitator)	Tongji University
	Monica Löfman (reporter)	Aalto University
	Theo-John Stergianos	Stellenbosch University
	Maria Ferentinou	Liverpool John Moores University
	Niel Marais	SRK Consulting
	Mark Green	Institute of Mine Seismology
	Francois Malan	Institute of Mine Seismology
	Mats Kahlström	Norwegian Geotechnical Institute
	Jiyang Ye	Norwegian Geotechnical Institute
4	María Megía (facilitator)	University of Granada
	Santiago Peña (reporter)	CEDEX-Research Institution
	Carlos Acosta	SAALG Geomechanics
	Alberto Troya	Orsted
	Daniel Taboas	Acciona
	Amadeu Deu	Igeotest
	Candela Sancho	Detektia
	Teresa Mateos	Acciona

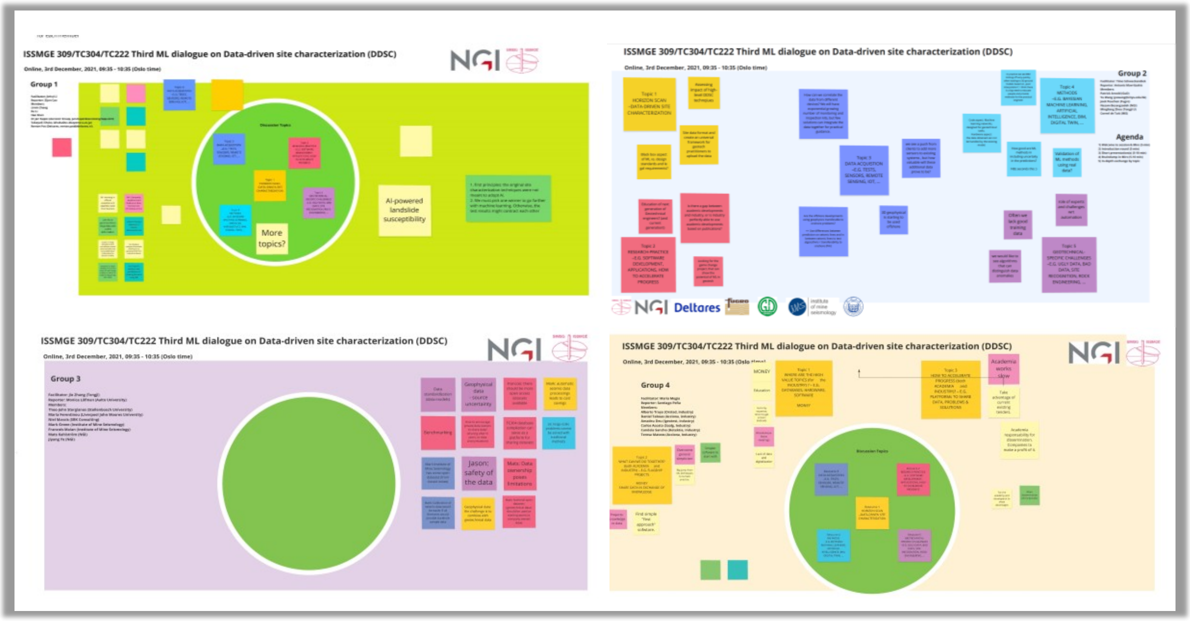

Figure 2. Micro boards for group discussions

TC corner

ISSMGE TC309/TC304/TC222 Third ML dialogue on "Data-Driven Site Characterization (DDSC)" (Con't)

The Miro Board (www.miro.com) was used to aid group discussions, as shown by Figure 2. A summary of the small group discussions, authored by the respective reporter and/or facilitator, is provided below.

Group 1 Discussion – *"Applications of ML in geotechnical engineering" (Zijun Cao & Jinhui Li)*
Facilitator (Jinhui Li, Harbin Institute of Technology, Shenzhen), Reporter (Zijun Cao, Wuhan University)

Group 1 members shared their experiences on ML applications and discussed the values of ML in geotechnical engineering. One member presented recent advances in terrain-based and image-based landslide identification using ML, AI-powered landslide susceptibility assessment and forecasting, an automated snow mapper powered by ML, called AutoSMILE, and an on-going study on earthquake-induced landslides on the Tibetan plateau via transfer learning. These applications and studies demonstrate the values of ML in dealing with advanced and difficult tasks in geotechnical and geological engineering.

Group members also discussed some challenges, such as combining multi-source information, reconciling contradictory data, change of mindset, and education.

Group 2 Discussion – *"Data-driven site characterization" (*Antonis Mavritsakis & Timo Schweckendiek*)*
Facilitator (Timo Schweckendiek, Deltares), Reporter (Antonis Mavritsakis, Deltares)

The discussion of the second group started with a presentation by a participant from the industry that demonstrated that the combination of high-level research and orientation to practice has provided promising ML solutions for offshore site investigation in the form of synthetic CPT generation using CPT and geophysical data. The presentation sparked a discussion focusing on the challenges met in adopting ML in geotechnical practice from broadly two categories: (1) implementation-related challenges encountered upon working with geotechnical data and (2) challenges met in bringing ML applications from theory to practice.

Group 3 Discussion – *"Data sharing between industry and research"* (Monica Löfman & Jie Zhang)
Facilitator (Jie Zhang, Tongji University), Reporter (Monica Löfman, Aalto University)

The discussion in Group 3 focused in data availability, applications of seismic data processing, and interfaces between big data (e.g., geophysical data) with limited data (e.g., borehole samples). Regarding datasets to be used in ML applications, three challenges were identified: (1) lack of open access datasets, (2) lack of data standardization, and (3) data safety.

The group concluded that ML applications will transform geotechnical engineering, partly because certain modern problems (e.g., high-speed railways) cannot be solved with traditional methods.

Group 4 Discussion – *"Acceleration of ML applications in industry" (*Santiago Peña & María Megía*)*
Facilitator (Maria Megia, University of Granada), Reporter (Santiago Peña, CEDEX)

This group consisted of two academics and six practitioners. Members opined that "high value" topics are typically measured by return to a company. Besides value, there is reluctance to apply ML because of a lack of education, lack of proven value, inability to meet deadlines, etc. To foster more collaboration between academia and industry, the recommendations are to leverage on supranational platforms of cooperative projects (e.g. https://project-geolab.eu/) and to incentivise companies to share data by showing how ML can lead to a competitive advantage. The group also suggested promoting workshops to showcase successful projects and to educate the community on usefulness of ML tools.

General Discussion and Recommendations

The number of practitioners and researchers who are exploring and developing digital technologies for geotechnical engineering remains a small minority and even their "blue sky" work is somewhat constrained

TC corner

ISSMGE TC309/TC304/TC222 Third ML dialogue on "Data-Driven Site Characterization (DDSC)" (Con't)

by existing data, existing paradigms, existing technologies, and existing rules of the game (somewhat different for practitioners and researchers), rather than completely "out of the box". The agenda for this data-centric geotechnics is emerging and certainly far from mature. There is a unanimous view that this small community can achieve greater impact and can hasten progress by fostering more collaborations and working more closely together.

Four basic issues on data access, standardization, quality, and protection were discussed. Data is a core asset in digital transformation. It is not surprising that most of the issues are related to data. Some recommendations could be gleaned from the small group and general discussions. The common theme is to find novel ways to foster greater connectivity between researchers, between practitioners, and between researchers and practitioners to leverage on the limited resources and talents at the disposal of this fledging "data-centric geotechnics" community.

1. The database sharing platform hosted by TC304, 304dB, should be further enhanced, because data access remains a critical gap (http://140.112.12.21/issmge/tc304.htm).
2. TC309/TC304/TC222 should host a contact list of people in academia/industry and their current interests to encourage people with common interests to connect and work together. An initial list consisting of participants in current 3MLIGD is shown in Table 1. Participants from past MLIGD could be added.
3. TC309/TC304/TC222 should host a list of ML projects to bring greater awareness of ongoing work and to connect projects/fundings with people or to learn from project outcomes.
4. Novel collaboration modes can be explored: "shadow project", "ML supremacy" project, "Learn and work" projects to grow a pipeline of talents. "ML supremacy" refers to important projects where physics-based approaches are impractical or even impossible to apply, but data-driven approaches can provide highly effective solutions.

Acknowledgments

The 3MLIGD was supported by ISSMGE TC309 (Chair, Dr Zhongqiang Liu), TC304 (Chair, Prof Jianye Ching), and TC222 (Chair, Dr Magnus Rømoen). It could not be held successfully without the enthusiastic participation and contributions from all the invited experts (refer to the list of 3MLIGD participants in Table 1). The organizers would like to thank Dr Mats Kahlström (TC222 secretary) for giving the presentation before group discussions and providing a summary of the presentation in this report. Special thanks go to group facilitators and reporters for their significant efforts in coordinating group discussions and contributing a summary of their discussions to this report. The organizers also extend our appreciation to Dr Jiyang Ye (Norwegian Geotechnical Institute) for making the online arrangements and contributing to the successful hosting of 3MLIGD.