



A Special Issue on Data Science for Geosciences

Vasily Demyanov¹ · Erwan Gloaguen² ·
Mikhail Kanevski³

Published online: 19 December 2019

© International Association for Mathematical Geosciences 2019

The role of data science has increased in the geosciences given the mounting recognition of the limitations associated with traditional physical model driven methods in describing complex phenomena and in explaining large amounts of conditioning Earth monitoring data, which have become readily available in the digital era. Data-driven approaches have the capacity to explain otherwise poorly understood, highly uncertain natural phenomena solely from the data, while physics-driven models lack precision in reproducing the data due to uncertainty in the governing equations and their parameters. Machine learning is currently used to explain complex phenomena in many scientific applications and in Earth sciences more specifically, and even to discover new phenomena (e.g., drug discovery) (Bergen et al. 2019). Machine learning suffers from well-known drawbacks, such as a lack of interpretability, since data-driven methods largely focus on statistical relations, rather than on causation. Significant recent efforts in computer science aim at tackling this challenge. Hybrid approaches that combine the predictive power of data-driven methods and the ease of communication of model-driven approaches are the most promising.

✉ Vasily Demyanov
v.demyanov@hw.ac.uk

Erwan Gloaguen
erwan.gloaguen@ete.inrs.ca

Mikhail Kanevski
mikhail.kanevski@unil.ch

¹ Heriot-Watt University, Edinburgh, UK

² Institut National de la Recherche Scientifique, Quebec City, Canada

³ University of Lausanne, Lausanne, Switzerland

This special issue provides a snapshot of emerging data science links with geoscience applications. It includes both the adaptation of data science technology to the domain content and challenging case studies in geosciences. This special issue takes its roots from the contributions to a related session at the IAMG 2018 annual conference, where many machine learning applications to different geoscience problems were presented.

A broad range of geoscience problems can be classified into one of the following three types: (1) discover, (2) describe, (3) predict.

1. *Discover* Discovering patterns in geoscience data is a natural application for machine learning, which empowers experts with a relatively unbiased way to interpret the data and help account for the associated uncertainty more accurately. The paper of Julie Halotel, Vasily Demyanov, and Andy Gardiner presents a comparative machine learning application to facies classification that demonstrates how using additional engineered features and geological knowledge can help improve pattern discovery.
2. *Describe* Describing geoscience phenomena in a mathematical/computational form is the ultimate goal of Earth modelling. Machine learning helps to avoid constraints and conceptual uncertainty of parametric modelling and the limitation of poorly justified assumptions such as linearity, stationarity, gaussianity, etc. This Special Issue includes several papers that show how machine learning can help describe prior information in seismic studies. The paper of Tue Holm-Jensen and Thomas Mejer Hansen suggests the use of PCA for full waveform compression and a ridge regression to represent a non-linear forward model incorporating prior information for tomographic data. This increases the resolution of the posterior probability density and is computationally faster than existing full waveform inversion. The paper of Lukas Mosser, Olivier Dubrulle, and Martin J. Blunt applies a deep learning method, specifically, name generative adversarial networks, to describe geological prior information to perform Bayesian inversion for acoustic wave propagation.
3. *Predict* The prediction of Earth properties or the Earth system has a paramount importance for all stages of decision making when interacting with nature. This Special Issue includes three papers that address different aspects of predictions with different machine learning techniques. The paper by Manik Singh and Sanjay Srinivasan uses a machine learning proxy to predict a seismic response from a hydraulic fracture propagation. The paper by Elena Baglaeva, Alexander Sergeev, Andrey Shichkin, and Alexander Buevich demonstrates how data sampling for training a multi-layer perceptron impacts the prediction of spatial distribution of heavy metals in soils with real data from urban locations. The final paper of the Special Issue by Milan Zukovic, Michal Borovsky, Matus Lach, and Dionissios Hristopoulos tackles the problem of speeding up the automatic spatial interpolation of massive million-point datasets with a modified planar rotator parallelised for a GPU.

Overall, the collection of papers in this Special Issue demonstrates how data science methods can help tackle computational and uncertainty quantification problems in geosciences with great success.

Reference

Bergen Karianne J, Johnson Paul A, de Hoop Maarten V, Beroza Gregory C (2019) Machine learning for data-driven discovery in solid Earth geoscience. *Science*. <https://doi.org/10.1126/science.aau0323>