GEOLOGICAL SURVEY OF NORWAY
- NGU -

NTNU

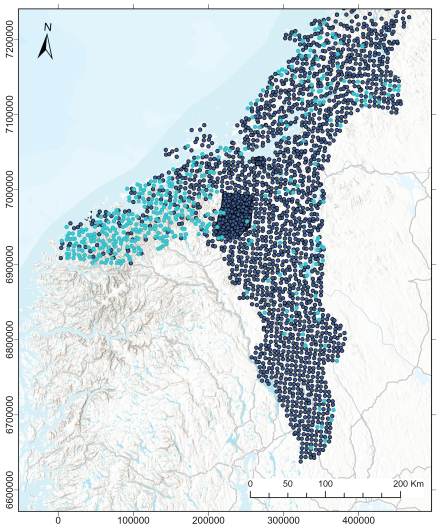
M.Sc Project

Can we “find” the “missing” data?

Soil sample

Location of soil samples

Samples with “censored” Arsenic



Input “missing”/ ”censored” values

Log-ratio transformation

Machine-learning/ Deep learning

In geoscience is common to have incomplete compositional datasets, i.e., datasets having “missing ” and/or “censored” values. Censored data refers to such low concentrations of elements that these cannot be quantified by our instruments, whereas “missing” data correspond with elements concentrations that were not measured. These incomplete data-sets pose a serious challenge to multivariate analysis, as log-ratio transformations and machine-learning techniques can only be applied to complete datasets.

In this project, the student’s goal will be to explore between several uni- and multivariate statistical approaches to replace “censored” and “missing” data with synthetically created values. Each approach can be evaluated on real geochemical datasets, where statistical robustness and geological suitability can be assessed. If you are curious about how your mathematical skills can help us to e.g., find new ore deposits, and identify environmentally vulnerable areas, come and join us at the Geological Survey of Norway!

Contact information:

Dr. Håkon Tjelmeland, haakon.tjelmeland@ntnu.no
Dr. Pedro Acosta-Gongora, pedro.acosta-gongora@ngu.no
Dr. Ana Carolina Rodrigues Miranda, ana.miranda@ngu.no