Near-surface systems can be complex and highly heterogeneous. The complex nature of these systems makes their numerical modelling a challenging problem in geosciences. Geophysical survey methods combined with direct measurements have been widely used to characterize the spatial distribution of the near-surface physical properties. Within this scope, geophysical inversion has been a preferable tool to predict quantitively the spatial distribution of the relevant near-surface properties. Ensemble-based data assimilation techniques are common geophysical inversion methods used in problems related to subsurface modelling and characterization. These methods allow the accurate prediction of the spatial distribution of the subsurface properties and have the ability to assess the uncertainty about the model predictions. However, these are computationally demanding inversion techniques, which makes their applicability to large data sets prohibitive.

This study presents the application of a computationally efficient ensemble-based data assimilation technique for inversion of a large-scale frequency-domain electromagnetic induction survey data set. The inversion method is based on randomized high-order singular value decomposition. We combine randomized linear algebra with high-order singular value decomposition, which allows to perform data assimilation in a low-dimensional model and data space. This inversion approach satisfies two objectives: it reduces the computational burden of the inversion and has the same characteristics as conventional ensemble-based data assimilation methods. The inversion method presented herein predicts the spatial distribution of subsurface electrical conductivity and magnetic susceptibility from frequency-domain electromagnetic induction data (as related to the in-phase and quadrature FDEM signal components).

The method is illustrated in a three-dimensional real case application where a set of geophysical and borehole data is available. The log-set composed by electrical conductivity and magnetic susceptibility is used as conditioning data to generate a prior ensemble of numerical three-dimensional models with geostatistical simulation. The predicted posterior distribution generates synthetic frequency-domain electromagnetic induction data that reproduces the observed data. The model predictions at a blind well location, not used in the generation of the prior ensemble, agree will with the observed log data, validating the quality of the applied method.