

Machine learning for controlled source RMT data selection

A. Platz¹ and U. Weckmann^{1,2}

¹Helmholtz Centre Potsdam - German Research Centre for Geosciences (GFZ),
Potsdam Germany, aplatz@gfz-potsdam.de

²University of Potsdam, Institute of Geosciences, Potsdam, Germany

SUMMARY

The Radio-Magnetotelluric (RMT) method is a geophysical near-surface imaging technique using the signal of radio transmitter as source. While in Europe and Central America there is a sufficient coverage of radio transmitters in the frequency range of 10kHz - 1MHz, in many other areas it is often insufficient. In these cases the use of either a horizontal magnetic or a horizontal electric dipole source is essential to generate the necessary source signal. As an additional benefit, the use of an artificial transmitter allows to extend the frequency range to lower frequencies. When using GFZ's horizontal magnetic dipole transmitter consisting of two perpendicular loops, eight different frequencies between 1kHz - 128kHz are emitted. As the different frequencies are transmitted in a sweep, only a smaller fraction of the entire recorded time series contains the required signal. Using statistical approaches for time series analysis results in a majority of events without usable signal for the target frequency. Picking and flagging events with signal in a manual fashion would be time consuming and a tedious approach; therefore we apply machine learning for the selection of suitable time segments (or events in frequency domain). A feasibility study using recurrent neural networks and auto- and cross-spectra as input data (features) was already successfully conducted based on data from Chile. However, in addition to the spectra, there exist other data that can be used as features, such as physically based quantities as e.g. coherences or single event transfer functions as well as different principal components. We performed a comprehensive principal component analysis and an unsupervised k-means clustering, to identify a set of suitable input data. The selected set is used to train different supervised machine learning algorithms as logistic regression, support vector machines, random forest and different kinds of neural networks. We will compare the performances of the trained algorithms and discuss possible limitations.

Keywords: Controlled source RMT, machine learning, principal component analysis, clustering, supervised learning
