Use of machine learning algorithms as a tool for interpretation of helicopter-borne electromagnetic data

O. Cortés Arroyo[1], E. Schönfeldt [1], B. Siemon[2], M. Fahle[1] and S. Janetz[1]

[1] Federal Institute for Geosciences and Natural Resources (BGR),
Sub-department Research and Development Centre for Post-Mining Areas (FEZB),
Olaf.CortesArroyo@bgr.de
[2] Federal Institute for Geosciences and Natural Resources (BGR), Sub-department Geophysical Exploration – Technical Mineralogy, Bernhard.Siemon@bgr.de

Summary

In order to rehabilitate former open-cast, lignite-mining areas, extensive information on aquifers is required, in particular on their mineralization, depth and thickness. Therefore, non-intrusive geophysical methods such as helicopter-borne electromagnetics (HEM) may provide meaningful results, while covering large areas in a short time window.

However, processing, analyzing and interpreting HEM data must be done very carefully and often remains a challenging and time-consuming task. However, current availability of ready-to-use, open-source Machine learning (ML) libraries may provide new, complementary tools for the geophysicist, due to their ability to detect previously unseen patterns and/or to transform high-dimensionality data into an equivalent result that is easier to interpret.

In this regard, the Federal Institute of Geosciences and Natural Resources (BGR) through the Research and Development Centre for Post-Mining Areas (FEZB) implemented the projects “D-AERO-Finsterwalde” and “FINA”.The Project “D-AERO-Finsterwalde” consisted of an HEM survey of a post-mining area located about 60 km southwest of the city of Cottbus, in the german federal state of Brandenburg. Using the helicopter-borne geophysical system of the BGR, electromagnetic data was recorded every 4 meters using a 10-meter-long flight probe (bird), which was towed on a rope at about 50 m above the ground. The helicopter survey acquired data on 85 NW-SE lines (250 m line spacing) and 23 NE-SW tie-lines (min. 625 m line spacing), as well as further 33 lines in three sub-areas at 100 m line spacing, with a total survey area of about 250 km² and approximately 1,300 km-length of flight lines. For each of the measuring points (more than 300,000 in total), the resulting six-frequency dataset was converted into 1D-resistivity models with 20 layers, using a Levenberg-Marquardt procedure that applies stronger vertical constraints, resulting in smoother models. Simultaneously, a magnetometer (in the bird) and a gamma-ray spectrometer (in the helicopter) were used. Information about local geology is included by means of a large database, consisting of more than 15,000 boreholes. HEM results show very good agreement with local geological features and a clear sensitivity of the data to the local underground-water and mineralization processes.

As part of project “FINA”, we make use of the collected information in “D-AERO-Finsterwalde” to estimate if interpretation of the results can be enhanced or complemented by use of ML-algorithms. After a first evaluation, three algorithms were selected: K-Means (K-M), Self-Organized Maps (SOM) and Random Forest (RF). The algorithms implemented are part of the Python Scikit-Learn library. K-M and SOM provided an automatic clustering and classification for all the resistivity models by using resistivity values, resistivity gradient, depths and location as parameters. Results show the presence of two separate regions in the area, each matching well with local properties regarding electrical resistivity, geology and mining history. The RF algorithm was trained to predict porosity three-dimensionally in order to gain information about the distribution of aquifers/aquitards. The model has a horizontal resolution of 50 x 50 m and an 1 m vertical discretization. Predictors for this analysis are resistivity, elevation and groundwater presence/absence. In total 58,000 measuring points were considered, showing a 90% accuracy by using 70% of the data as training data and 30% as test data.

Despite the large amount of processed information, the ML-algorithms provides results in only a few minutes, using a laptop computer. These results encourage us to further explore the application of ML in the HEM processing scheme.

Keywords: HEM, machine learning, interpretation, mineralization, lignite