A Machine Learning Approach to Estimate Plume Discharge from Electrical Geophysical Measurements

Neil Terry (nterry@usgs.gov), Frederick D. Day-Lewis, and John W. Lane Jr. (U.S. Geological Survey, Storrs, Connecticut, USA)
Jared Trost (U.S. Geological Survey, Mounds View, Minnesota, USA)
Isabelle Cozzarelli (U.S. Geological Survey, Reston, Virginia, USA)
Barbara Bekins (U.S. Geological Survey, Menlo Park, California, USA)

Background/Objectives. Geophysical methods for monitoring the distribution of contaminant plumes and their breakdown products offer the possibility of improved risk assessment at reduced cost, provided that the plume exhibits a sufficient contrast in physical or electrical properties with the surrounding material. However, geophysical interpretations are non-unique and generally require support from additional direct measurements and/or spatially extensive data. In this work, we demonstrate a machine learning approach applied to estimating discharge zones from a contaminant plume resulting from the biodegradation of crude oil into lake water at the National Crude Oil Spill Fate and Natural Attenuation Research Site in Bemidji, MN based on frequency domain electromagnetic induction (FDEM), ground penetrating radar (GPR), and physical point-scale electrical conductivity (EC) measurements. Monitoring activities since a 1979 crude oil spill at the site indicate a plume of partial breakdown products of hydrocarbons is now discharging into a nearby lake. An increase in the EC of the groundwater caused by production of carbon dioxide during biotransformation reactions represents a target for detection with electrical geophysical methods.

Approach/Activities. Ground- or water-based FDEM and GPR are geophysical techniques that can be used to rapidly across large areas to acquire data from the near subsurface at sub-meter (m) scale lateral resolution. We collected several line-kilometers of geophysical data over a lake ~325 m downgradient of the spilled oil zones, as well as several vertical EC profiles within the lake-water and lakebed sediments. FDEM data were used to detect variations in bulk EC below the sensor, and GPR data were used to estimate water column and organic lake sediment thickness.

Results/Lessons Learned. Initial plotting of the spatial distribution of FDEM raw quadrature data (a proxy for electrical conductivity below the sensor) showed large anomalies in regions associated with elevated EC and NVDOC profiles measured along the shoreline. However, elevated quadrature values also appeared in unexpected portions of the lake, indicating that FDEM data alone may not be sufficient for identifying plume discharge zones. Lakebed EC profiles, however, indicated two distinct groupings – (1) areas with EC that was relatively consistent throughout the organic layer and (2) areas with EC that increased with depth. The areas that showed increasing lakebed EC with depth also showed elevated FDEM quadrature values and were interpreted as zones of plume discharge. Lakebed EC measurements were thus grouped into two binary classes (class 0 = 'no plume' and class 1 = 'plume', respectively). Points with collocated EC profiles and geophysical data were used to train a logistic regression classification framework to predict areas of plume discharge from the geophysical data alone. Additionally, multiple regression models for each predicted class were developed to estimate profiles of EC vs. depth. Compared to conventional geophysical analysis (i.e., inversion), the approach yielded (1) directly interpretable information for decision making, e.g., contaminated vs. uncontaminated regions; and (2) quantification of prediction uncertainty. Machine learning frameworks are well-suited to handling a variety of data types and have the potential to be updated as new information becomes available.