Geophysical Research Abstracts
Vol. 20, EGU2018-9789, 2018
EGU General Assembly 2018

Geospatial Machine Learning Applied to Predict Seafloor Total Organic Carbon

Taylor Lee and Warren Wood

U.S. Naval Research Laboratory, Geology and Geophysics, United States (taylor.lee@nrlssc.navy.mil)

Total organic carbon (TOC) in seafloor sediment is an indicator of carbon flux (sequestration) from the ocean environment, as well as a precursor to shallow methanogenesis. TOC observations are sparse as result of the general inaccessibility of the seafloor, and expense associated with analysis. As a result, the vast majority of the ocean remains unsampled. Recently, machine learning techniques have been proven successful for similar global estimations of geological properties where little data exists. We have applied a non-parametric (i.e. data-driven) machine learning algorithm, specifically k-nearest neighbors (KNN), to estimate global distribution of seafloor TOC at 5-arc minute resolution. Observed (i.e. training) data uses $\sim$5000 data points and a built-in parameter selection uses a quantitative approach, whereby the number of parameters selected is based upon prediction error of a ranked noise grid. Prediction validation between observed and predicted values exhibited a correlation of $\sim$0.7. Standard deviation and absolute error values per validation were averaged over bins with width 0.5 % dry weight observed TOC. We find that lower seafloor TOC values show similar trends in absolute error and standard deviation suggesting that standard deviation in the nearest neighbors may approximate prediction error (i.e. uncertainty).

We also compute a new metric, parametric isolation, for each prediction. Parametric isolation reflects how well parameter (i.e. feature) space is sampled by observations and is calculated for each prediction value as the distance in parameter space to its single nearest neighbor. Parametric isolation specifically indicates where (geographically) to sample data next to most improve the sample density in parameter space. Data acquisition at locations with high inexperience are most optimal in furthering future predictions, not only for the sampled geographic area, but also for areas that are geologically similar (i.e. proximal in parameter space).

Our prediction provides data-driven feature selection and estimates of seafloor TOC with uncertainty in areas where previously large data deficiencies exist. Our complete seafloor TOC prediction with uncertainty contributes to furthering global and regional modeling efforts in the subsurface marine environment. Current efforts are focused on a subsurface time/depth component to model organic matter degradation into subsequent methane and/or methane hydrate accumulation.