

Applying machine learning to quantify CaCO_3 and TOC data from XRF core scanning spectra of sediment records

*Weng-Si Chao¹, An-Sheng Lee^{3,2}, Ralf Tiedemann¹, Bernd Zolitschka², Sofia Ya Hsuan Liou³, Lester Lembke-Jene¹

1. Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany, 2. University of Bremen, Institute of Geography, Bremen, Germany, 3. National Taiwan University, Department of Geosciences and Research Center for Future Earth, Taipei, Taiwan

Obtaining quantitative data of e.g. marine sedimentary calcium carbonate (CaCO_3) content requires time-consuming laboratory measurements of total carbon (TC) and total organic carbon (TOC), ultimately restricting high temporal resolution and dense spatial representation for sediment records. A common way to gain higher resolution and less time-consuming data is calibrating calcium counts acquired by X-ray fluorescence (XRF) core scanning. This is a rapid and non-destructive method producing semiquantitative elemental variations. For calibration, only a limited number of discrete and quantitatively measured conventional samples are necessary. However, this calibration method depends on a number of uncertainties including the choice of an adequate regression algorithm and software-based elemental intensity extraction from XRF spectra, which require careful manual settings and experience. Hence, the consecutive calibration becomes subjective and biased. We present a novel method to obtain quantitative CaCO_3 and TOC data by using machine learning (ML) techniques directly on Avaatech-XRF scanner-derived spectra and in combination with quantitative data. This avoids the manual bias of software fine-tuning as well as selecting an appropriate regression algorithm. Our dataset consists of samples from marine sediment cores recovered from high- to mid-latitudes of the Pacific Ocean (northwest Pacific to the South Pacific Ocean), representing a regional coverage. Several ML algorithms were tested to find the best combination for model building and result in this optimal combination: Non-negative matrix factorization (NMF) is integrated as pre-process with the support vector machine (SVM) to build calibration models. The optimal models are carefully evaluated in the training set, test sets and the case study. Resulting performances are notable (R^2 of CaCO_3 : 0.96 and of TOC: 0.78). This allows to generate high-resolution bulk chemistry records by applying XRF scanning data without losing accuracy and documents the large potential of applying ML techniques in the geosciences.

Keywords: Avaatech-XRF, calcium carbonate, TOC, machine learning, quantification