

EGU22-4495

<https://doi.org/10.5194/egusphere-egu22-4495>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



## Machine learning-based modeling of deep-sea polymetallic nodules spatial distribution: spatial autocorrelation and model transferability at local scales

Iason - Zois Gazis<sup>1</sup> and Jens Greinert<sup>1,2</sup>

<sup>1</sup>DeepSea Monitoring Group, GEOMAR Helmholtz Centre for Ocean Research Kiel, DeepSea Monitoring group, Kiel, Germany (igazis@geomar.de)

<sup>2</sup>Institute of Geosciences, Christian Albrechts University Kiel, Ludwig-Meyn-Str. 10-12, 24098 Kiel, Germany (jgreinert@geomar.de)

The spatial distribution of deep-sea polymetallic nodules (PMN) is of high interest due to increasing global demand in metals (Ni, Co, Cu), and their significant contribution to deep-sea ecology as hard-substrate. The spatial mapping is based on a combination of multibeam echosounders and underwater images in parallel to traditional ground-truth sampling by box coring. The combined analysis of such data has been advanced by using machine learning approaches, especially for automated image analyses and quantitative predictive mapping. However, the presence of spatial autocorrelation (SAC) in PMN distribution has not been extensively studied. While SAC could provide information regarding the patchy distribution of PMN and thus enlighten the variable selection before machine learning modeling, it could also result in an over-optimistic validation performance when not treated carefully. Here, we present a case study from a geomorphologically complex part of the Peru Basin. The local Moran's I analysis revealed the presence of SAC of the PMN distribution, which can be linked with specific seafloor acoustic and geomorphological characteristics such as aspect and backscatter intensity. A quantile regression forests (QRF) model was developed using three cross-validations (CV) techniques: random-, spatial-, and feature space cluster-blocking. The results showed that spatial block cross-validation is the least unbiased method. Opposite the commonly used random-CV overestimates the true prediction error. QRF predicts well in morphologically similar areas, but the model uncertainty is high in areas with novel feature space conditions. Therefore, there is the need for dissimilarity analysis and transferability assessment even at local scales. Here, we used the recently proposed method "Area of Applicability" to map the geographical areas where feature space extrapolation occurs.