Application of machine learning to map the global distribution of deep-sea sediments

Markus Diesing

Geological Survey of Norway
markus.diesing@ngu.no

GEOLOGICAL SURVEY OF NORWAY
- NGU -

mareano
collecting marine knowledge

https://mareano.no/en

Abstract

EGU2020-6755
https://doi.org/10.5194/egusphere-egu2020-6755
EGU General Assembly 2020

https://doi.org/10.5194/egusphere-egu2020-6755

Application of machine learning to map the global distribution of deep-sea sediments

Markus Diesing
Geological Survey of Norway, Trondheim, Norway, (markus.diesing@ngu.no)

The deep-sea floor accounts for >90% of seafloor area and >70% of the Earth's surface. It acts as a receptor of the particle flux from the surface layers of the global ocean, is a place of biogeochemical cycling, records environmental and climate conditions through time and provides habitat for benthic organisms. Maps of the spatial patterns of deep-sea sediments are therefore a major prerequisite for many studies addressing aspects of deep-sea sedimentation, biogeochemistry, ecology and related fields.

A new digital map of deep-sea sediments of the global ocean is presented. The map was derived by applying the Random Forest machine-learning algorithm to published sample data of seafloor lithologies and environmental predictor variables. The selection of environmental predictors was initially based on the current understanding of the controls on the distribution of deep-sea sediments and the availability of data. A predictor variable selection process ensured that only important and uncorrelated variables were employed in the model. The three most important predictor variables were sea-surface maximum salinity, sea-floor maximum temperature and bathymetry. The occurrence probabilities of seven seafloor lithologies (Calcareous sediment, Clay, Diatom ooze, Lithogenous sediment, Mixed calcareous-siliceous ooze, Radiolarian ooze and Siliceous mud) were spatially predicted. The final map shows the most probable seafloor lithology and an associated probability value, which may be viewed as a spatially explicit measure of map confidence. An assessment of the accuracy of the map was based on a test set of observations not used for model training. Overall map accuracy was 69.5% (95% confidence interval: 67.9% - 71.1%). The sea-floor lithology map bears some resemblance with previously published hand-drawn maps in that the distribution of Calcareous sediment, Clay and Diatom ooze are very similar. Clear differences were however also noted: Most strikingly, the map presented here does not display a band of Radiolarian ooze in the equatorial Pacific.

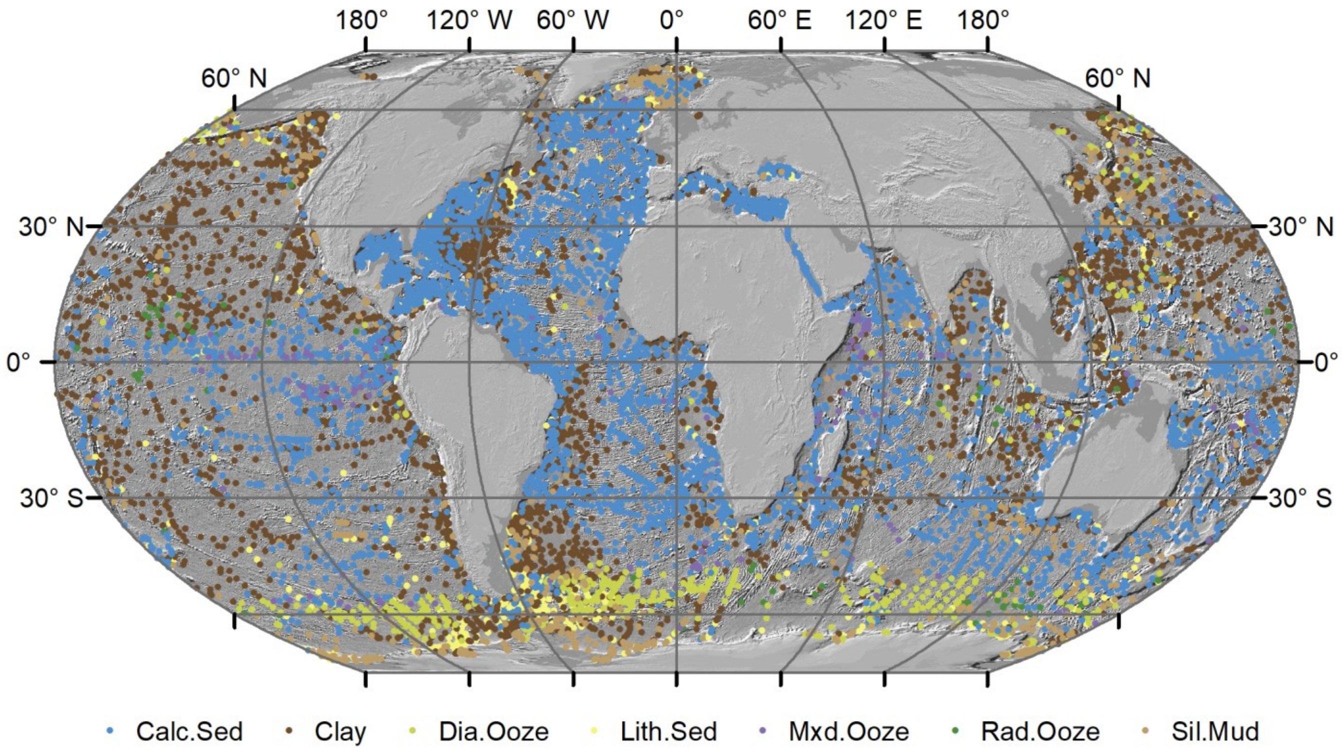
The probability surfaces of individual seafloor lithologies, the categorical map of the seven mapped lithologies and the associated map confidence will be made freely available. It is hoped that they form a useful basis for research pertaining to deep-sea sediments.

Seafloor lithologies

Table 2: Seafloor lithology classes used in this study, their abbreviations, their relationships to classes in Dutkiewicz et al. (2015) and the number and percentage of samples in the training and test datasets. Not included are Ash and volcanic sand/gravel, Sponge spicules and Shells and coral fragments of the original classification.

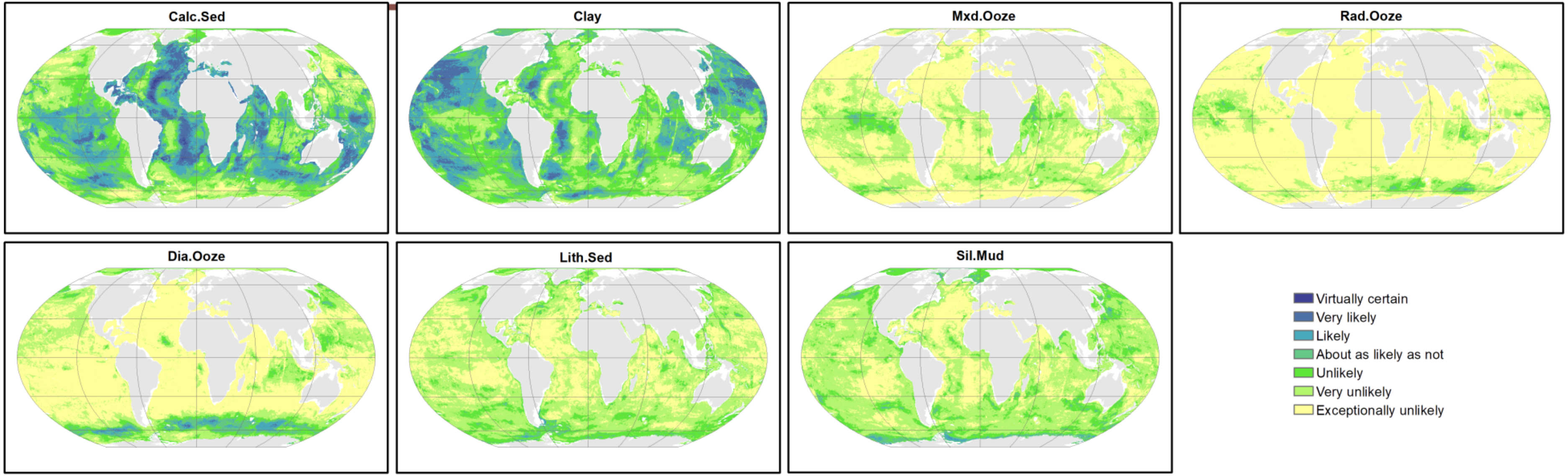
Lithology class	Abbreviation	Relation to Dutkiewicz et al. (2015)	Training dataset	Test dataset
Calcareous sediment	Calc.Sed	Calcareous ooze Fine-grained calcareous sediment	3029 (44.8 %)	1512 (44.8 %)
Clay	Clay	Clay	2219 (32.8 %)	1108 (32.8 %)
Diatom ooze	Dia.Ooze	Diatom ooze	361 (5.3 %)	180 (5.3 %)
Lithogenous sediment	Lith.Sed	Gravel and coarser Sand Silt	438 (6.5 %)	218 (6.4 %)
Mixed calcareous-siliceous ooze	Mxd.Ooze	Mixed calcareous-siliceous ooze	123 (1.8 %)	62 (1.8 %)
Radiolarian ooze	Rad.Ooze	Radiolarian ooze	60 (0.9 %)	30 (0.9 %)
Siliceous mud	Sil.Mud	Siliceous mud	539 (8.0 %)	269 (8.0 %)

Samples



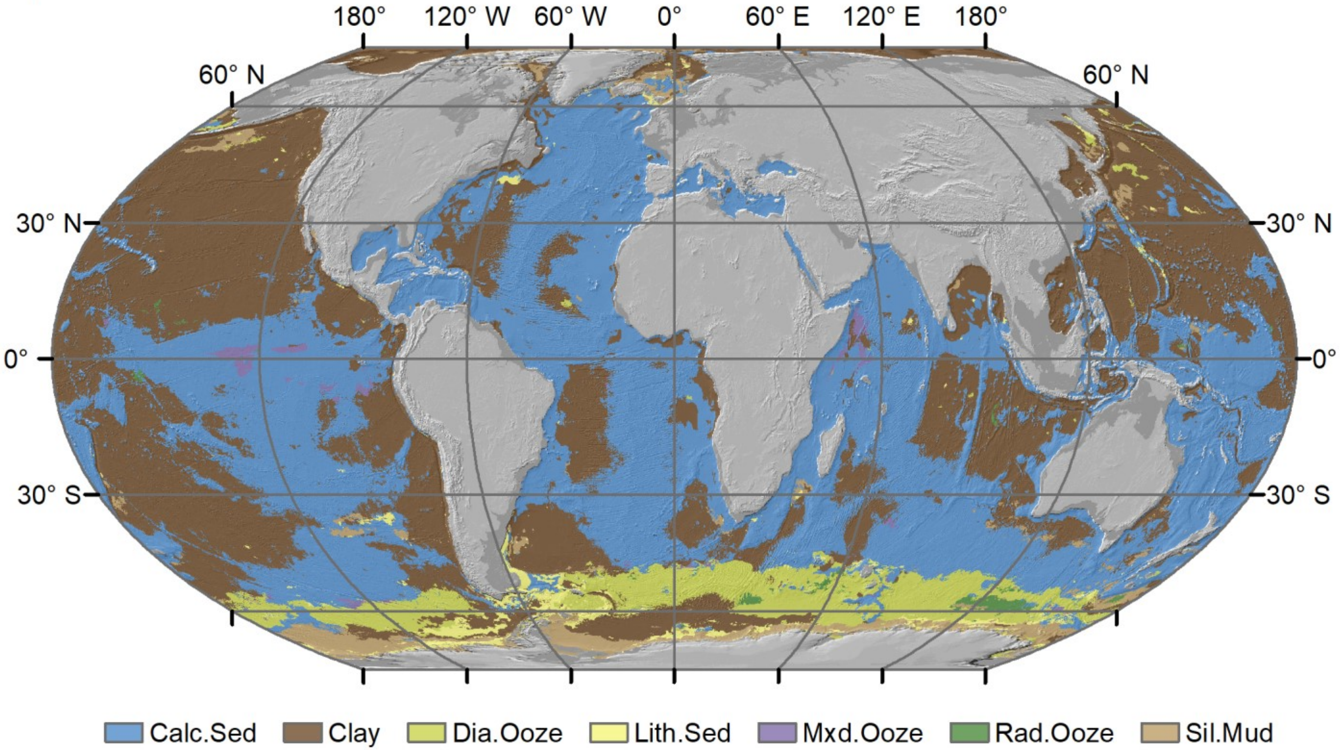
Source: Dutkiewitcz et al. (2015): https://doi.org/10.1130/G36883.1

Predicted probabilities

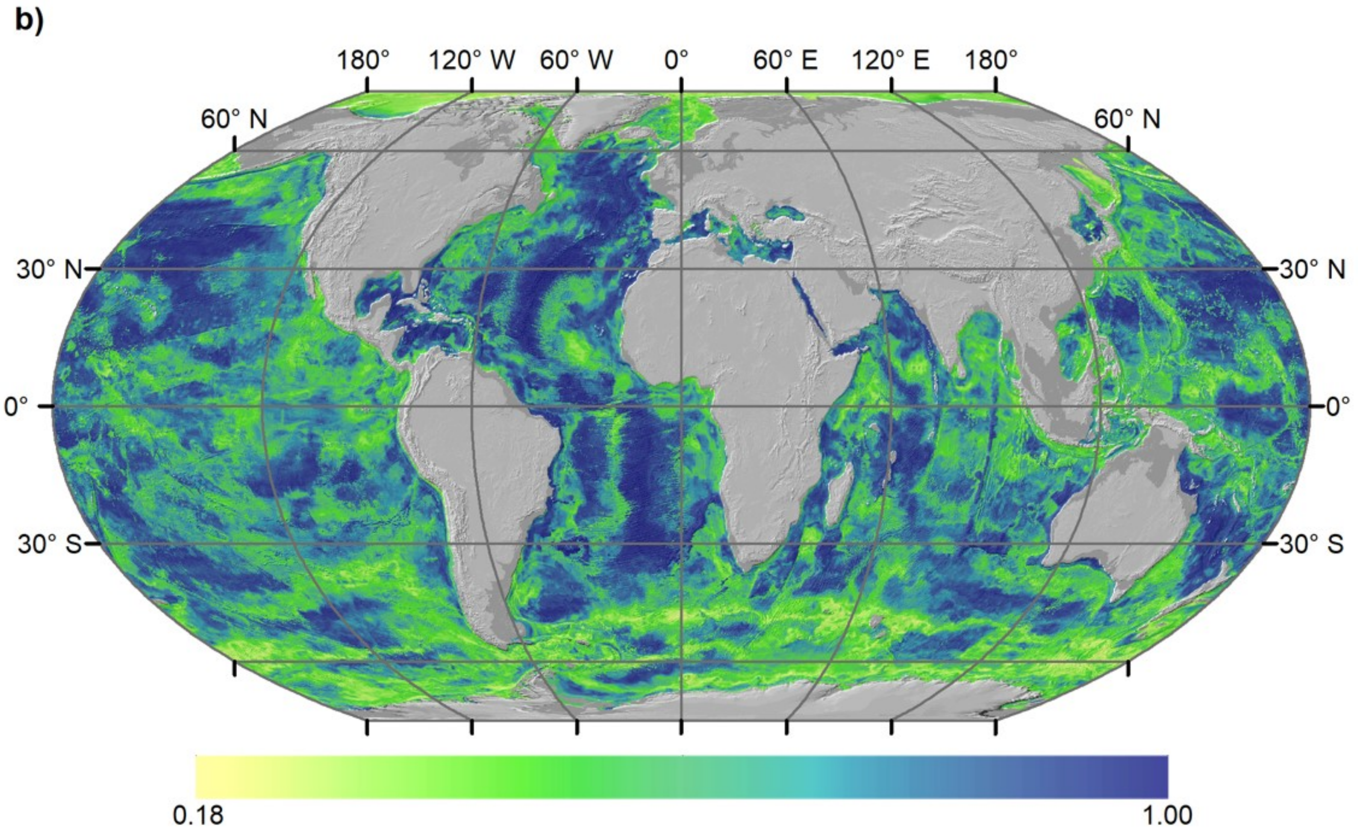


Seafloor lithology

a)



Map confidence



Confusion matrix

Table 3: Confusion matrix. Observed (reference) classes are in columns, predicted classes in rows.

	Calc.Sed	Clay	Dia.Ooze	Lith.Sed	Mxd.Ooze	Rad.Ooze	Sil.Mud	Row total	Error of commission
Calc.Sed	1316	229	11	64	47	5	52	1724	0.237
Clay	136	789	41	71	5	11	87	1140	0.308
Dia.Ooze	17	27	106	19	3	7	17	196	0.459
Lith.Sed	12	27	10	36	0	0	18	103	0.650
Mxd.Ooze	10	3	4	0	6	0	1	24	0.750
Rad.Ooze	0	1	0	0	1	3	1	6	0.500
Sil.Mud	21	32	8	28	0	4	93	186	0.500
Column total	1512	1108	180	218	62	30	269		
Error of omission	0.130	0.288	0.411	0.835	0.903	0.900	0.654		

Preprint

Earth System Science Data
The data publishing journal

https://doi.org/10.5194/essd-2020-22

Submitted as: data description paper

16 Apr 2020

Review status
This preprint is currently under review for the journal ESSD.

Deep-sea sediments of the global ocean

Markus Diesing
Geological Survey of Norway (NGU), Postal Box 6315 Torgarden, 7491 Trondheim, Norway

Received: 04 Feb 2020 – Accepted for review: 16 Apr 2020 – Discussion started: 16 Apr 2020

Abstract. Although the deep-sea floor accounts for more than 70 % of the Earth's surface, there has been little progress in relation to deriving maps of seafloor sediment distribution based on transparent, repeatable and automated methods such as machine learning. A new digital map of the spatial distribution of seafloor lithologies in the deep sea below 500 m water depth is presented to address this shortcoming. The lithology map is accompanied by estimates of the probability of the most probable class, which may be interpreted as a spatially-explicit measure of confidence in the predictions, and probabilities for the occurrence of seven lithology classes (Calcareous sediment, Clay, Diatom ooze, Lithogenous sediment, Mixed calcareous-siliceous ooze, Radiolarian ooze and Siliceous mud). These map products were derived by the application of the Random Forest machine learning algorithm to a homogenised dataset of seafloor lithology samples and global environmental predictor variables that were selected based on the current understanding of the controls on the spatial distribution of deep-sea sediments. The overall accuracy of the lithology map is 69.5 %, with 95 % confidence limits of 67.9 % and 71.1 %.
It is expected that the map products are useful for various purposes including, but not limited to, teaching, management, spatial planning,

Short summary
A new digital map of the sediment types covering the bottom of the ocean has been created....
Read more

https://doi.org/10.5194/essd-2020-22

Data availability

Markus Diesing

PANGAEA.

Data Publisher for Earth & Environmental Science

SEARCH SUBMIT ABOUT CONTACT

Citation: **Diesing, Markus (2020):** Deep-sea sediments of the global ocean mapped with Random Forest machine learning algorithm. *PANGAEA*, https://doi.org/10.1594/PANGAEA.911692

Always quote above citation when using data! You can download the citation in several formats below.

RIS Citation | BibTeX Citation | Copy Citation | Facebook | Twitter

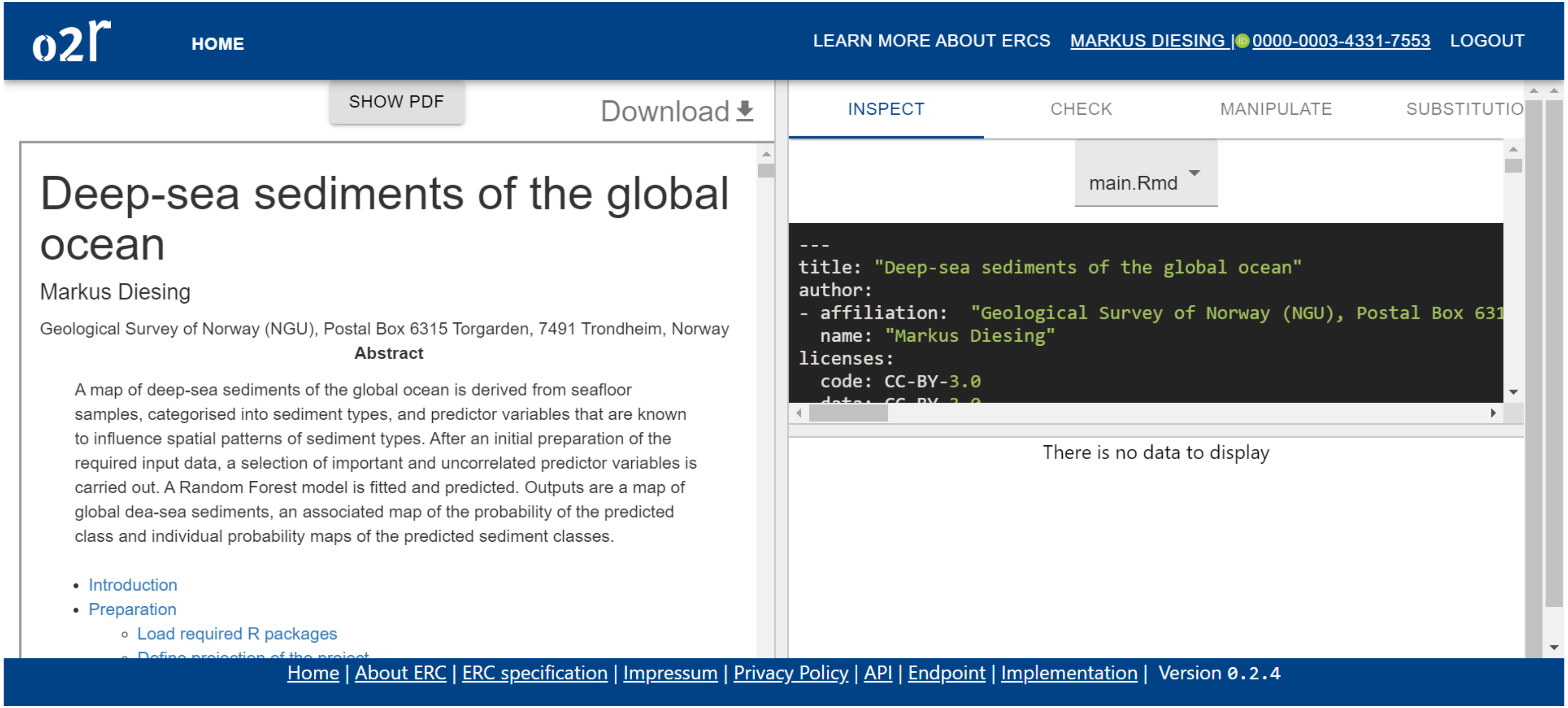
Abstract: The seafloor lithology of deep-sea sediments of the global ocean was spatially predicted. Seven lithology classes were predicted: Calcareous sediment, Clay, Diatom ooze, Lithogenous sediment, Mixed calcareous-siliceous ooze, Radiolarian ooze and Siliceous mud. The dataset contains probability surfaces of the seven seafloor lithologies, the probability of the most probable class (maximum probability) and the predicted seafloor lithology. The results are presented as geo-referenced floating-point TIFF-files with a spatial resolution of 10 km and Wagner IV equal-area projection as spatial reference.

Seafloor lithologies were mapped by building a predictive spatial model. This entails a two-step approach: Initially, the relationship between a set of predictor variables and a response variable is modelled from observations (samples). The established model is then employed to predict the response variable at unsampled locations for which values of the predictor variables are known.

The response variable is seafloor lithology, a qualitative multinomial variable. Seafloor lithology data were sourced from Dutkiewicz et al. (2015) and pre-processed in the following way: Only samples deeper than 500 m were used and a minimum distance between sample locations of

https://doi.pangaea.de/10.1594/PANGAEA.911692

Code availability



https://o2r.uni-muenster.de/#/erc/GWME2voTDb5oeaQFuTWMCEMveKS1MiXm

markus.diesing@ngu.no | www.ngu.no | @markus_diesing

Thank you for your attention!

mareano
collecting marine knowledge

https://mareano.no/en