GOSPODARKA SUROWCAMI MINERALNYMI – MINERAL RESOURCES MANAGEMENT

2023 Volume 39 Issue 2 Pages 5–36

DOI: 10.24425/gsm.2023.145886

MONIKA WASILEWSKA-BŁASZCZYK[1], JACEK MUCHA[2]

Regression methods in predicting the abundance of nodules from seafloor images – a case study from the IOM area, Pacific Ocean

Introduction

On the deep seabed of the Equatorial North Pacific Ocean within the Clarion-Clipperton Fracture Zone (CCZ) (Figure 1) 16 ISA (International Seabed Authority) contractors explore large deposits of polymetallic nodules. Polymetallic nodules resting on the bottom of the ocean are of interest due to their metal contents such as Ni, Cu, Co, and Mn and Rare Earth Elements (REE) (Hein et al. 2020). The exploration area allocated to each contractor is 75,000 square kilometers. The deposits are located at a depth of 4 to 6 km and are sampled directly (physical samples), usually using a box corer (hereinafter referred to as BC) (Sterk and Stein 2015).

✉ Corresponding Author: Monika Wasilewska-Błaszczyk; e-mail: wasilews@agh.edu.pl
[1] AGH University of Science and Techology, Kraków, Poland; ORCID iD: 0000-0003-0042-9782; Scopus ID: 36118314800; Researcher ID: T-8405-2018; e-mail: wasilews@agh.edu.pl
[2] AGH University of Science and Technology, Kraków, Poland; ORCID iD: 0000-0001-6366-083X; Scopus ID: 7005581861; Researcher ID: T-8460-2018; e-mail: jacekm@agh.edu.pl

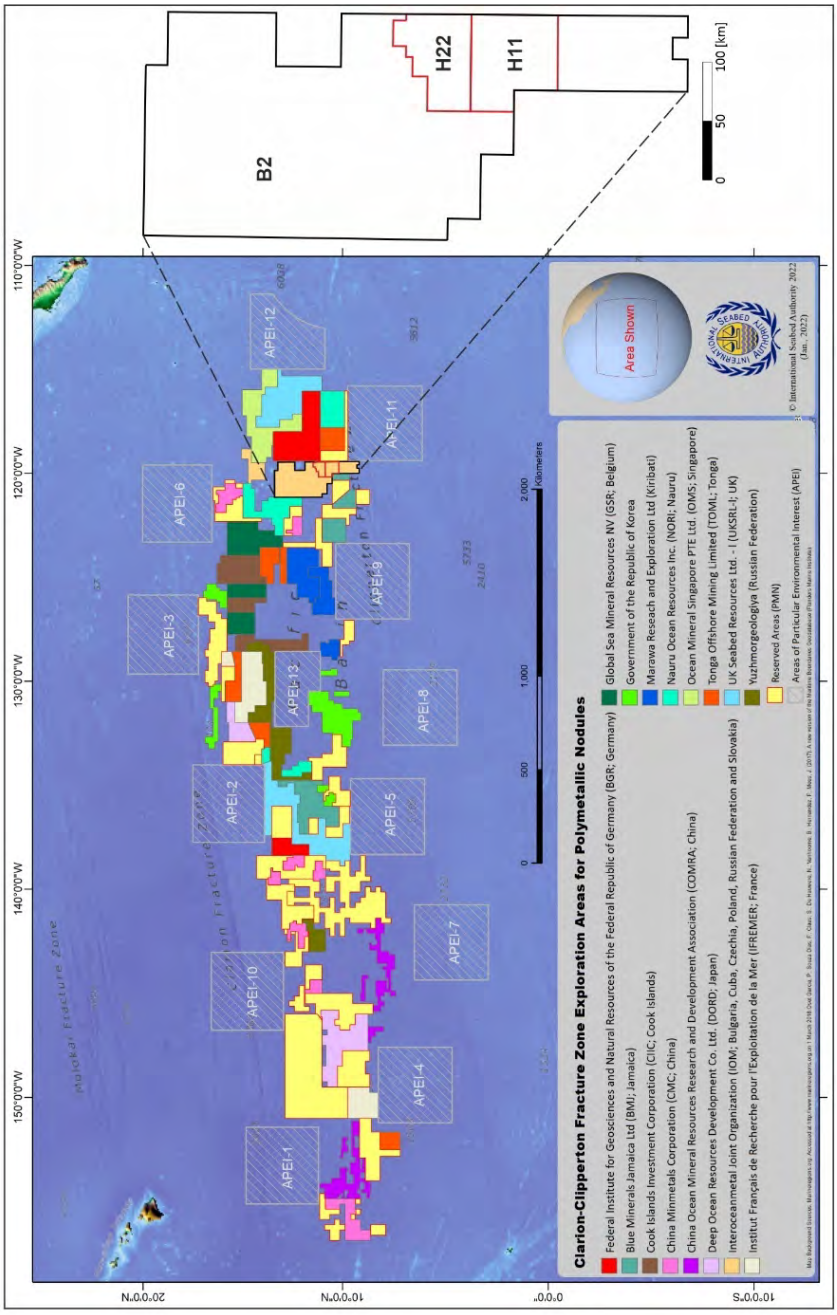

Fig. 1. Location of the B2 Interoceanmetal Joint Organization (IOM) exploration area for polymetallic nodules against the background of the Clarion-Clipperton Fracture Zone exploration areas (Clarion-Clipperton Fracture Zone Exploration Areas for Polymetallic Nodules 2022) and H22 and H11 exploration blocks

Rys. 1. Lokalizacja obszaru poszukiwań konkrecji polimetalicznych B2 Interoceanmetal Joint Organization (IOM) na tle obszarów poszukiwawczych Clarion-Clipperton Fracture Zone (Clarion-Clipperton Fracture Zone Exploration Areas for Polymetallic Nodules 2022) oraz bloków eksploracyjnych H22 i H11

BC sampling is performed in a relatively sparse grid due to the large surface area of the recognized deposits and the high cost of sampling. For example, in the B2 sector administrated by the Interoceanmetal Joint Organization (IOM), the distances between adjacent BC, where the abundance of polymetallic nodules (APN) is determined in situ (i.e. the mass of nodules per unit area), ranges from 3.3 up to 15 km depending on the stage of exploration. The abundance of nodules in this area is highly variable with a coefficient of variation of 60–70% (Mucha et al. 2013). For this reason, in order to recognize the continuity of deposits and to estimate resources, it is necessary to use indirect sampling methods and measurements (acoustic profiling, backscatter, photo, and video profiling). In recent years, several publications have indicated great possibilities but also significant limitations related to the use of data from the indirect sampling method of polymetallic nodules for resource estimation. Information from the backscatter helps with, among other things, the better resource assessment of polymetallic nodules and provides the data needed for planning the optimal mining path (Alevizos et al. 2018; Knobloch et al. 2017; Machida et al. 2021; Wong et al. 2020, 2019; Yang et al. 2020; Yoo et al. 2018). The large number of bottom images taken during photo-profiling conducted along the course of the research vessel are also of great importance. They are the source of information on the continuity of the nodule deposit. However, their role in the estimation of nodule resources still seems to be not fully appreciated.

The usefulness of photography is related to the well-known and very strong relationship linking the weight of individual nodules with their surface (or longer axis) (Handa and Tsurusaki 1981; Wasilewska-Błaszczyk and Mucha 2020; Felix 1980; Kuhn and Rathke 2017; Lipton et al. 2021; Yu and Parianos 2021). This encouraged us to search for a relationship linking the nodule abundance (APN) in samples collected using a box-corer with the percentage coverage of the bottom with nodules (NC), determined in the photo taken at the sampled site. The effects of these studies have been documented in numerous articles and reports (e.g. Kuhn and Rathke 2017; Park et al. 1999, 1996; Sharma 1989; Technical Report Summary, TOML Mineral Resource, Clarion Clipperton Zone, Pacific Ocean 2021; Yoo et al. 2015). Examples of regression models and their descriptions documented in articles and reports are presented in Table 1. The proposed regression models often do not have an intercept or are non-linear, and their strength is assessed using the determination coefficient R^2, which is not methodologically correct (Anderson-Sprecher 1994; Hahn 1973). Therefore, the often very high values of the coefficients of determination (R^2) (of the order of 90% or even 95%) given there are overestimated to an unknown degree. This fact, as well as the issue of different sizes of data sets and the range of variability of nodule abundance (APN) and the use of different regression models (linear and nonlinear) make it difficult to compare the quality of the models presented in the articles for different deposit areas. However, these relationships, described by the model with an intercept where the application of R^2 is correct, are weaker than expected, taking into account the strong correlation of the geometrical features of individual nodules with their masses (Kuhn et al. 2011; Kuhn and Rathke 2017; Mucha and Wasilewska-Błaszczyk 2020; Wasilewska-Błaszczyk and Mucha 2021).

Table 1. Examples of models of the dependence of the nodule abundance (APN) (or weight of nodule W) on the coverage of the ocean floor with nodules (NC) and different parameters describing nodule deposits or individual nodules presented in articles and reports

Tabela 1. Przykładowe modele zależności zasobności konkrecji (APN) (lub masy konkrecji W) od pokrycia dna oceanu konkrecjami (NC) oraz od różnych parametrów opisujących złoża konkrecji lub pojedyncze konkrecje prezentowane w artykułach i raportach

Area (References)	Regression method	Parameters	Model	R^2 or R	Comments
North Pacific (Felix 1980)	Linear	La	$Log_{10}W = 2.71log_{10}La - 0.18$	–	applicable to the flattened oblate-shaped nodules
Northern part of Central Pacific Basin (Handa and Tsurusaki 1981)	Linear (no intercept)	NCxMLa	APN = 7.7NCxMLa	R = 0.89	probable error – 2.5 kg/m^2
(Sharma 1989)	Linear	NC × MLa	APN = 7.74(NC × MLa)/100 + 3.78	R^2 = 0.56	N = 285, or only selected photographs for different classes on nodule burial:
			APN = 8.41(NC × MLa)/100 – 0.01	R^2 = 0.98	0–20%
			APN = 10.15(NC × MLa)/100 + 0.67	R^2 = 0.96	20–40%
			APN = 13.91(NC × MLa)/100 + 0.37	R^2 = 0.98	40–60%
			APN = 21.84(NC × MLa)/100 + 1.20	R^2 = 0.90	60–80%
			APN = 10.15(NC × MLa)/100 + 0.67	R^2 = 0.48	80–100%
Federal Institute for Geosciences and Natural Resources license area (BGR) (Kuhn et al. 2011)	Linear (no intercept)	NC	APN = 0.28NC	R^2 = 0.86	only for small to medium-sized nodules because larger nodules are buried in the sediment to a considerable degree (from 20 to 70%)
Federal Institute for Geosciences and Natural Resources license area (BGR) (Kuhn and Rathke 2017)	Multiple linear regression (no intercept)	A	$W = 0.042A^2 + 2.408A$	R^2 = 0.973	

Area (References)	Regression method	Parameters	Model	R^2 or R	Comments
TOML Mineral Resource, Clarion Clipperton Zone, Pacific Ocean; DeepGreen Metals Inc. (Technical Report Summary, TOML Mineral Resource, Clarion Clipperton Zone, Pacific Ocean 2021)	Linear (no intercept)	NC	NC = 2.99APN	$R^2 = 0.47$	about 70 samples
	Linear	La	$Log_{10}W = 2.8log_{10}La - 0.18$	–	Area B, N = 75; scatterplot of actual and estimated abundance (R = 0.736)
			$Log_{10}W = 2.71log_{10}La - 0.27$	–	Area C, N = 69; scatterplot of actual and estimated abundance (R = 0.866)
NORI Area D (CCZ), Deep Green Metals Inc. (Lipton et al. 2021)	Multiple regression	NC, MLa	APN = –15.20 + 0.24NC + 5.19MLa	–	N = 6
Korea Ocean Research & Development Institute (Park et al. 1999)	Linear (no intercept)	NC	APN = 0.28NC*	$R^2 = 0.95$	N = 251; total weight of nodules with the longest axis of more than 10 cm being greater than 1.5 kg, or the weight of each nodule with the longest axis of more than 10 cm being greater than 1 kg
			APN = 0.55NC*	$R^2 = 0.88$	N = 46; total percentage of nodules with the longest axis of more than 6 cm is greater than 30% in a sample
Korea license area (Yoo et al. 2015)	Linear	NC	APN = 1.666 + 0.456NC	$R^2 = 0.42$	N = 536
Interoceanmetal license area (Wasilewska-Błaszczyk and Mucha 2021)	Linear	NC	APN = 7.55 + 0.150NC	$R^2 = 0.154$	N = 66
	General linear model	NC, GT, SC	APN = 20.02 2.10I1(1) 0.60I1(2) 0.83I1(3) 4.10I2(1) 0.61I2(2) + 8.90ln(NC])	$R^2 = 0.704$	

R^2 – coefficient of determination, R – correlation coefficient, NC – nodule coverage (%), * NC was obtained from a photograph of the onboard free-fall grab sampler, La – long axes of nodule [cm], MLa – mean long axes of nodules (cm), W – weight of nodule (g), A – area of nodules (cm^2), I – the values of indicator variables in GLM regression for genetic type of nodules (GT) and level of sediment coverage of nodules (SC).

There are several factors that influence this state of affairs (Sharma 2017, 1993; Tsune and Okazaki 2014; Wasilewska-Błaszczyk and Mucha 2020). The quality of the bottom photos, together with the increasingly better cameras, seems to be of marginal importance here. Therefore, the very process of extracting nodules from photos, whether done manually or automatically, does not significantly reduce the strength of the relationship between the parameters of nodules determined from the photos and their abundance (Kuhn and Rathke 2017; Park et al. 1999; Schoening et al. 2017). Two natural factors will have a significant impact on the weakening of the strength of this dependence – nodule coverage with sediment and the features of the empirical distribution of the size of the nodules. Tsune (Tsune 2021, 2015) presented theoretical considerations on the effect of covering nodules with sediments and the size distribution of nodules on the estimation of nodule abundance from seafloor photographs. Tsune indicated several parameters describing the shape of the size distribution of nodules and affecting the size of errors in the estimation of nodule abundance from photographs. These include skewness, symmetry, and the number of modal values.

The average length of the longer nodule axe (MLa) can be frequently used in regression models as an expression of the general size of the nodules within the seafloor images (Handa and Tsurusaki 1981; Lipton et al. 2021; Piper et al. 1979; Sharma 1993, 1989; Sharma et al. 2013; Technical Report Summary, TOML Mineral Resource, Clarion Clipperton Zone, Pacific Ocean 2021). Kuhn and Rathke (Kuhn and Rathke 2017) applied the formula (area and weight of individual nodule, Table 1) for separated groups of nodules (small-sized nodules and medium or large-sized nodules) but the results, in particular for medium or large-sized nodules, turned out to be unsatisfactory as the nodules were covered with sediments. This natural factor (the coverage of nodules with sediments and their immersion in sediments, which is difficult to quantify from the images), has an equally large impact on reducing the strength of the relationship between the mass and geometric parameters of the nodules (Sharma 1989, 1993; Felix 1980; Kuhn et al. 2011; Kuhn and Rathke 2017; Tsune 2021).

Benthic activities (e.g. burrowing and movement traces) may be responsible for covering the nodules with sediment (Tsune and Okazaki 2014). The nodules lie on the sea-bottom sediment, generally half-buried, but some nodules are completely covered by sediment and are thus not visible in photographs (Polymetallic Nodules | International Seabed Authority 2022).

In the IOM (CCZ) area, approximately 70% of its surface area is partially covered with sediments of varying intensity (Felix 1980; Kotliński 2009). The reason for the varied coverage of nodules with sediments is the ongoing sedimentation and the activity of ocean currents. The buried nodules (Kotliński and Stoyanova 2007), which occur in sediments at a depth of more than 10 cm, are not prospective for exploitation. Sharma established individual regression models for computing APN in different seabed settings: five classes of Sediment-Water Interface Boundary thickness (SWIB) and five classes of nodule burial (Sharma 1993, 1989). The established linear regression models with intercept, taking into account the

product of NC and MLa for various classes of nodule burial, were mostly characterized by significantly higher coefficients of determination R^2 (48 to 98%) compared to R^2 = 56% for the regression model established for all data together.

When estimating the APN in part of B2 sector (the IOM area), Wasilewska-Błaszczyk and Mucha (Wasilewska-Błaszczyk and Mucha 2021) used not only NC, but also information about the degree of nodule coverage with sediment encoded in the form of a qualitative variable, which is possible in the general linear models (GLM). In addition, the authors included in GLM, as the second categorical variable, the sizes of the nodules previously indicated as an important feature of the nodules, significantly reducing the quality of the APN assessment based on the seafloor images.

CCZ polymetallic nodules are classified into three types based on their morphology, size and texture: smooth (S – type), rough (R – type), and smooth – rough mixed (S – R – type), which can be related, respectively, to their genetic classification: hydrogenetic (H), diagenetic (D) and mixed genesis – hydrogenetic-diagenetic (HD) (A geological model of polymetallic nodule deposits in the Clarion-Clipperton Fracture Zone 2010).

Certain features of the size fraction distribution of nodules in a part of the ocean floor (e.g. shape, the strength of asymmetry) are indicators of their belonging to particular genetic types of nodules. The authors coded them as ordinal variables – the increasing codes were assigned to H, HD and D types along with their increasing general sizes. The established GLM regression model combining the nodule abundance and the nodule seafloor coverage supported by qualitative variables was characterized by a significantly stronger relationship with the adjusted coefficient of determination (R^2_{adj}) equal to 70.4% compared to simple linear regression (SLR) combining only APN and NC with R^2_{adj} equal to 15.4%; in addition, a significant decrease in the standard error of estimation, from 4.2 to 2.5 kg/m^2, was found. In the article by Wasilewska-Błaszczyk and Mucha (Wasilewska-Błaszczyk and Mucha 2021), the verification of the model was performed for a separate data set lying within the research area. The promising results of these studies prompted the authors to make further attempts aimed at developing this methodology to support regression relationships with qualitative information and to verify regression models on a separate data set to confirm their universality in adjacent deposit areas.

The main direction of the development of the methodology for estimating the abundance of nodules from seabed images is the introduction of information on the distribution of nodule fractions within them. However, the consideration of this information in the regression model is significantly limited by the coverage of the nodule with sediments. This factor in a significant part of ocean floor images taken during photo profiling and, to a greater or lesser extent, falsifies the actual nodule fraction distribution. Thus, the geometric measures of individual nodules (area and the lengths of nodules) are underestimated. This is often confirmed by significantly different nodule fraction distributions determined for the box core sample compared to the distributions obtained for the seafloor photos taken just before sampling (Technical Report Summary, TOML Mineral Resource, Clarion Clipperton Zone, Pacific Ocean 2021; Wasilewska-Błaszczyk and Mucha 2020). Automation of the

process of assessing these parameters is justified only in the case of a lack of nodule coverage with sediments. Otherwise, such images should be excluded from the study; alternatively, attempts to include qualitative variables with coded information on the extent of nodule coverage with sediment in regression models appear promising (Wasilewska-Błaszczyk and Mucha 2021).

The assessment of the abundance of nodules based on seafloor images has potential, but its technical possibilities are also limited. This results, for example, from incomplete information about the geometrical dimensions of nodules presented only in the form of its projection onto the plane. The lack of information about the vertical axis of nodules and limitations in reconstructing the shape of nodules horizontally due to the natural seafloor conditions (e.g., immersion in the sediment and nodule coverage with sediment) means that one should take into account the risk of certain errors in estimating the abundance of nodules based on images.

Regression models documented in the literature are often based on a relatively small statistical sample, which makes them uncertain due to the lack of confirmation of their statistical significance. The R^2 values reported in the literature only indicate the strength of the relationships combining APN with nodule parameters obtained from images, while there are no measures expressing the accuracy of the prediction of abundance, for example, in the form of standard estimation error (SEE). When using the APN calculated from models to estimate resources, it is also extremely important to confirm the reliability of models on a separate data set.

1. The aim of the study

The main purpose of the study, the results of which are presented in the article, was to model the regression relationships of the abundance of polymetallic nodules (APN) with the various parameters (features) of the nodules obtained based on the analysis of the seafloor photos in the H22 exploration block (IOM area) (Figure 1). The modeling used the following quantitative parameters (variables): percentage nodule coverage of the seafloor (NC), mean and median lengths of the longer axes of nodules (MLa, MeLa), mean and median surface of nodules (MA, MeA). The following qualitative parameters were also used: the type of size fraction distribution (FD) and the degree of nodule coverage with sediments (SC). Depending on the number and type of these predictor variables, various variants of regression models were used (STATGRAPHICS 19® Centurion 2022): simple linear regression (SLR), multiple regression (MR), general linear models (GLM).

The second objective of the research was to verify and compare the correctness and effectiveness of the obtained models from the point of view of the accuracy of the prediction of nodule abundance made on a separate data set.

2. Initial modeling assumptions – quantitative and qualitative regression variables

The regression dependency between the nodule abundance and the nodule coverage of the seafloor in the photos is obvious. It results from a laboratory established strong non-linear (power) relationship between the mass of single nodules (W) and their surfaces (A). The coefficients of determination of the regression models expressing the strength of this relationship are high and amount to 91% for the area of the horizontal section of nodules ≤40 cm^2 and 83% for the area >40 cm^2 (Wasilewska-Błaszczyk and Mucha 2020). This is also confirmed by the graph of the median mass of nodules within individual fractions (Figure 2). Their fast growth accompanied by a significant increase in the mass range is

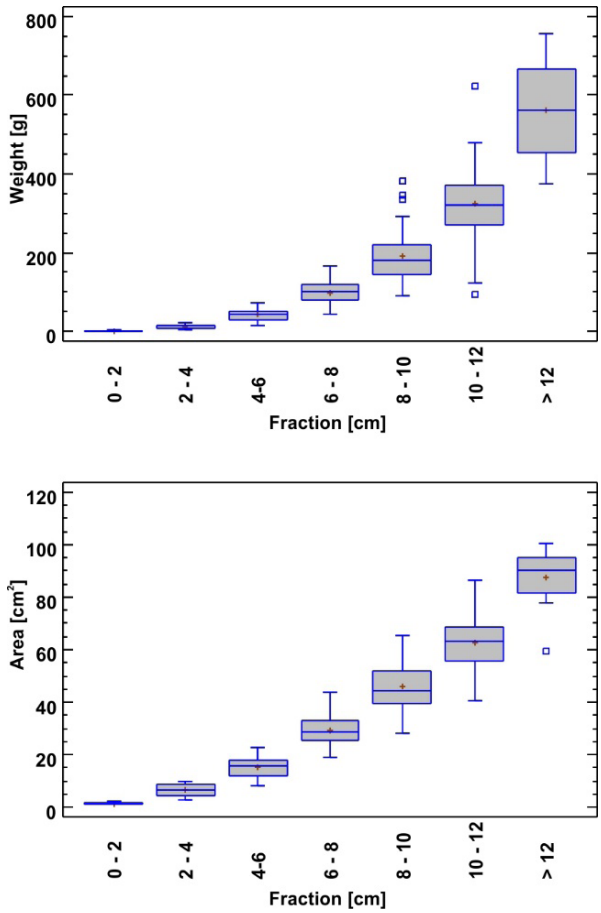

Fig. 2. Box and whisker plots of the weight and area of individual nodules in intervals of fraction (for 322 nodules from the trawl sample, IOM area)

Rys. 2. Wykresy ramka–wąsy masy i powierzchni poszczególnych konkrecji w przedziałach frakcji (dla 322 konkrecji z próbki włoka, obszar IOM)

observed especially for fractions larger than 6 cm. Therefore, it seems reasonable to distinguish more detailed groups of nodule sizes representing different fraction distributions considered in GLM. This need is also evidenced by the noticeable smoothing (underestimation) of the upper range of the abundance of nodules estimated on the basis of the GLM model (Wasilewska-Błaszczyk and Mucha 2021) in relation to the value of this parameter found in direct sampling.

To distinguish different types of size fraction distribution of nodules, an analysis of sixty-six samples collected with the use of the box corer in the H22 exploration block and associated with the genetic types (H, HD, and D) was performed. This analysis was supported by the results of the identification of size fraction distributions of nodules on seafloor images

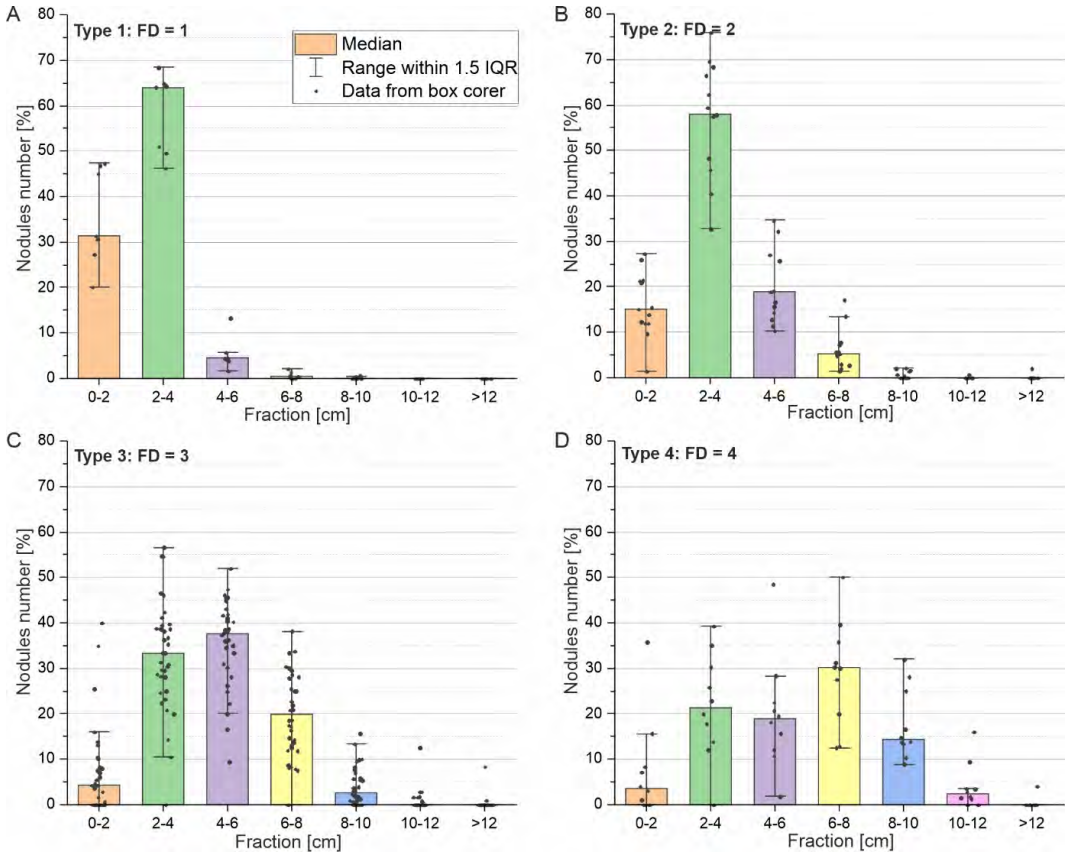

Fig. 3. Box chart with data points overlapped for nodules number in the fractions based on sixty-six box core samples from the H22 exploration block for four types of fraction distribution FD (A–D).
IQR – interquartile range

Rys. 3. Wykres pudełkowy z nakładającymi się punktami danych dla liczby konkrecji we frakcjach na podstawie sześćdziesięciu sześciu próbek czerpaka skrzynkowego z bloku eksploracyjnego H22 dla czterech typów rozkładu frakcji FD (A–D).
IQR – rozstęp międzykwartylowy

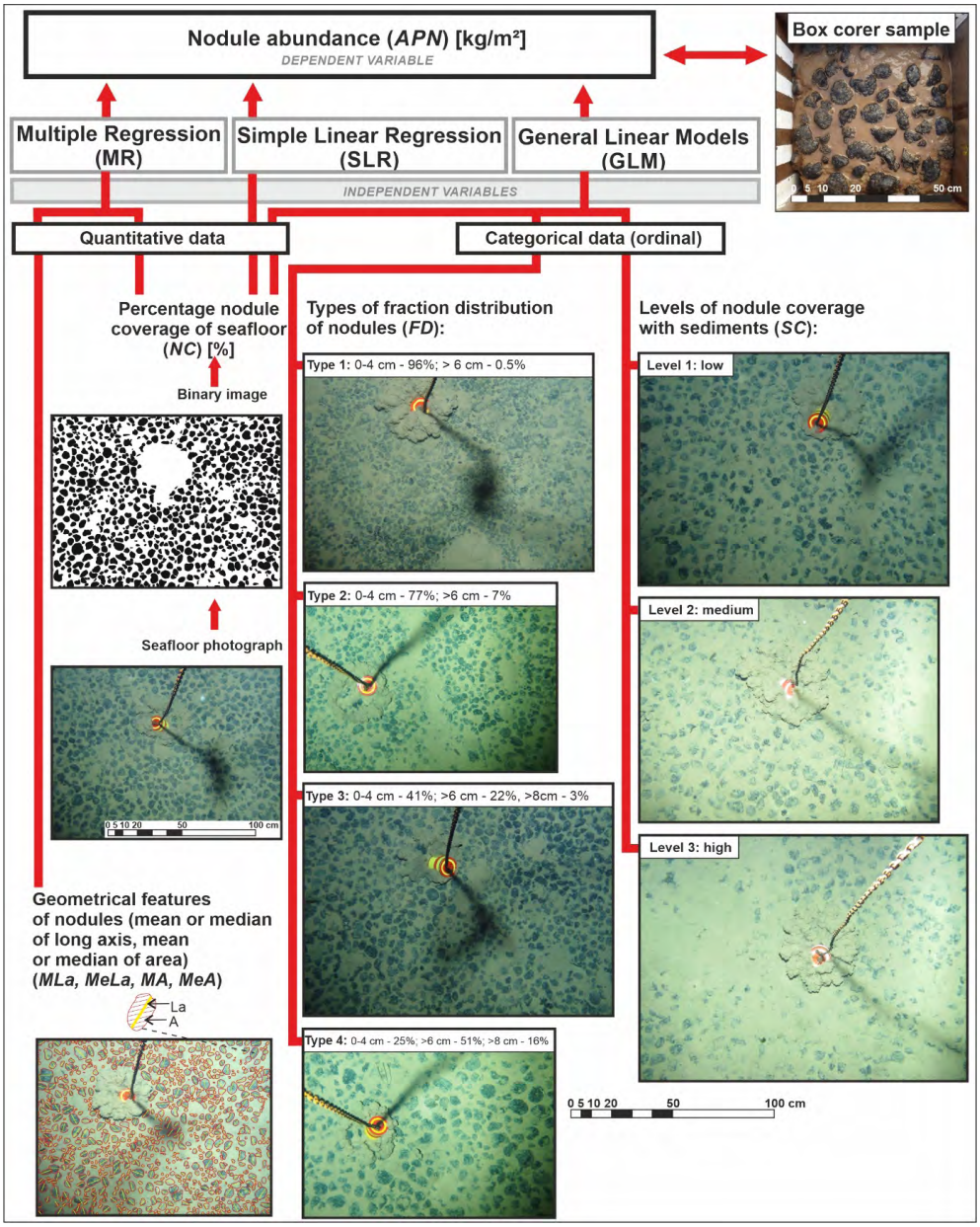

Fig. 4. Scheme of independent variables determined from seafloor photographs used in regression models (SLR – simple linear regression, MR – multiple regression, GLM – general linear models) to predict the abundance of nodules (dependent variable)

Rys. 4. Schemat zmiennych niezależnych wyznaczonych na podstawie zdjęć dna morskiego wykorzystywanych w modelach regresji (SLR – prosty model liniowy, MR – regresja wielokrotna, GLM – ogólne modele liniowe) do przewidywania zasobności konkrecji (zmienna zależna)

before sampling. Within the box corer samples classified as genotype D, the occurrence of two distinctly different distributions of nodule fractions with varied sizes of nodules was found, which was a premise for their separate treatment. For this reason, four types of empirical distributions were finally distinguished for modeling regression relationships, which are presented as box plots (Figure 3). Examples of typical ocean floor images with nodules representing particular types of distribution are presented in Figure 4, while the medians of the relative shares (percentages) of the number of nodules in the samples within different size fraction intervals are included in Table 1. These quantitatively define the types of size fraction distribution (further marked with the symbol FD).

The four featured types of size fraction distribution of nodules (FD) are characterized below (Figure 4):

- **Type 1 (FD 1)**: usually hydrogenetic (H) nodules; nearly 100% of the nodules are fractions up to 6 cm, and about 95% are nodules of fractions up to 4 cm; fraction distribution is characterized by a strong positive skewness.
- **Type 2 (FD 2)**: usually hydrogenetic-diagenetic (HD) nodules; nodule fractions up to 4 cm constitute approximately 77%, nodules from 6 to 8 cm constitute an average of 7%; fraction distribution is characterized by a moderate positive skewness.
- **Type 3 (FD 3)**: nodules usually belonging to the diagenetic type (D), less often hydrogenetic-diagenetic (HD); nodules of fractions up to 4 cm constitute about 40%, and large nodules (larger than 6 cm) constitute about 23%; the fraction distribution is characterized by a weak positive skewness.
- **Type 4 (FD 4)**: diagenetic nodules (D); fractions larger than 6 cm constitute much more than 50% and the number of nodules larger than 8 cm is greater than 10% (even up to 30–50%); flattened quasi-symmetrical fraction distribution.

Quantitative statistics of nodules in size fraction intervals (Table 2), as well as a set of typical images for four types of size fraction distribution (Figure 4), are sufficient to qualify nodules in the ocean floor image for one of the size fraction distribution types based on

Table 2. Medians of the share (%) of nodules number within different size fraction intervals (for sixty-six box core samples from the H22 exploration block)

Tabela 2. Mediany udziału (%) liczby konkrecji w różnych przedziałach frakcji wielkości (dla sześćdziesięciu sześciu próbek czerpaków skrzynkowych z bloku eksploracyjnego H22)

Type of fraction distribution	Fraction interval [cm]					
	0-4	0-6	2-6	> 6	2-8	>8
	Median of the share of nodules number per intervals [%]					
Type 1: FD = 1 (N=7)	95.5	99.5	68.1	0.5	68.6	0.0
Type 2: FD = 2 (N=13)	77.3	93.4	79.8	6.6	85.0	0.0
Type 3: FD = 3 (N=36)	41.0	77.5	67.4	22.5	91.6	3.0
Type 4: FD =4 (N=10)	25.4	49.0	43.8	51.0	78.6	15.8

N – number of box core samples.

expert assessment. This procedure has an advantage over the automation of the process of reading the contours of nodules from photos (automatic image processing) when there is a clear coverage of the nodules with sediment. Such a qualification of the distributions can be made based on a fragment of the photo that does not raise any doubts because it often happens that increased nodule coverage with sediment occurs only in a part of the photo.

In the case of nodule coverage with sediment (SC), three intensity levels were distinguished (Figure 4):

- Level 1 – no or negligible coverage with sediments;
- Level 2 – moderate coverage with sediments in small fragments of the ocean floor covered by the photo;
- Level 3 – intensive nodule coverage with sediments on most of the photo.

To emphasize the difference in the intensity of nodule coverage with sediments, three box corer samples illustrating different levels of SC (Figure 4), were selected. The size fraction distributions of nodules were assigned to the same type (FD = 3), and the abundance of nodules (APN) showed very similar values (from 13.9 to 14.1 kg/m^2).

3. Materials

The research area includes the H22 and H11 exploration blocks located in the central-eastern part of the B2 sector administrated by the Interoceanmetal Joint Organization (IOM) (Figure 1 and 5) (Baláž 2021; Kotliński et al. 2008). They are distinguished by a high abundance of nodules and seem to be the most promising in the perspective of their profitable exploitation (Abramowski et al. 2021). The study used data from box corer samples (polymetallic nodule abundance APN) and information obtained from seafloor images taken before sampling (with an area of about 1.5 m^2). The information from the photos was both quantitative (percentage nodule coverage of the seafloor NC) and qualitative (expert assessment of the level of sediment coverage SC and the types of size fraction distribution FD). The number of items of data used from both exploration blocks was sixty-six (H22) and forty-four (H11), respectively (Figure 5). The information from the H22 area was the basic data set, which was the training set for regression analysis. However, the information from the H11 area was a validation data set (test data). The statistical characteristics of the total data sets and data grouped into types of size fraction distribution (FD) are summarized in Table 3.

Furthermore, for thirty-seven seafloor images, the mean and median values of the two geometric measures of nodules were determined: their longer axis (MLa, MeLa) and their surface (MA, MeA) (Table 2).

The mean abundance of nodules (APN) in both the training (H22) and test (H11) data sets is characterized by close values of 13.5 and 13.0 kg/m^2. Based on statistical tests performed at the significance level of 0.05 there are no grounds to reject the hypotheses on the identity of distributions (Kolmogorov-Smirnov test) and the equality of medians (Mann-Whitney test)

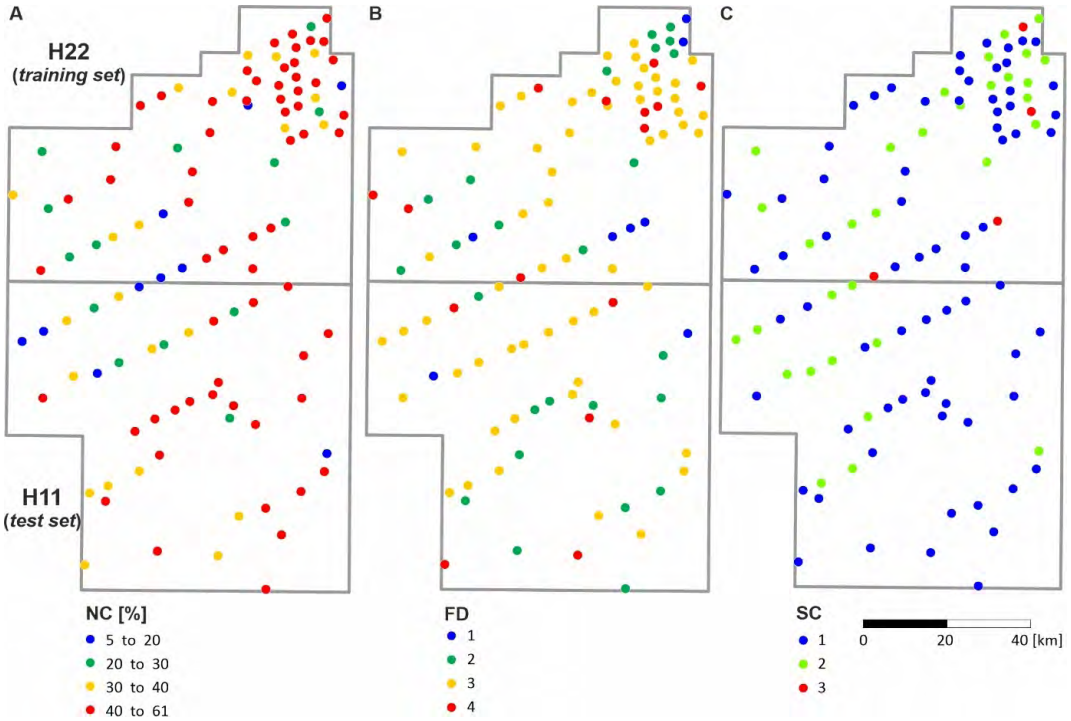

Fig. 5. The location of the box corer sampling sites and seafloor photographs in the H22 (training set) and H11 (test set) exploration blocks
A – information on nodule coverage of seafloor (NC), B – types of size fraction distribution (FD), C – levels of sediment coverage of nodules (SC)

Rys. 5. Lokalizacja miejsc pobierania próbek z czerpaka skrzynkowego i zdjęć dna oceanicznego w blokach eksploracyjnych H22 (zbiór treningowy) i H11 (zbiór testowy)
A – informacja o pokryciu konkrecjami dna oceanicznego (NC),
B – rodzaje rozkładu frakcji wielkościowej (FD), C – poziomy pokrycia konkrecji przez osady (SC)

(STATGRAPHICS 19® Centurion 2022) of the abundance of nodules (APN) in both exploration blocks. Analogous results of statistical testing of distributions and medians were obtained for APNs grouped into types of size fraction distribution. The variability of the abundance of nodules can be described as moderate with the coefficients of variation of around 35% (H22 block) and 30% (H11 block). A comparable or slightly lower variability in the range of 22–29% (H22) and 14–32% (H11) is characteristic for nodule abundance (APN) within the distinguished types of FD. Thus, the exploration blocks under consideration are characterized by a much lower relative variability of the nodule abundance than the entire B2 area.

Such a large differentiation of means is not observed between NCs for different FDs, which range from 34 to 44% (H22) and from 33 to 53% (H11). In general, for NC, the tendency is generally decreasing in the transition from FD = 1 to FD = 4.

4. Methods

Three regression methods (STATGRAPHICS 19® Centurion 2022) were used to assess the nodule abundance (APN) based on the quantitative and qualitative parameters of the nodule deposits:

- simple linear regression (SLR) for continuous predictor variables where only one variable is used (APN = f(NC));
- multiple regression (MR) where more than one predictor variable is continuous (APN = f(NC, MLa, MeLa, MA, MeA));
- general linear models (GLM) in cases where, next to continuous predictor variables (NC), the APN variability is additionally explained by categorical (ordinal) predictor variables (SC, FD): APN = f(NC, SC, FD).

The equations of the mentioned regression models are presented in Table 4.

Each of the determined regression models was evaluated for its statistical significance by calculating the so-called *p*-value. When the *p*-value is ≤0.05, the model can be considered statistically significant with a risk of error not greater than 5%. The analysis also eliminated those variables whose individual contribution to the explanation of the dependent variable was statistically insignificant.

The explanatory power of the regression models were compared on the basis of the value of the adjusted coefficient of determination , which is recommended for the comparison of the regression models that contain different numbers of independent variables:

$$R^2_{adj} = \left[1 - \left(\frac{n-1}{n-p}\right) \cdot \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}\right] \cdot 100\% \quad (1)$$

- n – count of data,
- p – the number of estimated model coefficients,
- $\hat{y}_i$ – theoretical value of the dependent variable Y determined from the model equation for the observation "i",
- y_i – empirical value of the dependent variable Y for the observation "i",
- $\bar{y}$ – arithmetic mean of the empirical values of the dependent variable Y.

The adjusted coefficient of determination R^2_{adj} expresses the percentage of the variability in the dependent variable, which was explained by the fitted model; compared to R^2, it is slightly lower. The use of the determination coefficient to compare the strength of the models is justified because all the regression models presented in this article have an intercept (Anderson-Sprecher 1994).

Table 3. Statistics of nodule abundance (APN) in the box core samples, and the percentage of seafloor nodule coverage (NC), mean area of nodules (MA) and mean long axes of nodules (MLa) determined from the photographs (1.5 m^2)

Tabela 3. Statystyki zasobności konkrecji (APN) w próbkach z czerpaka skrzynkowego oraz procent pokrycia konkrecjami dna oceanicznego (NC), średnia powierzchnia konkrecji (MA) i średnia długa oś konkrecji (MLa) wyznaczona ze zdjęć (1,5 m^2)

Data set (exploration block, number of data)	Variable	Types of size fraction distribution (FD)	Count	Average	Median	Coeff. of ariation	Minimum	Maximum	Skewness
Training set (H22, N = 66)	APN (kg/m^2)	1–4	66	13.5	13.8	34.8	1.5	23.1	–0.40
		1	7	8.1	6.9	25.3	6.3	11.6	1.16
		2	13	10.1	10.8	22.4	5.3	12.9	–1.06
		3	36	14.8	15.8	29.4	1.5	21.1	–1.65
		4	10	17.2	18.0	23.1	10.7	23.1	–0.13
	NC (%)	1–4	66	38.8	41.5	29.9	7.0	59.0	–0.85
		1	7	44.0	45.0	24.3	25.0	59.0	–0.67
		2	13	39.6	42.0	23.6	23.0	51.0	–0.63
		3	36	39.1	41.0	31.3	7.0	57.0	–1.01
		4	10	33.5	37.0	35.7	12.0	47.0	–0.75
Test set for GLM method (H11, N = 44)	APN (kg/m^2)	1–4	44	13.0	13.3	30.3	3.9	19.1	–0.43
		1	2	8.4	8.4	31.3	6.5	10.2	–
		2	12	12.5	12.7	23.2	8.2	16.6	–0.13
		3	25	12.8	13.0	32.3	3.9	18.7	–0.62
		4	5	17.2	18.7	14.0	13.7	19.1	–0.97
	NC (%)	1–4	44	39.7	41.0	38.3	5.0	61.0	–0.53
		1	2	46.5	46.5	44.1	32.0	61.0	–
		2	12	52.8	55.5	15.9	32.0	61.0	–1.45
		3	25	34.2	37.0	44.0	5.0	57.0	–0.44
		4	5	33.4	38.0	24.1	22.0	41.0	–0.80
Training set for MR (H22, N = 37)	APN (kg/m^2)	1–4	37	13.5	13.4	26.7	5.3	19.3	–0.48
	NC (%)	1–4	37	39.5	42.0	25.8	12.0	54.0	–0.89
	MA (cm^2)	1–4	37	8.5	8.5	33.6	4.0	15.5	0.29
	MLa (cm)	1–4	37	3.8	3.9	18.4	2.6	5.1	–0.12

The second key goodness-of-fit measure of the regression model to the data is the standard error of estimation (SEE), which informs about the average value of empirical deviations of the dependent variable from the theoretical values calculated from the model:

$$\mathrm{SEE} = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n - p - 1}} \tag{2}$$

The quality assessment of the regression models was performed using the two measures defined above, SEE and R^2_{adj}.

The predictive ability of APN from the regression models determined for the training set (H22) was checked on the test set (H11). For this purpose, four other measures based on residuals between the measured and theoretical values of the dependent variable determined from the regression equations were calculated and compared:

- the mean error (ME) characterizing the mean deviation of the measured Y values from the values indicated by the model:

$$\mathrm{ME} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i) \tag{3}$$

- the mean absolute error (MAE) characterizing the mean absolute deviation of the measured Y values from the values indicated by the model:

$$\mathrm{AE} = -\sum |y_i - \hat{y}_i| \tag{4}$$

- the mean percentage error (MPE):

$$\mathrm{MPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{y_i} \tag{5}$$

- the mean absolute percentage error (MAPE):

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{6}$$

An important assumption of regression analysis is the non-correlation of independent variables, as reported by the variance inflation factor (VIF). Values greater than 10.0 usually

indicate serious multicollinearity among the predictor variables, which leads to imprecise estimates of the model coefficients.

In GLM, the NC variable was used in the form of a natural logarithm because such a transformation ensures a significantly better fit of the model to the empirical relationship (Wasilewska-Błaszczyk and Mucha 2021).

5. Results and discussion

The first stage of the study was aimed at verifying the strength of the relationship between the percentage coverage of the bottom with nodules (NC) and the nodule abundance (APN) within the groups that have similar distributions of size fraction of nodules (FD) within the bottom images. For this purpose, individual simple regression models (SLR) showing the relationship between APN and NC were determined for the selected types of size fraction distributions (Figure 6, Table 4).

The variability of nodule abundance (APN) estimated based on individual simple linear regression models (SLR) is explained for FD: 2, 3, and 4 by NC in 56%, 62%, and 81%, respectively, as indicated by the coefficients of determination (R^2_{adj}). Standard estimation errors (SEE) expressed in % of the mean nodule abundance amount to 15.0%, 15.2%, and 9.9%, respectively. The regression model was insignificant for significance level $\alpha = 0.05$ (p-value = 0.6579) only for FD = 1 (H type of nodules); no correlation was found ($R^2_{adj} = 0$).

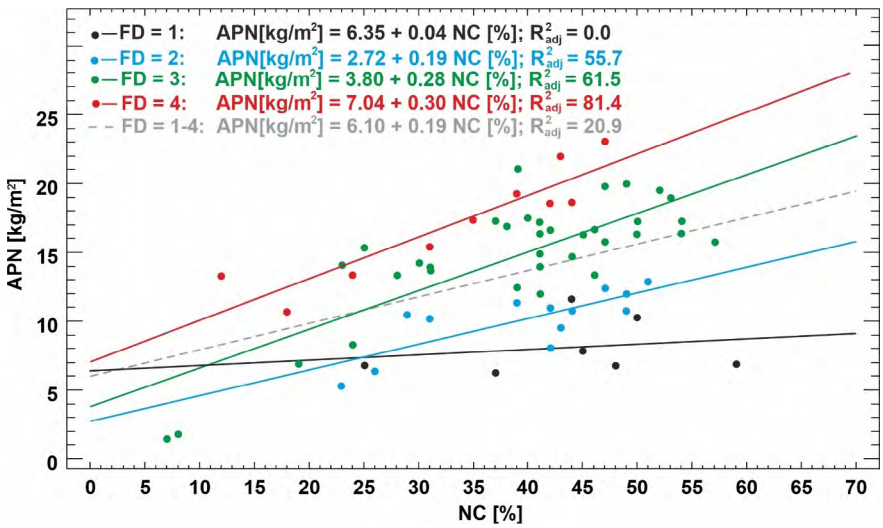

Fig. 6. Simple linear models between APN and NC for different types of size fraction distribution (FD)
R^2_{adj} – the adjusted coefficient of determination

Rys. 6. Proste modele liniowe zależności między APN i NC
dla różnych typów rozkładu frakcji wielkościowych konkrecji (FD)
R^2_{adj} – skorygowany współczynnik determinacji

The magnitude of the MAPE error of estimating the nodule abundance from these individual regression models for all data together is 21.1% and the ME error is 1.82 kg/m^2. The values of these errors are nearly half as large as the corresponding errors of the APN estimation from a simple regression model (characterized by R^2 = 20.9%) made without separating 4 types of size fraction distributions; the errors in this case are MAPE = 39.8% and MAE = = 3.48 kg/m^2. A visual comparison of the APN estimated from regression models (SLR) and the actual APN found in the box core samples is shown in Figure 7 (A – model for all data, B – individual models for different FD). Therefore, considering information on the types of nodule fraction distribution, despite at relatively small size of the separated data groups, significantly increases the quality of the APN estimation based on the NC. The conclusion is fully in line with expectations due to the strong power dependence of the mass of individual nodules and their area.

The observed increase in the strength of the relationship linking APN with NC along with the general increase in the size of nodules (FD: from 1 to 4) does not agree with the observation of Ellefmo and Kuhn (Ellefmo and Kuhn 2020), where a strong correlation is noted only for small nodules, while for large nodules, this correlation disappears due to their greater coverage with sediment.

A further significant increase in the coefficients of determination is observed for the GLM regression models (performed individually for FD types) that, in addition to NC, include the visually evaluated qualitative variable, namely the level of nodule coverage with sediment (SC) (Table 4, Figure 7D). The greatest increase in R^2_{adj} from 0.0% to 60.1% is observed for the smallest fractions of nodules (FD = 1). This GLM model, probably due to the small amount of data, remains statistically insignificant for the significance level $\alpha = 0.05$ (*p*-value = 0.1418), but it is characterized by values of the SEE that are approximately half the size (1.29 kg/m^2), MAE (0.73 kg/m^2) and MAPE (9.7%).

In the next step, the GLM model was made for all data in total, considering NC, SC, and information on size fraction distributions encoded in the form of a qualitative variable (FD). This is characterized by a very high coefficient of determination (R^2_{adj} = 87.2), while the MAE (1.24 kg/m^2 versus 1.12 kg/m^2) and MAPE errors (10.4% versus 8.9%) are only slightly higher than for individual GLM models, taking into account NC and SC (Table 4, Figure 7C). The advantage of this model is its statistical significance (*p*-value = 0.0000). Therefore, from a theoretical point of view, it can be practically used to assess the APN without the need to use individual models for groups of nodule fraction distributions, which are often calculated for numerically modest datasets.

The quality of the APN estimation made on the basis of the presented regression models specified for data from the H22 exploration block was also verified on the test set that did not participate in the creation of models (data from the H11 exploration block) (Figure 1 and 5). Referring to the statistics calculated for APN and NC, it can be concluded that the adjacent areas H22 (training set) and H11 (test set) with a similar area are characterized by similar deposit parameters, which may have a positive impact on the obtained results (Table 3). The MAE and MAPE errors for the SLR model for both data sets (training and test) are similar,

Table 4. Regression models between nodule abundance (APN) and seafloor nodule coverage (NC) (simple linear regression SLR) supported by the level of sediment coverage of nodules (SC) (general linear models GLM) based on photographs for individual types of size fraction distribution (FD) against the background of results for all data. Training and test sets included data from the H22 and H11 exploration blocks, respectively

Tabela 4. Modele regresji między zasobnością konkrecji (APN) a pokryciem konkrecjami dna oceanicznego (NC) (prosta regresja liniowa SLR) poparte poziomem pokrycia konkrecji osadami (SC) (ogólne modele liniowe GLM) na podstawie fotografii dla poszczególnych typów wielkości rozkład frakcji (FD) na tle wyników dla wszystkich danych. Zestawy treningowe i testowe obejmowały odpowiednio dane z bloków eksploracyjnych H22 i H11

	Training set (data from H22 exploration block)										Test set (data from H11 exploration block)					
Regression model (independent variables)	Type of size fraction distribution	Count of data	Equation of Estimated Model	R^2_{adj} (p-value)	SEE		ME	MAE	MPE	MAPE	Code of size fraction distribution	Count of data	ME	MAE	MPE	MAPE
					kg/m^2	% *	kg/m^2	kg/m^2	%	%			kg/m^2	kg/m^2	%	%
Simple linear regression (NC)	**Individual models for data classified to a given size fraction distribution**															
	FD = 1	7	APN = 6.35 + 0.04 NC	0.0 (0.6579)	2.20	27.2	0.00	1.57	–4.6	18.7	FD = 1	2	–0.17	1.28	–1.5	15.6
	FD = 2	13	APN = 2.72 + 0.19 NC	55.7	1.51	15.0	0.00	1.19	–2.6	13.4	FD = 2	12	0.03	1.95	–4.0	17.3
	FD = 3	36	APN = 3.80 + 0.28 NC	61.5 (0.0000)	2.70	15.2	0.00	2.22	–14.7	27.9	FD = 3	25	0.62	2.00	–8.8	20.1
	FD = 4	10	APN = 7.04 + 0.30 NC	81.4 (0.0002)	1.71	9.9	0.00	1.36	–0.8	8.6	FD = 4	5	–0.06	0.21	0.4	1.1
	FD = 1–4	66	Individual models above	–	–	–	0.00	1.82	9.2	21.1	FD = 1–4	44	0.35	1.75	–6.1	17.0
	Model for all data															
	FD = 1–4	66	APN = 6.10 + 0.19 NC	20.9 (0.0001)	4.18	31.0	0.00	3.48	–19.8	39.8	FD = 1–4	44	0.66	2.85	–13.6	27.0

Regression model (independent variables)	Training set (data from H22 exploration block)										Test set (data from H11 exploration block)					
	Type of size fraction distribution	Count of data	Equation of Estimated Model	R^2_{adj} (*p*-value)	SEE		ME	MAE	MPE	MAPE	Code of size fraction distribution	Count of data	ME	MAE	MPE	MAPE
					kg/m^2	% *	kg/m^2	kg/m^2	%	%			kg/m^2	kg/m^2	%	%
General linear models (NC, SC)	**Individual models for data classified to a given size fraction distribution supported by level of sediment coverage**															
	FD = 1	7	APN (kg/m^2) = = –11.37 – 2.33 I2(1) + + 0.61 I2(2) + 5.38 ln(NC)	60.1 (0.1418)	1.29	15.9	0.00	0.73	–1.0	9.7	FD = 1	2	–0.19	1.59	–2.1	19.5
	FD = 2	13	APN (kg/m^2) = = –30.21 – 2.42 I2(1) – – 0.058 I2(2) + 11.36ln(NC)	77.4 (0.0008)	1.08	10.7	0.00	0.71	–0.8	7.0	FD = 2	12	–0.07	1.86	–2.95	16.22
	FD = 3	36	APN (kg/m^2) = = –17.06 – 2.43 I2(1) + + 0.16 I1(2) + 9.21ln(NC)	80.8 (0.0000)	1.90	10.7	0.00	1.40	–1.3	10.6	FD = 3	25	–0.05	1.80	1.68	19.97
	FD = 4	10	APN (kg/m^2) = = –20.91 – 2.64 I2(1) – – 2.66 I2(2) + 11.63ln(NC)	87.2 (0.0017)	1.48	8.6	0.00	0.90	–0.2	4.9	FD = 4	5	–0.26	0.51	2.00	3.29
	FD = 1–4	66	Individual models above	–	–	–	0.00	1.12	1.00	8.94	FD = 1–4	44	–0.09	1.66	0.29	17.0
General linear models (NC, FD, SC)	**Model for all data**															
	FD = 1–4	66	APN (kg/m^2) = = –20.05 – 6.40 I1(1) – – 2.88 I1(2) + 2.55 I1(3) – – 2.40 I2(1) – 0.21 I2(2) + + 9.37 ln(NC (%))	87.2 (0.0000)	1.68	12.4	0.00	1.24	–0.5	10.4	FD = 1–4	44	–0.28	1.67	2.5	16.9
					–	–	0.00	0.90	0.2	12.8	FD = 1	2	–0.60	0.60	7.7	7.7
					–	–	0.00	0.91	1.9	9.7	FD = 2	12	–0.63	1.96	1.4	16.2
					–	–	0.00	1.41	–0.2	11.0	FD = 3	25	–0.09	1.89	2.8	21.0
					–	–	0.00	1.32	1.3	7.7	FD = 4	5	–0.30	0.30	1.8	1.8

R^2_{adj} – the adjusted coefficient of determination, * SEE related to the mean value of the parameter. The values of indicator variables in GLM were determined based on the values of categorical (ordinal) variables according to the following scheme: I1(1) = 1 if FD = 1, –1 if FD = 4, 0 otherwise; I1(2) = 1 if FD = 2, –1 if FD = 4, 0 otherwise; I1(3) = 1 if FD = 3, –1 if FD = 4, 0 otherwise; I2(1) = 1 if SC = 1, –1 if SC = 3, 0 otherwise; I2(2) = 1 if SC = 2, –1 if SC = 3, 0 otherwise.

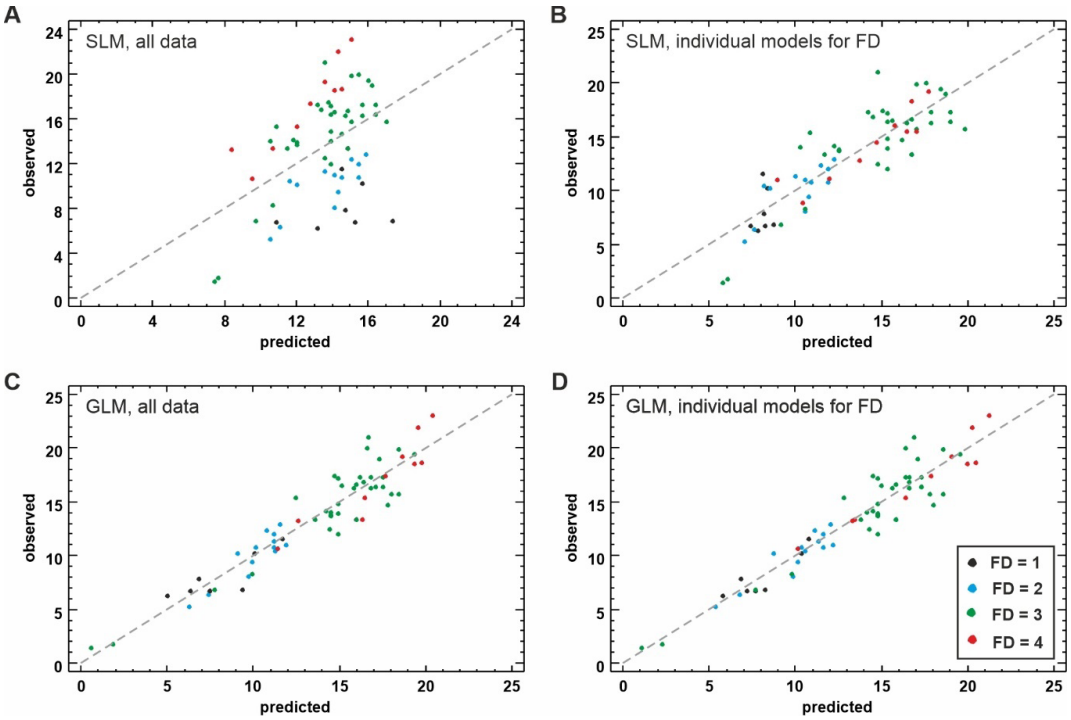

Fig. 7. Scatter plots of the relationship between predicted and observed APN based on simple linear regression (SLR: NC) and general linear models (GLM: NC, SC) for all data (A and C, respectively) and for individual types of size fraction distribution FD (B and D, respectively)
(for the training set, regression models are summarized in Table 4)

Rys. 7. Wykresy punktowe zależności między przewidywaną a obserwowaną APN na podstawie prostej regresji liniowej (SLR: NC) i ogólnych modeli liniowych (GLM: NC, SC) dla wszystkich danych (odpowiednio A i C) oraz dla poszczególnych typów rozkładu frakcji wielkości FD (odpowiednio B i D)
(dla zestawu treningowego modele regresji podsumowano w tabeli 4)

and sometimes even smaller for the test set, which may be related to the less intense effect of the nodule coverage with sediment in the H11 area in relation to the H22 area (Table 4).

The MAE and MAPE errors for the GLM model (NC, SC) are generally twice as high for the test set as they are for the training set (the MAPE errors for different FDs range from 3.3% to 20% and 4.9–10.6%, respectively); however, taking into account the distinctiveness of these sets, this seems fully acceptable. This can be confirmed by the MAPE errors for the GLM model (NC, FD, and SC), which for blocks H22 and H11 are 10.4 and 16.9%, respectively. Therefore, the results of the quality assessment of the GLM model for the test dataset may prove its universality. The obtained results of estimating the nodule abundance (APN) based on a photographic survey seem extremely good, considering the different geometrical bases of the photos (photo area of approximately 1.5 m^2) and the box core sample (area 0.25 m^2).

Table 5. Regression models between nodule abundance (APN) and seafloor nodule coverage (NC) supported by the categorical (ordinal) variables: the level of sediment coverage of nodules (SC) and size fraction distribution (FD) or supported by quantitative variables: median of nodule area (MeA) and median of the long axis (MeLa) based on photographs (count of data = 37)

Tabela 5. Modele regresji między zasobnością konkrecji (APN) a pokryciem konkrecjami dna morskiego (NC) wsparte zmiennymi jakościowymi (porządkowymi): poziomem pokrycia konkrecji osadami (SC) i rozkładem wielkości frakcji (FD) lub wsparte zmiennymi ilościowymi: medianą powierzchni konkrecji (MeA) i medianą osi długiej (MeLa) na podstawie zdjęć (liczba danych = 37)

Regression Method	Independent Variables	Equation of Estimated Model	R^2_{adj} (p-value)	SEE	MAE	MPE	MAPE
SLR	NC	APN (kg/m^2) = = 7.91 + 0.14(NC (%))	13.7 (0.0140)	3.35	2.78	–8.2	25.2
	ln(NC)	APN (kg/m^2) = = –1.92 + 4.25ln(NC (%))	12.0 (0.0203)	3.38	2.82	–8.4	25.5
	NCxMeLa	APN (kg/m^2) = = 5.55 + 0.055 MeLaxNC	44.9 (0.0000)	2.68	2.13	–5.4	19.0
MR	NC, MeA	APN (kg/m^2) = = –0.24 + 0.18(NC (%)) + 0.80MeA	52.3 (0.0000)	2.49	1.95	–4.6	17.2
GLM	NC, MeA, SC (I2)	APN (kg/m^2) = = –2.42 – 1.61 I2(1) + 0.03 I2(2) + + 0.23(NC (%)) + 0.90MeA	55.5 (0.0000)	2.41	1.75	–3.9	14.8
MR	NC, MeLa	APN (kg/m^2) = = –5.64 + 0.17(NC (%)) + 3.26MeLa	51.9 (0.0000)	2.50	1.94	–4.6	17.1
GLM	NC, MeLa, SC (I2)	APN (kg/m^2) = = –8.02 – 1.44 I2(1) – 0.11 I2(2) + + 0.22(NC (%)) + 3.60MeLa	54.0 (0.0000)	2.45	1.74	–3.9	14.7
GLM	ln(NC), FD (I1)	APN (kg/m^2) = = –12.69 – 5.09 I1(1) – 2.40 I1(2) + + 1.77 I1(3) + 6.95 ln (NC (%))	73.3 (0.0000)	1.87	1.51	−2.4	12.9
	ln(NC), FD (I1), SC (I2)	APN (kg/m^2) = = –21.17 – 6.14 I1(1) – 2.65 I1(2) + + 2.40 I1(3) – 2.32 I2(1) – – 0.65 I2(2) + 9.59 ln(NC (%))	84.8 (0.0000)	1.41	1.04	−0.9	7.9

SLR – simple linear regression, MR – multiple regression, GLM – general linear models, R^2_{adj} – adjusted coefficient of determination, SEE – standard error of estimation, MAE – mean absolute error, MPE – mean percentage error, MAPE – mean absolute percentage error. The values of indicator variables in GLM were determined based on the values of categorical (ordinal) variables according to the following scheme: I1(1) = 1 if FD = 1, −1 if FD = 4, 0 otherwise; I1(2) = 1 if FD = 2, −1 if FD = 4, 0 otherwise; I1(3) = 1 if FD = 3, −1 if FD = 4, 0 otherwise; I2(1) = 1 if SC = 1, −1 if SC = 3, 0 otherwise; I2(2) = 1 if SC = 2, −1 if SC = 3, 0 otherwise.

In the case of using images from photo profiling (Neptune), over the larger area (in the IOM area it is on average about 4.5 m^2), a positive result in determining the quality parameters (FD and SC) is expected. This is due to fact that there is a greater risk of incorrect classification of nodules to types of size fraction distribution or mistakes in determining the level of sediment coverage in the case of smaller images due to the influence of local variations on these parameters.

In the case of the models listed in Table 4 and Table 5, the variance inflation factor (VIF) is always less than 2, so there is no serious multicollinearity of the variables used.

In the next stage of the study, the quality of the GLM regression models (APN = f(NC, FD, SC)) and MR models, were compared (Table 5) with consideration to the geometric features of the nodules (in the form of mean areas and longer nodule axes, which are an expression of nodule fraction distributions). Because of the different sizes of the qualitative data sets (FD and SC) and the geometric features of the nodules (MA, MLa), the GLM model (APN = f(NC, FD, SC)) was repeated to obtain the comparability of the results. The statistics of the used geometrical features of nodules for the dataset containing thirty-seven samples from the H22 exploration block are presented in Table 3.

The results of the research in the form of estimation errors presented in Table 5 suggest that the mean area or the mean longer axis of nodules within the seafloor photo is less useful information (with lower predictive ability) because it has a lesser impact on the accuracy of the APN estimation than the information on the type of nodule fraction distribution included in the GLM method. This is proved by a similar R^2_{adj} of 52% and SEE equal to 2.5 kg/m^2 for the MR (APN = f (NC, MA)), and MR (APN = f (NC, Mla) models, significantly lower than the obtained R^2_{adj} equal to 73% for GLM (APN = f (NC, FD)). However, these results are noticeably better compared to the SLR model (APN = f (NC)), and slightly better than the SLR model, which includes information about Mla in the NcxMLa form of the product (R^2_{adj} 12% and 45%, respectively). The obtained results can be associated with a different degree of nodule coverage with sediment, which results in biasing the estimates of the mean geometric features of the nodules with an unknown error. The levels of errors in the APN estimation for both consideration of the geometric variables (MR (APN = f(NC, MeA)) and MR (APN = f(NC, MeLa)) are comparable (Table 5); however, they are slightly better for the median nodule area. Taking into account the MA and Mla variables in the GLM method and supporting them with the qualitative variable SC, did not significantly strengthen the correlation relationship and did not eliminate errors in the assessment of the average geometric parameters resulting from the nodule coverage with sediments (an increase in R^2_{adj} to the value of 54–55% from the value of 52% and a reduction of the SEE error to the value of 2.41 kg/m^2 from of 2.5 kg/m^2).

In addition to the presented MR models (Table 5), other characteristics of the nodules were also taken into account in the model specification and estimation process, namely mean short axis and mean perimeter. However, their inclusion in the regression model was prevented due to collinearities between them. Moreover, simple correlations linking the mass of individual nodules separately to their short axis and circumference turned out to be much

weaker than the analogous correlations with the long axis and the nodule area. Summarizing the presented research results, the regression model may only include the mean area of the nodules or their mean long axis.

The best results of the APN assessment were obtained for the GLM model (APN = f (NC, FD, and SC)), where quantitative and qualitative parameters explain as much as 87% of the variability of APN in the data set from the H22 area, with an acceptable SEE error of 1.68 kg/m^2 (Table 4). Figure 8 is a graphical presentation of the dependence of the estimated APNs based on the GLM model (APN = f (NC, FD, SC)) for various assumed values of quantitative and qualitative parameters (NC, FD, SC). However, attention should be paid to the limited applicability of the model depending on the FD and NC values. For example, when FD = 1, the model can be used for NC > 21%, and when FD = 3, the model can be used for NC > 8%. However, this limitation of the model seems to be irrelevant because it applies to low nodule abundances (APN) that are much less than 5 kg/m^2, i.e. parts of the nodule deposit that are unlikely to be exploited.

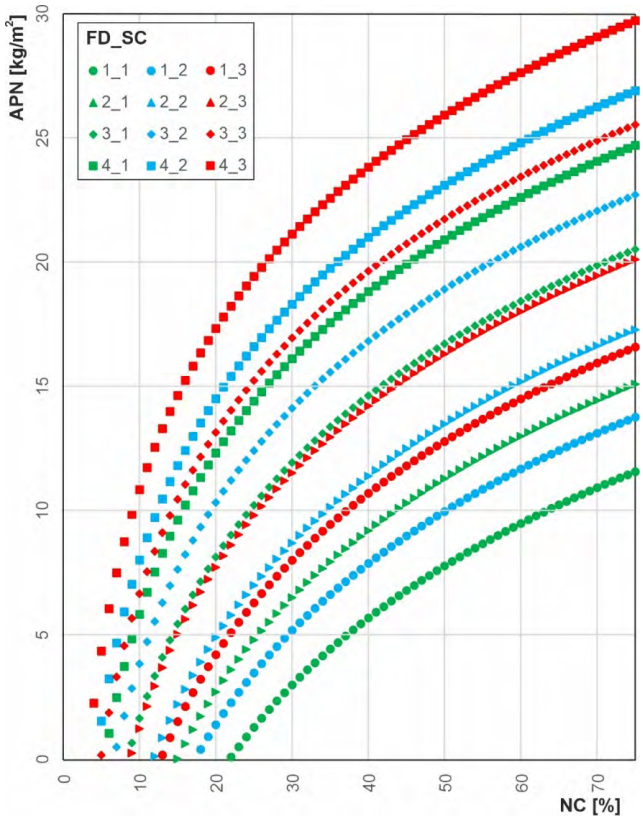

Fig. 8. Plot of the predicted APN values based on the GLM model (APN = f (NC, FD, SC)) for different values of quantitative and qualitative parameters (NC, FD, and SC)

Rys. 8. Wykres przewidywanych wartości APN na podstawie modelu GLM (APN = f (NC, FD, SC)) dla różnych wartości parametrów ilościowych i jakościowych (NC, FD i SC)

6. Summary

A prediction of polymetallic nodule abundance based solely on data obtained from seafloor images gives much better results using the general linear model (GLM) compared to the simple linear regression model (SLR) (Wasilewska-Błaszczyk and Mucha, 2021) and the multiple regression model (MR). The advantage of GLM over SLR and MR lies in the simultaneous consideration of quantitative variables (seafloor nodule coverage in images (NC), mean lengths of the main axes or nodule areas) and qualitative variables (the level of nodule coverage with sediments (SC), and the type of size fraction distribution of nodules (FD)). The codes (variants) of the qualitative variables are assigned to seafloor images on an ordinal scale based on expert judgment. An issue that needs to be investigated in advance is the determination of the correct number of SC and FD codes. The analysis of these factors, conducted in the presented study, showed that the optimal solution is to adopt 3 (SC) and 4 (FD) variants. Under these assumptions, the lowest abundance prediction errors in the areas of the seafloor covered by the photographs were obtained from the GLM equations. The results of the error assessment for different variants of regression models, separately for the training and test datasets, are summarized in Table 6.

Table 6. Comparison of mean absolute and relative errors in estimating the abundance of nodules using various regression models

Tabela 6. Porównanie średnich błędów bezwzględnych i względnych w szacowaniu zasobności konkrecji za pomocą różnych modeli regresji

Regression model, number of data	Training dataset		Test dataset	
	MAE (kg/m^2)	$\frac{MAE}{\overline{APN}}\cdot 100\%$	MAE (kg/m^2)	$\frac{MAE}{\overline{APN}}\cdot 100\%$
SLR (APN = f(NC)), n = 66	3.48	25.8%	2.85	21.9%
SLR (APN = f(NC) separately for FD, n = 66	1.82	13.5%	1.75	13.5%
SLR (APN = f(NCxMeLa), n = 37	2.13	15.8%	np	np
MR (APN = f(NC, MeA) n = 37	1.95	14.4%	np	np
MR (APN = f(NC, MeLa) n = 37	1.94	14.4%	np	np
GLM (APN = f(NC, FD, SC), n = 66	1.24	9.2%	1.67	12.8%
GLM (APN = f(NC, FD, SC), n = 37	1.04	7.7%	np	np
GLM (APN = f(NC, MeA, SC), n = 37	1.75	13.5%	np	np
GLM (APN = f(NC, MeLa, SC), n = 37	1.74	13.5%	np	np

MAE – mean absolute error, – arithmetic mean nodule abundance in the data set (training or test), $\overline{APN}$ – mean nodule abundance, SLR – simple linear regression, MR – multiple regression, GLM – general linear model, NC – nodule coverage, SC – nodule coverage with sediments, FD – the type of size fraction distribution, MeA and MeLa – medians of the area and longer nodule axes, n – number of data in the training set, np – not performed.

The mean absolute error (MAE) was adopted as the measure of the average absolute error in the prediction of the abundance of nodules. This error related to the average abundance of nodules in the data set and was expressed as a percentage and treated as a measure of the relative error.

The smallest errors in estimating nodule abundance (depending on the data set) were obtained from the regression models for GLM using NC, FD, and SC as independent variables (in the range of 1.0–1.7 kg/m^2, which is 7–13% of the mean abundance in the data set). The same model, using the geometrical features of nodules (median area or longer axis of nodules) instead of FD, gives worse estimation results with an error of about 1.75 kg/m^2 (13.5%). Similar error values of about 1.8 kg/m^2 (13.5%) but determined separately for different types of fractions (FD) also give simple linear models; this proves a greater influence of this ordinal variable on the results of nodule abundance estimation than SC. Less accurate estimates of the nodule abundance are provided by multiple regression models (MR) based solely on quantitative variables (NC, MeA, and MeLa) with an error of approximately 1.95 kg/m^2 (14.4%) and a simple linear model, where the independent variable is the product of NC and MeLa with an error of 2.13 kg/m^2 (15.8%). The least accurate estimate of the abundance of nodules is obtained using a simple linear model with NC as the independent variable with errors in the range of 2.85–3.48 kg/m^2 (21–26%).

Conclusions

The variables determining the effectiveness of GLM to predict the abundance of nodules based on bottom images are the seafloor nodule coverage and the type of size fraction distribution. The second ordinal variable, namely the coverage of the nodule with sediments, plays a minor role. This is confirmed by the relatively satisfactory abundance prediction results based on simple linear models linking the abundance of nodules with the seafloor nodule coverage, provided that they are determined separately for different types of nodule fraction distribution. The simple linear regression model (SLR), determined for a full data set without breaking it down into types of size fraction distribution, gives the prediction of the abundance with the greatest error compared to other models (GLM, MR).

The best prediction results obtained using the GLM, with an average error in the range of 1.0–1.7 kg/m^2, can be considered acceptable to estimate the abundance of nodules based on the results of the photographic survey.

The possibilities of obtaining smaller prediction errors of nodule abundance by using other variants of ordinal variables in GLM models seem to be exhausted. These may seem a bit high on a case-by-case basis, but one should be consider that within the calculation blocks, there is a large set of data determined from the photos. Therefore, the error in determining the mean abundance decreases, albeit only to a certain level due to the autocorrelation of abundance at a smaller scale of observation.

A certain reduction of the error, which is difficult to quantify, can be expected if an additional ordinal variable related to the dominant morphological form of the nodules is included in the model. The main morphological types of nodules present in the CCZ are discoidal, ellipsoidal, tabular, polynucleic aggregate, irregular shaped, and fragments (Szamałek et al. 2016). These terms are not strictly defined, but it can be assumed that they differ in the length of the vertical axis. With identical nodule areas determined from the photographs, they differ in mass depending on the morphotype.

This research was financed from AGH University of Science and Technology grant no. 16.16.140.315.

REFERENCES

Abramowski et al. 2021 – Abramowski, T., Urbanek, M., Baláž, P. 2021. Structural Economic Assessment of Polymetallic Nodules Mining Project with Updates to Present Market Conditions. *Minerals* 11(3), DOI: 10.3390/min11030311.

A geological model of polymetallic nodule deposits in the Clarion-Clipperton Fracture Zone (Technical Study: No. 6), 2010. , ISA Technical Study. ISA (International Seabed Authority), Kingston, Jamaica.

Alevizos et al. 2018 – Alevizos, E., Schoening, T., Köser, K., Snellen, M. and Greinert, J. 2018. Quantification of the fine-scale distribution of Mn-nodules: insights from AUV multi-beam and optical imagery data fusion. Biogeosciences Discussions 1–29, DOI: 10.5194/bg-2018-60.

Anderson-Sprecher, R. 1994. Model Comparisons and R 2. *The American Statistician* 48, pp. 113–117, DOI: 10.1080/00031305.1994.10476036.

Baláž, P. 2021. Results of the first phase of the deep-sea polymetallic nodules geological survey in the Interoceanmetal Joint Organization licence area (2001–2016). *Mineralia Slovaca* 53, pp. 3–36.

Clarion-Clipperton Fracture Zone Exploration Areas for Polymetallic Nodules 2022. International Seabed Authority, January 2022 [Online:] https://www.isa.org.jm/map/clarion-clipperton-fracture-zone [Accessed: 2022-03-29].

Ellefmo, S.L. and Kuhn, T. 2020. Application of Soft Data in Nodule Resource Estimation. *Natural Resources Research* 30, pp. 1069–1091, DOI: 10.1007/s11053-020-09777-2.

Felix, D. 1980. Some Problems in Making Nodule Abundance Estimates from Seafloor Photographs. *Marine Mining* 2, pp. 293–302.

Hahn, G. 1973. The coefficient of determination exposed! *Chemical Technology* 3, pp. 609–612.

Handa, K. and Tsurusaki, K. 1981. *Manganese Nodules: Relationship between Coverage and Abundance in the Northern Part of Central Pacific Basin*. (No. 15), Deep Sea Mineral Resources Investigation in the Northern Part of Central Pacific Basin, January–March 1979 (GH79-1 Cruise), Ed. by Atsuyuki Mizuno. Geological Survey of Japan.

Hein et al. 2020 – Hein, J.R., Koschinsky, A. and Kuhn, T. 2020. Deep-ocean polymetallic nodules as a resource for critical materials. *Nature Reviews Earth & Environment* 1, pp. 158–169, DOI: 10.1038/s43017-020-0027-0.

Knobloch et al. 2017 – Knobloch, A., Kuhn, T., Rühlemann, C., Hertwig, T., Zeissler, K.-O. and Noack, S. 2017. *Predictive Mapping of the Nodule Abundance and Mineral Resource Estimation in the Clarion-Clipperton Zone Using Artificial Neural Networks and Classical Geostatistical Methods*. [In:] Sharma, R. (ed.), Deep-Sea Mining: Resource Potential, Technical and Environmental Considerations. Springer International Publishing, Cham, pp. 189–212, DOI: 10.1007/978-3-319-52557-0_6.

Kotliński, R., 2009. *Relationships Between Nodule Genesis And Topography In The Eastern Area Of The C-C region*. [In:] *Establishment of a geological model of polymetallic nodule deposits in the clarion-clipperton fracture zone of the equatorial North Pacific Ocean*. Proceedings of the International Seabed Authority's Workshop Held 13–20 May, 2003 in Nadi, Fiji. International Seabed Authority, Kingston, Jamaica, pp. 203–221.

Kotliński et al. 2008 – Kotliński, R., Mucha, J. and Wasilewska, M. 2008. Deposits of polymetallic nodules in the Pacific: problems of their reserve estimation. *Gospodarka Surowcami Mineralnymi – Mineral Resources Management* 24, pp. 257–266.

Kotliński, R. and Stoyanova, V. 2007. *Buried and surface polymetallic nodule distribution in the eastern Clarion-Clipperton Zone: main distinctions and similarities*. [In:] *Advances in Geosciences, Advances in Geosciences*. World Scientific Publishing Company, pp. 67–74. https://doi.org/10.1142/9789812708946_0006

Kuhn, T. and Rathke, M. 2017. *Report on visual data acquisition in the field and interpretation for SMnN (No. Blue Mining Deliverable D1.31.), Blue Mining project. European Commission Seventh Framework Programme*.

Kuhn et al. 2011 – Kuhn, T., Rühlemann, C. and Wiedicke-Hombach, M.M. 2011. Development of Methods And Equipment For the Exploration of Manganese Nodules In the German License Area In the Central Equatorial Pacific. Presented at the Ninth ISOPE Ocean Mining Symposium, 19–24 June, *The International Society of Offshore and Polar Engineers*, Maui, Hawaii, USA, pp. 174–177.

Lipton et al. 2021 – Lipton, I., Nimmo, M. and Stevenson, I. 2021. *Technical Report: NORI Area D Clarion Clipperton Zone Mineral Resource Estimate* (No. 319002). Deep Green Metals Inc.

Machida et al. 2021 – Machida, S., Sato, T., Yasukawa, K., Nakamura, K., Iijima, K., Nozaki, T. and Kato, Y. 2021. Visualisation method for the broad distribution of seafloor ferromanganese deposits. *Marine Georesources & Geotechnology* 39, pp. 267–279, DOI: 10.1080/1064119X.2019.1696432.

Minerals: Polymetallic Nodules | International Seabed Authority [Online:] https://www.isa.org.jm/exploration-contracts/polymetallic-nodules [Accessed: 2022-04-04].

Mucha, J. and Wasilewska-Błaszczyk, M. 2020. Estimation Accuracy and Classification of Polymetallic Nodule Resources Based on Classical Sampling Supported by Seafloor Photography (Pacific Ocean, Clarion-Clipperton Fracture Zone, IOM Area). *Minerals* 10, DOI: 10.3390/min10030263.

Mucha et al. 2013 – Mucha, J., Wasilewska-Błaszczyk, M., Kotliński, R.A. and Maciąg, L. 2013. Variability and Accuracy of Polymetallic Nodules Abundance Estimations in the IOM Area – Statistical and Geostatistical Approach. [In:] *Proceedings of Tenth ISOPE Ocean Mining and Gas Hydrates Symposium*. Presented at the Tenth ISOPE Ocean Mining and Gas Hydrates Symposium, International Society of Offshore and Polar Engineers, Szczecin, Poland, pp. 27–31.

Park et al. 1996 – Park, C.-Y., Chon, H.-T. and Kang, J.-K. 1996. Correction of nodule abundance using image analysis technique on manganese nodule deposit. *Economic and Environmental Geology* 29, pp. 429–437.

Park et al. 1999 – Park, C.-Y., Park, S.-H., Kim, C.-W., Kang, J.-K. and Kim, K.-H. 1999. An Image Analysis Technique for Exploration of Manganese Nodules. *Marine Georesources & Geotechnology* 17, pp. 371–386, DOI: 10.1080/106411999273684.

Piper et al. 1979 – Piper, D.Z., Leong, K. and Cannon, W.F. 1979. *Manganese Nodule and Surface Sediment Compositions: DOMES Sites A, B, And C*. [In:] Bischoff, J.L., Piper, D.Z. (eds.), *Marine Geology and Oceanography of the Pacific Manganese Nodule Province,* Marine Science. Springer US, Boston, MA, pp. 437–473, DOI: 10.1007/978-1-4684-3518-4_13.

Polymetallic Nodules | International Seabed Authority 2022. Polymetallic Nodules | International Seabed Authority. URL [Online:] https://www.isa.org.jm/documents/polymetallic-nodules [Accessed: 2022-04-26].

Schoening et al. 2017 – Schoening, T., Jones, D.O.B. and Greinert, J. 2017. Compact-Morphology-based poly-metallic Nodule Delineation. *Scientific Reports* 7, DOI: 10.1038/s41598-017-13335-x.

Sharma, R. 2017. *Assessment of Distribution Characteristics of Polymetallic Nodules and Their Implications on Deep-Sea Mining.* [In:] Sharma, R. (ed.), *Deep-Sea Mining: Resource Potential, Technical and Environmental Considerations*. Springer International Publishing, Cham, pp. 229–256, DOI: 10.1007/978-3-319-52557-0_8.

Sharma, R. 1993. Quantitative estimation of seafloor features from photographs and their application to nodule mining. *Marine Georesources & Geotechnology* 11, pp. 311–331, DOI: 10.1080/10641199309379926.

Sharma, R., 1989. Computation of Nodule Abundance From Seabed Photos. *Presented at the 21st Annual Offshore Technology Conference*, Offshore Technology Conference, Houston, Texas, pp. 201–212, DOI: 10.4043/6062-MS.

Sharma et al. 2013 – Sharma, R., Khadge, N.H. and Jai Sankar, S. 2013. Assessing the distribution and abundance of seabed minerals from seafloor photographic data in the Central Indian Ocean Basin. *International Journal of Remote Sensing* 34, pp. 1691–1706, DOI: 10.1080/01431161.2012.725485.

STATGRAPHICS 19® Centurion 2022. Program for statistical analysis, data visualization and predictive analytics.

Sterk, R. and Stein, J.K. 2015. *Seabed Mineral Deposits: An Overview of Sampling Techniques and Future Developments*. [In:] *Deep Sea Mining Summit. Presented at the Deep Sea Mining Summit*, Aberdeen, Scotland, p. 29.

Szamałek et al. 2016 – Szamałek, K., Damrat, M., Frydel, J., Kaulbarsz, D., Kramarska, R., Relisko-Rybak, J., Mucha, J. and Wasilewska-Błaszczyk, M. 2016. *Technical report on the Interoceanmetal Joint Organization polymetallic nodules project in the Pacific Ocean Clarion-Clipperton Fracture Zone (No. 501-2.4/2-15)*. Polish Geological Institute – National Research Institute, Warsaw, Poland.

Technical Report Summary, TOML Mineral Resource, Clarion Clipperton Zone, Pacific Ocean (No. AMC Project 321012), 2021. DeepGreen Metals Inc.

Tsune, A. 2021. Quantitative Expression of the Burial Phenomenon of Deep Seafloor Manganese Nodules. Minerals 11, DOI: 10.3390/min11020227.

Tsune, A. 2015. *Effects of Size Distribution of Deep-Sea Polymetallic Nodules on the Estimation of Abundance Obtained from Seafloor Photographs using Conventional Formulae*. [In:] *Proceedings of Eleventh Ocean Mining and Gas Hydrates Symposium*. Presented at the Eleventh Ocean Mining and Gas Hydrates Symposium, International Society of Offshore and Polar Engineers, Kona, Big Island, Hawaii, USA, p. 7.

Tsune, A. and Okazaki, M. 2014. *Some Considerations about Image Analysis of Seafloor Photographs for Better Estimation of Parameters of Polymetallic Nodule Distribution*. [In:] *Proceedings of The Twenty-Fourth International Ocean and Polar Engineering Conference, International Society of Offshore and Polar Engineers*, 5–20 June 2014. Busan, Korea, pp. 72–77.

Wasilewska-Błaszczyk, M. and Mucha, J. 2020. Possibilities and Limitations of the Use of Seafloor Photographs for Estimating Polymetallic Nodule Resources – Case Study from IOM Area, Pacific Ocean. *Minerals* 10, DOI: 10.3390/min10121123.

Wasilewska-Błaszczyk, M. and Mucha, J. 2021. Application of General Linear Models (GLM) to Assess Nodule Abundance Based on a Photographic Survey (Case Study from IOM Area, Pacific Ocean). *Minerals* 11, DOI: 10.3390/min11040427.

Wong et al. 2019 – Wong, L.J., Pallayil, V., Muthuvel, P., Amudha, K., Kalyan, B., Vishnu, H., Atmanand, M.A. and Chitre, M., 2019. Acoustic Backscattering Properties of Polymetallic Nodules from the Indian Ocean Basin: Results from a Laboratory Measurement. *Presented at the 2019 IEEE Underwater Technology* (UT). DOI: 10.1109/UT.2019.8734442.

Wong et al. 2020 – Wong, L.J., Kalyan, B., Chitre, M. and Vishnu, H., 2020. Acoustic Assessment of Polymetallic Nodule Abundance Using Sidescan Sonar and Altimeter. *IEEE J. Oceanic Eng.* 1–11, DOI: 10.1109/JOE.2020.2967108.

Yang et al. 2020 – Yang, Y., He, G., Ma, J., Yu, Z., Yao, H., Deng, X., Liu, F. and Wei, Z., 2020. Acoustic quantitative analysis of ferromanganese nodules and cobalt-rich crusts distribution areas using EM122 multibeam backscatter data from deep-sea basin to seamount in Western Pacific Ocean. Deep Sea Research Part I: *Oceanographic Research Papers* 161, DOI: 10.1016/j.dsr.2020.103281.

Yoo et al. 2015 – Yoo, C.M., Hyeong, K., Kim, H.J., Chi, S.-B., Kang, J.K., 2015. Image analyses for estimation of polymetallic nodule abundance in Korean Contract Area., in: 44th Underwater Mining Conference (Expanded Abstract). Presented at the 44th Underwater Mining Conference, Florida, USA, pp. 1–2.

Yoo et al. 2018 – Yoo, C.M., Joo, J., Lee, S.H., Ko, Y., Chi, S.-B., Kim, H.J., Seo, I. and Hyeong, K. 2018. Resource Assessment of Polymetallic Nodules Using Acoustic Backscatter Intensity Data from the Korean Exploration Area, Northeastern Equatorial Pacific. *Ocean Science Journal* 53, pp. 381–394, DOI: 10.1007/s12601-018-0028-9.

Yu, G. and Parianos, J. 2021. Empirical Application of Generalized Rayleigh Distribution for Mineral Resource Estimation of Seabed Polymetallic Nodules. *Minerals* 11, DOI: 10.3390/min11050449.

REGRESSION METHODS IN PREDICTING THE ABUNDANCE OF NODULES FROM SEAFLOOR IMAGES – A CASE STUDY FROM THE IOM AREA, PACIFIC OCEAN

Keywords

nodule abundance, regression methods, nodule coverage of seafloor,
fraction distribution, Clarion-Clipperton Zone (CCZ)

Abstract

The main source of information on the abundance of polymetallic nodules (APN) is the results of direct seafloor sampling, mainly using box corers. Due to the vast spread of nodule occurrence in the Pacific, the distances between successive sampling sites are significant. This makes it difficult to reliably estimate the nodule resources, especially in parts of the deposit with small areas corresponding to the areas scheduled for extraction in the short term (e.g. within one year). It seems justified to try to increase the accuracy of nodule resource estimates through the use of information provided by numerous photos of the ocean floor taken between sampling stations. In particular, the percentage of nodule coverage of the ocean floor (NC), the data on fraction distribution of nodules (FD) and the coverage of nodules with sediments (SC) are important here. In the presented study, three regression models were used to predict the nodule abundance from images: simple linear regression (SLR), multiple regression (MR), and general linear model (GLM). The GLM provides the most accurate prediction of nodule abundance (APN) due to the ability of this model to simultaneously take into account both quantitative variable (NC) and qualitative variables (FD, SC). The mean absolute errors of APN prediction are in the range of 1.0–1.7 kg/m^2, which is 7–13% of the average nodule abundance determined for training or testing data sets. This result can be considered satisfactory for predicting the abundance in ocean floor areas covered only by photographic survey.

METODY REGRESJI W SZACOWANIU ZASOBNOŚCI KONKRECJI POLIMETALICZNYCH ZE ZDJĘĆ DNA OCEANICZNEGO – STUDIUM PRZYPADKU Z OBSZARU IOM, OCEAN SPOKOJNY

Słowa kluczowe

zasobność konkrecji, metody regresji, pokrycie konkrecjami dna oceanicznego,
rozkład frakcji konkrecji, strefa Clarion-Clipperton (CCZ)

Streszczenie

Podstawowym źródłem informacji o zasobności oceanicznych konkrecji polimetalicznych (APN) są wyniki bezpośredniego opróbowania dna najczęściej za pomocą próbników skrzynkowych. Z uwagi na ogromne rozprzestrzenienie wystąpień konkrecji w strefie Clarion-Clipperton na Pacyfiku odległości między kolejnymi stacjami opróbowania są znaczne. Utrudnia to wiarygodne oszacowanie zasobów konkrecji w oparciu o uśrednione zasobności konkrecji stwierdzone w próbnikach

skrzynkowych, szczególnie w obszarach o relatywnie małych powierzchniach odpowiadających przykładowo obszarom planowanej, przyszłej eksploatacji w okresach rocznych. W tej sytuacji uzasadnione wydają się próby zwiększenia dokładności oszacowań zasobów konkrecji przez wykorzystanie informacji jakich dostarczają liczne zdjęcia dna oceanicznego wykonywane między stacjami opróbowania. Istotne są tu w szczególności procentowe pokrycie dna oceanicznego konkrecjami (NC), możliwe do ustalenia ze zdjęć dane dotyczące liczby, dane dotyczące rozkładu frakcji konkrecji (FD) oraz przysypanie konkrecji osadem (SC). W prezentowanych badaniach zastosowano trzy modele regresji: prostą regresję liniową (SLR), regresję wieloraką (MR) oraz ogólny model liniowy (GLM). GLM zapewnia najdokładniejsze przewidywanie zasobności konkrecji (APN) ze względu na zdolność tego modelu do jednoczesnego uwzględniania zarówno zmiennych ilościowych (NC), jak i zmiennych jakościowych (FD, SC). Średnie absolutne błędy predykcji mieszczą się w przedziale 1,0–1,7 kg/m^2, co stanowi 7–13 % średniej zasobności konkrecji określonej na podstawie opróbowania bezpośredniego w zbiorze danych (treningowym lub testowym). Wynik ten można uznać za satysfakcjonujący w praktyce prognozowania zasobności konkrecji w miejscach dna objętych jedynie rejestracją fotograficzną.