

Application of Soft Data in Nodule Resource Estimation

Steinar Løve Ellefmo & Thomas Kuhn

Natural Resources Research
Official Journal of the International
Association for Mathematical
Geosciences

ISSN 1520-7439

Nat Resour Res
DOI 10.1007/s11053-020-09777-2



Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.



Original Paper

Application of Soft Data in Nodule Resource Estimation

Steinar Løve Ellefmo^{1,3} and Thomas Kuhn²

Received 27 April 2020; accepted 31 October 2020

Minerals and metals are of uttermost importance in our society, and mineral resources on and beneath the deep ocean floor represent a huge potential. Deciding whether mining from the deep ocean floor is financially, environmentally and technologically feasible requires information. Due to great depths and harsh conditions, this information is expensive and time and resource consuming to obtain. It is therefore important to use every piece of data in an optimum way. In this study, data retrieved from images and expert knowledge were used to estimate minimum and maximum nodule abundances at image locations from an area in the Clarion-Clipperton-Zone of the equatorial North East Pacific. From the minimum and maximum values, box cores and the spatial correlation quantified through variogram, a conditional expectation and associated uncertainty were obtained through the Gibbs sampler. The conditional expectation and the uncertainty were used with the assumed certain abundance data from the box cores in a kriging exercise to obtain better informed estimates of the block by block abundance. The quality assessment of the estimations was done based on distance criterion and on kriging quality indicators like the slope of regression and the weight of the mean. From the original image locations, alternative image configurations were tested, and it was shown that such alternatives produce better estimates, without extra costs. Future improvements will focus on improving the estimation of the minimum and the maximum values at image locations.

KEY WORDS: Ordinary kriging, Conditional expectation, Gibbs sampler, Variance of measurement error, Resource classification.

INTRODUCTION

Society needs minerals. Minerals and metals constitute important ingredients in everything from cars, over toothpaste to mobile phones. To replace petroleum-based energy production with renewable energy and e-mobility requires significant amounts of certain metals such as copper, nickel, cobalt and rare earth elements (REEs) to name a few (Mån-

berger and Stenqvist 2018; Teske 2019). Although the remaining metal resources onshore seems to be abundant for, e.g., copper (Singer 2017), the uneven global distribution of minerals and metals as well as the growing ecological pressure on terrestrial deposits call for alternative sources due to high future demand (Elshkaki et al. 2018).

The deep sea offers a great potential for mineral resources (Cathles 2011; Hannington et al. 2011; Hannington 2013; Singer 2014; Ellefmo et al. 2019), but there is still a significant amount of uncertainty associated with any future mining of the deep sea. The uncertainties are linked to social, legal, ecological, technical and geological factors. Mineral resources on the deep ocean floor can roughly be categorized into seafloor massive sulfides, cobalt-

¹Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway.

²Federal Institute for Geosciences and Natural Resources (BGR), Stilleweg 2, 30655 Hannover, Germany.

³To whom correspondence should be addressed; e-mail: steinar.ellefmo@ntnu.no

rich ferromanganese crusts and manganese nodules (Sharma 2017). These deposits contain significant and varying amounts of copper, zinc, iron, gold, silver, manganese, cobalt, nickel and REEs. The deposits are situated on or beneath the seafloor at depths ranging from 800 to 6000 meters.

According to increasing geological confidence, an identified mineral resource can be classified as an inferred, an indicated or a measured resource (JORC 2012). The classification is performed by a competent or qualified person, and it is dependent on the amount, quality and characteristic of available geodata. The competent person must have at least 5 years of experience relevant to both the style of mineralization, the type of deposit and the performed task. A *task* can be resource estimation and classification or planning and execution of exploration activities and presentations of the results. Obtaining large amounts of high-quality data from the deep ocean floor is expensive and time- and resource-intensive. It requires exploration cruises with highly skilled scientists and operators and expensive equipment. The necessary support vessel needs to be fitted with accommodation, laboratories and systems to handle launch and recovery of remotely operated and autonomous underwater vehicles used in geodata collection and site inspections. It is of uttermost importance that the time on site is spent efficiently and that the collection of high-quality geodata is maximized. The deep ocean floor is vast and underexplored. Geodata utilization and uncertainty quantification is therefore a key to prioritize exploration efforts and do resource estimations, making the right decision (Eidsvik et al. 2015). This means exploiting all data for potential information.

This study focuses on resource estimation of manganese nodules. Geodata necessary to estimate the abundance of nodules have been collected traditionally with box cores. These squared cores cover roughly an area of 1/4th square meter and are pushed by gravity down into the mud on the ocean floor. A mechanism closes the core to contain the material on the ocean floor covered by the core. One box core takes only one increment per launch and recovery, and with water depths down to 6000 m, such sampling is time consuming. Once the increment is recovered onboard the support vessel, it can be processed and analyzed. A box core increment is associated with low uncertainty where most of the uncertainty is associated with the positioning and the small area of the sampling site. To reduce the required amount of box core data, this study investi-

gates the potential benefit of image data where it is possible to estimate a minimum and maximum nodule abundance based on a combination of expert knowledge and information in the image. Abundance expectations derived through such an approach are termed “soft” and are proxies associated with significant uncertainty. However, image data are relatively cost efficient, and it is possible to collect huge amount of image data from a large area just during few dives. The question that arises is therefore whether a large amount of uncertain data (images) can replace a low amount of certain data (box cores).

Initial tests were run by Ellefmo and Kuhn (2018), but, in that study, it was possible to exploit a direct correlation between the estimated abundance from the images and real abundance measured by the box cores. A similar recent study focused on small nodules and exploited similar correlations (Mucha and Wasilewska-Błaszczuk 2020). The correlation approach, however, collapses once larger nodules are included because the larger nodules are covered with more sediments than the smaller ones (Kuhn and Rathke 2017). Gazis et al. (2018) attempted to use hydroacoustic data in combination with optical imagery and artificial intelligence. They managed to estimate the abundance by using acoustic and optical data and a predictive random forests machine learning model. Yoo et al. (2018) used acoustic backscatter data and reported a good correlation between backscatter intensities and mean nodule abundances.

In the present study, the minimum and maximum abundances are merged with available box core data into the kriging with inequalities algorithm to calculate a conditional expectation at each image location and associated uncertainty (cf. Chilès and Delfiner 2012). The conditional expectations and uncertainties are then plugged into an ordinary kriging algorithm along with the available box core data. This paper illustrates the methodology on a case from the Clarion-Clipperton-Zone of the equatorial North East Pacific. Resource estimation quality indicators were used to assess the effect of introducing the data retrieved from the images. Different image configurations with images at their original location, at random locations and along transects covering the whole area of interest were tested. No attempts were made to classify the resource since more work is needed to improve the quantification of the minimum and maximum abundance values from the images.

Application of Soft Data in Nodule Resource Estimation

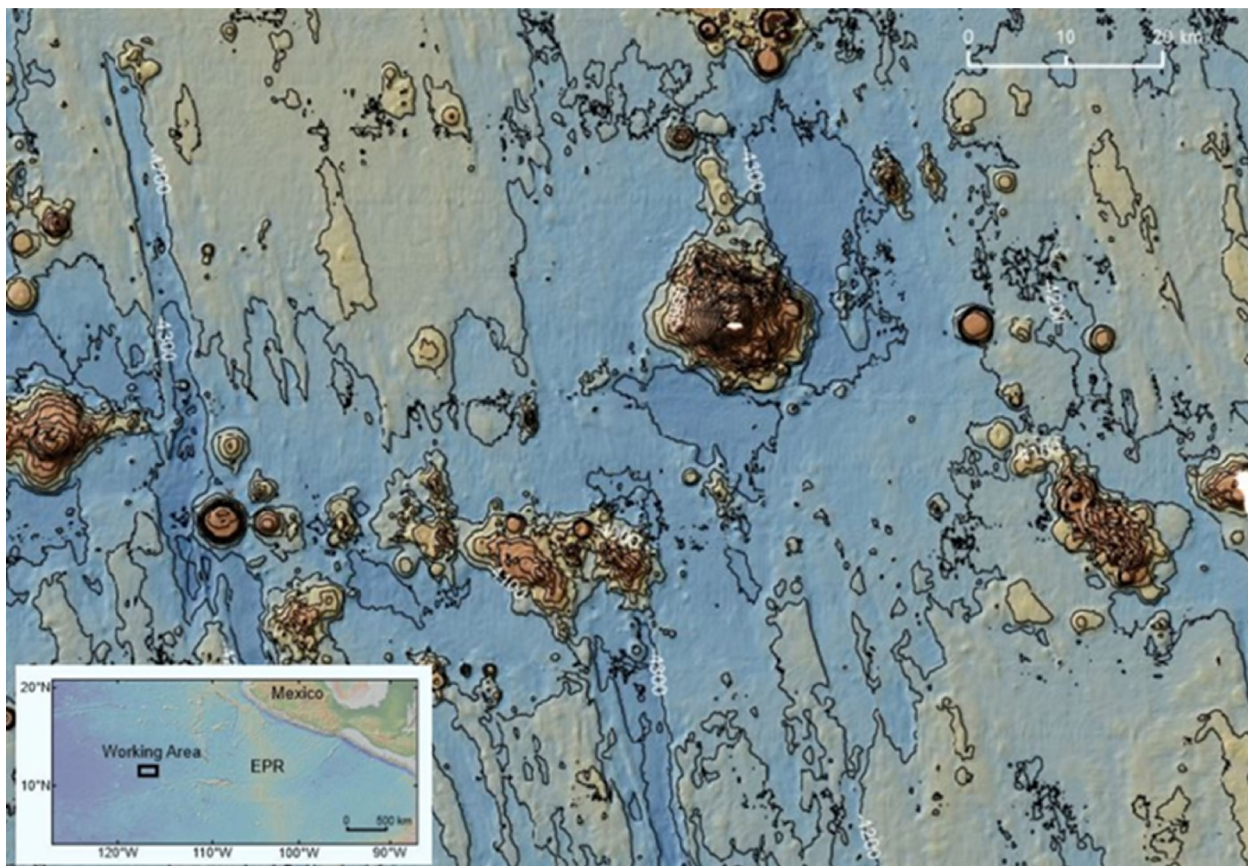


Figure 1. General bathymetry of the working area (large figure) and location of the working area in the NE Pacific (small figure). The exemplary data for this study were derived from the plateau area in the NE corner of the large figure.

BACKGROUND

Geological Background

The working area is located about 900 nautical miles (~ 1.700 km) southwest of Manzanillo, Mexico, in the equatorial NE Pacific Ocean (Fig. 1). The seafloor depth of this area ranges between 4000 and 4300 m in general and consists of plateau-like areas, numerous seamounts rising between a few hundred to more than 2000 m above their surroundings as well as NNW-SSE-oriented ridge and graben structures (Fig. 1) (Kuhn et al. 2020). These ridge and graben structures seem to be bounded by faults, and some of those faults may still be active (Kuhn et al. 2017a, b). The seafloor of the plateau-like areas can be smooth and flat, and these are the regions, which are of interest for Mn nodules exploration (Kuhn et al. 2020). The near-surface sediments consist of pelagic clay and siliceous ooze with trace

amounts of coarser-grained detrital and volcanic material (Kuhn and Shipboard Scientific Party 2015) and (Heller et al. 2018). The surface sediments contain up to 0.6 wt.% organic carbon associated with sedimentation rates of 0.35–0.5 cm/kyr (Mewes et al. 2014; Kuhn et al. 2017a, b).

Manganese nodules mainly occur on the sediment surface or within the upper 10 cm of the sediments in the abyssal plains of the working area. They form concretions of different shape and size, but the main shapes are spheroidal or ellipsoidal and the main size class is, in the present case, bimodal with mean values of the long nodule axes being about 3 cm and 6 cm, respectively (Kuhn et al. 2020).

There are two principal processes leading to the formation of manganese nodules in deep-sea abyssal plains, namely hydrogenetic and diagenetic. Hydrogenetic formation means direct precipitation of Fe oxyhydroxides and Mn oxides from oxygen-rich

near-bottom seawater (Koschinsky and Hein 2003). Diagenetic precipitation occurs within pore-space of deep-sea sediments because of Mn oxide precipitation from almost oxygen-free (suboxic) pore water upon contact with oxic near-bottom water (Hein et al. 2020). Both processes lead to the formation of different layers around a nucleus forming the Mn nodule. Their alternation is mainly controlled by paleo-climatic conditions in the upper ocean. Hydrogenetic precipitation enriches metals like Co and rare earth elements in the nodules, whereas copper and nickel are enriched mainly by diagenetic processes (Kuhn et al. 2017b).

Mineral Resource Estimation and Classification

Geostatistics is a branch of spatial statistics, and it is a framework to model a spatial or temporal phenomenon. It is the preferred methodology used in mineral resource estimation and classification (Journel and Huijbregts 1978; Rossi and Deutsch 2013), but it has found extensive uses in environmental studies, soil sciences to model nutrients, meteorology and public health (Goovaerts 1997). It is used to estimate values at unsampled locations and the associated uncertainty, and it includes both univariate and multivariate implementations. A mineral resource can be classified as measured, indicated or inferred dependent on the amount and quality of available geodata (JORC 2012). The

International Seabed Authority has published their own code (ISA 2015) for publishing nodule estimation results.

In a mining and mine planning context, mineral resource estimation can be used for two purposes (Blom et al. 2019): (1) short-term mine planning or (2) long-term mine planning. In short-term mine planning, the challenge is to estimate the most accurate value in a block in order to decide whether a block that is inside the ultimate pit or mining area should be mined and deposited or mined and sent to the processing plant. Minimizing conditional bias is important in this case (Isaaks 2005), and any estimate of global resource will be significantly smoothed if the conditional bias is minimized. Assuring a minimized conditional bias is achieved by adjusting the kriging search neighborhood until the slope of regression between the estimate and the unknown true value is close to 1 (Armstrong 1998; Isaaks 2005). In long-term mine planning, which is the focus of this work, the challenge is not to find the most accurate value of a specific block, but rather render it possible to estimate some tonnage and grade above some cutoff. The focus on minimizing the conditional bias is not important in this case as argued by Nowak and Leuangthong (2017). Further, Armstrong (1998) warned about using a smoothed block model for long-term mine planning. Kriging quality indicators that can be used to assess the performance of the estimation are summarized in Table 1. Knobloch et al. (2017) used neural net-

Table 1. Kriging quality indicators.

Kriging quality indicators	Explanation	Goal	References
Lagrange multiplier	Parameter that makes sure that ordinary kriging is unbiased. It increases with increasing extrapolation, poor data coverage and/or clustering	Low	Snowden (2017)
Mean distance	The average distance between the center of the blocks and the samples that are used in the estimation of the blocks	Shorter the better	Rossi and Deutsch (2013)
Slope	The slope in the regression between the estimate and the unknown value at a given location and estimate	Close to 1 to minimize the conditional bias	Armstrong (1998)
Kriging standard deviation	The square root of the kriging variance that is minimized when the optimal weights are calculated	Low	Rossi and Deutsch (2013)
Weight of mean (WoM)	It quantifies how much the estimation relies on the estimated global mean (in the ordinary kriging implementation). The lower the weight of mean, the better because the estimate is then controlled by the data, not the global mean	Close to zero	Rivoirard (1987)
Relative prediction error	Twice (1.96) the ratio between the kriging standard deviation and the estimate. Can be given as a percentage or a decimal number	Low	Knobloch et al. (2017)

Application of Soft Data in Nodule Resource Estimation

works and classical kriging to estimate nodule resources and used the relative prediction error (RPE) explained in Table 1 to classify the resource.

Mineral resource classification is a subjective endeavor. The competent or qualified person (CP/QP) assesses the data quality and the amount of data, applies a fit for purpose estimation technique and classifies the mineral resource according to some criteria. Rossi and Deutsch (2013) discussed the use of such criteria summarized in Table 1, and examples are given in Benndorf (2015) and Mucha and Wasilewska-Błaszczuk (2020). Armstrong (1998) argued against the use of the kriging variance as a kriging quality indicator because it is reported to be relatively insensitive to the data configuration. What might add to the use of kriging variance is the co-use of the Lagrange multiplier (Table 1) (Snowden 2017). Armstrong (1998) argued in favor of the weight of mean and the slope, both explained briefly in Table 1, especially when optimizing the search neighborhood to minimize conditional bias. Another option presented by Rossi and Deutsch (2013) is to estimate through more than one kriging pass and vary the requirements from pass to pass, normally from more to fewer requirements. The differences from kriging pass to kriging pass can materialize itself in different number of sectors in the search neighborhood, different number of maximum consecutive empty sectors, the size of the neighborhood and minimum number of samples inside the search neighborhood. Such requirements can be combined with requirements on the slope, the weight of mean and the kriging variance in the final resource classification. In addition, conditional simulation represents an approach that provides a model of local and global uncertainty and which can be used to assess the *probability* of having a block grade above some cutoff (Boucher and Dimitrakopoulos 2012). Conditional simulation can also be used to quantify the conditional bias (Isaaks 2005).

Using Auxiliary Information in Resource Estimation

Auxiliary data are extra or secondary geodata that have the potential to improve the estimation precision and accuracy in geostatistics. Wackernagel et al. (2002) presented approaches to take advantage of auxiliary information including kriging with external drifts where a primary variable typically at borehole location is combined with a secondary

variable providing low(er) frequency/resolution information about the variable(s) under study. Abrahamsen and Benth (2001) presented a methodology to estimate trend coefficients given both exact data and inequality constraints. Omre (1987) developed Bayesian kriging as a generalized form of kriging with an external drift. Cokriging and collocated cokriging is presented in Wackernagel (2010) and Chilès and Delfiner (2012). Cokriging is an approach used if the dataset consists of correlated variables where one variable dominates the other in terms of the number of data points or if the variables must be estimated simultaneously due to relationships between the variables. Collocated cokriging is used if one of the variables is known at the target points. Eidsvik and Ellefmo (2013) looked at the use of uncertain auxiliary data in resource estimation.

DATA

“Hard data” for this study were collected using a so-called box corer (Fig. 2). With this device, a block of deep-sea sediment measuring 50 cm by 50 cm of seafloor area and 30–40 cm thickness was sampled at 41 locations. The typical distance between neighboring sampling locations was between 2500 and 3000 m. See Figure 5. The Mn nodules lying on top of the sediments or within the first 10 cm were collected manually from the box corer, and the sum of their weight was measured immediately after recovery of the box core on board the vessel. This way the so-called wet nodule abundance in kg/m² was calculated. This nodule abundance and the metal content of the nodules are the major parameters that control the economic value of the Mn nodule deposit (Knobloch et al. 2017). However, taking box core samples is time consuming and thus expensive. It takes about 4 h to sample one box core increment in about 4200 m water depth. Thus, only about six increments can be taken per day. Kuhn et al. (2016) showed that the average spacing of box corer increments should be in the range of about 1300 to 4700 m in order to reach the indicated resource level of a Mn nodule deposit given that box cores are the only source of information.

Box core sites can be linked by video mapping of the seafloor. For this study, the German video sledge STROMER was used; it is equipped with several high-resolution video and photo cameras (Fig. 3). The sledge is permanently connected with a surface vessel via a fiber optical cable during the



Figure 2. Box corer used for sampling of seafloor sediments and nodules (left panel) as well as surface (upper right) and downcore profile (lower right) of the sediment block sampled.

deployment. The vessel sailed with a speed of about 1 knot along pre-defined transects, with the STROMER being towed behind the vessel and kept about 3 m above the seafloor. Images were taken automatically every 5 or 10 s, and underwater positions of the video sled were calculated using a combination of hydroacoustic and inertial navigation systems. In total, 5504 images were taken and included in this analysis. The accuracy of the positioning for each picture is around ± 5 m (Rühlemann and Shipboard Scientific Party 2018).

For the analysis of underwater images, BGR together with the University of Bielefeld developed an automated image analysis software called “Mangan Analyzer.” The objective of this software is to detect manganese nodules automatically and reliably in many seafloor images. The core algorithm

for the analysis of seafloor images with special focus on the estimation of manganese nodule coverage is the so-called hyperbolic, self-organizing map (HSOM) as a neural network approach (Schöning et al. 2012). The process to map single nodules within an image requires several steps including illumination correction, feature transformation, machine learning, quantification, and post-processing (Fig. 4).

Figure 5 shows the box core and the image locations. The area covered by the images followed a slightly skewed distribution that varied between 0.6 m^2 and about 12 m^2 with a median area of 5.5 m^2 and a mean area equal to 5.7 m^2 . This differed from the 0.25 m^2 covered by the box cores data, but both images and box core data were, in this study, as-

Application of Soft Data in Nodule Resource Estimation

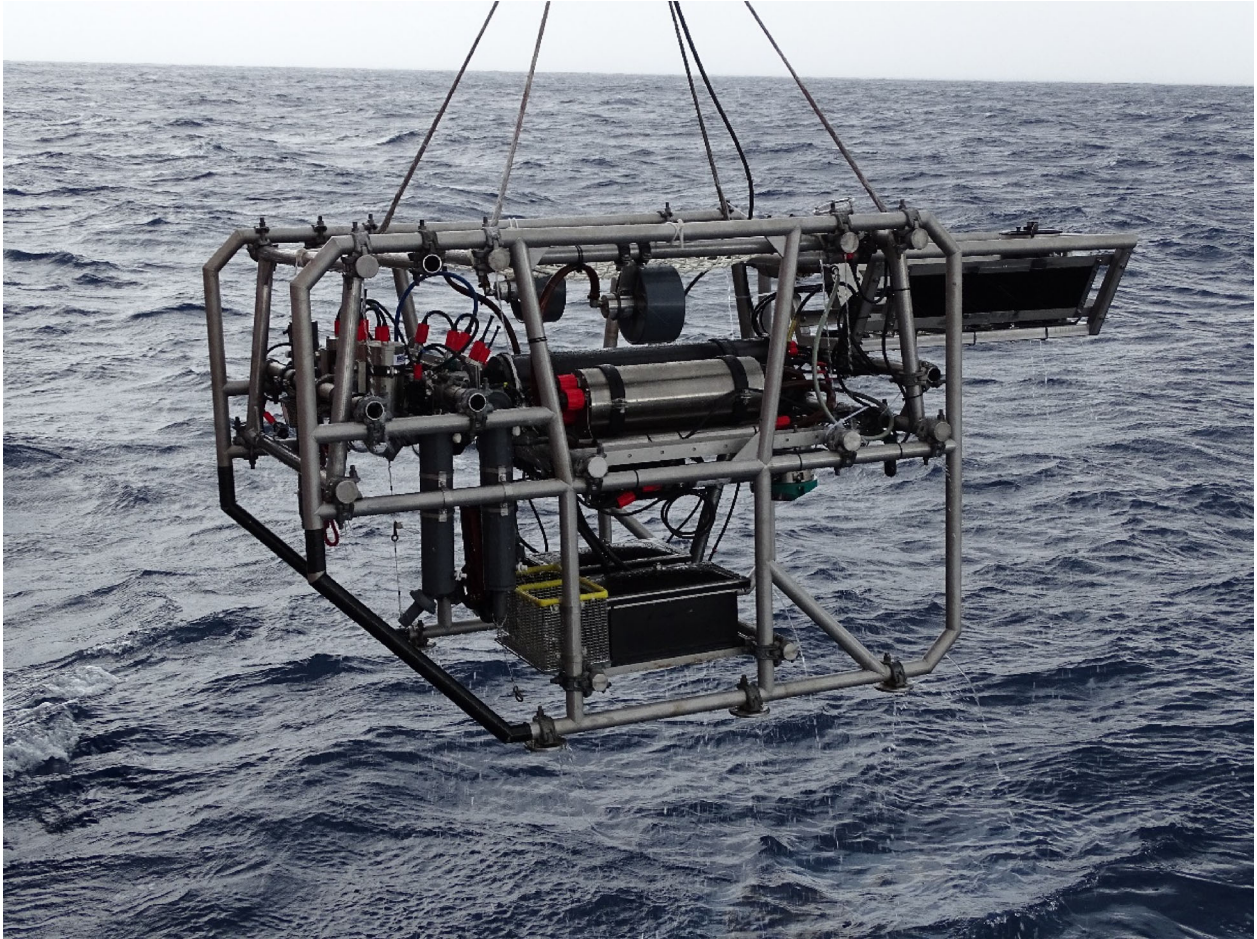


Figure 3. BGR video sled STROMER used for this study (Photo: S. Sturm).

sumed as point data compared to the 1 km² large areas/blocks being estimated.

The area of each nodule in each image was calculated, and the nodules were divided into different size classes. A mathematical relationship between the number of nodules per size class and their weight was established based on box core stations, and thus the total weight per image in kg/m² can be calculated. The number of nodules per size class and not the measured nodule area was used as a basis for this relationship because it turned out that nodules were arbitrarily covered by sediments and that there was only a weak statistical relationship between coverage and abundance. It was assumed that some part of every nodule was visible in the images. The nodule abundance from images calculated this way still was by factor of 1.1–3.7 lower than the nodule abundance measured in box cores along the video

transects. Therefore, the abundance calculated from images was considered minimum abundance only. In this study, the maximum abundance was calculated by multiplying the obtained minimum value by the correction factor 2.21 and 3.71 for the northern and the southern group of images in Figure 5, respectively. These correction factors were the maximum ratios between nodule abundance derived from box core stations and the abundance from the image analysis at the position of each box core station. Summary statistics of the parameter under study is given in Figure 6. There were 41 box cores with an average abundance of 22.8 kg/m², with minimum and maximum values of 10.2 and 36.1 kg/m², respectively.

The minimum and maximum values estimated from the 5504 images are given in Figures 7 and 8. The lowest minimum value was 2.2 kg/m², and the

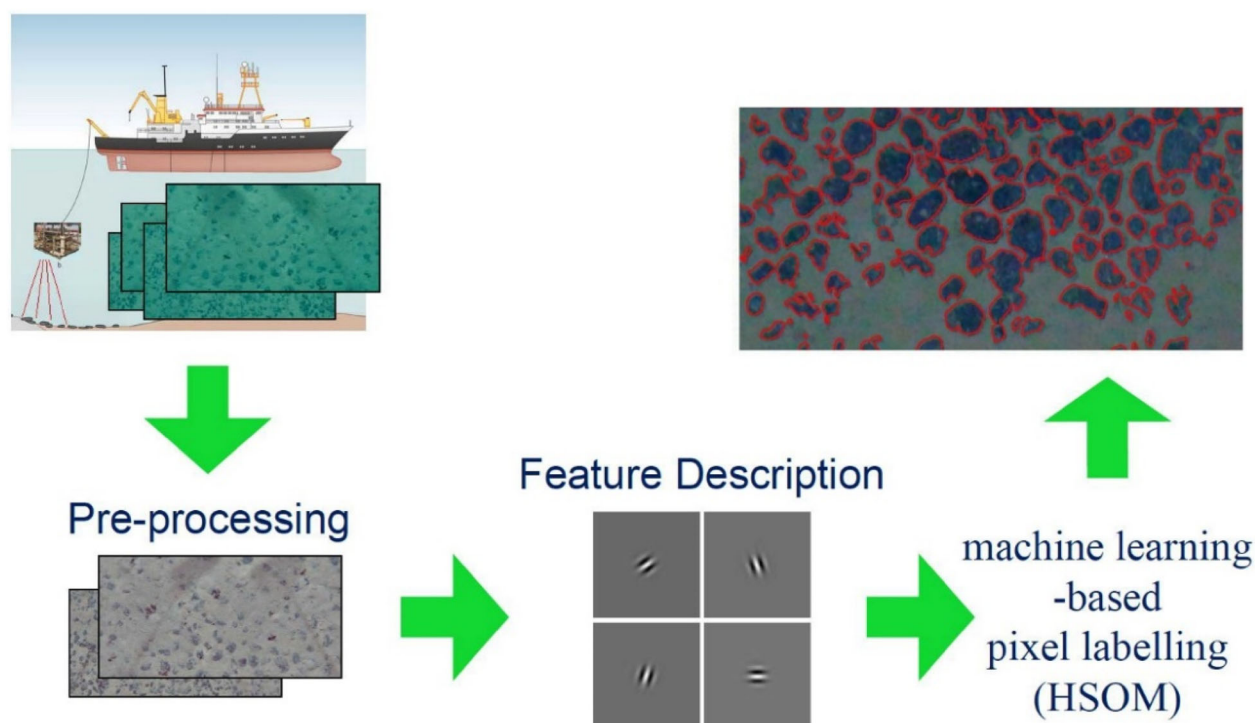


Figure 4. Working steps of the image processing software MANGAN ANALYZER.

largest minimum value was 17.1 kg/m^2 . Corresponding maximum values were 7.3 kg/m^2 and 57.3 kg/m^2 . The mean minimum and maximum values were 10 and 28.2 kg/m^2 , respectively. The maximum values varied more than the minimum values due to the use of the two different multipliers (2.21 and 3.71). Figure 9 shows a scatterplot revealing the two populations of minimum and maximum values originating from the application of the two multipliers. The sub-population characterized by the lower slope of the two corresponds to the northernmost (131STR) transect in Figure 5. This indicates that the northernmost transect has a lower maximum abundance for the same minimum abundance and thereby a lower uncertainty due to a smaller nodule size.

Given the minimum and the maximum values in Figures 7 and 8, a pseudo-standard deviation of a distribution satisfying the minimum and maximum boundaries can be estimated from the range formula (Hozo et al. 2005), thus:

$$\text{StdV}_{\text{Pseudo}} = \frac{\text{Max} - \text{Min}}{4} \quad (1)$$

The pseudo-variance is the squared pseudo-standard deviation calculated from Eq. (1). The histogram of this pseudo-variance is given in Figure 10. The mean was $24.1 (\text{kg/m}^2)^2$, and the most probable value (the mode) was close to 12 due to the positively skewed distribution. This value was used to validate the quantified uncertainty at images location.

There are studies that reported good correlations between nodule abundance and areal characteristics of nodules retrieved from images (e.g., Felix 1980; Lipton et al. 2016; Mucha and Wasilewska-Błaszczuk 2020). For the area under study, it was difficult to find such reliable correlations due to the multimodal nodule size distribution in the area (Kuhn et al. 2020). Figures 11 and 12 show the relationship between the total nodule coverage in area-% and the calculated minimum abundance in kg/m^2 . Two distinctly different relationships can be identified per transect. The reason is regional differences in nodule size distribution, and to what extent the nodules are covered by sediments. Generally speaking, larger-sized nodules (long nodule axis $> 4 \text{ cm}$ diameter) are covered by sediments to a higher degree, and if they dominate a location, they

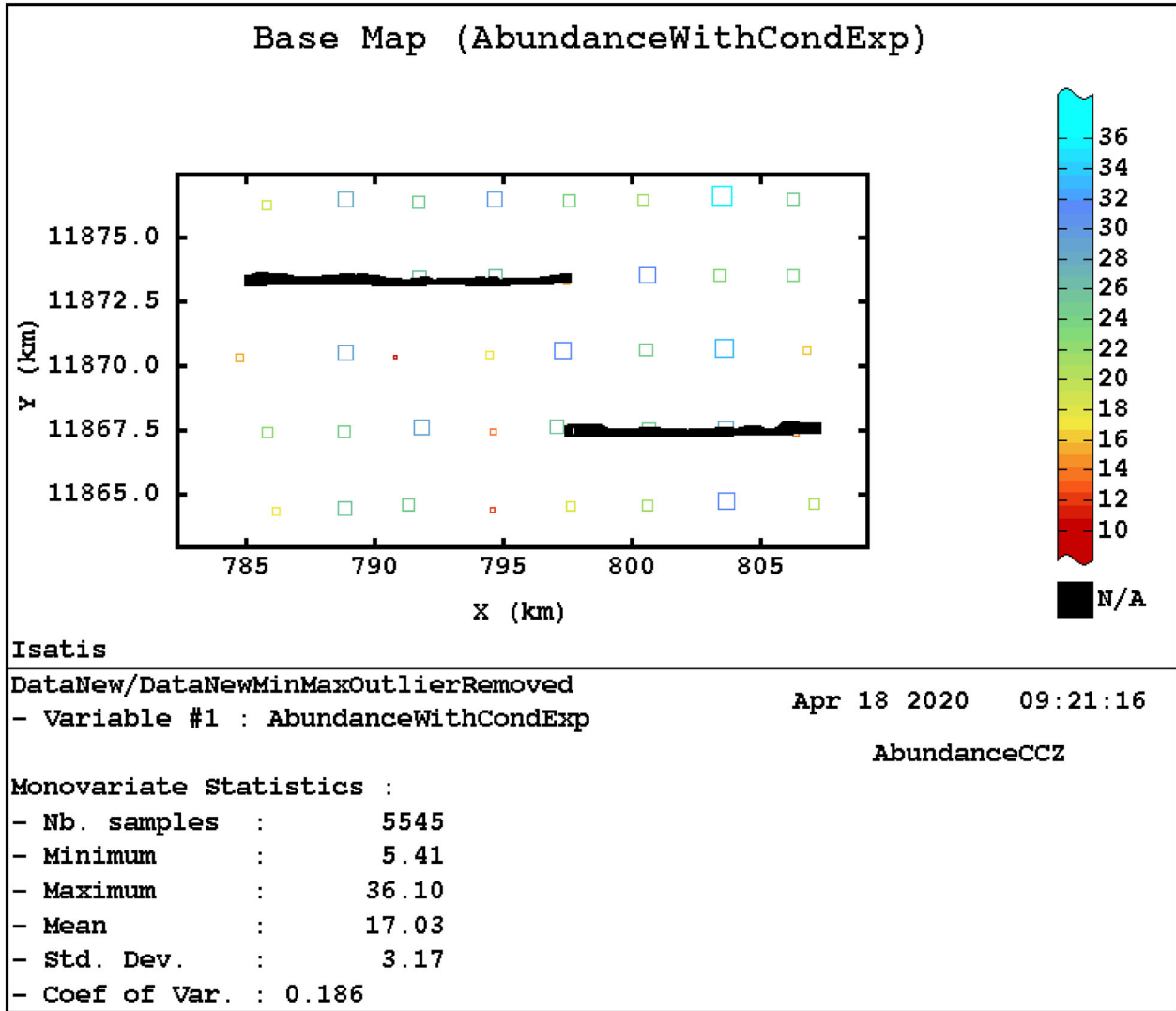


Figure 5. Base map showing the box core and original image location. Box core increments taken systematically at all squares. Color and size of the squares indicate the abundance. Images were collected along the southeastern (102STR) and the northwestern (131STR) transects. The univariate statistics show the number of samples and the nodule abundance given in kg/m². Number of samples consists of 5504 images and 41 core box increments.

may obliterate the relationship between areal nodule coverage and nodule abundance (Kuhn and Rathke 2017).

To study the effects of different data configurations, tests were executed where the images have been placed randomly in the area covering the box core data ± 1000 m and in transects. Images at a random location (Fig. 13) represented an operationally unlikely scenario but can be used to assess how good in terms of the kriging quality indicators given in Table 1 an estimation can be. The scenario with all the images distributed along transects

(Fig. 14) that covered the whole area was operationally a more realistic scenario that conceptually can be compared to onshore in-fill drilling in exploration.

The minimum and maximum values at images points were kept although the image positions have changed, and new conditional expectations and variances of measurement errors at image locations were estimated using the Gibbs sampler explained in the next section. This is not realistic because new minimum and maximum values should have been estimated at the new locations. Therefore, a com-

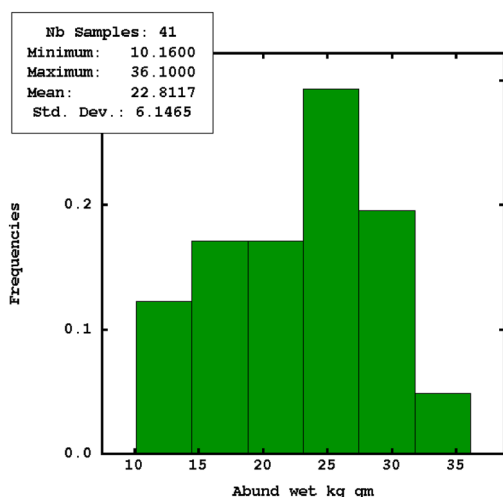


Figure 6. Summary statistics and histogram showing the characteristics of the abundances collected with the box cores.

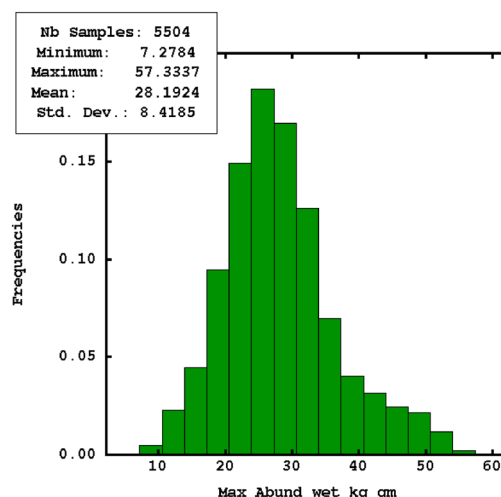


Figure 8. Summary statistics characterizing the maximum abundances in kg/m² obtained from images.

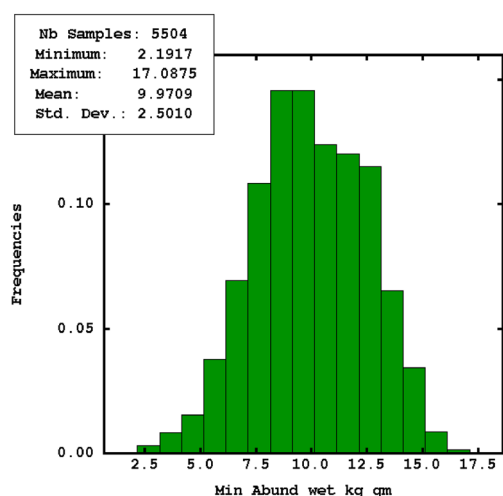


Figure 7. Summary statistics characterizing the minimum abundances in kg/m² obtained from images.

parison between the estimates and the RPE obtained with only box core data and those obtained with images at random location or in transects was not relevant. The slope, the weight of mean, the distances and the kriging standard deviation that do not depend on the prediction can, however, be compared to assess the effects of different data configurations.

METHODOLOGY

Linear Geostatistics

The necessary theoretical framework for geostatistics was developed in the 1960s by Matheron (1963) based on empirical work performed by the South-African mining engineer Daniel Krige in 1951. Geostatistics and its application to mining has been described several textbooks, including Journel and Huijbregts (1978), Goovaerts (1997), Armstrong (1998), Grunwald (2005), Buyong (2007) and Chilès and Delfiner (2012). In his publication in 1963, Matheron introduced the term and concept of the regionalized variable and he defined it as “*sensu stricto, an actual variable, taking a definite value in each point of space*”. The regionalized variable $Z(x_0)$ at location x_0 can be described conceptually by a structured component ($m(x_0)$), a spatially correlated random component $\varepsilon'(x_0)$ and a spatially uncorrelated random component $\varepsilon''(x_0)$, thus:

$$Z(x_0) = m(x_0) + \varepsilon'(x_0) + \varepsilon''(x_0) \quad (2)$$

The structured component $m(x_0)$ is called the *drift* (Matheron 1963), and it represents the mean of the regionalized variable at location x_0 . The spatially correlated random component comprises fluctuations around this drift. Matheron (1963) termed Eq. 2 the *universal model of spatial variation*. Geostatistics has a broad area of use and both the drift, the spatially correlated fluctuations and the spatially

Application of Soft Data in Nodule Resource Estimation

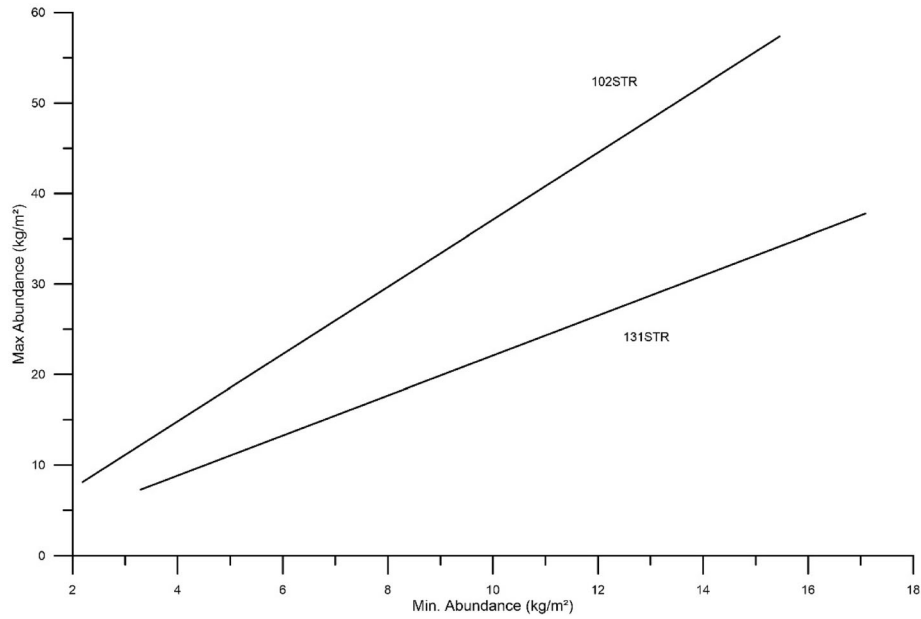


Figure 9. Scatterplot showing two populations of minimum and maximum values.

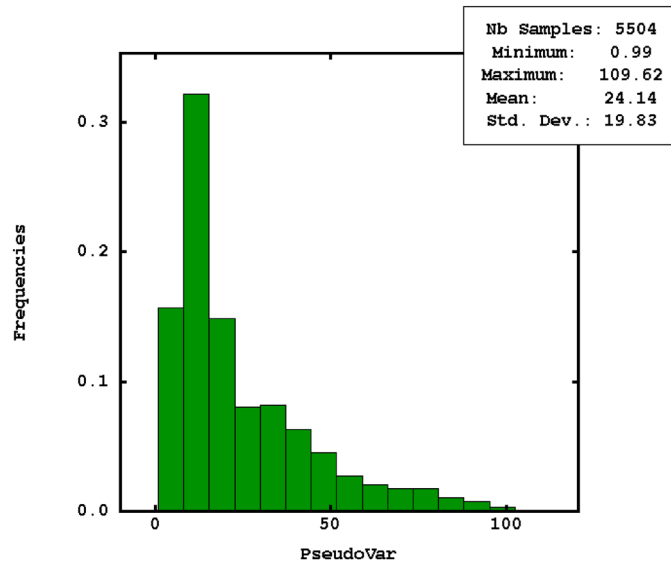


Figure 10. Summary statistics of the pseudo-variance (the pseudo-standard deviation squared) calculated from the minimum and maximum abundance values (kg/m²).

uncorrelated random component, are properties that are specific and unique for the investigated phenomenon whether it is geology, meteorology or social sciences.

The regionalized variable is normally so complex that a deterministic formulation is not feasible. A probabilistic framework is required (Hans Wackernagel 2010). In most applications of geostatistics, this probabilistic framework is based on a

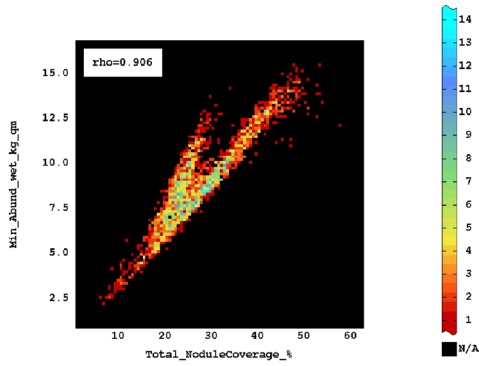


Figure 11. Relationship between total nodule coverage and calculated minimum abundance in the southeastern transect (102STR). Two distinctly different populations and relationships can be identified. Legend indicates point density in the scatter plot. Coverage varies between 6% and 57.5% with average of 27.5%. Minimum abundance varies from 2.2 to 15.5 with average of 9 kg/m².

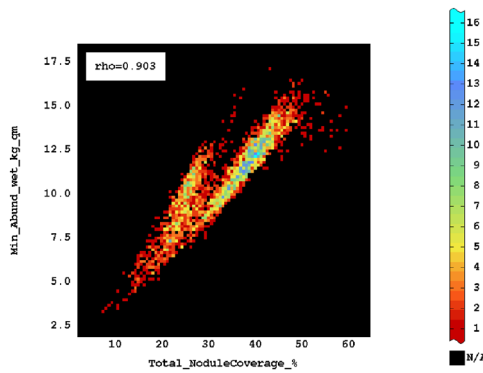


Figure 12. Relationship between total nodule coverage and calculated minimum abundance in the northwestern transect (131STR). Two distinctly different populations and relationships can be identified. Legend indicates point density in the scatter plot. Coverage varies between 7.1% and 59.2% with average of 33.2%. Minimum abundance varies from 3.3 to 17.1 with average of 10.8 kg/m².

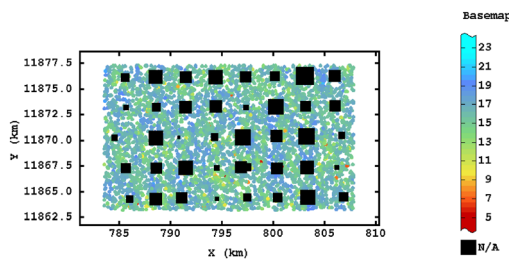


Figure 13. Images located randomly in the area of the box core data (black squares). Colors indicate the conditional expectation at each image location. Size of black squares is proportional to abundance.

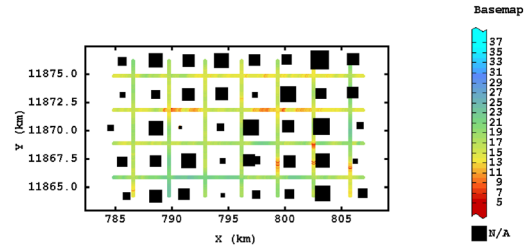


Figure 14. Images located in transects covering the area of the box core data (black squares). Colors indicate the conditional expectation at image location. Size of black squares is proportional to abundance.

modeling of the spatial correlation using the variogram, γ . The variogram quantifies the average squared difference between realizations of the regionalized variable, and it is used to find the weights λ_i that minimize the estimation variance of the linear estimator $Z^*(x)$, thus:

$$Z^*(x) = \sum_i \lambda_i * Z(x_i) \quad (3)$$

$Z(x_i)$ represents data at locations x_i and λ_i are the associated weights. Different geostatistical implementations make different assumptions on the properties of the drift, $m(x_0)$. Ordinary kriging assumes that the drift is unknown, but constant inside the study area. By introducing the Lagrange multiplier μ that ensures that the estimator $Z^*(x)$ is unbiased, the kriging system used to calculate the optimal weights that minimized the estimation variance can be expressed by Eqs. 4 and 5 (Armstrong 1998):

$$\sum_{j=1}^N \lambda_j \gamma(x_i, x_j) + \mu = \bar{\gamma}(x_i, V), \quad i = 1, 2, \dots, N \quad (4)$$

$$\sum_{i=1}^N \lambda_i = 1 \quad (5)$$

The $\gamma(x_i, x_j)$ is the variogram value between the data points at locations x_i and x_j that are used in the prediction of the volume or block V , and the $\bar{\gamma}(x_i, V)$ is the average variogram value between the data-points and the V . The kriging variance is given by Eq. (6). The kriging standard deviation is the square root of this kriging variance.

$$\sigma_K^2 = \sum_{j=1}^N \bar{\gamma}(x_i, V) - \bar{\gamma}(V, V) + \mu \quad (6)$$

Application of Soft Data in Nodule Resource Estimation

Gibbs Sampler

The Gibbs sampler is a variant of the Metropolis–Hastings algorithm (Robert 2015); it is used to sample from multivariate distributions by sampling systematically from the conditional distribution at each sample location. The applications of the Gibbs sampler vary widely from the use in meteorology (Onibon et al. 2004), environmental studies (Michalak 2008) to social sciences (Lynch 2007) and specifically in psychology (Yildirim 2012) and geosciences (Hansen et al. 2012). It was introduced to Bayesian statistics with the work done by Gelfand and Smith (1990).

In the present work, the datasets consist of hard data represented by the box core analyses and soft data represented by the images and the estimated minimum and maximum values. The algorithm used in this work comprises the following the workflow implemented in the geostatistical software Isatis™:

1. Transform hard data and n minimum and n maximum values defining the inequalities into Gaussian space $(N(0, I))$. Store the anamorphosis for back-transformation into original data space and subsequently the calculation of the conditional expectation and associated uncertainty.
2. Calculate the experimental variogram of the Gaussian hard data and fit a variogram model.
3. Initialize a vector of length n with values satisfying the minimum and the maximum limits at each soft data location; $x^t = x_1^t, \dots, x_n^t$, $t = 0$.
4. For $t = 0, 1, 2, \dots$ draw on random an index i between 1 (one) and n .
5. Use simple kriging (the expectation is known) to estimate a value (z_i) and the kriging standard deviation (σ_i) at location i from the hard data and $x_1^t, \dots, x_{i-1}^t, x_{i+1}^t, \dots, x_n^t$. This would be all hard data and all soft data except the value at location i . A unique neighborhood is used to ensure that all data are included in the simple kriging step.
6. Draw a value s from the conditional distribution $s \sim N(z_i, \sigma_i)$ and assign it to location i .
7. Assign $x^{t+1} = x_1^t, x_2^t, x_{i-1}^t, s, x_{i+1}^t, \dots, x_n^t$.
8. Back-transform x^{t+1} to real data space.
9. Store the back-transformed version of x^{t+1} .
10. Repeat steps 4–10 until convergence or the fixed number of simulations has been reached.
11. Calculate the *conditional expectation* as the arithmetic mean of the stored back-transformed

values and the associated *standard deviation* at each soft data location. The conditional expectation is the expected abundance at the image locations given the minimum and maximum constraints, the box core data and the variogram model quantifying the spatial correlation.

In our case, the early values are within boundaries because they have been sampled to satisfy the minimum and maximum limits, but these early values might not be good representatives of the final distribution. Three Gibbs sampler runs were executed to check for consistency. The conditional expectation and the associated standard deviation were used as input with the hard data in kriging with variance of measurement error following the workflow implemented in Isatis™.

Kriging with Variance of Measurement Error

Kriging with variance of measurement error is a procedure that enables the incorporation of data associated with different levels of uncertainty. The ordinary kriging system as used in this study on matrix form is given as

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1N} & 1 \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2N} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \dots & \gamma_{NN} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(x_1, V) \\ \bar{\gamma}(x_2, V) \\ \vdots \\ \bar{\gamma}(x_N, V) \\ 1 \end{bmatrix} \quad (7)$$

Including uncertain data, the extra nugget effect (σ_x) associated with the (more) uncertain data is added to the diagonal of the variance–covariance matrix (Heuvelink et al. 2016):

$$\begin{bmatrix} \gamma_{11} + \sigma_{x1} & \gamma_{12} & \dots & \gamma_{1N} & 1 \\ \gamma_{21} & \gamma_{22} + \sigma_{x2} & \dots & \gamma_{2N} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \dots & \gamma_{NN} + \sigma_{xN} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(x_1, V) \\ \bar{\gamma}(x_2, V) \\ \vdots \\ \bar{\gamma}(x_N, V) \\ 1 \end{bmatrix} \quad (8)$$

Table 2. Variogram model for the original abundance data.

	Sill	Directional scale, N70E/N250E	Directional scale, N160E/N340E
S1, nugget	3.8		
S2, spherical	17	700 m	700 m
S3, spherical	17	3400 m	13,000 m
Total sill	37.8		

Kang et al. (2017) used a similar approach in their study when they introduce measurement errors in the spatial prediction of soil moisture. Adding the extra nugget effect as indicated by Eq. (8) can be compared to have more data further away from the target block because these points are associated with a higher variogram value.

Kriging Parameterization

Variogram Model

An experimental variogram was estimated for both the original abundance data and its Gaussian transformation. Both were rotated N70E to account for anisotropy. The sills and direction scales are given in Table 2. A graphical representation of Table 2 is shown in Figure 15. The variogram model of the Gaussian transformed hard data is given in Table 3. Given only 41 box core data, the anisotropy can be questioned, but the identified directions correspond well with the NNW–SSE trending ridge and graben structures shown in Figure 1.

Block Size and Search Neighborhood

The block size was set to slightly smaller than the distance between the hard data, i.e., block size of 1 km × 1 km. This is in accordance with recommendations by Armstrong (1998) and Hekmat et al. (2013), and it considered out of scope for this study to optimize the block size and the search neighborhood configuration. The search neighborhood used in the kriging step was rotated according to the anisotropies with the long-axis-oriented N160E. The short axis and long axis of the search ellipsoid were set to 2900 m and 11,250 meters, respectively. This corresponds roughly to the ratio between the directional scales for variogram structure 3 in Table 2. The search neighborhood was divided into

eight sectors, and the maximum number of consecutive empty sectors was set to 2. This ensured interpolation rather than extrapolation. The optimum number of samples per sector was set to 2 giving an optimum number of samples of 16. The minimum number of samples was fixed to 3.

RESULTS

Conditional Expectations at Image Location

From Figures 16 and 17, it can be seen that the difference between the maximum and the conditional expectation was slightly larger than the difference between the conditional expectation and the minimum. This means that the distribution at each soft data point was slightly skewed with a tail toward higher values.

Figures 18 and 19 show the conditional expectation and the associated estimated variance, respectively. It can be seen from Figure 18 that the average conditional expectation was 17 kg/m². The variance of measurement error in Figure 19 at image location was bimodal, and it shows an average value of 9.2 (kg/m²)². This is comparable to the pseudo-variance in Figure 10 calculated from Eq. (1). The two modes in Figure 19 represent the variance of measurement error in the northern and the southern transects with image points shown in Figure 5. The variance of measurement error varied from 1 to 21.8 (kg/m²)², which is 58% of the total sill of 38 (kg/m²)² given in Table 2. The three independent Gibbs sampler runs gave no significant differences in conditional expectations and variance of measurement errors.

Resource Estimation

Resource estimation results using box core data only and box core data and images in different

Application of Soft Data in Nodule Resource Estimation

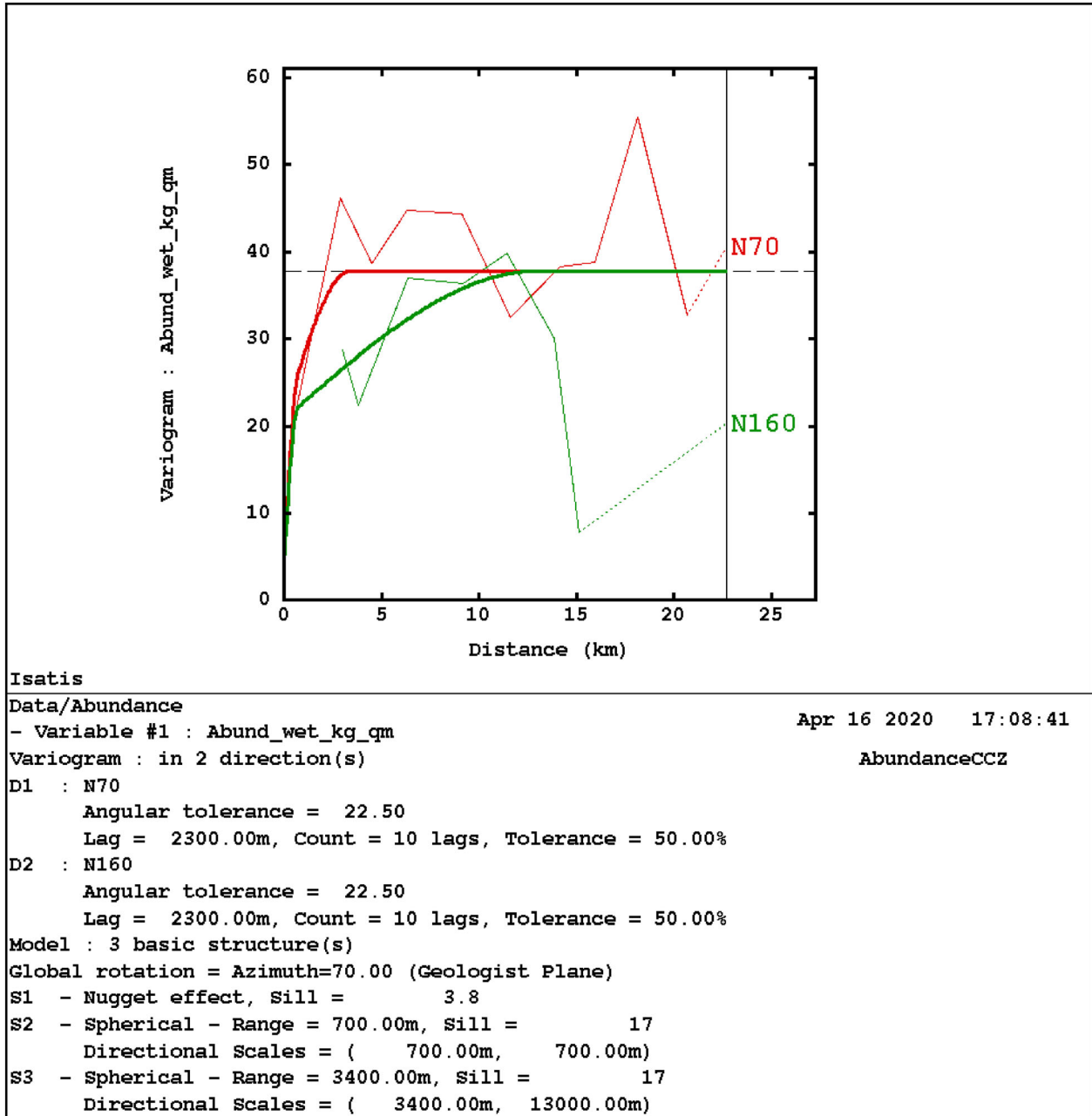


Figure 15. Variogram model for the original abundance data (in kg/m²).

configurations are given in Table 4. The estimated value was higher if only box core data were used. The kriging standard deviation was lower for the estimation that included image data and improved significantly in the alternative image configurations. RPE was higher when images were included because the estimated value on average was significantly

lower. The Lagrange multiplier was larger when images were included in their original locations. This indicates that the data points were clustered and/or there were more cases of extrapolation. For this search neighborhood, the slopes and the weight of the mean were practically the same with or without images. This would indicate the same conditional

Table 3. Variogram model for the Gaussian transformed hard data.

	Sill	Directional scale, N70E/N250E	Directional scale, N160E/N340E
S1, nugget	0.22		
S2, spherical	0.33	1000 m	2500 m
S3, Spherical	0.45	2500 m	10,000 m
Total sill	1		

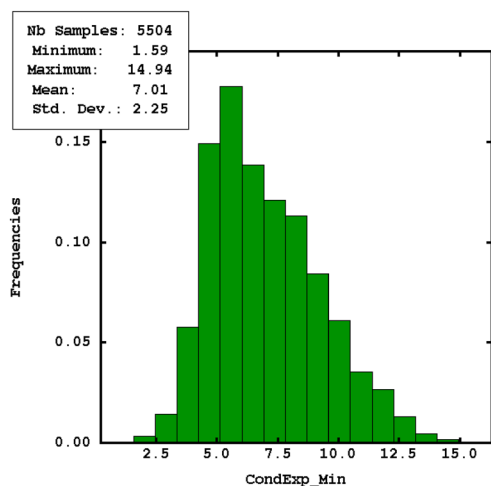


Figure 16. Histogram showing the difference between the conditional expectation and the minimum value at each image location.

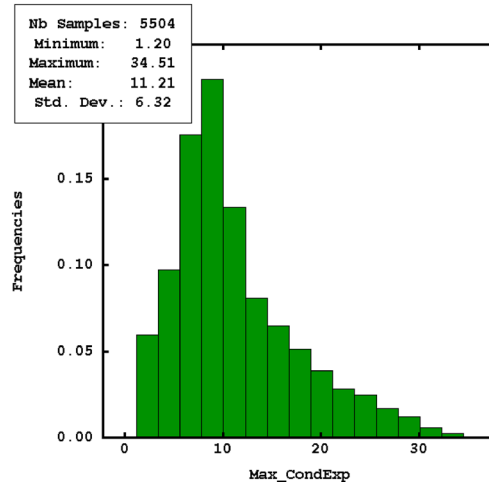


Figure 17. Histogram showing the difference between the maximum value and the conditional expectation at each image location.

bias. The same parameters were significantly better in the two alternative data configurations. At the same time, the variance of the estimated block values was higher when the image data were included. This is a result of more low-grade blocks.

The mean distance to data points included in the estimation was naturally lower when the image data were included and, from a pure data configuration perspective, the estimates with image data were therefore better. The same was the case if one looks at the number of values inside the neighborhood. This indicator was larger for the case where the images were included. In other words, the estimate was “better informed.” Given the requirements on the data points inside the search neighborhood and the sector division of the neighborhood, the number of estimated blocks increases as expected with the introduction of the image points.

The slope that indicates the degree of conditional unbiasedness shows similar results with and without images, with an average slope of 0.76 with images at the original locations and 0.78 without.

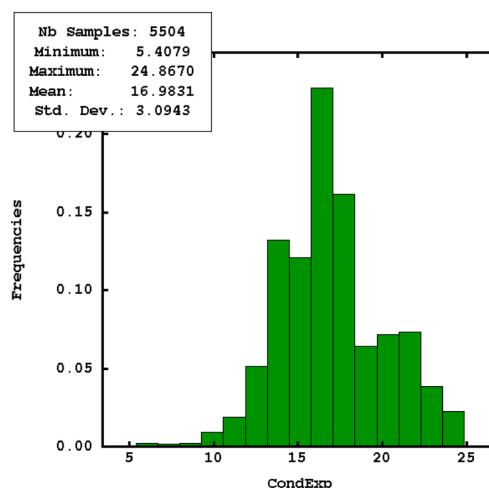


Figure 18. Estimated conditional expectation at image location.

The kriging standard deviation shows a small improvement from 3.44 without to 3.32 with images. The weight of mean shows significant improvements

Application of Soft Data in Nodule Resource Estimation

in the alternative image configurations, dropping down to 0.07 for the operationally unlikely event of having the images points spread at random across the area. The more likely case where the images are placed along transects covering the area of interest,

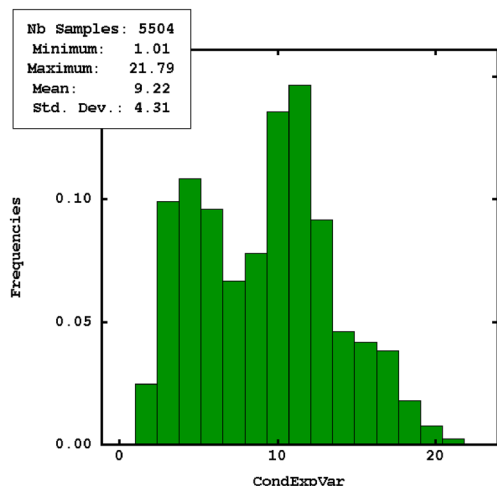


Figure 19. Estimated variance at image location.

also shows a significant improvement with a drop down to 0.15 from 0.29.

Figures 20 and 21 show estimation results with box core data only and with images in transects covering the whole area. Comparing the results in these figures, the results in Figure 21 appear more noisy indicating that the estimation was to a greater extent controlled by data rather than the global model defined by the variogram. For the slope, the standard deviation was lower when image data were included because the slope was constantly higher. Looking at the number of blocks that were estimated, there was an increase from 215 blocks when only box core data were used to 236 when data from images in transects were included. This was a 10% increase without compromising the estimation quality, just using the available data. Assuming an average abundance of 17.5 kg/m² and the block size of 1 km², the increase in unclassified resource amounted to 367.500 tonnes of nodules.

DISCUSSION

How can we define a resource and how can the inclusion of uncertain image data influence the re-

Table 4. Resource estimation using only box cores (hard data) and box core data and image data in different configurations.

	Estimate	Stdv	Slope	WoM	Lagrange	RPE	Mean distance	Data points in estimation
<i>Box cores only</i>								
Number of blocks	215							
Min	16.18	2.26	0.62	0.16	- 3.84	0.18	2919	5.00
Max	30.17	3.95	0.92	0.42	- 1.17	0.44	4854	10.00
Mean	23.22	3.44	0.78	0.29	- 2.28	0.30	3954	7.40
Variance	8.34	0.06	0.00	0.00	0.36	0.00	115,443	1.02
<i>Images and box cores, images at original location</i>								
Number of blocks	221							
Min	12.97	2.04	0.59	0.15	- 7.28	0.19	306	6.00
Max	29.54	3.93	0.92	0.44	- 1.23	0.52	4741	14.00
Mean	20.89	3.32	0.76	0.29	- 3.03	0.32	2983	9.49
Variance	10.66	0.18	0.00	0.00	0.90	0.00	1,019,208	4.66
<i>Images and box cores, images in transects</i>								
Number of blocks	236							
Min	12.11	1.61	0.76	0.10	- 5.50	0.17	392	7.00
Max	23.01	3.04	0.93	0.25	- 1.23	0.42	1610	16.00
Mean	17.54	2.36	0.88	0.15	- 2.01	0.27	1135	15.34
Variance	4.88	0.14	0.00	0.00	0.20	0.00	87,843	1.70
<i>Images and box cores, images with random location</i>								
Number of blocks	327							
Min	13.17	0.98	0.85	0.04	- 3.16	0.12	244	11.00
Max	19.52	1.90	0.96	0.16	- 0.64	0.22	682	16.00
Mean	16.5	1.30	0.94	0.07	- 1.18	0.16	479	15.73
Variance	0.87	0.02	0.00	0.00	0.14	0.00	5547	1.01

The estimate and the RPE should not be compared

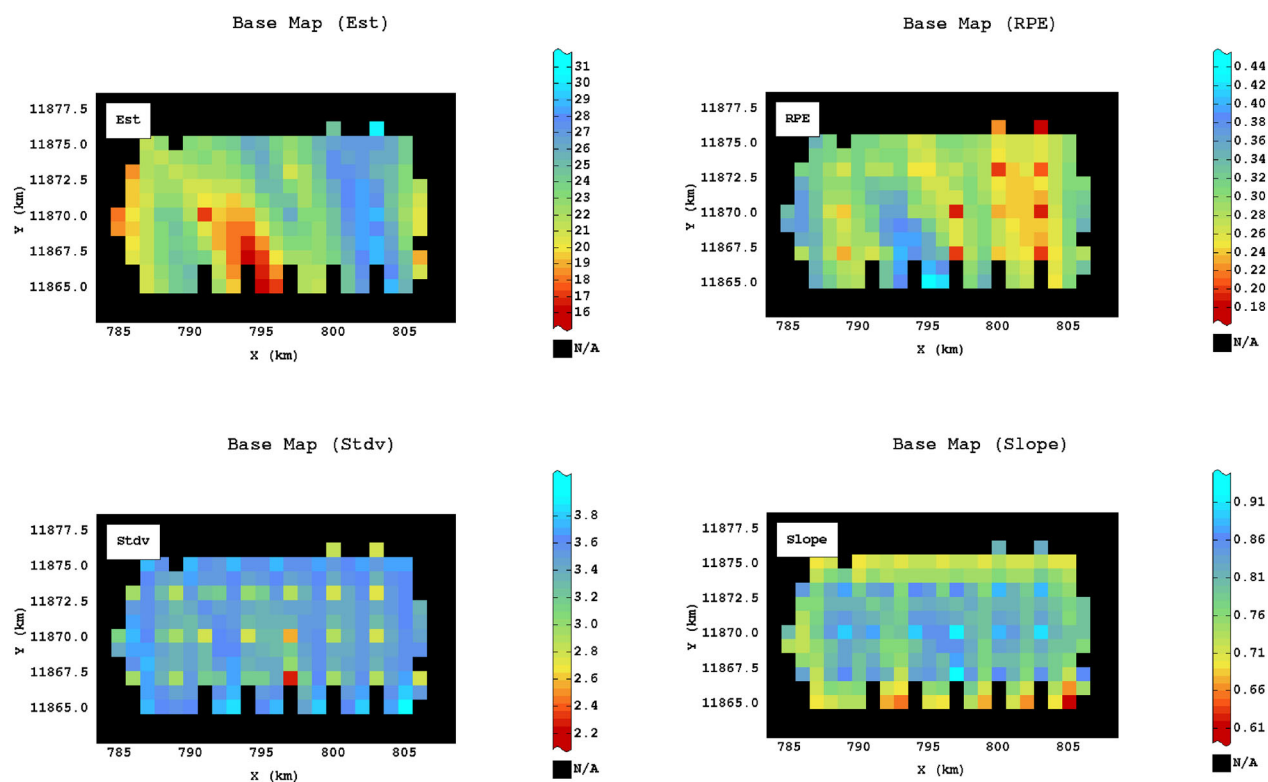


Figure 20. Estimation and quality indicators with only box core data. From upper left and clockwise: the estimated value, the RPE, the slope and the kriging standard deviation.

source classification? What parameter should we use to inform the classification? How well suited is the relative prediction error (RPE) as a kriging quality indicator if it depends on the estimated value and uncertain image data are incorporated? Answers to these and other related and relevant questions must be assessed and found by the competent person (CP) who does the resource estimation and the resource classification. The reporting codes must be read and understood. Geological continuity is not necessarily the same as continuity in kriging quality indicators that are dependent on a series of assumption made by the CP. In this work, a methodology was explored where information was retrieved from images, combined with expert knowledge and conditional expectations at image points were calculated. Other methodologies discussed here are available, some of which could potentially have been used for the presented purpose, e.g., kriging with inequalities presented by Abrahamsen and Benth (2001).

None of the kriging indicators can alone be used directly as a classification indicator. The Lagrange multiplier can be used to assess clustering

and the amount of extrapolation and can be, in co-use with the kriging variance, an efficient indicator of data configuration. Since the neighborhood was defined using eight sectors and maximum two empty consecutive sectors were allowed, there was little extrapolation. The data coverage was good. Given that the image data were densely organized along rather short transects, we knew that there was clustering because the image points were grouped close together. This explains the higher Lagrange multiplier when the images were included. The kriging variance was also the basic tool in developing confidence intervals and was naturally dependent on the block size. The larger the block size was, the lower was the kriging variance. The selection of block sizes is subjective, but it is often linked to some annual or semi-annual production tonnages. Here, it was fixed according to data density and it was detached from any operational constraints.

Pure distance criteria can be effective, assuming they are derived from the variogram model, but this approach collapses when data of different quality are combined. Slope and weight of mean were good

Application of Soft Data in Nodule Resource Estimation

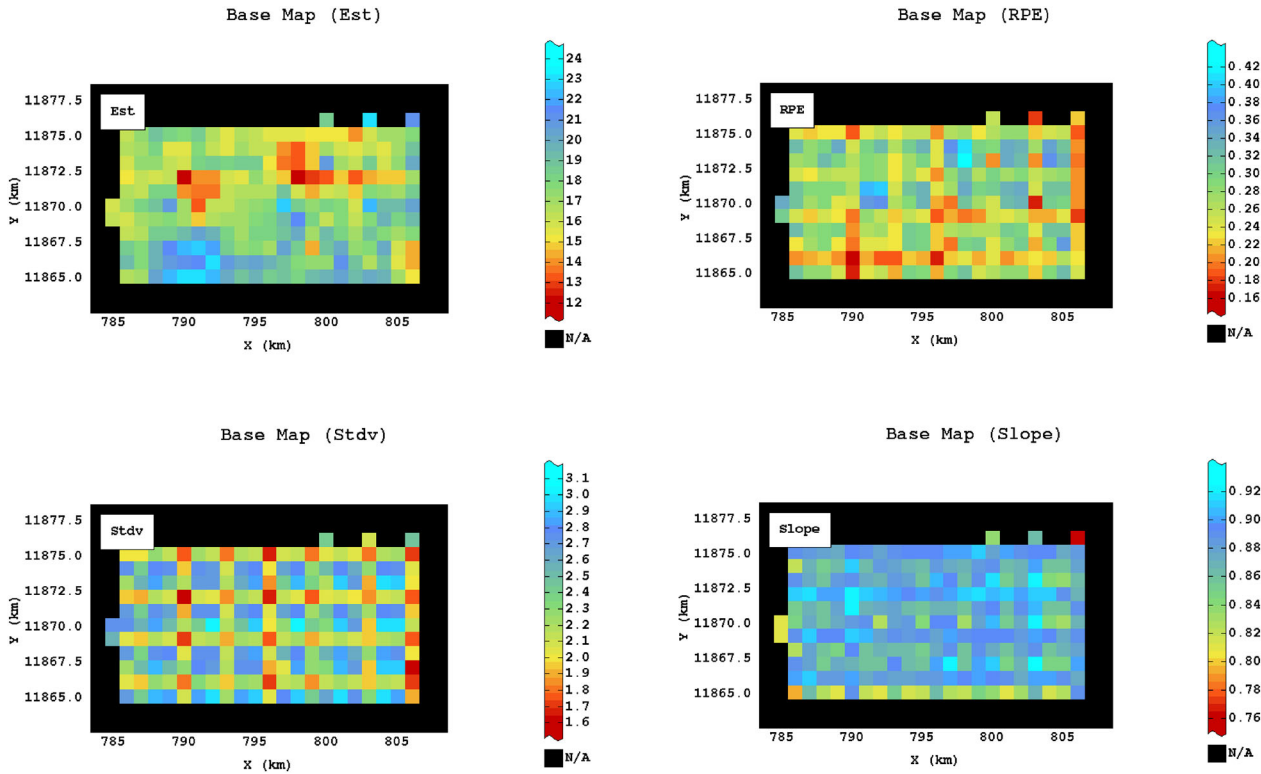


Figure 21. Estimation and quality indicators with box core and image data. From upper left and clockwise: the estimated value, the RPE, the slope and the kriging standard deviation.

indicators to use in assessing the degree of conditional bias, but the relevance of conditional bias as a criterion was dependent on the purpose of the estimation. The kriged estimates that included the image data were lower than estimates based only on box core data. The reason is that the conditional expectation was lower than the average abundance from the box core data. This is something future activities on predicting the minimum and maximum abundances must focus on. The minimum and maximum values influence naturally the conditional expectations at image location and one could expect that, given representative box core and image sampling, the average conditional expectation should be similar to the average abundance in the material collected with the box cores. In the data presented here, this was not the case; the conditional expectations were lower. Looked from another perspective, box cores might not be able to capture the variability of abundance in the area and the available box cores might seriously over-estimate the abundance.

The degree of improvement when image retrieved information is incorporated is dependent on the uncertainty associated with the conditional expectations at image location. This uncertainty is directly dependent of the defined minimum and maximum values at image locations and the variogram model. There are interesting approaches that can be exploited to improve the minimum and maximum values. These include, for example, the use of backscatter intensity or potentially links between bathymetric expressions. The area under study shows a variable nodule size distribution. In such a case, the simple correlations between coverage and abundance are no longer valid and other approaches must be assessed. The reason for the collapse was primarily because large nodules tend to be covered with more sediments than smaller nodules and it is very difficult to assess the amount of sediment cover. A simple correlation approach followed by ordinary kriging also failed to incorporate the extra uncertainty associated with abundance that was estimated from the correlation. The global extra uncertainty derived from the correlation coefficient

can be built-in into the methodologies presented here, but that would not be a location-dependent uncertainty that was achieved with the min/max approach presented here.

The areas covered by the images and the box cores varied, and the support associated with each data point was therefore not consistent. In this study, all data were assumed point data compared to the blocks that were estimated. An extension to what has been tested in this study can be to calculate the dispersion variance within the image areas similar to the study of Castrignanò et al. (2019). This dispersion variance would then affect the total variance of measurement error associated with each image point.

Nodules are formed through both hydrogenetic and diagenetic processes with a dominance of the diagenetic process in almost all nodules. However, in small nodules, the hydrogenetic fraction seems to be higher (Heller et al. 2018), and because hydrogenetic process means direct precipitation from near-bottom seawater it can only be realized under reduced or no sedimentation. This might be the reason why small sized nodules are located on the seafloor with less sediment cover than the larger, more diagenetically formed nodules.

Looking away from the fact that the soft data are associated with higher uncertainty, the estimates using the image data are better informed. The average distance between the target block and the data was smaller and the number of data points involved in the estimation was larger. At the same time, the slope, the weight of mean and the kriging variance were practically the same for the original data configuration with images along northern and southern transects only covering parts of the area of interest. The kriging quality parameters improve significantly if other data configurations are used. Here, tests were made distributing the images randomly and along transects in the area of the box cores. The former is an operationally unlikely scenario, but running photo transects at different orientations covering different parts of the study area is more likely. Both configurations show significant improvements in the kriging quality indicators while having averages of number of points included in the estimation close to the maximum of 16 stemming from the definition of the search neighborhood.

It is out of scope for this study to assess and to incorporate the effect of differences in support between the box cores and the images. The box cores covered an area of about 0.25 m^2 , while the images

were taken at different altitudes and they covered an area between 0.6 m^2 to roughly 12 m^2 with median area of 5.5 m^2 and mean area of 5.7 m^2 . All these supports were assessed as being point data compared to the 1000 m^2 block area. Variograms calculated from the minimum and maximum data (not presented here) showed similar range structures, but naturally both lower (for the minimum data) and higher (for the maximum data) sills. This corresponded well with the fact that the standard deviations of the box core data, the minimum and maximum data presented in Figures 6, 7 and 8, were 6.15 , 2.5 and 8.4 kg/m^2 , respectively.

CONCLUSIONS

Conditional expectations and associated uncertainty retrieved from images of manganese nodules using the Gibbs sampler have been incorporated with box core data. The difference in the soft and hard data's ability to quantify correctly the nodule abundance is used in the calculation of the kriging weights. Given the uncertainty in the image data and their original location, the incorporation of the image data does not improve the kriging quality indicators, but it does not worsen it either. From a data configuration point of view, the estimations with images are better informed due to more and closer data points being involved in the estimation. Also given the restricted neighborhood with requirements on the minimum number of data points, the number of sectors and the maximum number of consecutive empty sectors, more target blocks are estimated when the image data are included. Another conclusion is that photo transects should be carried out in different directions covering different parts of the study area rather than following one direction over already known box core locations. Future developments will focus on improving the estimates of the minimum and the maximum values from the images.

ACKNOWLEDGMENTS

This research has been supported by the Federal Institute for Geosciences and Natural Resources (BGR) under the Project Number A-0203002.A. We wish to thank the captains and crews of research vessels RV Kilo Moana and RV Sonne for their effective cooperation during the explo-

Application of Soft Data in Nodule Resource Estimation

ration campaigns at sea. Furthermore, we acknowledge Nils Jeisecke from Saltation GmbH for his advice with the image software MANGAN Analyzer.

FUNDING

Open Access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital).

OPEN ACCESS

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Abrahamsen, P., & Benth, F. E. (2001). Kriging with inequality constraints. *Mathematical Geology*, 33, 719–744. <https://doi.org/10.1023/A:10111078716252>.
- Armstrong, M. (1998). *Basic linear geostatistics*. Berlin: Springer.
- Benndorf, J. (2015). Vorratsklassifikation nach internationales Standards - Anforderungen und Modellenansätze in der Lagerstättenbearbeitung. *Markscheidewesen*, 122(2–3), 6–14.
- Blom, M., Pearce, A. R., & Stuckey, P. J. (2019). Short-term planning for open pit mines: a review. *International Journal of Mining, Reclamation and Environment*, 33(5), 318–339.
- Boucher, A., & Dimitrakopoulos, R. (2012). Multivariate block-support simulation of the Yandi iron ore deposit, Western Australia. *Mathematical Geosciences*, 44(4), 449–468.
- Buyong, T. (2007). *Spatial data analysis for geographic information science*. Penerbit UTM, Skudai, Johor Bahru. <http://eprints.utm.my/id/eprint/30021/>.
- Castrignanò, A., Quarto, R., Venezia, A., & Buttafuoco, G. (2019). A comparison between mixed support kriging and block cokriging for modelling and combining spatial data with different support. *Precision Agriculture*, 20(2), 193–213.
- Cathles, L. M. (2011). What processes at mid-ocean ridges tell us about volcanogenic massive sulfide deposits. *Mineralium Deposita*, 46(5–6), 639–657.
- Chilès, J. P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty* (Vol. 713). London: Wiley.
- Eidsvik, J., & Ellefmo, S. L. (2013). The value of information in mineral exploration within a multi-gaussian framework. *Mathematical Geosciences*, 45(7), 777–798.
- Eidsvik, J., Mukerji, T., & Bhattacharjya, D. (2015). *Value of information in the earth sciences: Integrating spatial modeling and decision analysis*. Cambridge: Cambridge University Press.
- Ellefmo, S. L., & Kuhn, T. (2018). Towards an improved nodule resource estimation and classification using hard and soft data. *Presented at the 47th Underwater Mining Conference, Bergen*.
- Ellefmo, S. L., Søreide, F., Cherkashov, G., Juliani, C., Panthi, K. K., Petukhov, S., et al. (2019). Quantifying the unknown. *Cappelen Damm Akademiske*. <https://doi.org/10.23865/noasp.81>.
- Elskhaki, A., Graedel, T. E., Ciacci, L., & Reck, B. K. (2018). Resource demand scenarios for the major metals. *Environmental Science and Technology*, 52(5), 2491–2497.
- Felix, D. (1980). Some problems in making nodule abundance estimates from seafloor photographs. *Marine Mining*, 2, 293–302.
- Gazis, I.-Z., Schoening, T., Alevizos, E., & Greinert, J. (2018). Quantitative mapping and predictive modeling of Mn nodules' distribution from hydroacoustic and optical AUV data linked by random forests machine learning. *Biogeosciences*, 15(23), 7347–7377.
- Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410), 398–409. <https://doi.org/10.2307/2289776>.
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford: Oxford University Press.
- Grunwald, S. (2005). What do we really know about space-time continuum of soil-landscapes? In S. Grunwald (Ed.), *Environmental soil-landscape modeling: Geographic information technologies and pedometrics* (pp. 3–36). Boca Raton: CRC Taylor & Francis.
- Hannington, M. D. (2013). The role of black smokers in the Cu mass balance of the oceanic crust. *Earth and Planetary Science Letters*, 374, 215–226.
- Hannington, M., Jamieson, J., Monecke, T., Petersen, S., & Beaulieu, S. (2011). The abundance of seafloor massive sulfide deposits. *Geology*, 39(12), 1155–1158. <https://doi.org/10.1130/G32468.1>.
- Hansen, T. M., Cordua, K. S., & Mosegaard, K. (2012). Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling. *Computational Geosciences*, 16(3), 593–611.
- Hein, J. R., Koschinsky, A., & Kuhn, T. (2020). Deep-ocean polymetallic nodules as a resource for critical materials. *Nature Reviews Earth & Environment*, 1(3), 158–169.
- Hekmat, A., Osanloo, M., & Moarefvand, P. (2013). Block size selection with the objective of minimizing the discrepancy in real and estimated block grade. *Arabian Journal of Geosciences*, 6(1), 141–155.
- Heller, C., Kuhn, T., Versteegh, G. J. M., Wegorzewski, A. V., & Kasten, S. (2018). The geochemical behavior of metals during early diagenetic alteration of buried manganese nodules. *Deep Sea Research Part I: Oceanographic Research Papers*, 142, 16–33.

- Heuvelink, G. B. M., Brus, D., Hengl, T., Kempen, B., Leenaars, J. G. B., & Ruiperez-Gonzalez, M. (2016). Uncertainty quantification of interpolated maps derived from observations with different accuracy levels. In *Proceedings of spatial accuracy 2016* (pp. 49–51). Presented at the 12th international symposium on spatial accuracy assessment in natural resources and environmental sciences, accuracy 2016, Montpellier, France: International Spatial Accuracy Research Association (ISARA). Retrieved April 15, 2020, from <https://research.wur.nl/en/publications/uncertainty-quantification-of-interpolated-maps-derived-from-obs>.
- Hozo, S. P., Djulbegovic, B., & Hozo, I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*, 5(1), 13. <http://doi.org/10.1186/1471-2288-5-13>.
- ISA. (2015). Reporting standard of the international seabed authority for mineral exploration results assessments, mineral resources and mineral reserves. International Seabed Authority. Retrieved April 17, 2020, from <https://ran-s3.s3.amazonaws.com/isa.org/jm/s3fs-public/documents/EN/Contrac ts/Templates/AnnexV.pdf>.
- Isaaks, E. (2005). The kriging oxymoron: A conditionally unbiased and accurate predictor. In O. Leuangthong, & C. V. Deutsch (Eds.), *Geostatistics Banff 2004* (2nd ed., Vol. 14, pp. 363–374). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-3610-1_37.
- JORC. (2012). Australasian code for reporting of exploration results, mineral resources and ore reserves. Joint Ore Reserves Committee: The Australasian Institute of Mining and Metallurgy, Australian Institute of Geoscientists and Minerals Council of Australia. Retrieved March 20, 2020, from http://www.jorc.org/docs/JORC_code_2012.pdf.
- Journel, A. G., & Huijbregts, C. (1978). *Mining geostatistics*. London: Academic Press.
- Kang, J., Jin, R., Li, X., & Zhang, Y. (2017). Block kriging with measurement errors: A case study of the spatial prediction of soil moisture in the middle reaches of Heihe River Basin. *IEEE Geoscience and Remote Sensing Letters*, 14(1), 87–91.
- Knobloch, A., Kuhn, T., Rühlemann, C., Hertwig, T., Zeissler, K.-O., & 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 R. Sharma (Ed.), *Deep-sea mining* (pp. 189–212). Cham: Springer. https://doi.org/10.1007/978-3-319-52557-0_6.
- Koschinsky, A., & Hein, J. R. (2003). Uptake of elements from seawater by ferromanganese crusts: Solid-phase associations and seawater speciation. *Marine Geology*, 198(3–4), 331–351.
- Kuhn, T., & Shipboard Scientific Party. (2015). *Low-temperature fluid circulation at seamounts and hydrothermal pits: Heat flow regime, impact on biogeochemical processes and its potential influence on the occurrence and composition of manganese nodules in the NE Pacific*, Cruise Report SO240/FLUM (No. https://doi.org/10.2312/cr_so240) (p. 185). Hannover, Germany: Bundesanstalt für Geowissenschaften und Rohstoffe. Retrieved April 15, 2020, from https://www.tib.eu/de/suchen/?tx_tibsearch_search%5Bdocid%5D=awi%3AAd oi~10.2312%252Fcr_so240&tx_tibsearch_search%5Bcont roller%5D=Download&cHash=12ef30b4e13ca8dc5fe5bab6 2a08bf26#download-mark.
- Kuhn, T., & Rathke, M. (2017). *Visual data acquisition in the field and interpretation for SMnN* (No. BLUE MINING D1.31b). Retrieved April 24, 2020, from https://bluemining.eu/download/project_results/public_reports/BLUE-MINING-D1.31b-Final-Report-on-visual-data-acquisition-in-the-field-and-interpretation-for-SMnN.pdf.
- Kuhn, T., Rühlemann, C., & Knobloch, A. (2016). Classification of manganese nodule estimates: Can we reach the “measured resource” level? In S. Hong (Ed.), *Resource and environmental assessments for seafloor mining development*. Presented at the Underwater Mining Conference, Incheon, South Korea. Retrieved April 16, 2020, from <http://oceanre p.geomar.de/43434/>.
- Kuhn, T., Uhlenkott, K., Martinez, P., Vink, A., & Rühlemann, C. (2020). Manganese Nodule Fields from the NE Pacific as Benthic Habitats. In P. Harris & E. Baker (Eds.), *Seafloor geomorphology as benthic habitat* (2nd ed., pp. 933–947). Amsterdam: Elsevier.
- Kuhn, T., Versteegh, G. J. M., Villinger, H., Dohrmann, I., Heller, C., Koschinsky, A., et al. (2017a). Widespread seawater circulation in 18–22 Ma oceanic crust: Impact on heat flow and sediment geochemistry. *Geology*, 45(9), 799–802.
- Kuhn, T., Wegorzewski, A., Rühlemann, C., & Vink, A. (2017b). Composition, formation, and occurrence of polymetallic nodules. In R. Sharma (Ed.), *Deep-sea mining: Resource potential, technical and environmental considerations* (pp. 23–63). Cham: Springer. https://doi.org/10.1007/978-3-319-52557-0_2.
- Lipton, I. T., Nimmo, M. J., & Parianos, J. M. (2016). *NI 43-101 technical report TOML Clarion Clipperton Zone project, Pacific Ocean* (p. 280).
- Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York: Springer.
- Månberger, A., & Stenqvist, B. (2018). Global metal flows in the renewable energy transition: Exploring the effects of substitutes, technological mix and development. *Energy Policy*, 119, 226–241.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8), 1246–1266. <https://doi.org/10.2113/gsecongeo.58.1246>.
- Mewes, K., Mogollón, J., Picard, A., Rühlemann, C., Kuhn, T., Nöthen, K., et al. (2014). Impact of depositional and biogeochemical processes on small scale variations in nodule abundance in the Clarion-Clipperton Fracture Zone. *Deep Sea Research Part I: Oceanographic Research Papers*, 91, 125–141.
- Michalak, A. M. (2008). A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling: Constrained Geostatistical Gibbs sampler. *Water Resources Research*. <https://doi.org/10.1029/2007WR006645>.
- Mucha, J., & Wasilewska-Błaszczuk, 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(3), 263. <https://doi.org/10.3390/min10030263>.
- Nowak, M., & Leuangthong, O. (2017). Conditional Bias In Kriging: Let’s keep it. In J. J. Gómez-Hernández, J. Rodrigo-Ilarri, M. E. Rodrigo-Clavero, E. Cassiraga, & J. A. Vargas-Guzmán (Eds.), *Geostatistics valencia 2016* (pp. 303–318). Cham: Springer. https://doi.org/10.1007/978-3-319-46819-8_2_0.
- Omre, H. (1987). Bayesian kriging–merging observations and qualified guesses in kriging. *Mathematical Geology*, 19, 15. <https://doi.org/10.1007/BF01275432>.
- Onibon, H., Lebel, T., Afouda, A., & Guillot, G. (2004). Gibbs sampling for conditional spatial disaggregation of rain fields. *Water Resources Research*. <https://doi.org/10.1029/2003WR002009>.
- Rivoirard, J. (1987). Two key parameters when choosing the kriging neighborhood. *Mathematical Geology*, 19(8), 851–856.
- Robert, C. P. (2015). The Metropolis-Hastings algorithm. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri, & J. L. Teugels (Eds.), *Wiley StatsRef: Statistics reference online* (pp. 1–15). Chichester: Wiley. <https://doi.org/10.1002/9781118445112.stat07834>.

Application of Soft Data in Nodule Resource Estimation

- Rossi, M. E., & Deutsch, C. V. (2013). *Mineral resource estimation*. New York: Springer.
- Rühlemann, C., & Shipboard Scientific Party. (2018). *Geology, Biodiversity and Environment of the German license area for the exploration of polymetallic nodules in the equatorial NE Pacific. Cruise Report of R/V SONNE Cruise MANGAN 2018* (p. 173). Hannover: Bundesanstalt für Geowissenschaften und Rohstoffe.
- Schöning, T., Kuhn, T., & Nattkemper, T. W. (2012). Estimation of polymetallic nodule coverage in benthic images. In H. Zhou & C. L. Morgan (Eds.), *Marine minerals: Finding the right balance of sustainable development and environmental protection* (p. 11). Shanghai: The Underwater Mining Institute. Retrieved April 15, 2020, from <http://www.timmschoening.de/assets/pdfs/Schoening-UMI2012.pdf>.
- Sharma, R. (2017). *Deep-sea mining: Resource potential, technical and environmental considerations*. New York, NY: Springer.
- Singer, D. A. (2014). Base and precious metal resources in seafloor massive sulfide deposits. *Ore Geology Reviews*, 59, 66–72.
- Singer, D. A. (2017). Future copper resources. *Ore Geology Reviews*, 86, 271–279.
- Snowden. (2017). Supervisor 8.7—Multi-block kriging neighbourhood analysis. *Snowden Group Mining and Technology Consultants*. Retrieved February 19, 2020, from <https://snowdengroup.com/news/supervisor-8-7-multi-block-kriging-neighbourhood-analysis/>.
- Teske, S. (Ed.). (2019). *Achieving the Paris climate agreement goals: Global and regional 100% renewable energy scenarios with non-energy GHG pathways for +1.5°C and +2°C*. Cham: Springer. <https://doi.org/10.1007/978-3-030-05843-2>.
- Wackernagel, H. (2010). *Multivariate geostatistics. Softcover version of original hardcover edition 2003* (3rd ed.). Berlin: Springer.
- Wackernagel, H., Bertino, L., Sierra, J. P., & Gonzales del Rio, J. (2002). Multivariate kriging for interpolating with data from different sources. In C. W. Anderson, V. Barnett, P. C. Chatwin, & A. H. El-Shaarawi (Eds.), *Quantitative methods for current environmental issues* (pp. 57–75). London: Springer, London. <https://doi.org/10.1007/978-1-4471-0657-9>.
- Yildirim, I. (2012). *Bayesian inference: Gibbs sampling* (p. 6). Retrieved February 1, 2020, from <http://www.mit.edu/~ilker/papers/GibbsSampling.pdf>.
- Yoo, C. M., Joo, J., Lee, S. H., Ko, Y., Chi, S.-B., Kim, H. J., et al. (2018). Resource assessment of polymetallic nodules using acoustic backscatter intensity data from the Korean exploration area, northeastern equatorial Pacific. *Ocean Science Journal*, 53(2), 381–394.