
Article

Empirical Application of Generalized Rayleigh Distribution for Mineral Resource Estimation of Seabed Polymetallic Nodules

Gordon Yu ^{1,*} and John Parianos ¹

¹ Nautilus Minerals Pacific Ltd, East Brisbane, 4169, Australia

* Correspondence: gordonyu2000@gmail.com

Abstract: An effective empirical statistical method is developed to improve the process of mineral resource estimation of seabed polymetallic nodules and is applied to analyse the abundance of seabed polymetallic nodules in the Clarion Clipperton Zone (CCZ). The newly proposed method is based on three hypotheses as the foundation for a model of “Idealized Nodules”, which was validated by analysing nodule samples collected from the seabed within the Tonga Offshore Mining Limited (TOML) exploration contract. Once validated, the “Idealized Nodule” model was used to deduce a set of empirical formulae for predicting the nodule resources, in terms of Percentage Coverage and Abundance. The formulae were then applied to analysing a total of 188 sets of nodule samples collected across the TOML areas, comprising box-core samples and towed camera images collected by one of the authors and detailed in [4]. The analysis also relies upon detailed box-core sample measurements from other areas reported by [7]. Numerical results for resource prediction were compared with field measurements, and reasonable agreement has been achieved. The new method has the potential to achieve more accurate mineral resource estimation with reduced sample numbers and sizes. They may also have application in improving the efficiency of design and configuration of mining equipment.

Keywords: polymetallic nodules; mineral resource estimation; statistical analysis; Generalized Rayleigh Distribution; Clarion Clipperton Zone

1. Introduction

Polymetallic nodules are mineral particles found in many of the world’s oceans [1]. A major deposit lies within the Clarion Clipperton Zone (CCZ) of the tropical North Pacific [2]. Nodules grow via precipitation in an organized manner in and on clay-ooze at the seabed [2] and they are often found with others of similar size and form [3–5]. Nodule ‘abundance’ is the kg (usually wet) weight of nodules per square metre of seabed and is used to estimate tonnage of nodules in a mineral resource estimation (as the surrounding clay-ooze should be able to be disregarded at the first step of mining [4]). Interest in the deposit, from the perspectives of development, marine environment and regulation, has increased over the last ten years [5,6].

In Section 2, three hypotheses are proposed as the basis for an idealized model of seabed polymetallic nodules. The hypotheses made for the “idealized nodule” model are based on analyses of nodule samples collected from the seabed at CCZ. One of the key hypotheses is based on numerical evidence that long axes of the seabed nodules follow the Generalized Rayleigh Distribution (GRD).

Section 3 presents the mathematical characteristics of the GRD pertaining to the analysis of nodules samples. The traditional statistical methods for estimating the parameters of the sample distribution, and for performing the Goodness-of-fit test for GRD are discussed. While they are found useful for analysing nodule samples, the traditional numerical procedure is too complex for practical applications.

In Section 4, based on the “idealized nodule” model discussed in Section 2, a simplified practical approach is developed to replace the complex numerical methods in Section 3. As a result, empirical formulae are derived to directly predict the parentage coverage and abundance of seabed nodules.

Section 5 shows the numerical results of testing the three hypotheses as the basis for “idealized nodule” model. Strong numerical evidence is found, providing validation to hypotheses and consequently the “idealized-nodule” model.

In Section 6, a total of 188 sets of samples of long axes of seabed polymetallic nodules from the CCZ are analysed using the empirical method developed in Section 4. The resource estimations in terms of percentage coverage and abundance are compared with the field measurements.

2. Three hypotheses for an idealized model of seabed polymetallic nodules

Polymetallic nodules from the CCZ are found in a wide range of forms [4], but within parts of the central and eastern CCZ covered by an exploration contract held by Tonga Offshore Mining Limited (TOML) they often form irregular slightly prolate spheroid-like forms ([4]; Figure 1). Growth around the horizontal axes (X and Y) is believed to be a function of horizontal space and mineral supply, and growth along the vertical axis is also a function of permissive layer of chemical conditions term the geochemically active layer [4]. Nodules have a very consistent density [4] and a relationship between the major horizontal axis and nodule weight (i.e. volume) has been long recognized [4,8,9].

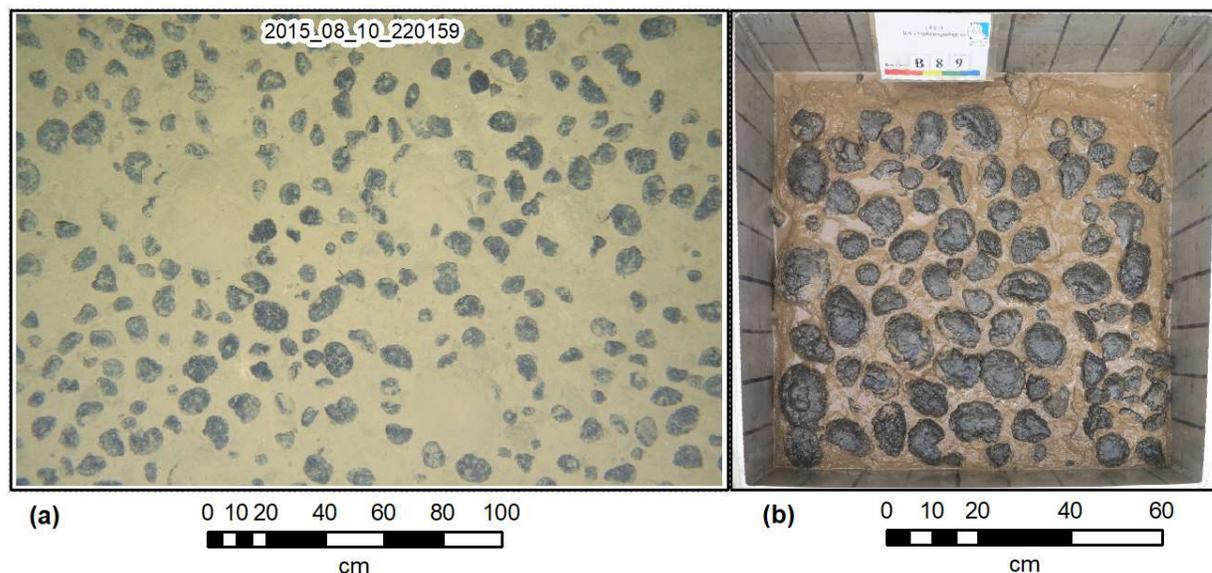


Figure 1: Example towed seabed photo (a) and box-core sample (b). Mounds of clay-ooze without nodules in the seabed photo are caused by bioturbation

Based on the above observation, to allow mathematical modelling of seabed polymetallic nodules, we have made the following three somewhat severe fundamental hypotheses:

1. Each nodules piece is of ellipsoidal shape (e.g., (a), (b) in Figure 2), which is defined by its three axes X_i , Y_i and Z_i , where $i = 1, 2, 3 \dots N$, with N being the number of nodules. Here X_i is the long or major axis, which is usually in the horizontal plane while Y_i and Z_i are the two typically shorter minor axes in the horizontal and vertical planes.

2. Within a certain boundary (domain) on the seabed, the ellipsoidal nodules are similar in shape, i.e., the ratio between two minor axes and the major axis, $\varepsilon_1 = \frac{Y_i}{X_i}$ and $\varepsilon_2 = \frac{Z_i}{X_i}$ are constant.
3. Within a certain boundary on the seabed, the long axis of nodule X_i follows a Generalized Rayleigh Distribution (GRD), which is defined by a pair of parameters α and β (See Section 3).

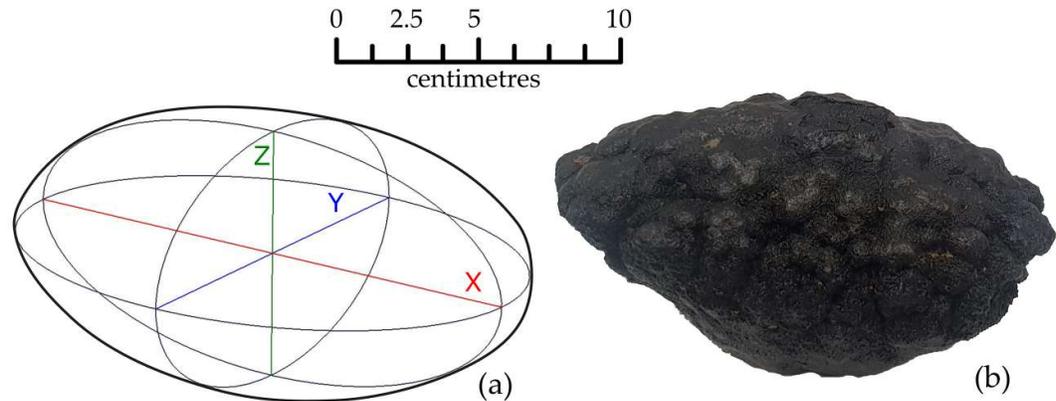


Figure 2: An idealized nodule of ellipsoidal shape (a) and an example from the TOML C1 area (b)

The above idealization is supported by analysis of nodule data and they were found accurate to certain degree. Specifically, the hypothesis 1 and 2 above will be justified using regression analysis of nodule dimensions and weights of seabed nodules samples in Section 5.1, which the hypothesis 3 will be validated by “Goodness-of-Fit” tests in Section 5.2 using nodule samples collected from TOML areas.

3. Generalized Rayleigh Distribution (GRD) and the traditional method

The Rayleigh distribution has been widely used to model phenomena in various technical fields. For instance, in the field of oceanography, Longuet-Higgins [10] showed the heights of narrow-banded random ocean waves follows the Rayleigh distribution. The *Generalized Rayleigh Distributions* (GRD), a family of two-parameter variations, have also been proposed although its practical application is limited. For a random variable X following the GRD, its probability density function (PDF) $f(x)$ is in the form:

$$f(x) = 2\alpha\beta^2 x e^{-(\beta x)^2} [1 - e^{-(\beta x)^2}]^{\alpha-1}, \quad x > 0, \quad (1)$$

where $\alpha > 0$ and $\beta > 0$ are shape and scale parameters, respectively.

The cumulative distribution function (CDF) $F(x)$ is given by

$$F(x) = [1 - e^{-(\beta x)^2}]^\alpha, \quad x > 0. \quad (2)$$

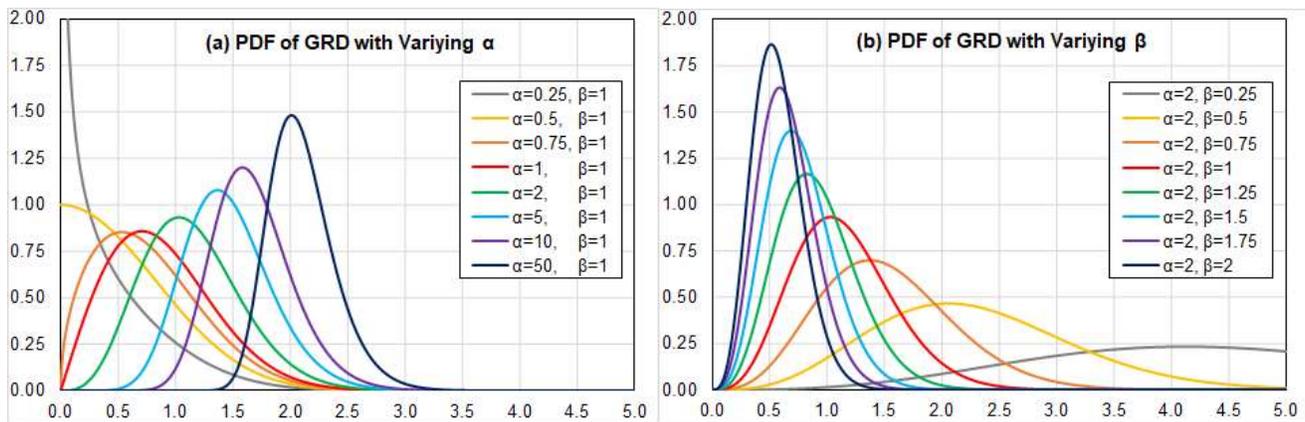


Figure 3: PDF of Generalized Rayleigh Distribution for Varying Parameters α (a) and β (b)

Figure 3 shows the PDF of Generalized Rayleigh Distribution for various values of parameters α and β . For the statistical analysis of seabed polymetallic nodules, the parameter range is $\alpha \geq 1$.

3.1 Mean and standard derivation of the Generalized Rayleigh Distribution

As derived in Appendix A, the mean μ of and the variance σ^2 of the Generalized Rayleigh Distribution can be written as

$$\begin{cases} \mu = \frac{\alpha}{\beta} F_1(\alpha) \\ \sigma = \mu \sqrt{G(\alpha)} \end{cases}, \quad (3)$$

where

$$\begin{cases} F_1(\alpha) = \int_0^{\infty} \sqrt{z} e^{-z} [1 - e^{-z}]^{\alpha-1} dz \\ F_2(\alpha) = \int_0^{\infty} z e^{-z} [1 - e^{-z}]^{\alpha-1} dz \\ G(\alpha) = \frac{F_2(\alpha)}{\alpha [F_1(\alpha)]^2} - 1 \end{cases}. \quad (4)$$

Formally, Eq.(3) can be used to estimate α and β when μ and σ are known. However, due to the complexity of functions $F_1(\alpha)$ and $F_2(\alpha)$ in Eq.(4), the solution process is rather tedious. Due to the complexity in evaluating $F_1(\alpha)$, $F_2(\alpha)$, an empirical method is developed below in Section 4 to simplify the calculation for practical applications.

3.2 Test of goodness of fit of Generalized Rayleigh Distribution

3.2.1 Parameter estimation by Maximum Likelihood Estimation (MLE)

For a random sample X_1, X_2, \dots, X_n of size n , to determine the two parameters α and β , which determine the Generalized Rayleigh Distribution (GRD), the Maximum Likelihood Estimation (MLE), which maximizes the log likelihood function, gives the following a pair of equations (Abd-Elfattah [11]):

$$\frac{\sum_{i=1}^n x_i^2 e^{-\beta^2 x_i^2} / (1 - e^{-\beta^2 x_i^2})}{\sum_{i=1}^n \ln(1 - e^{-\beta^2 x_i^2})} + \frac{\sum_{i=1}^n x_i^2 / (1 - e^{-\beta^2 x_i^2})}{n} = \frac{1}{\beta^2}, \quad (5)$$

$$\alpha = \frac{-n}{\sum_{i=1}^n \ln[1 - e^{-\beta^2 x_i^2}]}. \quad (6)$$

Eq.(5) is first solved *iteratively* by the Newton-Raphson Method to yield β , and Eq.(6) is then used to calculate α . It is obvious that the solution process for Eq.(6) is quite tedious. An empirical alternative is, therefore, devised in Section 4 below to simplify the process for practical application.

3.2.2 The Anderson–Darling Test Statistics

Once the parameters α and β are estimated as above, it is important to test whether they will yield a Generalized Rayleigh Distribution which give a “good-fit” for the sample. For computational purpose, the Anderson–Darling (AD) test statistics A_n^2 and V_n^2 can be written as:

$$\begin{cases} A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \{ \ln[z_i + \ln(1-z_{n+1-i})] \} \\ V_n^2 = \frac{n}{2} - 2 \sum_{i=1}^n z_i - \sum_{i=1}^n [2 - \frac{2i-1}{n}] \ln(1-z_i) \end{cases} \quad (7)$$

Here $z_i = F(x_i)$, where $F(x_i)$ is the empirical Probability Density Function (PDF), calculated using Eq.(2) above, and arranged into ascending order.

3.2.3 The Test Criteria for Hypothesis

The value of A_n^2 and V_n^2 calculate above are then compared with their corresponding critical values y_γ and u_γ respectively. If

$$\begin{cases} A_n^2 < y_\gamma \\ V_n^2 < u_\gamma \end{cases}, \quad (8)$$

the null hypothesis that sample data follow generalized Rayleigh distribution, is accepted at the particular significance level γ (or at $1 - \gamma$ confidence level). The critical values y_γ and u_γ are simulated by the Monte Carlo Method, as discussed in Appendix B.

4. A new empirical method and its application to nodule resources

As shown above in Section 3, the traditional formulation can be used for (1) analysing whether a numerical sample follows a Generalized Rayleigh Distribution, and (2) if it does, for estimating the parameters which yield the best fit. However, the complexity of the numerical process makes it difficult to be applied in practice. In this Section, we introduce a simplified numerical procedure for predicting the mineral resource of seabed nodules.

4.1 Empirical estimation of parameters of the Generalized Rayleigh Distribution

As shown in Section 3.2.1, Eq.(3) and Eq.(4) can formally be used to estimate α and β when μ and σ are known. To simplify the calculation of $F_1(\alpha)$ and $F_2(\alpha)$, using the technic of nonlinear regression, it can be shown, for the interested range of $1 \leq \alpha \leq 9$, the following empirical formulae:

$$\begin{cases} F_1(\alpha) = \frac{1.26364}{(\alpha + 0.51265)^{0.85786}} \\ F_2(\alpha) = \frac{1.73929}{(\alpha + 1.11917)^{0.73864}} \\ G(\alpha) = \frac{0.19249}{(\alpha - 0.41497)^{0.67379}} \end{cases}, \text{ for } 1 \leq \alpha \leq 9 \quad (9)$$

are accurate to an accuracy of 10^{-4} (See Figure 4).

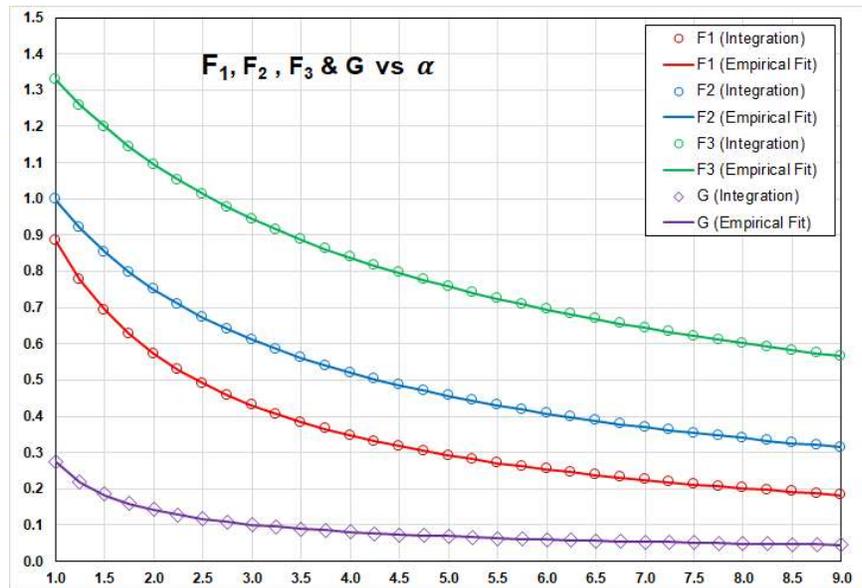


Figure 4: $F_1(\alpha)$, $F_2(\alpha)$ and $F_3(\alpha)$: Numerical Integration vs. Empirical Formulas

Also included in Figure 4 is $F_3(\alpha)$, which will be discussed further in Section 4.2 below. The empirical form of $F_3(\alpha)$ is

$$F_3(\alpha) = \frac{2.48274}{(\alpha + 1.71977)^{0.62304}} \quad (10)$$

Combining Eq.(3) and Eq.(9), it is straightforward to derive:

$$\begin{cases} \alpha = 0.41497 + 0.08669 \left(\frac{\mu}{\sigma}\right)^{2.96828} \\ \beta = \frac{1.26364 \alpha}{(\alpha + 0.51265)^{0.85786}} \left(\frac{1}{\mu}\right) \end{cases} \quad (11)$$

For a random sample X_1, X_2, \dots, X_n of size n , the mean μ and standard deviation σ are first estimated from the sample, and then the two formulae in Eq.(11) can be used in sequence to estimate parameters α and β .

It is evident from the first equation of Eq.(11) that shape parameter α is determined by the ratio of μ and σ (in effect the reciprocal of the coefficient of variation or signal-to-noise-ratio). The scale parameter β , on the other hand, is related to the mean μ . This is consistent with the observation in Figure 3. It can be posited that geologically these two parameters may in turn relate to the stability and thickness of the geochemically active layer in which the nodules grow, as described in [4].

4.2 Resource Estimation for Seabed Polymetallic Nodules Using Coverage and Abundance

To estimate the percentage coverage and abundance of seabed polymetallic nodules using the measurements of their long axes, the first two hypothesis, as in Section 2 above, are applied:

1. The nodule is assumed to be in an idealized ellipsoid shape.
2. Nodules within a certain boundary are assumed to be "similar" in shape, the ratios between the lengths of the two minor axes and the major axis (denoted by ε_1 and ε_2) are constant.

4.2.1 Perdition of Nodule Coverage: Idealized Nodules

Assuming X_1, X_2, \dots, X_N are samples of long axes of ellipsoid nodules, with N being the number of nodules in the photo, the total area S_c being covered by nodules can be calculated by adding the projections of nodules in the photo image:

$$S_c = \sum_{i=1}^N S_i = \sum_{i=1}^N \frac{\pi}{4} X_i Y_i \quad . \quad (12)$$

Assuming the projections of nodules in the photo are *elliptical* with $Y_i = \varepsilon_1 X_i$ being the short axes (constant $\varepsilon_1 \leq 1$), Eq.(12) can be re-arranged as:

$$S_c = \sum_{i=1}^N \frac{\pi}{4} X_i Y_i = \frac{\pi}{4} \sum_{i=1}^N X_i (\varepsilon_1 X_i) = \frac{\pi \varepsilon_1}{4} \sum_{i=1}^N X_i^2 \quad . \quad (13)$$

Using (35) in Appendix A and taking $m = 2$, Eq.(13) becomes:

$$S_c = \frac{\pi \varepsilon_1}{4} \sum_{i=1}^N X_i^2 = \frac{\pi \varepsilon_1}{4} N \overline{X^2} = \frac{\pi \varepsilon_1}{4} N \frac{\alpha}{\beta^2} F_2(\alpha) \quad . \quad (14)$$

Assuming S_p is the area of the photo, the nodule percentage coverage C_N becomes:

$$C_N = \frac{S_c}{S_p} = \frac{\pi \varepsilon_1 N \alpha}{4 S_p \beta^2} F_2(\alpha) = \frac{\varepsilon_1 N}{S_p \beta^2} \frac{1.36603 \alpha}{(\alpha + 1.11917)^{0.73864}} \quad . \quad (15)$$

4.2.2 Prediction Nodule Abundance I: Based on "Idealized Nodule" Model

Similarly, the total Weight W_a of nodules in the photo can be calculated by adding the weights of nodules in the photo:

$$W_a = \sum_{i=1}^N W_i = \sum_{i=1}^N \frac{\pi}{6} \rho X_i Y_i Z_i \quad . \quad (16)$$

Assuming the two minor axes Y_i and Z_i are related to the major axis X_i by $Y_i = \varepsilon_1 X_i$ and $Z_i = \varepsilon_2 X_i$, Eq.(16) can be re-arranged as:

$$W_a = \sum_{i=1}^N \frac{\pi}{6} \rho X_i Y_i Z_i = \frac{\pi}{6} \rho \sum_{i=1}^N X_i (\varepsilon_1 X_i) (\varepsilon_2 X_i) = \frac{\pi \rho \varepsilon_1 \varepsilon_2}{6} \sum_{i=1}^N X_i^3 \quad . \quad (17)$$

Using (35) in Appendix A and taking $m = 3$, Eq.(17) becomes:

$$W_a = \frac{\pi \rho \varepsilon_1 \varepsilon_2}{6} \sum_{i=1}^N X_i^3 = \frac{\pi \rho \varepsilon_1 \varepsilon_2}{6} N \overline{X^3} = \frac{\pi \rho \varepsilon_1 \varepsilon_2}{6} N \frac{\alpha}{\beta^3} F_3(\alpha) \quad . \quad (18)$$

Assuming S_p is the area of the photo, the nodule abundance A_N becomes:

$$A_N = \frac{W_a}{S_p} = \frac{\pi \rho \varepsilon_1 \varepsilon_2 N \alpha}{6 S_p \beta^3} F_3(\alpha) = \frac{\rho \varepsilon_1 \varepsilon_2 N}{S_p \beta^3} \frac{1.29996 \alpha}{(\alpha + 1.71977)^{0.62304}} \quad . \quad (19)$$

4.2.3 Prediction Nodule Abundance II: With Empirical Long-Axis-Weight Relationship

According to [9], [4] and Eq. (3) (also see Section 5.1 below), the weight of the nodule is correlated to its long axis by:

$$\log_{10} W_i = k \log_{10} X_i + b \quad , \quad (20)$$

where k and b are constants and k is usually smaller and close to 3.0. Eq.(20) can be re-arranged as

$$W_i = 10^b X_i^k \quad . \quad (21)$$

Then, the total weight W_a of nodules in the photo can be calculated by adding the weights of nodules in the photo:

$$W_a = \sum_{i=1}^N W_i = \sum_{i=1}^N 10^b X_i^k = 10^b \sum_{i=1}^N X_i^k \quad . \quad (22)$$

Using (35) in Appendix A and taking $m = k$, Eq.(20) becomes:

$$W_a = 10^b \sum_{i=1}^n X_i^k = 10^b N \bar{X}^k = 10^b N \frac{\alpha}{\beta^k} F_k(\alpha) \quad . \quad (23)$$

$F_k(\alpha)$, with $2 < k < 3$, can be approximated by interpolating $F_2(\alpha)$ and $F_3(\alpha)$:

$$F_k(\alpha) = [F_3(\alpha) - F_2(\alpha)] k + [3F_2(\alpha) - 2F_3(\alpha)] \quad . \quad (24)$$

Assuming S_p is the area of the photo, the nodule abundance A_N becomes:

$$A_N = \frac{W_a}{S_p} = \frac{10^b N \alpha}{S_p \beta^k} F_k(\alpha) = \frac{10^b N \alpha}{S_p \beta^k} \{ [F_3(\alpha) - F_2(\alpha)] k + [3F_2(\alpha) - 2F_3(\alpha)] \} \quad . \quad (25)$$

4.2.4 Relation between Nodule Coverage and Abundance

The abundance A_N are related to the coverage C_N by dividing Eq.(19) by Eq.(15):

$$\frac{A_N}{C_N} = \frac{2 \rho \varepsilon_2 F_3(\alpha)}{3 \beta F_2(\alpha)} \quad . \quad (26)$$

1. Combining with the first equation of Eq.(3), Eq.(26) above becomes

$$\frac{A_N}{C_N} = \frac{2}{3} \rho \varepsilon_2 \mu \frac{F_3(\alpha)}{\alpha F_1(\alpha) F_2(\alpha)} \quad . \quad (27)$$

Using the using the technique of nonlinear regression, Eq.(21) can be rewritten into the form:

$$\frac{A_N}{C_N} = \rho \varepsilon_2 \mu \frac{0.70834(\alpha + 0.41373)^{0.99261}}{\alpha} \quad . \quad (28)$$

For the range of $1 < \alpha < 4$, of interest to nodules, Eq.(28) can be further approximated, within 0.2% error, by

$$\frac{A_N}{C_N} \approx \rho \varepsilon_2 \mu \left[1 - 0.3 \left(1 - \frac{1}{\alpha} \right) \right] \quad . \quad (29)$$

The above formulation, based on the "Idealized-Nodule" model, provides two efficient ways to estimate the nodule resources:

1. Eq.(15) and Eq.(19) can be used independently to calculate the nodule percentage coverage C_N and the abundance A_N , respectively.
2. If an estimation of the C_N is already computed using digitization technique from seabed imagery, then Eq.(28) or Eq.(29) can be used to compute A_N .

5. Test of hypotheses of the Idealized Nodule Model

The three fundamental hypotheses that define the "idealized nodule" in Section 2 find considerable support from numerical measurements of samples of seabed nodule.

5.1 Test of hypothesis 1 and 2: linear regression analyses on nodule dimensions and weights

In order to examine the validity of the first two hypotheses for the "idealized nodules", linear regression analyses were carried out on published nodule major and minor axes, volume and weight data from two sites in the central and eastern CCZ [7] (Figure 9). Specifically, the following cases of linear regressions were carried between:

1. Case 1: Nodule long axis and its horizontal minor axis;
2. Case 2: Nodule long axis and its vertical minor axis;
3. Case 3: Nodule weight and its volume; and

4. Case 4: Nodule long axis and its weight.

The numerical results are shown in Figure 6 and are summarized in Table 1. The linear regression was performed for 1-3 with intercept forced be zero to match the fact that the two variables vanish at the origin.

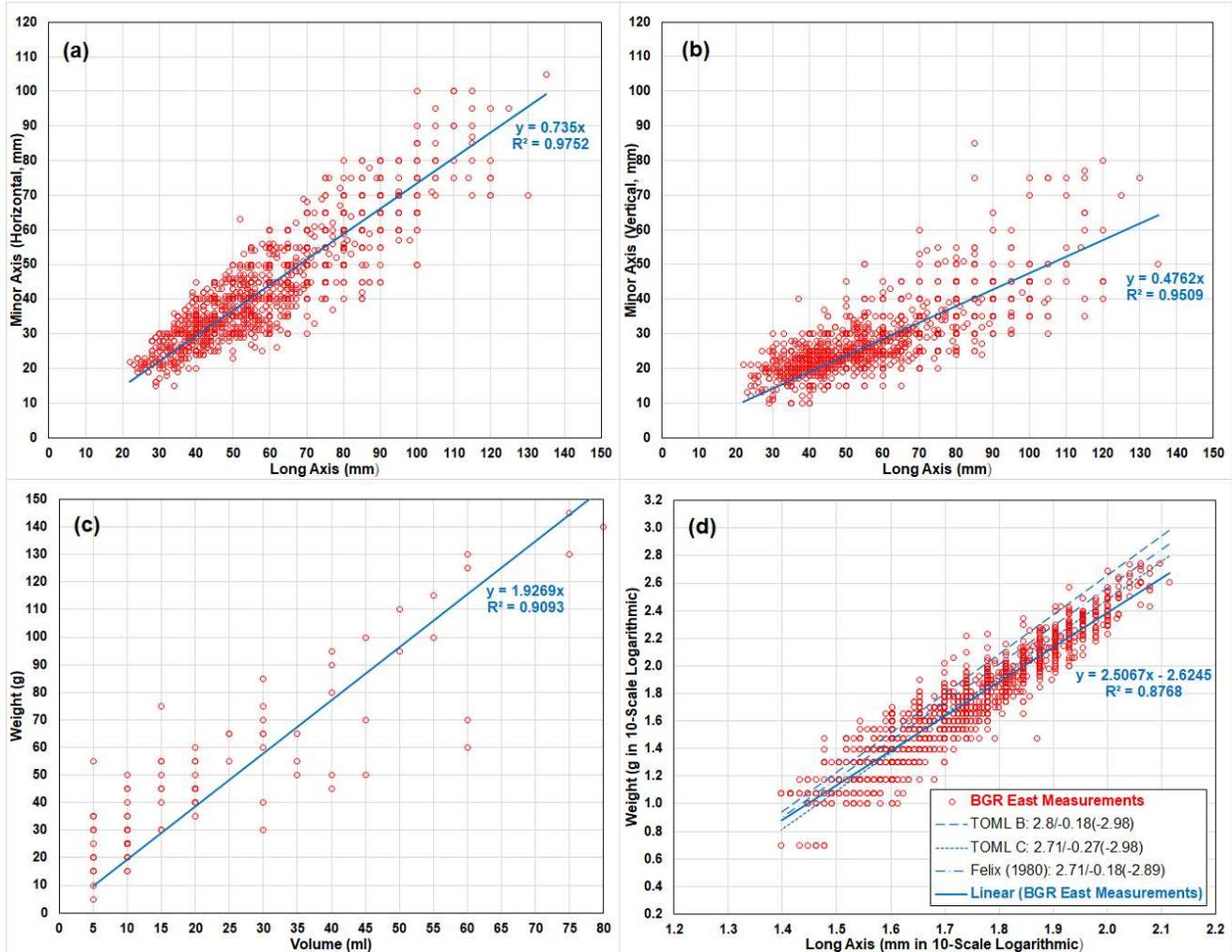


Figure 5: Linear Regression between Nodule Dimensions and Weights for BGR East Region. Data source [7]. For location refer also Figure 9. Note that intercepts of Felix [9] (3) and TOML [4];1,2) based long-axis formulae are converted from cm to mm (in logarithmic scale).

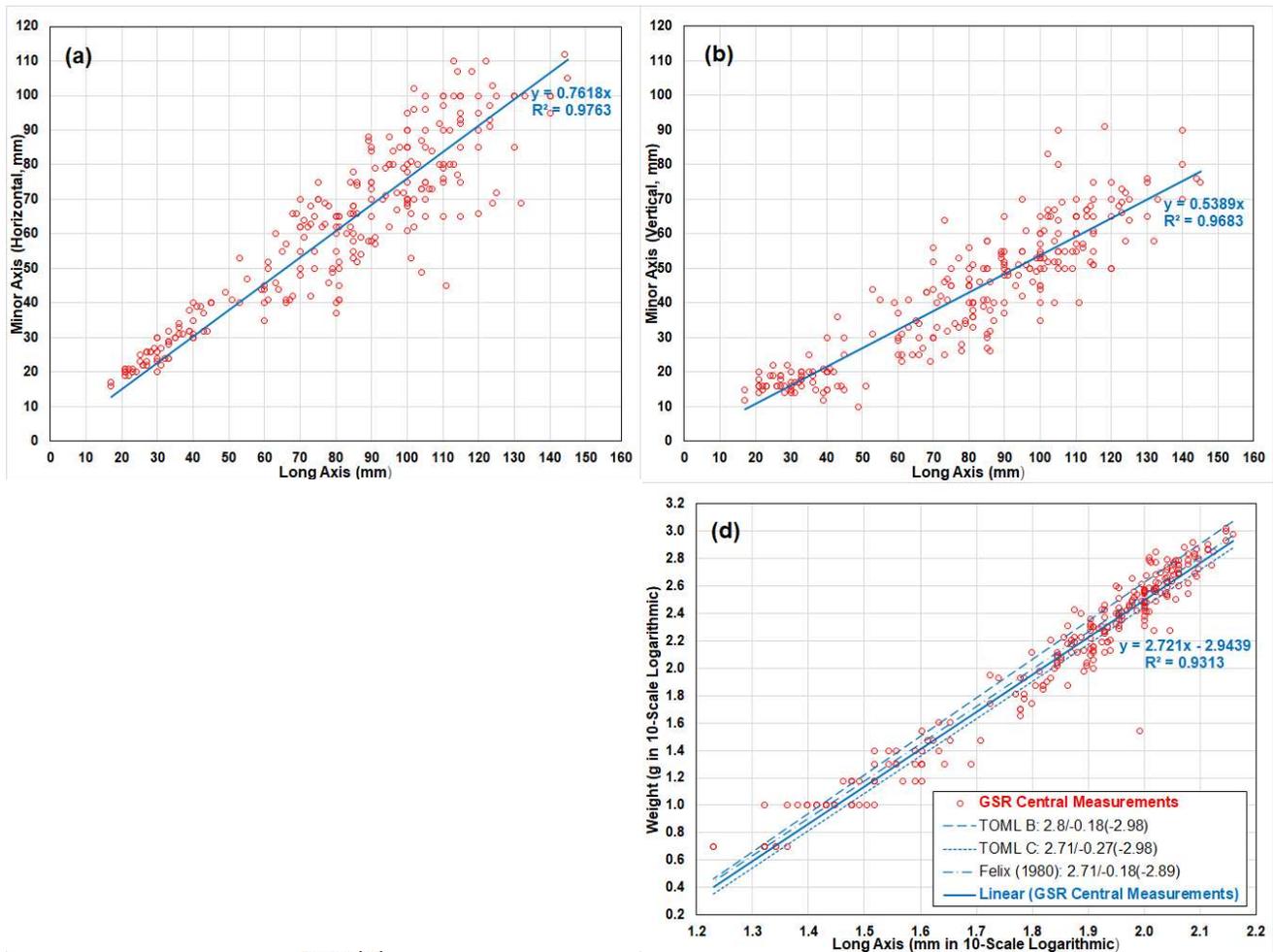


Figure 6: Linear Regression between Nodule Dimensions and Weights for GSR Central Region. Data source [7]. For location refer also Figure 9. Note that intercepts of Felix [9] (3) and TOML [4];1,2) based long-axis formulae are converted from cm to mm (in logarithmic scale).

Table 1: Results of Linear Regression for Nodule Dimensions and Weights

Case	Regression Parameters	Regions	Sample Size	Estimated Slope	Estimated Intercept	Correlation (R ²)	
1	Minor Axis Y (Horizontal, mm)	Long Axis X (mm)	BGR East	1376	0.7350	0 (Forced)	97.52%
			GSR Central	259	0.7618		97.63%
2	Minor Axis Z (Vertical, mm)	Long Axis X (mm)	BGR East	1376	0.4762	0 (Forced)	95.09%
			GSR Central	259	0.5389		96.83%
3	Weight (g)	Volume (cm ³)	BGR East	99	1.9269	0 (Forced)	90.93%
			GSR Central	No Data	/		/
4	Weight (Logarithmic, g)	Long Axis X (Logarithmic, mm)	BGR East	1376	2.5067	-2.6245	87.68%
			GSR Central	259	2.7210		-2.9439

The results from Cases 1 and 2 show the two minor axes are correlated to the long axis in statistically significant ways. With $R^2 > 95\%$, a great majority (>95%) of the data points supports the hypothesis that, with a certain boundary, the ratios between the length of the major axes X and the lengths of the horizontal and the vertical minor axes Y and Z can be considered as constants. It is also noted the ratios vary between regions. The result from the 3rd Case indicates a nodule density of 1.93 g/cm³ although it is only supported by

a small sample. The Case 4 indicates the nodule weight is strongly correlated with the 2.5056 and 2.7210 power of the long axes (versus 3 for the original “idealized nodule” model), supported by about 88% and 93% of the data points from the two regions, respectively.

While the above results give broad, yet strong support to the first two hypotheses for “idealized nodules” in Section 2 (these being on ellipsoid form and similarity of form within a domain), we use additional analysis in Section 2 to validate the 3rd hypothesis regarding a GRD distribution.

5.2 Test of hypotheses 3: Goodness-of-Fit test of Generalized Rayleigh Distribution for Nodule Long Axes

In this Section, the traditional “Goodness-of-Fit” tests, as outlined in Section 3, are carried out to check the validity of the 3rd hypothesis for the “idealized nodules”. It is to check whether the long axes of seabed nodules follow the Generalized Rayleigh Distribution (GRD) in a statistically significant way. A total of 9 samples of nodule long axes were analysed. While Table 2 shows the key statistical properties of the 9 data sets, Table 3 presents the results of “Goodness-of-Fit” tests.

Table 2: Summary of 9 Sets of Nodule Samples used for Goodness-of-Fit Test

No	Sample ID	TOML Area	Type	Sample Size	Mean	Standard Deviation
1	2015_08_10_172643	B	Towed Photo	336	3.091	2.512
2	2015_08_10_220159	B	Towed Photo	153	7.767	2.387
3	2015_08_11_121357	B	Towed Photo	403	5.978	1.732
4	2015_08_29_131349	C	Towed Photo	440	5.425	1.995
5	2015_09_02_185307	C	Towed Photo	113	3.827	1.270
6	CCZ15-B51	D	Washed Sample	67	7.486	2.404
7	CCZ15-B102	F	Washed Sample	278	4.318	1.705
8	CCZ15-B106	F	Washed Sample	559	3.681	1.298
9	CCZ15-B110	F	Washed Sample	135	6.910	2.602

Table 3: Results of Goodness-of-Fit Test for 9 Sets of Samples of Nodule Long Axes

No	Sample ID	α	β	A_n^2	γ_γ	V_n^2	u_γ	Conclusion
1	2015_08_10_172643	0.623	0.211	9.957	0.833	5.436	0.43	Not Generalized Rayleigh Dist. at 5% Level of Significance
2	2015_08_10_220159	2.714	0.162	0.66	0.784	0.257	0.414	
3	2015_08_11_121357	3.598	0.226	0.69	0.770	0.226	0.410	
4	2015_08_29_131349	1.965	0.210	0.541	0.777	0.305	0.408	
5	2015_09_02_185307	2.690	0.327	0.159	0.789	0.076	0.419	
6	CCZ15-B51	1.701	0.144	0.376	0.804	0.200	0.423	Generalized Rayleigh Dist. at 5% Level of Significance
7	CCZ15-B102	1.396	0.243	1.435	0.790	0.605	0.412	
8	CCZ15-B106	2.410	0.321	0.738	0.778	0.314	0.410	
9	CCZ15-B110	1.890	0.171	0.361	0.791	0.192	0.418	

For each set of data in Table 3, α and β are solved iteratively by Eqs.(5) and (6), using the MLE method described in Section 3.2.1. The Anderson–Darling (AD) test statistics A_n^2 and V_n^2 are then calculate by Eq.(7) and they are in turn compared with their critical values γ_γ and u_γ , which are computed by the Monte Carlo Simulation in Appendix B. According to the test criteria in Eq.(8), among the 9 sets of samples, 7 of them have passed the AD tests at 95% confidence level. It indicates a strong probability that the samples of

long axes of polymetallic nodules do follow Generalized Rayleigh Distribution. This conclusion may be conditional (e.g. by geological domain), and more research may be needed to identify the conditions.

Figure 7 and Figure 8 show the visual comparison of distribution of nodule long axes from raw data, computed by the traditional method, and by the new empirical method. Probability density functions (PDF) and cumulative distribution functions (CDF) are plotted in Figs. 7 and 8, respectively, using Sample ID: 2015_08_29_131349 as an example. For the raw data, a bin size of 0.25 was selected to create a histogram of the sample of long axis, and the PDF and the CDF are computed. The dark blue line shows the raw counts of the original data set, with the light blue line showing the smoothed data using the Savitzky-Golay filter [12] for the PDF in Figure 7. The green line shows PDF and CDF based on parameters α and β calculated iteratively by MLE method described in Section 3.2.1 above. The red line shows PDF and CDF based on parameters α and β calculated by the empirical formulas in Eq.(11). The empirical formulas, while much more straightforward to use, do give reasonably accurate results for the statistical distributions in practical applications.

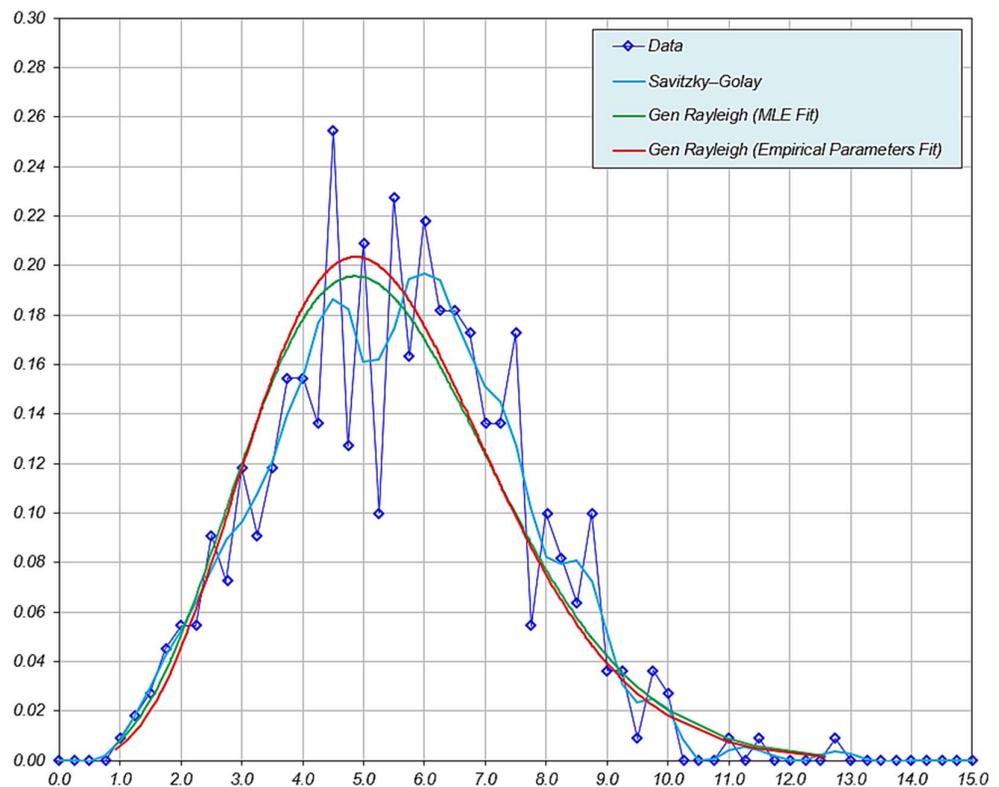


Figure 7: Probability Density Function (PDF) of the example towed seabed photograph (2015_08_29_131349; Table 2, Table 3)

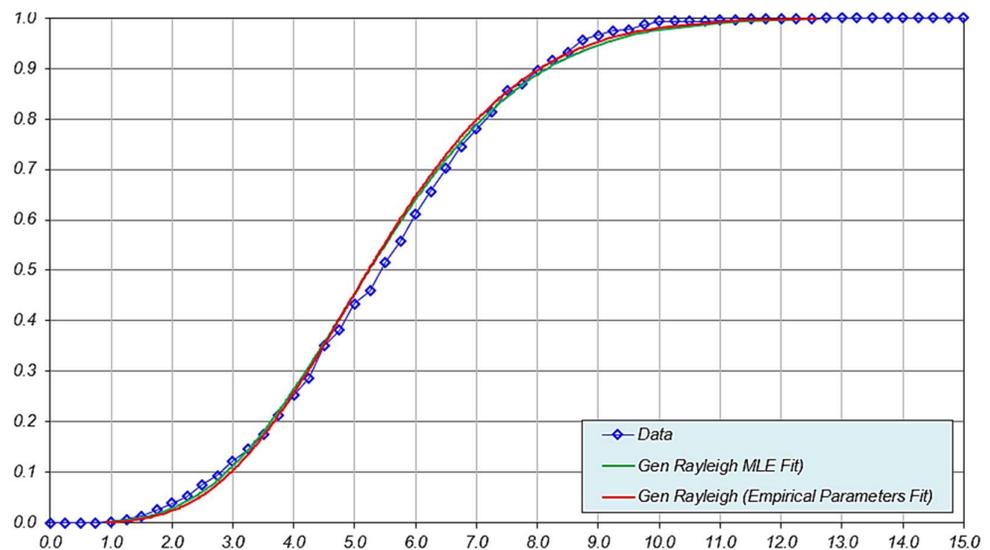


Figure 8: Cumulative Distribution Function (CDF) of the of the example towed seabed photograph (2015_08_29_131349)

5.3 Comments on the level of support

Significantly, the Linear Regression Analysis in Section 5.1 and the Goodness-of-Fit Test in Section 5.2 do support the three seemingly drastic fundamental hypotheses made in Section 2 for “idealized nodules”. However, more analyses are needed to check the generality. Nonetheless, as the hypotheses have been validated, the empirical method developed in Section 4 can be used for nodule resource prediction in the next Section.

Possible reasons for achieving the statistically significant validations, include that the samples are located within a particular growth domain (the CCZ), and that the conditions of growth within this domain are remarkably consistent (nodules grow in effect from the ocean’s epibenthos, and slowly enough to “average out” short term (millennia scale) variances in growth conditions).

5. Numerical results of nodule resource prediction using the new empirical method

Once newly proposed “idealized nodules” model has been validated in the previous Section, the empirical formulation as developed in Section 3, particularly the $Eq.(19)$ (or $Eq.(25)$) for nodule abundance A_N are applied to a total of 188 samples of seabed polymetallic nodule collected in 2015 as part of the TOML CCZ15 marine expedition in this Section.

6.1 Sample datasets

The 188 samples of seabed polymetallic nodules used for the empirical analyses can be grouped in three datasets:

1. Dataset 1: regional scale box-core sample dataset (physical weights). This involves 4 TOML exploration contract areas (TOML B, C, D, F; Figure 9) spanning some 2,000 km of longitude and 700 km of latitude. The dataset thus allows for examination of a general relationship.
2. Dataset 2: local scale box-core sample dataset (physical weight) of two distinct facies types but only within the TOML F area (~200 x 200 km). Type 1 nodules are smaller and often densely packed, type 2 nodules are significantly larger and more variable (cf [5]). The dataset thus allows for differences in nodule types from an area where the distinction between type is simple and straightforward.
3. Dataset 3: two local scale towed photo sample datasets (long-axis abundance estimate) between the TOML B and C areas (~300 km apart). The dataset is limited in

that actual nodule weights cannot be compared, but it allows for larger datasets from two distinctly different areas to be compared.

Coverage was also measured for datasets 2 and 3 from seabed photographs (boxcore mounted and towed respectively) using Image J software. Dataset 1 was not able to be measured due to a lack of images (the box-core camera had frequently malfunctioned).

A summary of the datasets used for analysis is shown in Table.4 and Figure 9.

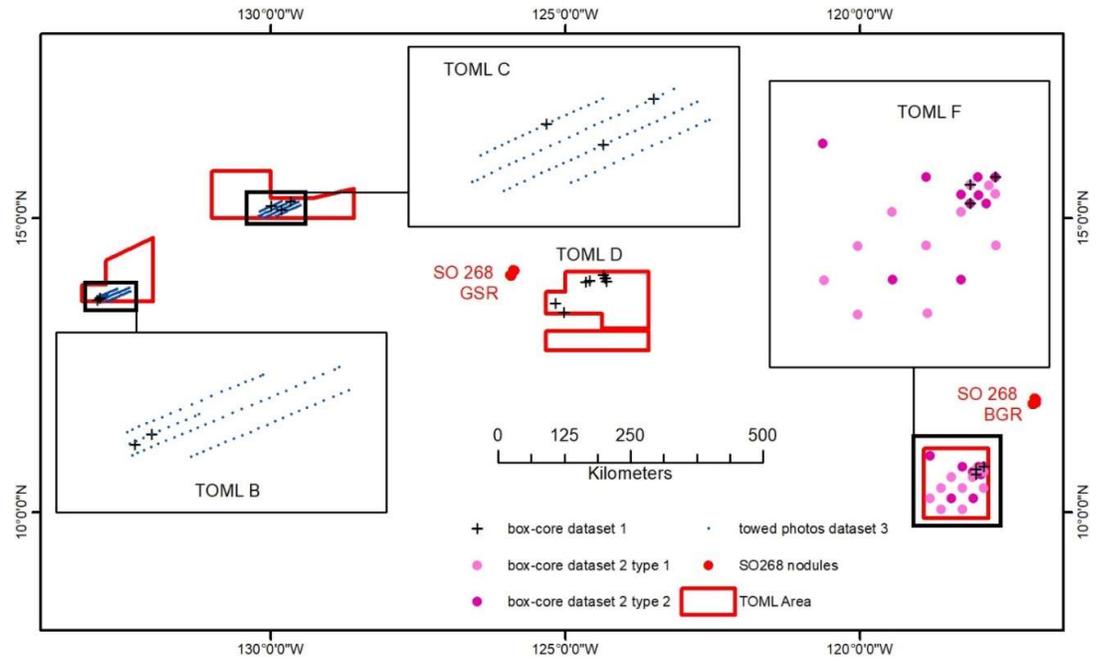


Figure 9: Sample datasets locations

Table 4: Sample set summary

Data-set	TOML Areas	Number of samples	Comparative data type	Range of abundances	Range of sample mean long axes*	Range of CoV of sample long axes*
1	B, C, D, F	2, 3, 7, 3	Washed sample weights	3.2 to 25.7 kg/m ²	2.2 to 7.6 cm	0.23 to 0.86
2	F	11 for Type 1 9 for Type 2	Washed sample weights	1.2 to 21.3 3.3 to 29.1	2.2 to 3.9 2.6 to 9.2	0.28 to 0.45 0.28 to 0.72
3	B C	68 85	Long axis estimates on individual nodule images	0.03 to 31 0.01 to 18	1.6 to 7.8 1.5 to 6.1	0.24 to 0.96 0.25 to 0.83

* For datasets 1 and 2 long axes measured from grid photos of the nodules after collection, separation from the host clay-ooze and washing. For dataset 3 long axes measured from photos of the seabed as detailed in [4].

6.2 Prediction of Abundance of Seabed Nodules

To make a thorough assessment of the accuracy of abundance prediction made by the new empirical method, particular Eq.19 (or Eq.25), Figure 10, 11 and 12 show the ratios between the abundance prediction and those from the actual box-cores measurements (for datasets 1 and 2) or the best available estimate from long-axis measurements (for dataset 3). In the Figures, ratios are plotted against the mean long axes of each sample. For each dataset there are three charts are presented:

1. Chart (a) showing ratios based on abundance calculated directly by the empirical formula Eq. (19), which is strictly based on the three hypotheses for Idealized Nodule model in Section 2. The axis ratios ε_1 and ε_2 used in the formula are extracted from the analyses of BGR Eat and GSR Central data (Table 1 in Section 5.1);

2. Chart (b) showing the ratios in Chart (a) corrected by a “linear adjustment”. As the results in Chart (a) display biases of slightly over-estimating abundance for larger nodules. The result of linear regression of the ratios in Chart (a) are used to adjust the ratios, and the corrected results are shown in Chart (b); and
3. Chart (c) showing ratios based on abundance calculated by the empirical formula Eq. (19), which has incorporated the long-axis-weight relationship observed by several researchers (e.g., Felix [9] (3)).

Figure 10 shows the results for Dataset 1, which is collected across the TOML areas. It is observed that empirical formula Eq. (19), which is strictly based on the three hypotheses for Idealized Nodule model in Section 2, gives a slight bias of over-predicting the abundance for larger nodules. Unbiased prediction can be achieved once either the “linear adjustment” or the long-axis-weight relationships are applied.

Figure 11 depicting Dataset 2 shows that nodule types can have an influence in the prediction. While empirical formula Eq. (19) results in a similar bias to that seen for Dataset 1, facies specific “linear adjustment” results in unbiased estimates with slightly higher levels of scatter. Use of empirical formula Eq. (25) based on the coefficients of [9] works better for Type 1 nodules than for Type 2, suggesting that different coefficients may work better.

Figure 12 with Dataset 3 shows that the accuracy of prediction made by empirical formula Eq. (19) may vary for nodules between different areas. In effect the nodules from TOML B1 appear to be “less ideal” per Section 2 than the nodules from TOML C1. This may be due to the fact that the TOML B1 nodules are likely older and more often formed from multiple generations of growth (i.e. fragments of nodules with younger concentric growth phases). This could predispose them to be more equant in shape. Again, the linear adjustment addresses the size-bias seen in the direct application of Eq.(19). Application of Eq.(25) gives broadly similarly agreeable results, but it is worth noting that the compared estimate for this dataset is from long-axis weight measurements.

The slight biases of over-estimation of abundance for larger nodules reveals a limitation of the empirical formula Eq. (19), which is directly based on the three fundamental hypotheses for the “idealized nodules”. While the first two hypothesis state that the nodules are in ellipsoidal shape and they are “similar” in shape, in reality, it is obvious that the nodule shapes are complex and for nodules of various sizes, the ratios between the minor axes and the long one may vary with nodule size. However, this bias seems much less severe while empirical formula Eq. (25) is applied, which is based on an empirical relationship between the nodule long axis and its weight (e.g., Felix [9] (3)).

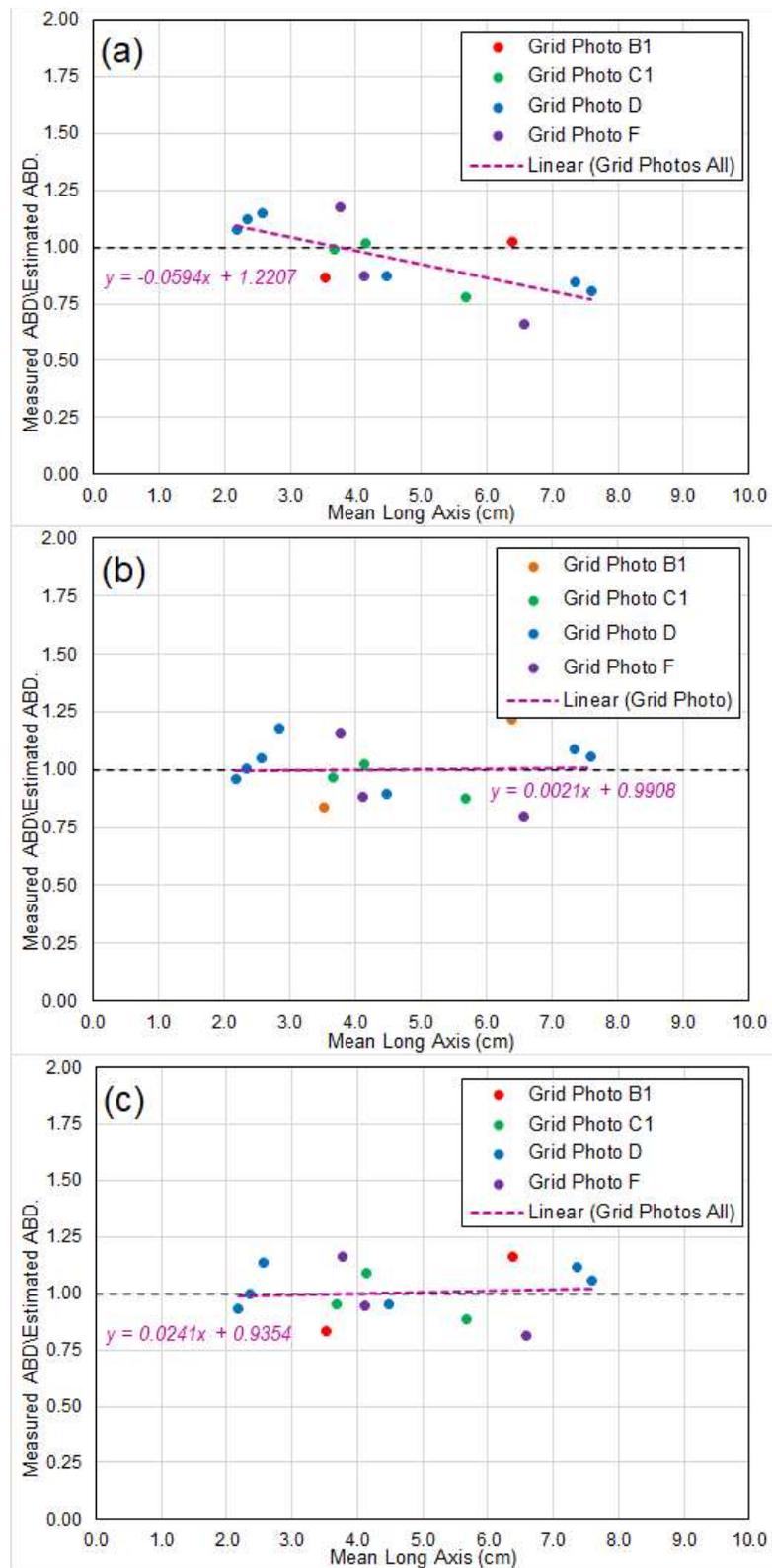


Figure 10: Comparisons of the measured Abundances and the predicted ones by the new empirical method with Dataset 1: (a) by empirical formula Eq.(19); (b) by empirical formula Eq.(19) with a linear adjustment; (c) by empirical formula Eq.(25).

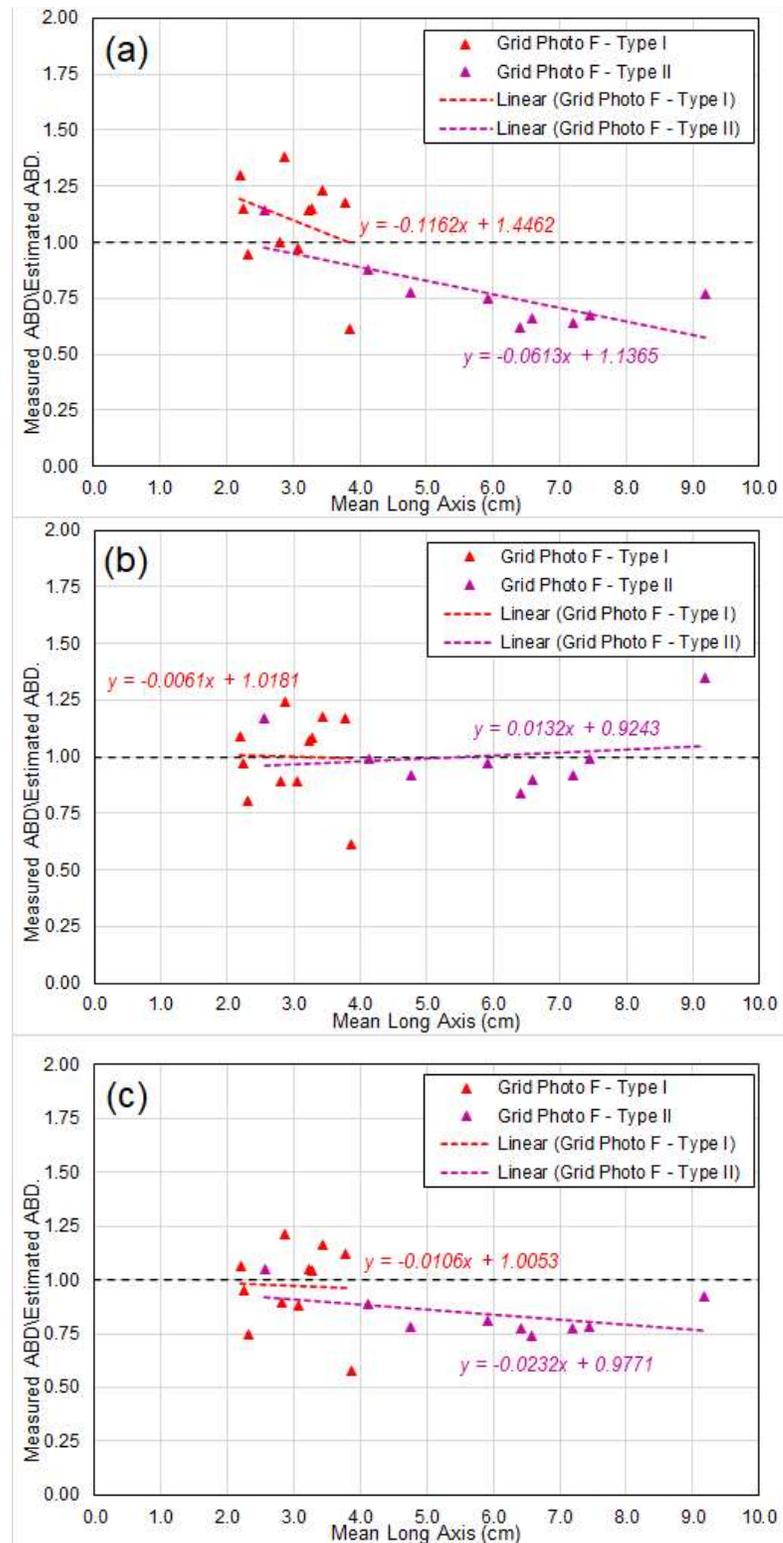


Figure 11: Comparisons of the measured Abundances and the predicted ones by the new empirical method with Dataset 2: (a) by empirical formula Eq.(19); (b) by empirical formula Eq.(19) with a linear adjustment; (c) by empirical formula Eq.(25).

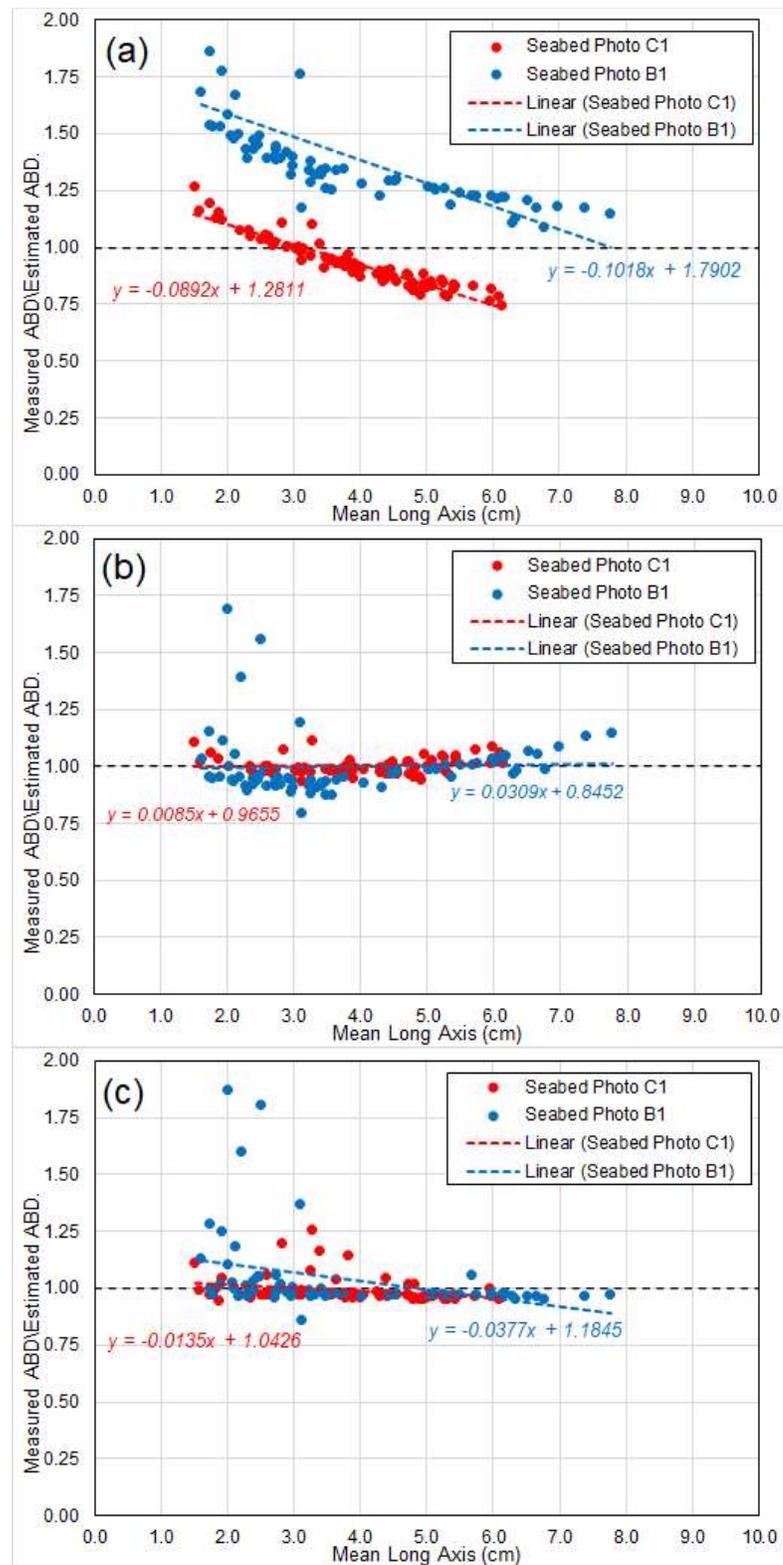


Figure 12: Comparisons of the measured Abundances and the predicted ones by the new empirical method with Dataset 3: (a) by empirical formula Eq.(19); (b) by empirical formula Eq.(19) with a linear adjustment; (c) by empirical formula Eq.(25).

Estimates of coverage using the new empirical method show mixed but encouraging results when compared with field measurements from three areas. Dataset 2 from TOML F shows a systematic bias independent of nodule facies types. In contrast, Dataset 3 from TOML B and C do not display any appreciable bias. This is likely related to the degree of sediment cover between the areas (Figure 14).

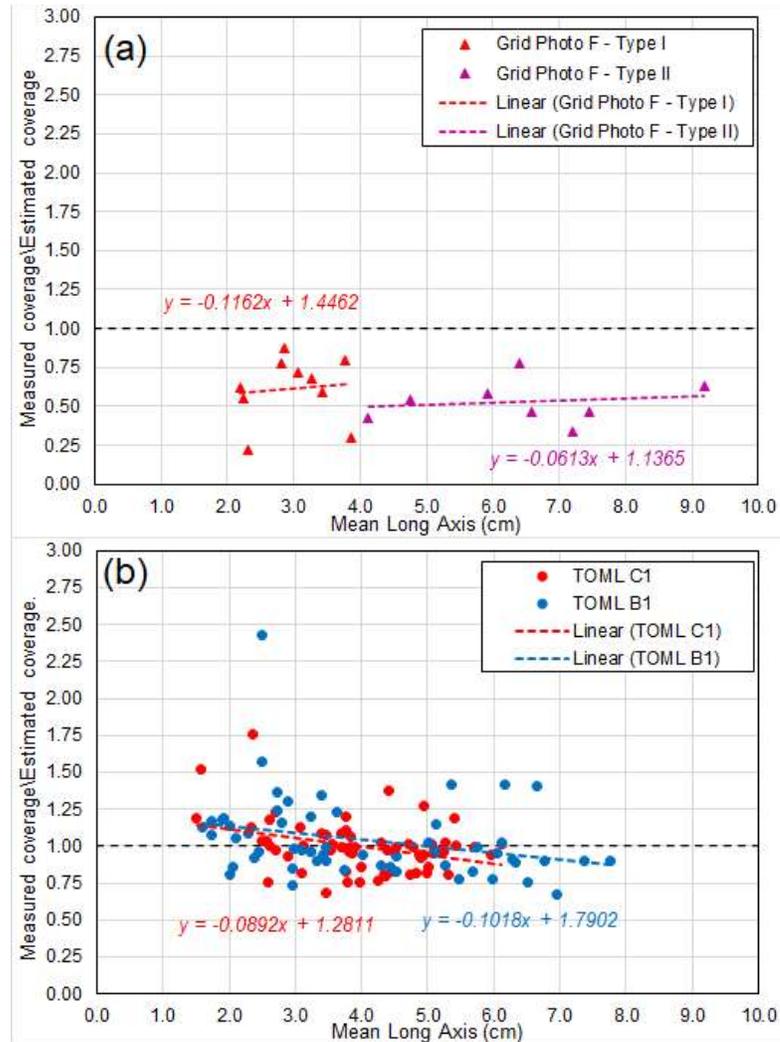


Figure 13: Comparisons of the measured Percentage Coverage and the predicted ones by the new empirical method for Datasets 2 in (a) and Dataset 3 in (b) with predictions by empirical formula Eq.(15).

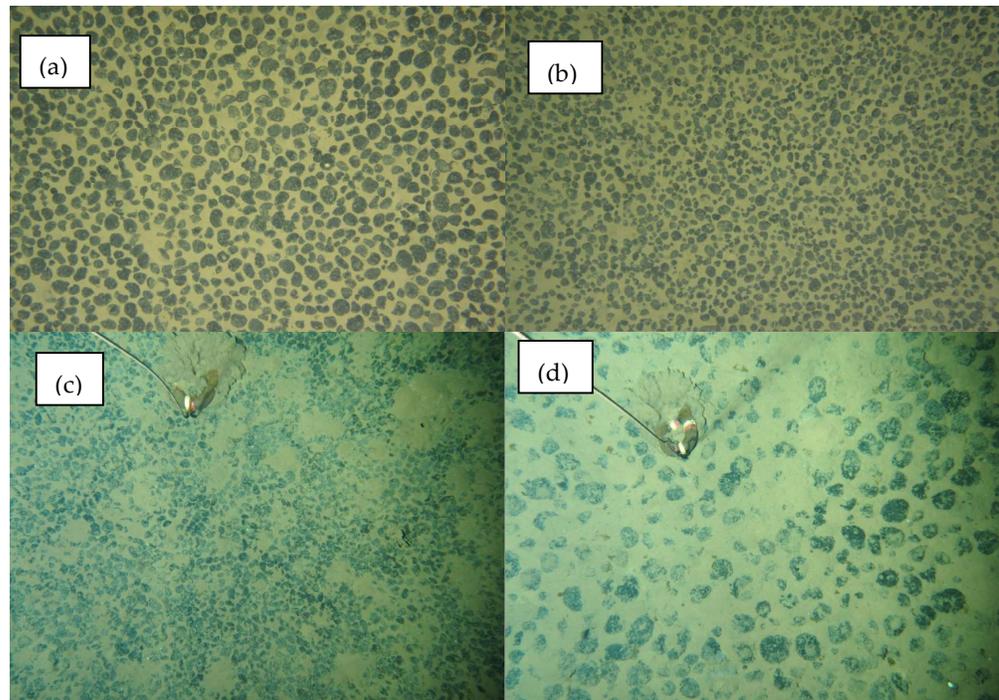


Figure 14: Seabed photos from TOML B, C and F. (a) TOML B, CCZ15-F02: 2015_08_17_032745 (b) TOML C, CCZ15-F05: 2015_08_29_071000 (c) TOML F type 1 nodules CCZ15-B105 (d) TOML F type 2 nodules CCZ15- B99. Images in (a) and (b) are 2.4 x 1.6 m in area. Trigger weight in images (c) and (d) is 28 x 16 cm long.

7. Conclusions

It is concluded that:

1. An idealized module for the dimensions of seabed polymetallic nodules is proposed for mineral resource prediction. This “idealized nodule” model is based on three hypotheses. One of the key hypotheses is, with in certain boundary, the nodule long axes follow a two-parameter Generalized Rayleigh Distribution (GRD). These three hypotheses were tested using field measurements from available nodule samples collected from CCZ. Strong numerical evidence was found, providing validation to the three hypotheses, possibly due to the relatively stable seabed environment and the long growth period of the nodules, removing short-term transient effects.
2. Once the “idealized nodule” model has been validated, a new empirical method has been developed based on the three hypotheses. Specifically, explicit empirical formulae have been derived for direct calculation of GRD parameter α and β (Eq.(11)), for Percentage Coverage C_N (Eq.(15)), and for Abundance A_N (Eq.(19) or Eq.(25)). These formulas are found accurate for mineral resource estimation and are much easier to use than the traditional analytical methods for GRD.
3. To check the effectiveness of the new empirical method for predicting nodule mineral resource, the predicted nodule Abundances are compared with field measurements for a total of 188 nodule samples collected across the TOML areas. While the general agreement is acceptable, the direct application of the “idealized nodule” module does display a slight bias of over-estimating Abundance for larger nodules. However, unbiased accurate prediction of nodule Abundance can be achieved by applying either a “linear adjustment” or a long-axis-weight relationship (Section 6).
4. Despite agreeable numerical results for nodule samples from parts of the TOML exploration contract area, analyses of samples from other regions will be needed to

better understand the generality of the empirical model and its derived formulae. Such analysis is needed in any event to calibrate the model in other areas.

5. Estimates of coverage using the new empirical method show mixed but encouraging results when compared with field measurements from three areas. For two areas the new empirical method provides close agreement but from the third area there is a consistent offset. This may be related to the degree of sediment cover in that area.
6. The new empirical method with derived explicit formulae have shown the potential of achieving more accurate mineral resource projection with reduced sample numbers and sizes. The new understanding of the nodule size distribution can also improve the efficiency of design and configuration of mining equipment.

Author Contributions: Conceptualization, GY and JP; derivation of formulae GY; sample collection and selection JP; software and modelling, GY; validation, GY and JP; formal analysis, GY; data curation, JP; Both authors drafted sections and figures and discussed and reviewed each other's contributions. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable

Acknowledgements: This research was completed by the authors while under the employment of Nautilus Minerals Pacific, and their allocation of time to spend on the subject is gratefully acknowledged.

The sample images used are property of Tonga Offshore Mining Limited and their permission to use the images for this research is also gratefully acknowledged.

The authors would also like to thank their wives, Yuet Terry Ting and Nicola Parianos for their encouragement and support during the course of this work.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A: Functions $F_1(\alpha)$, $F_2(\alpha)$ and $F_3(\alpha)$

For a random sample X_1, X_2, \dots, X_N of size N , following the Generalized Rayleigh Distribution (GRD) with its probability density function (PDF) $f(x)$ in the form of

$$f(x) = 2\alpha\beta^2 x e^{-(\beta x)^2} [1 - e^{-(\beta x)^2}]^{\alpha-1}, \quad x > 0, \quad (1)$$

its mean value \bar{X} or μ can be expressed in the integral form below

$$\bar{X} = \mu = \frac{1}{N} \sum_{i=1}^N X_i = \int_0^{\infty} x f(x) dx = 2\alpha\beta^2 \int_0^{\infty} x^2 e^{-(\beta x)^2} [1 - e^{-(\beta x)^2}]^{\alpha-1} dx \quad . \quad a)$$

Similarly, the means of the sums of square and cubic \bar{X}^2 and \bar{X}^3 be written respectively as:

$$\bar{X}^2 = \frac{1}{N} \sum_{i=1}^N X_i^2 = \int_0^{\infty} x^2 f(x) dx = 2\alpha\beta^2 \int_0^{\infty} x^3 e^{-(\beta x)^2} [1 - e^{-(\beta x)^2}]^{\alpha-1} dx \quad , \quad (30)$$

and

$$\bar{X}^3 = \frac{1}{N} \sum_{i=1}^N X_i^3 = \int_0^{\infty} x^3 f(x) dx = 2\alpha\beta^2 \int_0^{\infty} x^4 e^{-(\beta x)^2} [1 - e^{-(\beta x)^2}]^{\alpha-1} dx \quad . \quad (31)$$

Eqs. a), (30) and (31) can then be combined formally as below

$$\bar{X}^m = \frac{1}{N} \sum_{i=1}^N X_i^m = \int_0^{\infty} x^m f(x) dx = 2\alpha\beta^2 \int_0^{\infty} x^{m+1} e^{-(\beta x)^2} [1 - e^{-(\beta x)^2}]^{\alpha-1} dx \quad , \quad (32)$$

where $m = 1, 2, 3$.

Assuming $z = (\beta x)^2$ and $dz = 2\beta^2 x dx$, Eq. (32) becomes

$$\begin{aligned} \bar{X}^m &= 2\alpha\beta^2 \int_0^{\infty} x^{m+1} e^{-(\beta x)^2} [1 - e^{-(\beta x)^2}]^{\alpha-1} dx \\ &= \alpha \int_0^{\infty} \frac{z^{\frac{m}{2}}}{\beta^m} e^{-z} [1 - e^{-z}]^{\alpha-1} dz = \frac{\alpha}{\beta^m} \int_0^{\infty} z^{\frac{m}{2}} e^{-z} [1 - e^{-z}]^{\alpha-1} dz \quad . \end{aligned} \quad (33)$$

Further defining

$$F_m(\alpha) = \int_0^{\infty} z^{\frac{m}{2}} e^{-z} [1 - e^{-z}]^{\alpha-1} dz \quad , \quad \text{where } m = 1, 2, 3, \quad (34)$$

the Eq. (33) can be written as

$$\bar{X}^m = \frac{\alpha}{\beta^m} F_m(\alpha) \quad . \quad (35)$$

For Eq. (35), when $m = 1$, the mean μ of the sample is

$$\mu = \frac{\alpha}{\beta} F_1(\alpha) \quad . \quad (36)$$

The variance σ^2 of the Generalized Rayleigh Distribution can be calculated as

$$\begin{aligned} \sigma^2 &= \int_0^{\infty} (x - \mu)^2 f(x) dx = \int_0^{\infty} (x^2 - 2\mu x + \mu^2) f(x) dx \\ &= \int_0^{\infty} x^2 f(x) dx - 2\mu \int_0^{\infty} x f(x) dx + \mu^2 \int_0^{\infty} f(x) dx \\ &= \int_0^{\infty} x^2 f(x) dx - 2\mu\mu + \mu^2 = \int_0^{\infty} x^2 f(x) dx - \mu^2 \quad , \end{aligned} \quad (37)$$

or

$$\sigma^2 + \mu^2 = \int_0^{\infty} x^2 f(x) dx = \overline{X^2} = \frac{\alpha}{\beta^2} F_2(\alpha) \quad (38)$$

Combining Eq. (36) into Eq. (38) gives

$$\frac{\sigma^2 + \mu^2}{\mu^2} = \frac{\frac{\alpha}{\beta^2} F_2(\alpha)}{\frac{\alpha^2}{\beta^2} [F_1(\alpha)]^2} = \frac{F_2(\alpha)}{\alpha [F_1(\alpha)]^2} \quad (39)$$

Defining

$$G(\alpha) = \frac{F_2(\alpha)}{\alpha [F_1(\alpha)]^2} - 1 \quad (40)$$

Eq. (39) can be rewritten as

$$\frac{\sigma^2}{\mu^2} = G(\alpha) \quad (41)$$

or

$$\sigma = \mu \sqrt{G(\alpha)} \quad (42)$$

From Eq. (34), functions $F_1(\alpha)$, $F_2(\alpha)$ and $F_3(\alpha)$ can be expressed as:

$$F_m(\alpha) = \int_0^{\infty} z^{\frac{m}{2}} e^{-z} [1 - e^{-z}]^{\alpha-1} dz \quad , \text{ where } m = 1, 2, 3 \quad (43)$$

By generalized Binomial Theorem, Eq. (43) gives

$$\begin{aligned} F_m(\alpha) &= \int_0^{\infty} z^{\frac{m}{2}} e^{-z} \left[\sum_{k=0}^{\infty} \binom{\alpha-1}{k} 1^{(\alpha-1)-k} (-e^{-z})^k \right] dz \\ &= \sum_{k=0}^{\infty} \binom{\alpha-1}{k} \int_0^{\infty} z^{\frac{m}{2}} e^{-z} [(-e^{-z})^k] dz = \sum_{k=0}^{\infty} \binom{\alpha-1}{k} (-1)^k \int_0^{\infty} z^{\frac{m}{2}} e^{-(k+1)z} dz \quad (44) \end{aligned}$$

Assuming $y = (k+1)z$, the integral in Eq. (43) becomes

$$\begin{aligned} \int_0^{\infty} z^{\frac{m}{2}} e^{-(k+1)z} dz &= \int_0^{\infty} \left(\frac{y}{k+1} \right)^{\frac{m}{2}} e^{-y} d \left(\frac{y}{k+1} \right) = \frac{1}{(k+1)^{\frac{m}{2}+1}} \int_0^{\infty} y^{\frac{m}{2}} e^{-y} dy \\ &= \frac{1}{(k+1)^{\frac{m}{2}+1}} \Gamma\left(\frac{m}{2} + 1\right) \quad (45) \end{aligned}$$

where $\Gamma\left(\frac{m}{2} + 1\right)$ is the Gamma Function. Inserting Eq. (43) into Eq. (43) gives

$$\begin{aligned} F_m(\alpha) &= \sum_{k=0}^{\infty} \binom{\alpha-1}{k} (-1)^k \frac{\Gamma\left(\frac{m}{2} + 1\right)}{(k+1)^{\frac{m}{2}+1}} = \Gamma\left(\frac{m}{2} + 1\right) \sum_{k=0}^{\infty} \frac{(\alpha-1)!}{k! (\alpha-1-k)! (k+1)^{\frac{m}{2}+1}} \\ &= \Gamma\left(\frac{m}{2} + 1\right) \sum_{k=0}^{\infty} \frac{(\alpha-1)(\alpha-2) \dots [(\alpha-1-k)+1]}{(k+1)!} \frac{(-1)^k}{(k+1)^{\frac{m}{2}}} \\ &= \Gamma\left(\frac{m}{2} + 1\right) \sum_{k=0}^{\infty} (-1)^k \frac{(\alpha-1)(\alpha-2) \dots (\alpha-k)}{(k+1)! (k+1)^{\frac{m}{2}}} \quad (46) \end{aligned}$$

For $m = 1, 2$ and 3 , noting $\Gamma\left(\frac{1}{2} + 1\right) = \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$, $\Gamma\left(\frac{2}{2} + 1\right) = \Gamma(2) = 1$ and $\Gamma\left(\frac{3}{2} + 1\right) = \Gamma\left(\frac{5}{2}\right) = \frac{3}{4}\sqrt{\pi}$ respectively, $F_1(\alpha)$, $F_2(\alpha)$ and $F_3(\alpha)$ can then be expressed as:

$$\begin{aligned} F_1(\alpha) &= \frac{\sqrt{\pi}}{2} \sum_{k=0}^{\infty} (-1)^k \frac{(\alpha-1)(\alpha-2) \dots (\alpha-k)}{(k+1)! \sqrt{k+1}} \\ &= \frac{\sqrt{\pi}}{2} \left[1 - \frac{(\alpha-1)}{\sqrt{2} * 2!} + \frac{(\alpha-1)(\alpha-2)}{\sqrt{3} * 3!} - \frac{(\alpha-1)(\alpha-2)(\alpha-3)}{\sqrt{4} * 4!} + \dots \right] \quad (47) \end{aligned}$$

and

$$F_2(\alpha) = \sum_{k=0}^{\infty} (-1)^k \frac{(\alpha-1)(\alpha-2)\dots(\alpha-k)}{(k+1)!(k+1)} \quad (48)$$

$$= \left[1 - \frac{(\alpha-1)}{2*2!} + \frac{(\alpha-1)(\alpha-2)}{3*3!} - \frac{(\alpha-1)(\alpha-2)(\alpha-3)}{4*4!} + \dots \right],$$

and

$$F_3(\alpha) = \frac{3}{4}\sqrt{\pi} \sum_{k=0}^{\infty} (-1)^k \frac{(\alpha-1)(\alpha-2)\dots(\alpha-k)}{(k+1)!(k+1)^{\frac{3}{2}}} \quad (49)$$

$$= \frac{3}{4}\sqrt{\pi} \left[1 - \frac{(\alpha-1)}{2\sqrt{2}*2!} + \frac{(\alpha-1)(\alpha-2)}{3\sqrt{3}*3!} - \frac{(\alpha-1)(\alpha-2)(\alpha-3)}{4\sqrt{4}*4!} + \dots \right].$$

Appendix B: Computation of Critical Values for Goodness-of-Fit Test of Generalized Rayleigh Distribution by Monte Carlo Simulations

The critical values y_γ and u_γ of the test statistics A_n^2 and V_n^2 can be computed as their percentage points using Monte Carlo simulations. Given the parameters α and β and sample size n , the numerical procedure consists the following steps:

1. A set of random numbers of size n is generated in the interval (0,1) as values of cumulative distribution function (CDF). Eq.2 is used to back-calculate a sample X_1, X_2, \dots, X_n of size n for given α and β .
2. For a sample X_1, X_2, \dots, X_n , Eq.16 and Eq.17, based on MLE method, are solved iteratively to estimate parameters $\hat{\alpha}$ and $\hat{\beta}$.
3. Parameters $\hat{\alpha}$ and $\hat{\beta}$ are used in Eq.2 to calculate $z_i = F(x_i)$, with values in ascending order.
4. Eq.18 is used to calculate test statistics A_n^2 and V_n^2 , using values of z_i calculated in step 3.
5. Steps 1. to 4. above are repeated to generate a sample for A_n^2 and V_n^2 .
6. The percentiles of A_n^2 and V_n^2 are calculated as critical values. The $(1-\gamma)^{\text{th}}$ percentile is taken as the critical value for the level of significance of γ .

For a set of critical value for given parameters α and β and sample size n , 250,000 Monte Carlo simulations are carried out to ensure the convergence to 2 digits. The simulated results are presented in Table 1, Figure 15, Figure 16 below. In Figure 15 and Figure 16 results for both 200,000 and 250,000 simulations are shown, and the relative error is within 0.1%.

Table 5: Critical values y_γ and u_γ for various shape parameters α and sample sizes n

Shape Parameter α		1.0			1.5			2.0			2.5			3.0			3.5			4.0			
Significance Level γ		10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	
Sample Size n	100	y_γ	0.66	0.79	1.10	0.65	0.78	1.08	0.64	0.77	1.07	0.64	0.76	1.06	0.64	0.76	1.05	0.64	0.76	1.05	0.64	0.76	1.05
		u_γ	0.34	0.41	0.58	0.34	0.41	0.57	0.33	0.41	0.57	0.33	0.40	0.57	0.33	0.40	0.57	0.33	0.41	0.57	0.34	0.41	0.57
	200	y_γ	0.66	0.79	1.11	0.65	0.78	1.08	0.65	0.77	1.07	0.64	0.77	1.05	0.64	0.76	1.05	0.64	0.76	1.06	0.64	0.76	1.05
		u_γ	0.34	0.41	0.58	0.34	0.41	0.58	0.34	0.41	0.57	0.34	0.41	0.57	0.34	0.40	0.57	0.34	0.41	0.57	0.34	0.41	0.57
	300	y_γ	0.66	0.79	1.10	0.65	0.78	1.08	0.65	0.77	1.08	0.64	0.77	1.06	0.64	0.76	1.05	0.64	0.76	1.05	0.64	0.76	1.06
		u_γ	0.34	0.41	0.58	0.34	0.41	0.58	0.34	0.41	0.57	0.34	0.41	0.57	0.34	0.41	0.57	0.34	0.41	0.57	0.34	0.41	0.57
	400	y_γ	0.66	0.80	1.11	0.65	0.78	1.08	0.65	0.77	1.07	0.64	0.77	1.06	0.64	0.76	1.06	0.64	0.76	1.06	0.64	0.76	1.05
		u_γ	0.34	0.41	0.59	0.34	0.41	0.58	0.34	0.41	0.58	0.34	0.41	0.57	0.34	0.41	0.57	0.34	0.41	0.57	0.34	0.41	0.57
	500	y_γ	0.66	0.80	1.11	0.65	0.78	1.09	0.64	0.77	1.07	0.64	0.77	1.06	0.64	0.77	1.06	0.64	0.76	1.06	0.64	0.76	1.05
		u_γ	0.34	0.41	0.59	0.34	0.41	0.58	0.34	0.41	0.57	0.34	0.41	0.57	0.34	0.41	0.57	0.34	0.41	0.57	0.34	0.41	0.57

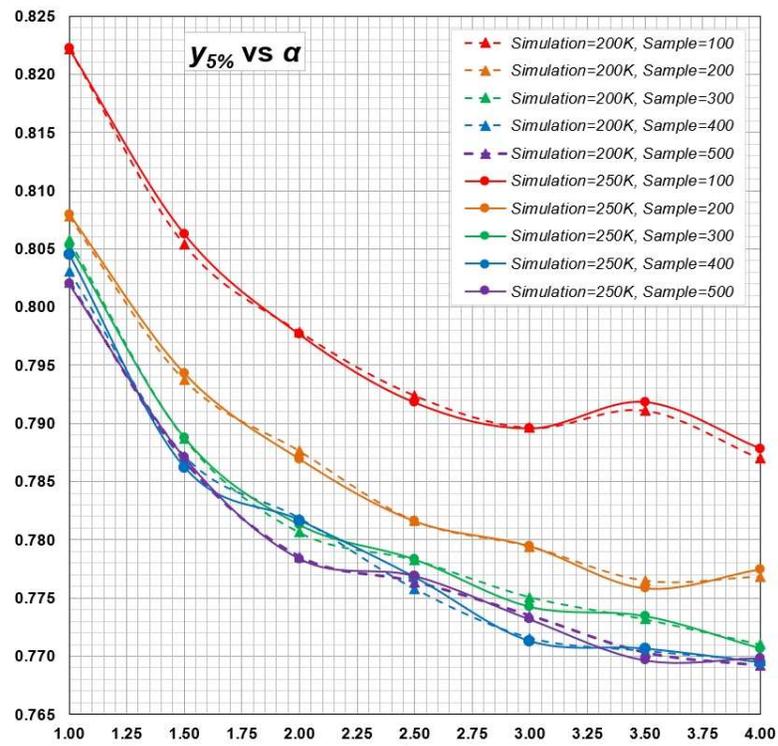


Figure 15: Critical values y_γ versus shape parameters α and sample sizes n with 200,000 and 250,000 simulations.

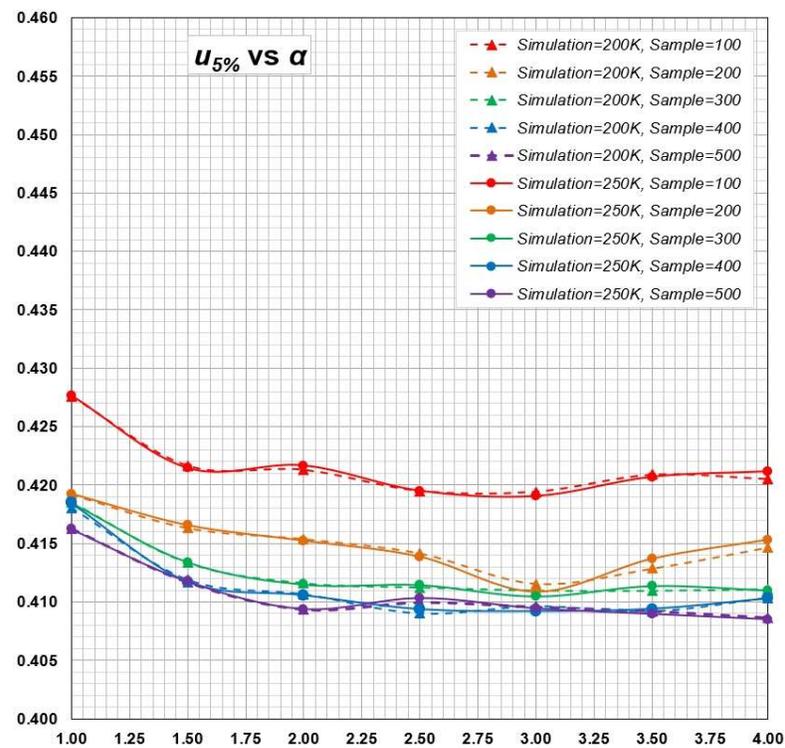


Figure 16: Critical values u_γ versus shape parameters α and sample sizes n with 200,000 and 250,000 simulations.

Appendix C: Determination of Minimal Sample Size for Estimation of Statistical Distribution Using Monte Carlo Simulations

It is of practical importance to estimate the minimal sample size necessary to achieve desired accuracy of estimation. To test whether a sample size of 100 is sufficient to yield an accurate estimation of statistical distribution, 500 sets of samples with $size = 100$ were randomly selected from the whole sample with $size = 440$, and for each set of selected samples, parameters α and β are computed empirically using Eq.(11), and a corresponding GRD is generated.

For a typical simulation, the mean values of α and β , calculated by averaging results from 500 simulations, are within 4%, compared with values of α and β calculated based on the whole sample ($size=440$). The maximum errors for α and β between a small-sample ($size=100$) and the whole sample ($size=440$) is about 10% and 5%, respectively. Figure 17 and Figure 18 show CDFs and PDFs from 10 simulations (randomly selected from a total of 500 simulations), compared with the CDF and PDF, calculated from the whole sample ($size=440$). The reasonably good agreement between the results from samples with $size=100$ and those from the whole sample ($size=440$) indicates that a smaller sample size of 100 can generally produce a good estimation of the statistical distribution.

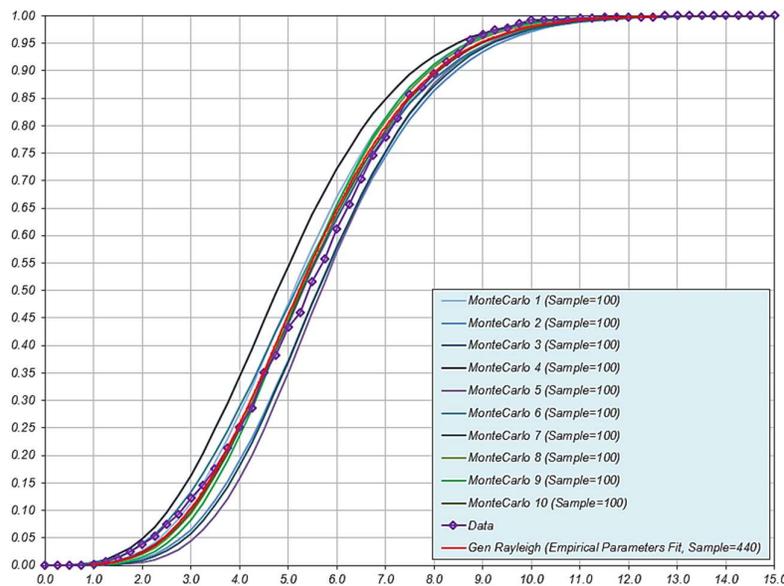


Figure 17: CDF simulated using subsets of sample with Size=100

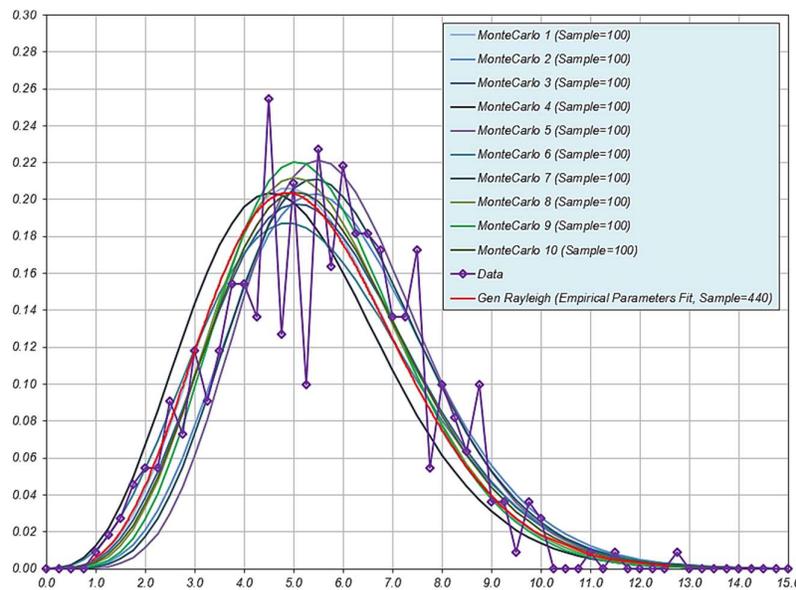


Figure 18: PDF simulated using subsets of sample with Size=100

References:

- Haynes, B.W.; Law, S.L.; Barron, D.C.; Kramer, G.W.; Maeda, R.; Magyar, M.. Pacific manganese nodules: characterisation and processing. *United States Geol. Surv. Bull.* **1985**, *679*, 44.
- International Seabed Authority *A geological Model of Polymetallic Nodule Deposits in the Clarion-Clipperton Fracture Zone*; International Seabed Authority: Kingston, Jamaica, 2010;
- Fouquet, Y.; Depauw, G.; GEMONOD Polymetallic Nodules Resource Classification. In Proceedings of the Workshop on Polymetallic Nodule Resources Classification; International Seabed Authority: Kingston, Jamaica, 2014.
- Lipton, I.; Nimmo, M.; Parianos, J. *TOML Clarion Clipperton Zone Project, Pacific Ocean*; AMC Consultants Pty Ltd: Brisbane, Australia, 2016;
- Lipton, I.; Nimmo, M.; Stevenson, I. *NORI Area D Clarion Clipperton Zone Mineral Resource Estimate - Update*; AMC Consultants Pty Ltd: Brisbane, Australia, 2021;
- International Seabed Authority *Secretary General Annual Report*; International Seabed Authority: Kingston, Jamaica, 2020;
- Schoening, T.; Gazis, I.-Z. Sizes, weights and volumes of poly-metallic nodules from box cores taken during SONNE cruises SO268/1 and SO268/2 Available online: <https://doi.pangaea.de/10.1594/PANGAEA.904962> (accessed on Feb 9, 2021).
- Kaufman, R.; Siapno, W.D. Future needs of deep ocean mineral exploration and surveying. *Offshore Technol. Conf. Prepr.* **1972**, *2*, 309–332.
- Felix, D. Some problems in making nodule abundance estimates from sea floor photographs. *Mar. Min.* **1980**, *2*, 293–302.
- Longuet-Higgins, M.S. On the statistical distribution of the heights of sea waves. *J. Mar. Res.* **1952**, *XI*, 245–266.
- Abd-Elfattah, A.M. Goodness of fit test for the generalized Rayleigh distribution with unknown parameters. *J. Stat. Comput. Simul.* **2011**, *81*, 357–366, doi:10.1080/00949650903348155.
- Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639, doi:10.1021/ac60214a047.