British Geological Survey
Expert | Impartial | Innovative

Geostatistical Modelling Work for East of Gannet and Montrose Fields NCMPA

GeoAnalytics and Modelling Programme

Commissioned Report CR/19/062

BRITISH GEOLOGICAL SURVEY

GEOANALYTICS and Modelling PROGRAMME

COMMISSIONED REPORT CR/19/062

Geostatistical Modelling Work for East of Gannet and Montrose Fields NCMPA

B P Marchant

Contributor/editor

D Dove

Bibliographical reference

MARCHANT, B.P.2019. Geostatistical Modelling Work for East of Gannet and Montrose Fields NCMPA. *British Geological Survey Commissioned Report,* CR/19/062. 18pp.

Keyworth, Nottingham British Geological Survey 2019

BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.

The British Geological Survey is a component body of UK Research and Innovation.

British Geological Survey offices

Environmental Science Centre, Keyworth, Nottingham NG12 5GG

Tel 0115 936 3100

BGS Central Enquiries Desk

Tel 0115 936 3143
email enquiries@bgs.ac.uk

BGS Sales

Tel 0115 936 3241
email sales@bgs.ac.uk

The Lyell Centre, Research Avenue South, Edinburgh EH14 4AP

Tel 0131 667 1000
email scotsales@bgs.ac.uk

Natural History Museum, Cromwell Road, London SW7 5BD

Tel 020 7589 4090
Tel 020 7942 5344/45 email bgslondon@bgs.ac.uk

Cardiff University, Main Building, Park Place, Cardiff CF10 3AT

Tel 029 2167 4280

Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB

Tel 01491 838800

Geological Survey of Northern Ireland, Department of Enterprise, Trade & Investment, Dundonald House, Upper Newtownards Road, Ballymiscaw, Belfast, BT4 3SB

Tel 01232 666595
www.bgs.ac.uk/gsni/

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU

Tel 01793 411500 Fax 01793 411501
www.nerc.ac.uk

UK Research and Innovation, Polaris House, Swindon SN2 1FL

Tel 01793 444000
www.ukri.org

Website www.bgs.ac.uk
Shop online at www.geologyshop.com

Foreword

This report is on work undertaken by BGS for the Joint Nature Conservation Committee (JNCC).

Contents

FIGURES

Summary

The background to this work as described in the JNCC specification document is as follows:

"The Joint Nature Conservation Committee (JNCC) and Marine Scotland Science (MSS) completed a survey in November 2015 to collect evidence to characterise and develop the first point in a marine monitoring time-series for the East of Gannet and Montrose (EGM) Nature Conservation Marine Protected Area (NCMPA). One of the objectives of the survey was: to "Describe the extent and distribution, structure and functions, and supporting processes of offshore deep-sea muds within EGM". Sediment sampling was undertaken and particle size analysis (PSA) conducted for these samples to determine sediment and, by extension, habitat designation. Three sedimentary benthic habitats were identified from PSA of 155 large (0.25 m2) Hamon grabs across the site: A5.2 Sublittoral sand, A5.3 Sublittoral mud and A5.4 Sublittoral mixed sediment. The report found that the sedimentary habitats of the site were fairly distinctive and differed from previous survey data acquired from the site, therefore JNCC requires new sedimentary habitat information to update the habitat mapping for the site."

The required new sedimentary habitat information consists of series of 100m resolution raster layers covering the site and containing:

(i) predictions of proportions of sand, gravel and mud

(ii) estimates of the uncertainty of each of the layers in (i) consisting of upper and lower limits and width of the confidence interval

(iii) a classification layer indicating membership of one of the four EUNIS level 3 sediment texture classes: Mud, Sand, Coarse Sediment and Mixed Sediment

(iv) a layer describing the proportion of model runs which indicate membership of the most probable class in (iii)

The EUNIS level 3 sediment class definitions (Long, 2006) are:

Mud and sandy mud: Ratio of sand to mud less than 4:1 and less than 5% gravel.

Sand and muddy sand: Ratio of sand to mud greater than 4:1 and less than 5% gravel.

Mixed sediment: Between 5 and 80% gravel and ratio of sand to mud less than 9:1.

Coarse sediment: sediment distributions satisfying none of the above.

These classes are illustrated on the ternary diagram shown in Figure 1.

This report describes the production of the habitat information by BGS.

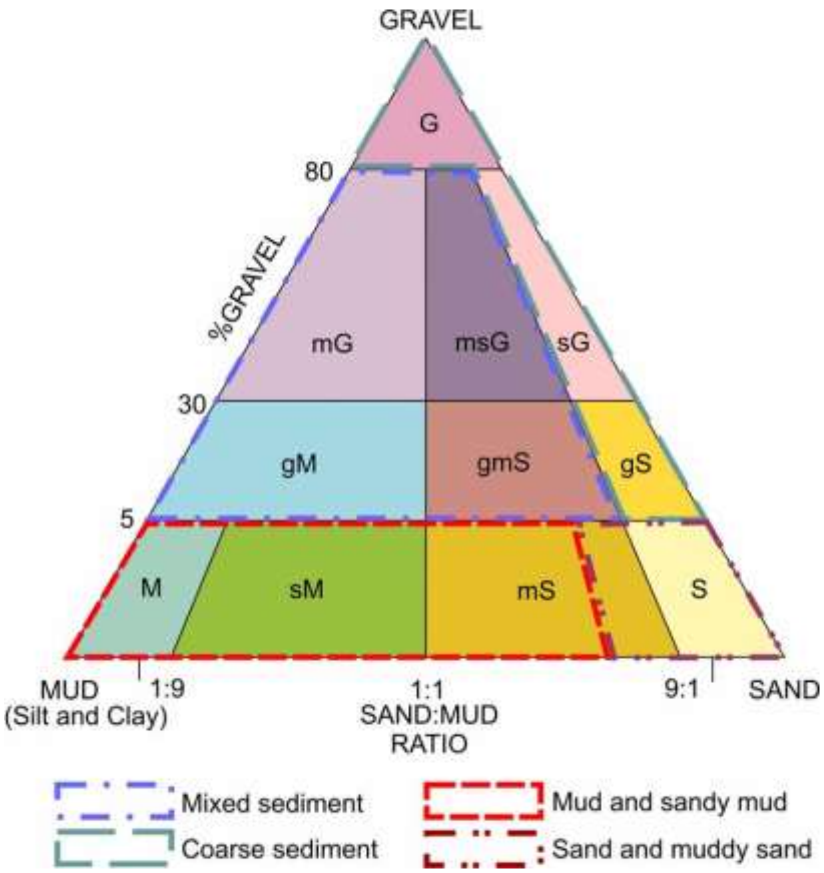

Figure 1 Classification of sediment classes reproduced from Long (2006).

1 Introduction

The methodology to be used is based on that introduced by Lark et al. (2012) and modified by Lark et al. (2015). A detailed account of the methodology is provided in these papers and this report should be read in conjunction with them to give a complete explanation of the approach. This report provides additional information where required.

2 Methodology

2.1 DATA

The data used are 155 particle size analyses from the 2015 survey of the EGM NCMPA conducted by the JNCC and MSS. The locations of the analysed samples are relatively evenly dispersed across the NCMPA (Figure 2). Each Hamon grab sample had a support of 0.25 m^2 and the data considered were percent by mass of gravel (particles diameter > 2mm), mud (particles diameter < 0.063 mm) and sand (particles 2mm > diameter > 0.063 mm). No continuous bathymetry or backscatter information were available to inform the modelling work.

2.2 STATISTICAL METHODOLGY

The JNCC required interpolated maps of the proportion of each of the three particle size classes and associated measures of uncertainty. Geostatistical methods such as kriging are required to quantify uncertainty in spatial variables such as these. The standard geostatistical approach first quantifies the degree of spatial correlation in measurements of the property of interest by means of a variogram. Then the best linear unbiased or *kriging* predictor is used to determine the expected value of that measurement at sites where it has not been measured and to quantify the uncertainty in these predictions. The kriged prediction at a site is a weighted sum of nearby measurements. The weights decrease with increasing distance between the measurement and prediction site. The rate of the decrease in weights is determined from the variogram.

In this study, predictions of three variables (% gravel, sand and mud) were required. These variables must be represented in single model to ensure that the predictions account for the correlation between each variable. The situation is further complicated by the particle size measurements being compositional variables (i.e. they must sum to 100% at each site). Thus, if the proportion of sand and gravel is known at a site then the proportion of mud can be inferred immediately. This compositional property can mean that standard exploratory analyses indicate erroneous negative correlations between the three variables. Pawlowsky-Glahn and Olea (2004) suggest that compositional co-kriging should be applied in this situation. In this approach, an additive log-ratio transformation is used to reduce the three variables to two. The transformed variables are the natural logarithm of the ratio of two of the components of the composition to the third. In this report we used the sand content (the most prevalent particle size class) as the denominator of the log-ratio so our two variables are $z_1$and $z_2$where:

$$z_1 = \ln\left(\frac{\%\ \text{gravel}}{\%\ \text{sand}}\right),$$

$$z_2 = \ln\left(\frac{\%\ \text{mud}}{\%\ \text{sand}}\right). \qquad \text{Eqn (1)}$$

These two variables are then represented in a linear model of coregionalization (LMCR). The LMCR includes a variogram for each of the two variables and a cross-variogram. Each variogram describes the expected squared difference between two measurements of the variable as a function of the distance between the measurements. The cross-variogram considers the covariance of the difference between two measurements of one variable and the difference between two measurements of the other variable made at the same pair of sites as the first. Again, this covariance is expressed as a function of the distance between the pair of measurement locations. The LMCR is then used within the co-kriging predictor to produce maps of each variable across the study area. The co-kriging prediction of a variable is a weighted sum of nearby measurements of each variable. The weights associated with each measurement are determined from the LMCR. The co-kriging approach can also be used to simulate multiple realizations of each variable across the study region. A set of simulated variables at each site can be back transformed to yield simulations of % gravel, % mud and % sand:

$$\%\ \text{gravel} = \frac{100\exp(z_1)}{1 + \exp(z_1) + \exp(z_2)},$$

$$\%\ \text{mud} = \frac{100\exp(z_2)}{1 + \exp(z_1) + \exp(z_2)},$$

$$\%\ \text{sand} = \frac{100}{1 + \exp(z_1) + \exp(z_2)}. \qquad \text{Eqn(2)}$$

If a sufficient number of simulated realisations are produced then these can be used to determine across the study region the expected proportion of each particle size class, the confidence limit on these proportions and the proportion of simulations for which each site lies in a specified EUNIS level 3 sediment texture class.

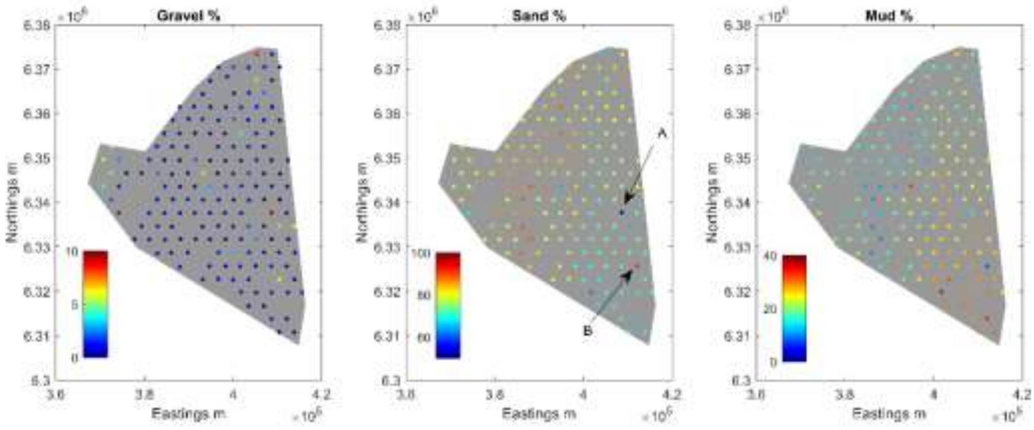

Figure 2 Spatial variation of measured percentages of gravel, sand and mud in Hamon grab samples. Note that different colour-scales are used in each plot.

3 Results

3.1 EXPLORATORY DATA ANALYSIS

The measurements of % gravel, % mud and % sand are shown in Figure 2 and their histograms in Figure 3. The percentages of gravel are generally small with an average of 1.3% and only nine of the 155 measurements exceeding 5%. The median % gravel is 0.44% but 13 values larger than 4% are scattered across the study region. There is no clear spatial pattern to these larger values except in the far-west of the study region where there appears to be a region of consistently large values. Large-scale patterns of spatial variation are evident in the % sand and % mud measurements. These variables average 78.9 and 21.0% respectively. Two potential outliers are highlighted in Figure 2. One, marked 'A', can be considered a global outlier since its gravel % of 35% is substantially larger than the values observed in the remainder of the dataset. The next largest value is 10.4%. The other potential outlier, marked 'B', could be a local outlier. The observed vales of 0.1% gravel, 91.8% sand and 8% mud are consistent with the wider dataset but not consistent with the measurements at neighbouring sites.

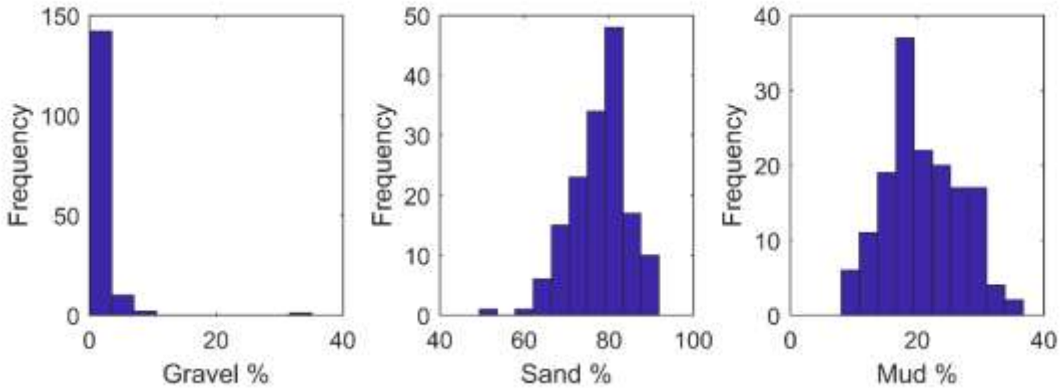

Figure 3 Histograms of measured percentages of gravel, sand and mud in Hamon grab samples.

3.2 GEOSTATISTICAL MODELLING

The maps and histograms of the additive log transformed data values are shown in Figures 4 and 5 respectively. The large-scale variation of the % mud to % sand ratio leads to a clear spatial pattern in z_2 whereas the spatial pattern in z_1 is more irregular except for larger values in the west of the region which correspond to the relatively large gravel % in this area. The distributions of each variable are approximately symmetric and we consider them to be consistent with the Gaussian assumption of the LMCR. Transformed variable z_1 has a variance of 1.77 whereas z_2 has a variance of 0.14.

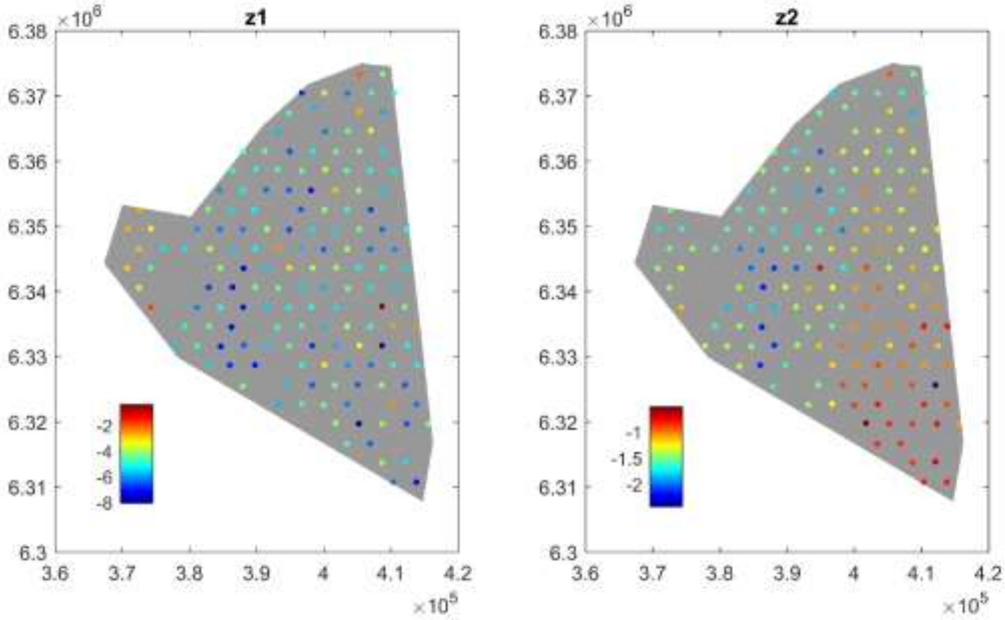

Figure 4 Spatial variation of additive log transformed variables.

We follow Lark et al. (2015) and estimate the parameters of the LMCR for these transformed variables by residual maximum likelihood (REML). The estimated variograms and cross-variograms of the LMCR are shown in Figure 6. There is spatial correlation apparent in the variograms up to and beyond 30 km. The proportion of spatial correlation is much less for z_1 than z_2 reflecting the smoother pattern of spatial variation in z_2 (Figure 4).

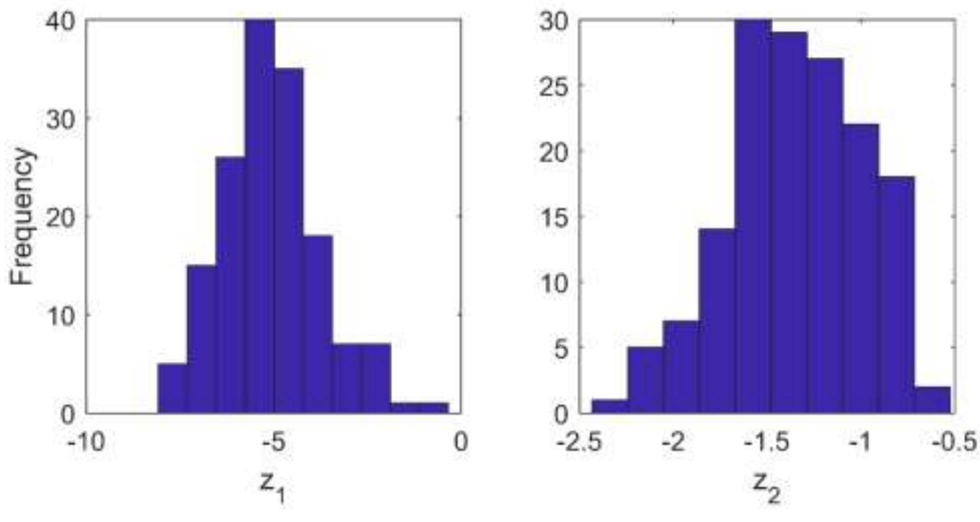

Figure 5 Histograms of additive log transformed variables.

We validate the estimated model by leave-one-out cross validation. In this procedure, a single measurement is removed from the dataset and then the remaining measurements and the co-kriging predictor are used to predict the removed measurement and its uncertainty by co-kriging. The process is then repeated, removing each measurement in turn. If the estimated LMCR appropriately represents the measured data then one would expect the mean squared standardised error to be 1.0 and the median squared standardised error to be 0.45. The standardised error is the difference between the predicted and measured values divided by the standard error of the prediction. When this procedure was applied to the data the mean standardised error was 1.00 and the median 0.39. Such values would be considered to be sufficiently close to the expected values. However, the two largest standardised errors were 14.6 and 32.9 which would be expected to occur with a probability of less than 1 in 10,000 and would be unlikely to occur in a dataset of 155 measurements. Given that these two measurements were previously identified as being visually inconsistent with the other data (Figure 2) we remove them from the dataset and repeat the model estimation procedure.

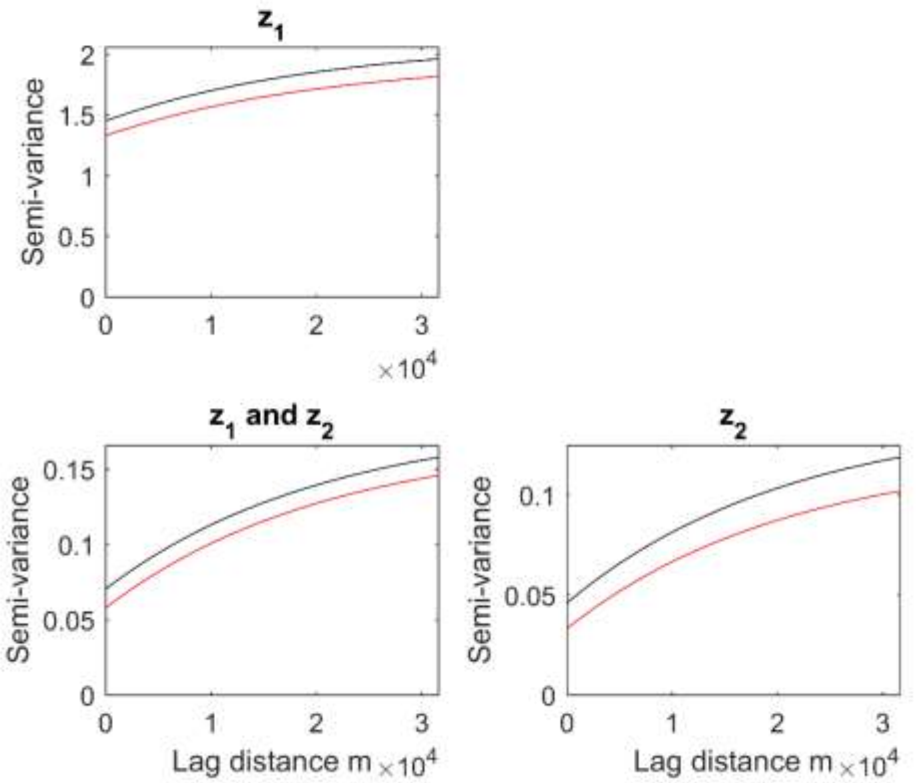

Figure 6 Variograms and cross-variograms for additive log transformed variables. REML estimated models for all data are shown in black and for data once outliers have been excluded in red.

The estimated LMCR once these observations have been deleted is shown in Figure 6 (red). The same pattern of spatial correlation remains but the semi-variances are decreased. Upon leave-one-out cross-validation the mean standardised prediction error is 0.99 and the median 0.43. The largest standardised prediction error has decreased to 9.74. Figure 7 shows the predicted percentages upon leave-one-out cross-validation plotted against the observed values. A strong correlation between predicted and observed values is evident for sand and mud. The small number of relatively large gravel values are not so well predicted.

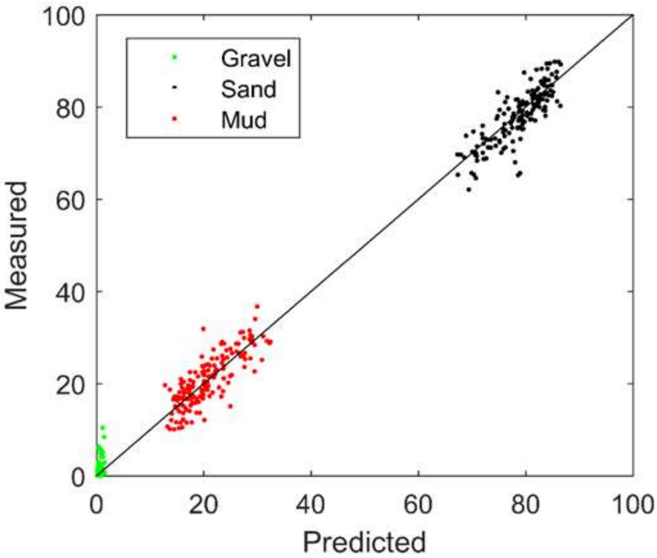

Figure 7 Predicted percentages of gravel, sand and mud upon leave-one-out cross validation against measured percentages.

The co-kriged maps of z_1 and z_2 are shown in Figure 8. These follow the same pattern of underlying variation as the observations shown in Figure 4. However, as is common with kriged predictions, the variation is smoothed with, for example, the largest predictions of z_1 being -4 compared with observed values close to zero. The maps indicate the relatively large values of the % gravel to % sand ratio in the west of the region and of the % mud to % sand ratio in the south east of the region.

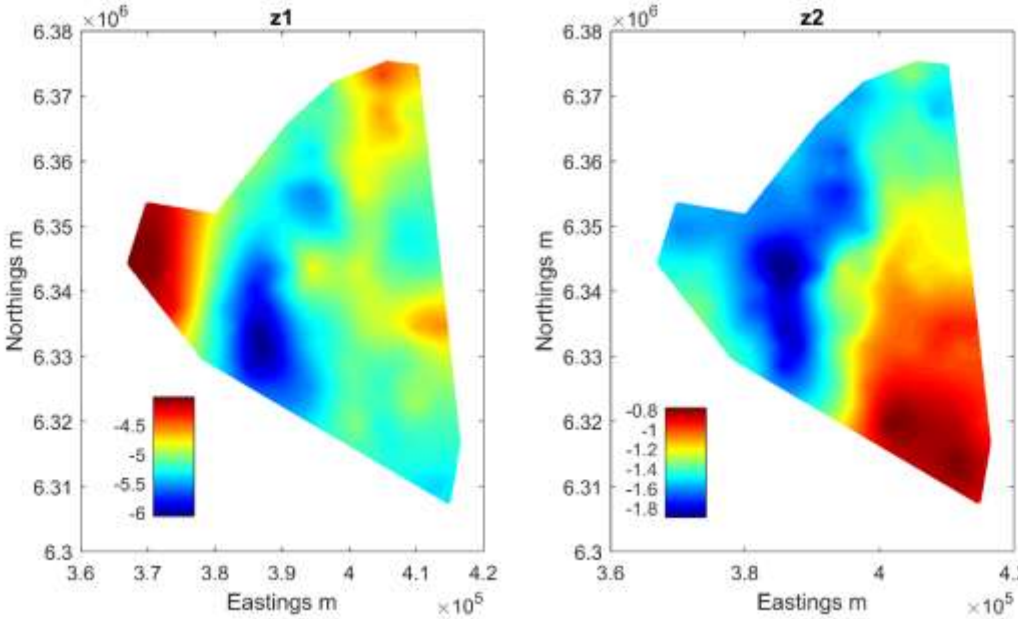

Figure 8 Spatial predictions of additive log transformed variables.

The LMCR was used to produce 1000 simulated realisations of the variation of z_1 and z_2 across the study area. Each realization was back transformed according to Equation 2 to produce 1000 realizations of percentage gravel, sand and mud. The averages of these 1000 realisations, (i.e. the expected gravel, sand and mud percentages) are shown in Figure 9. The pattern of variation of these variables are consistent with the observations shown Figure 2. Figure 10 shows the proportion of the realisations that are members of each of the EUNIS level 3 sediment classes. None of the realisations include an example of the Coarse Sediment class. The probability of Mixed Sediment is small everywhere. It averages 0.03 across the study area with the largest values of 0.15 occurring in the extreme west of the region. The Mud and Sandy Mud class dominates the south and east of the NCMPA whereas the Sand and Muddy Sand class dominates the north and west. Figure 11 compares the EUNIS class memberships of the observed samples to the most probable class according to the LMCR. These are in reasonable agreement although the isolated examples of Coarse Sediment amongst the observations are not evident in the predicted map. The map of the proportion of realisations being within the most probable class (Figure 12) confirms that in the east and south of the region that membership of the Mud and Sandy Mud class is likely and membership of the Sand and Muddy Sand class is likely in the north and west of the region. There is large uncertainty in the class membership at the boundary between these areas.

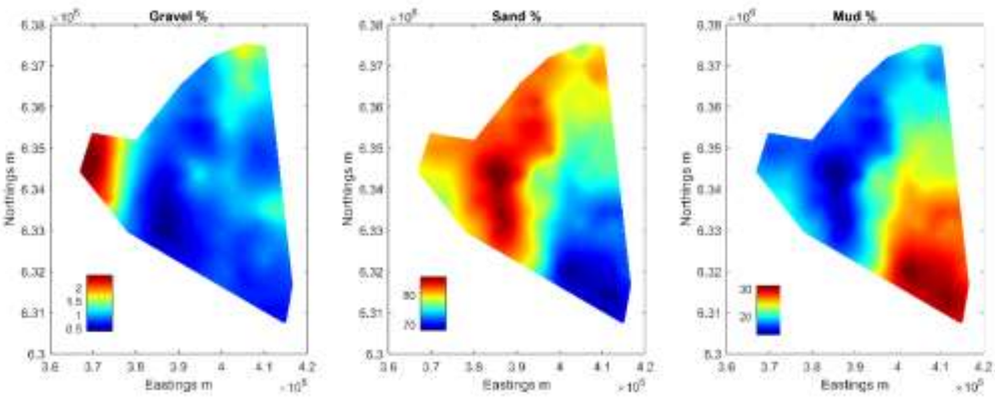

Figure 9 Spatial variation of predicted percentage gravel, sand and mud. Note that different colour-scales are used for each plot. Predictions are included in output file as columns Gmean, Smean and Mmean.

4 Discussion

The cross-validation results suggest that this geostatistical modelling exercise has led to an LMCR which is generally consistent with the observed data. Thus the model can be used to predict the sediment particle size distribution and to quantify its uncertainty. However, two points should be highlighted when interpreting the resultant maps. First, two outliers were removed from the observed data prior to producing the maps. Had these observations been included they would have led to large predictions of gravel or sand content in their immediate vicinity. Such predictions would appear to be inconsistent with the underlying variation of the sediment across the region. The maps should be interpreted as being indicative of this underlying variation but it should be remembered that at the spatial-scale of a grab sample it is possible for other texture classes to occur.

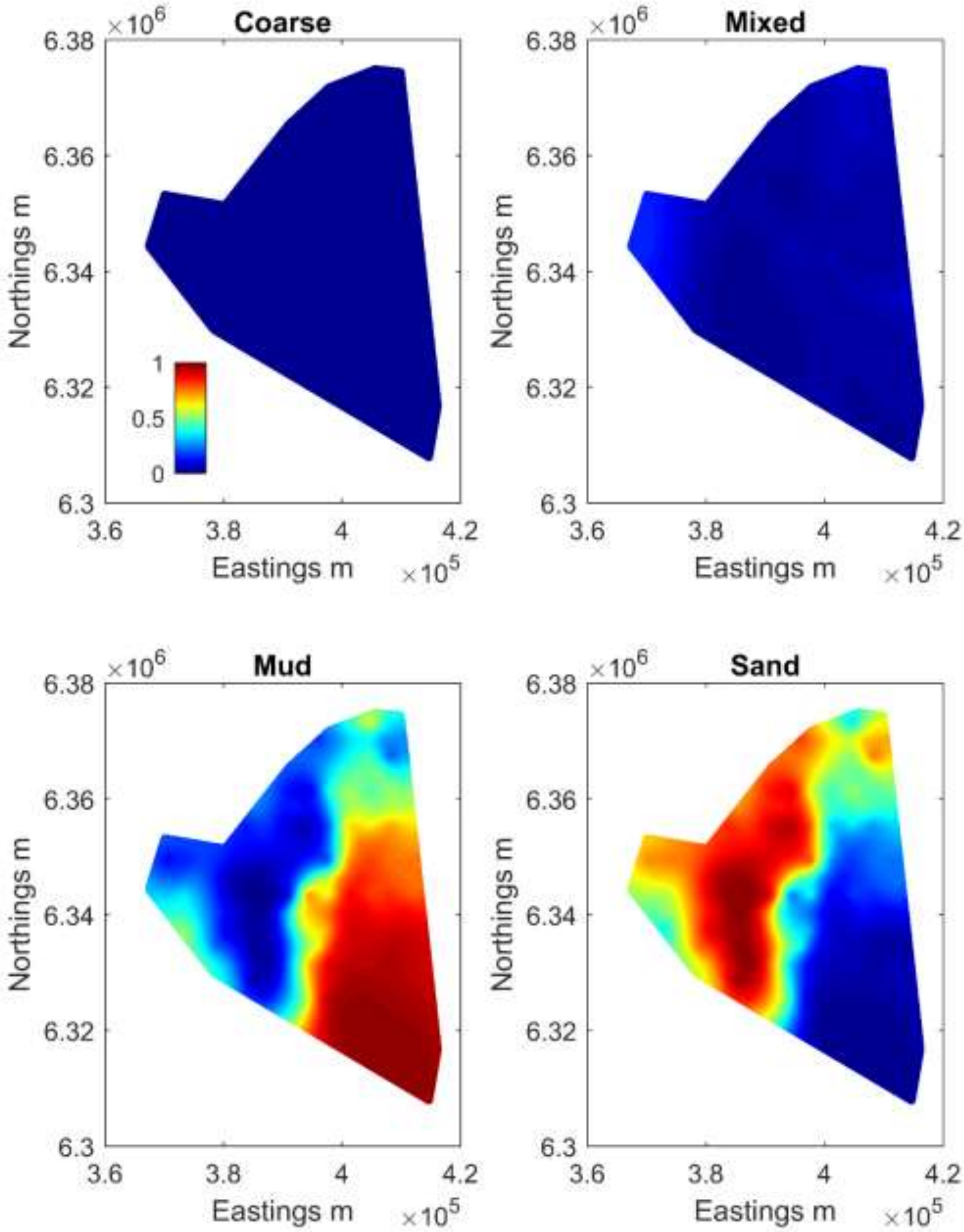

Figure 10 Proportion of simulated realizations that are members of each of the level 3 EUNIS sediment texture classes (Coarse Sediment, Mixed Sediment, Mud and Sandy Mud, Sand and Muddy Sand). Data are included output file as columns "Cp", "MIp", "MUp" and "Sp".

Second, gravel and hence the Mixed Sediment class does appear to be slightly under-represented in the model predictions. Five percent of the observed samples were Mixed Sediment whereas this class was not the most likely at any location (Figure 11). This is to be expected for a class that occurs sporadically across the region. Figure 10 indicates that this class is possible at any location but that the proportion of occurrences amongst the simulations rarely increases beyond 0.1. The average proportion of occurrences of 0.03 is rather less than the observed proportion of 0.05. We suspect that this discrepancy is largely because the model underestimates the degree to which gravel has elevated proportions in the extreme west of the region. The model assumes that the same model of spatial correlation applies across the region (*i.e.* the model is stationary). Generally, there is little spatial correlation amongst the gravel observations and the model reflects this. However, some spatial correlation appears to occur in the west where relatively large gravel values consistently occur. This is an indication of non-stationary variation. In this circumstance, additional spatial information such as swath bathymetry or backscatter data might be useful in identifying the extent of this non-stationarity and modifying the model.

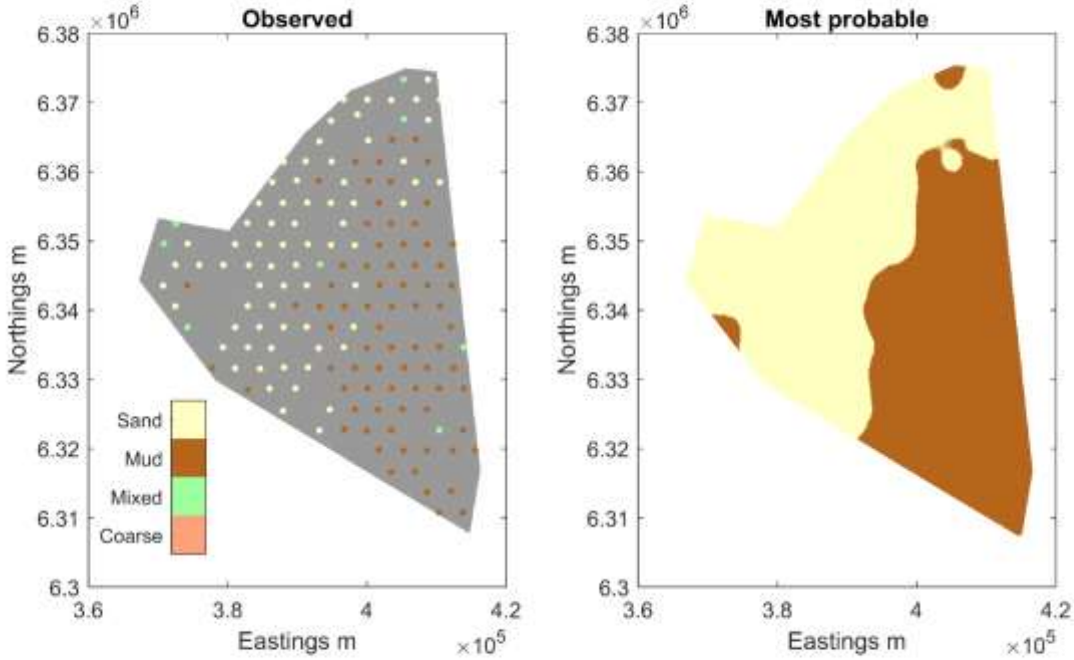

Figure 11 Level 3 EUNIS sediment texture class membership for the observed samples (left) and most likely class according to the LMCR (right). Most likely class is included in output file as column "EUNISmax"

5 Supplementary Material

The outputs of this work are provided as a csv file "EGM_psd.csv". This consists of predicted quantities at 183,876 locations on a grid of spacing 100m which covers the EGM NCMPA. The variables included in this file are:

x: Eastings from rectilinear coordinates (UTM 31N [WGS84])

y: Northings from rectilinear coordinates (UTM 31N [WGS84])

Gmean: Expected proportion of gravel

Smean: Expected proportion of sand

Mmean: Expected proportion of mud

G025: Lower limit of 95% confidence interval for gravel

S025: Lower limit of 95% confidence interval for sand

M025: Lower limit of 95% confidence interval for mud

G975: Upper limit of 95% confidence interval for gravel

S975: Upper limit of 95% confidence interval for sand

M975: Upper limit of 95% confidence interval for mud

Cp: Proportion of simulated realisations indicating Coarse Sediment

MIp: Proportion of simulated realisations indicating Mixed Sediment

MUp: Proportion of simulated realisations indicating Mud and Sandy Mud

Sp: Proportion of simulated realisations indicating Sand and Muddy Sand

EUNISmax: most probable EUNIS class

EUNISp: proportion of simulated realisations indicating most probable EUNIS class.

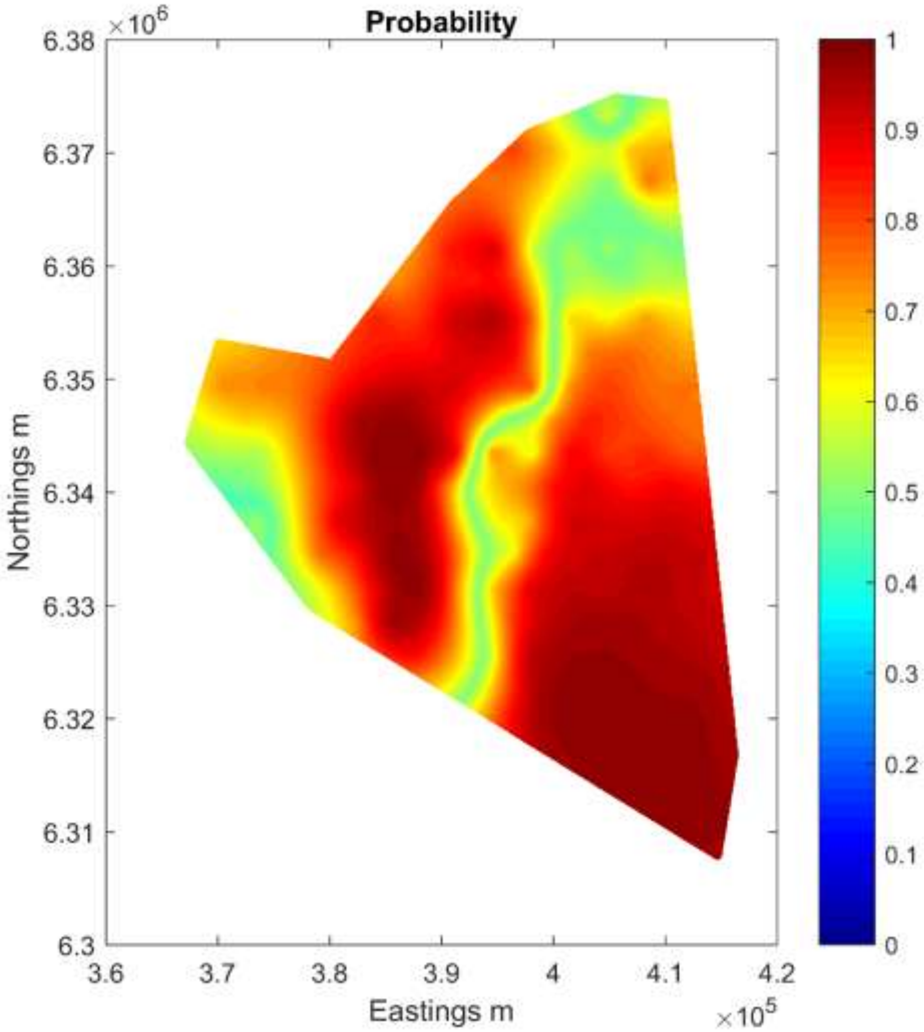

Figure 12 Proportion of simulated realizations with membership of the most likely EUNIS sediment texture class shown in Figure 11. Included in output file as column "EUNISp".

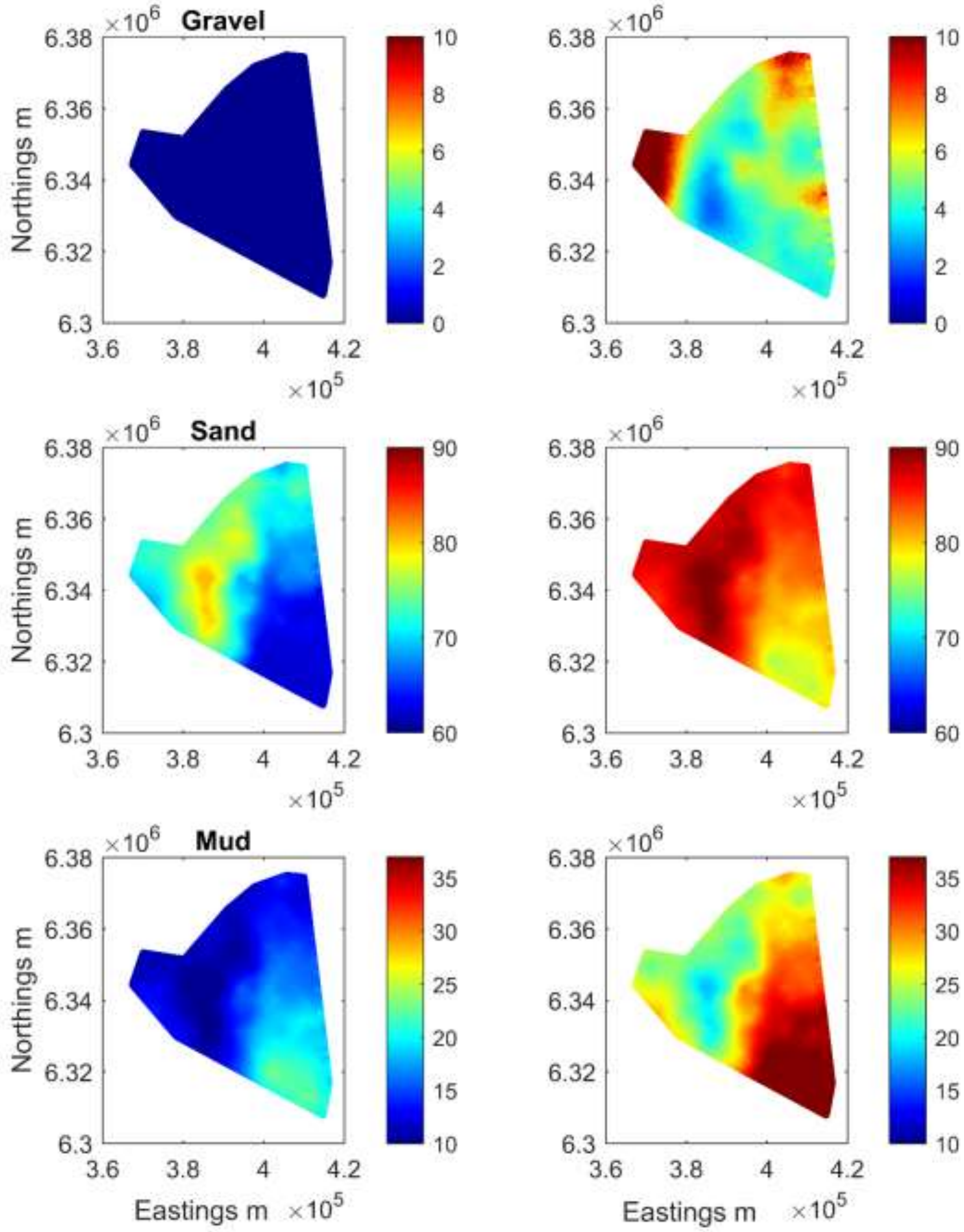

Figure 13 Lower (left) and upper (right) limits of the 95% confidence interval for predictions of gravel (top), sand (middle) and mud (bottom). Note that different colour-scales are applied for each variable. Data are included in the output file as "G025", "S025", "M025", "G975", "S975" and "M975".

References

Lark, R.M., D. Dove, S.L. Green, A.E. Richardson, H. Stewart, A. Stevenson. 2012. Spatial prediction of seabed sediment texture classes by cokriging from a legacy database of point observations, Sedimentary Geology, Volume 281, 15 December 2012, Pages 35-49, ISSN0037-0738.

Lark RM, Marchant, BP, Dove D, Green SL, Stewart H and Diesing M (2015) Combining observations with swath bathymetry and backscatter to map seabed sediment texture classes; the empirical best linear unbiased predictor. Sedimentary Geology, 328: 17-32.

Long, D. 2006. BGS detailed explanation of seabed sediment modified Folk classification. MESH (Mapping European Seabed Habitats) available at https://www.researchgate.net/profile/David_Long19/publication/284511408_BGS_detailed_explanation_of_seabed_sediment_modified_folk_classification/links/56545ab308aeafc2aabbc3c9/BGS-detailed-explanation-of-seabed-sediment-modified-folk-classification.pdf

Marchant BP (2018) Model-based geostatistics. In: McBratney, A.B., Minasny, B., Stockmann, U. (eds), Pedometrics: A system of quantitative soil information. Springer.

Pawlowsky-Glahn, Vera; Olea, Ricardo A. (2004). Geostatistical Analysis of Compositional Data, Oxford University Press.