Principal Component and Cluster Analysis for determining diversification of bottom morphology based on bathymetric profiles from Brepollen (Hornsund, Spitsbergen)*

Mateusz Moskalik[1]
Jarosław Tęgowski[2]
Piotr Grabowiecki[1]
Monika Żulichowska[3]

[1] Institute of Geophysics,
Polish Academy of Sciences,
Księcia Janusza 64, 01–452 Warsaw, Poland;

e-mail: mmosk@igf.edu.pl

e-mail: graba@igf.edu.pl

[2] Department of Oceanography and Geography,
University of Gdańsk,
al. Marszałka J. Piłsudskiego 46, 81–378 Gdynia, Poland;

e-mail: j.tegowski@ug.edu.pl

[3] Department of Biology and Earth Sciences,
Jagiellonian University,
Gronostajowa 7, 30–387 Kraków, Poland;

e-mail: m_zulik@interia.pl

Abstract

Navigation charts of the post-glacial regions of Arctic fjords tend not to cover regions from which glaciers have retreated. Whilst research vessels can make detailed bathymetric models using multibeam echosounders, they are often too large to enter such areas. To map these regions therefore requires smaller boats carrying single beam echosounders. To obtain morphology models of equivalent quality to those generated using multibeam echosounders, new ways of processing data from single beam echosounders have to be found. The results and comprehensive analysis of such measurements conducted in Brepollen (Hornsund, Spitsbergen) are presented in this article. The morphological differentiation of the seafloor was determined by calculating statistical, spectral and wavelet transformation, fractal and median filtration parameters of segments of bathymetric profiles. This set of parameters constituted the input for Principal Component Analysis and then in the form of Principal Components for the Cluster Analysis. As a result of this procedure, three morphological classes are proposed for Brepollen: (i) steep slopes (southern Brepollen), (ii) flat bottoms (central Brepollen) and gentle slopes (the Storebreen glacier valley and the southern part of the Hornbreen glacier valley), (iii) the morphologically most diverse region (the central Storebreen valley, the northern part of the Hornbreen glacier valley and the north-eastern part of central Brepollen).

1. Introduction

The widespread use of multi-beam echosounders in scientific research permits the collection of complex information in a short time. Much work has been done in recent years in the Spitsbergen region using this technology, which has delivered very detailed maps as well as information on the area's morphological characteristics (e.g. Ottesen & Dowdeswell 2006, 2009, Ottesen et al. 2007, 2008, Forwick et al. 2009, Dowdeswell et al. 2010). But such work requires the use of large vessels; this increases the costs of exploration and it also has its limitations. For reasons of safety, data recording is usually performed in areas already covered by marine publications and charts (e.g. The Norwegian Hydrographic Service and Norwegian Polar Research 1990, United Kingdom Hydrographic Office 2007, Statens Kartverk 2008). It is often the case, however, that existing maps do not show areas from which glaciers have retreated and are insufficiently detailed (Pastusiak 2010). Small boats with a shallow draught then have to be employed, as they provide a safer working environment when sailing in unexplored areas. In such difficult measuring conditions it is usually only single-beam echosounders that can be used. Direct interpolation of the profiles obtained enables geographical regionalisation in that individual bays, once influenced by glaciers, can be identified (Moskalik et al. 2013a) and their shapes characterised (Moskalik et al. 2013b). But again, these properties describe pre-glacial valleys in their entirety but not in fine detail.

In the present work, the bathymetric profiles were analysed under the assumption that areal diversity is expressed by the diversity of regional profiles. Moreover, the density of depth measurements being far greater than that of the inter-profile distances, additional information can be obtained on the nature of the bottom forms.

2. Study area; selecting profiles for analysis

Brepollen, the region where this research was carried out, is the inner part of the Hornsund Fjord, which itself is the most southerly fjord in
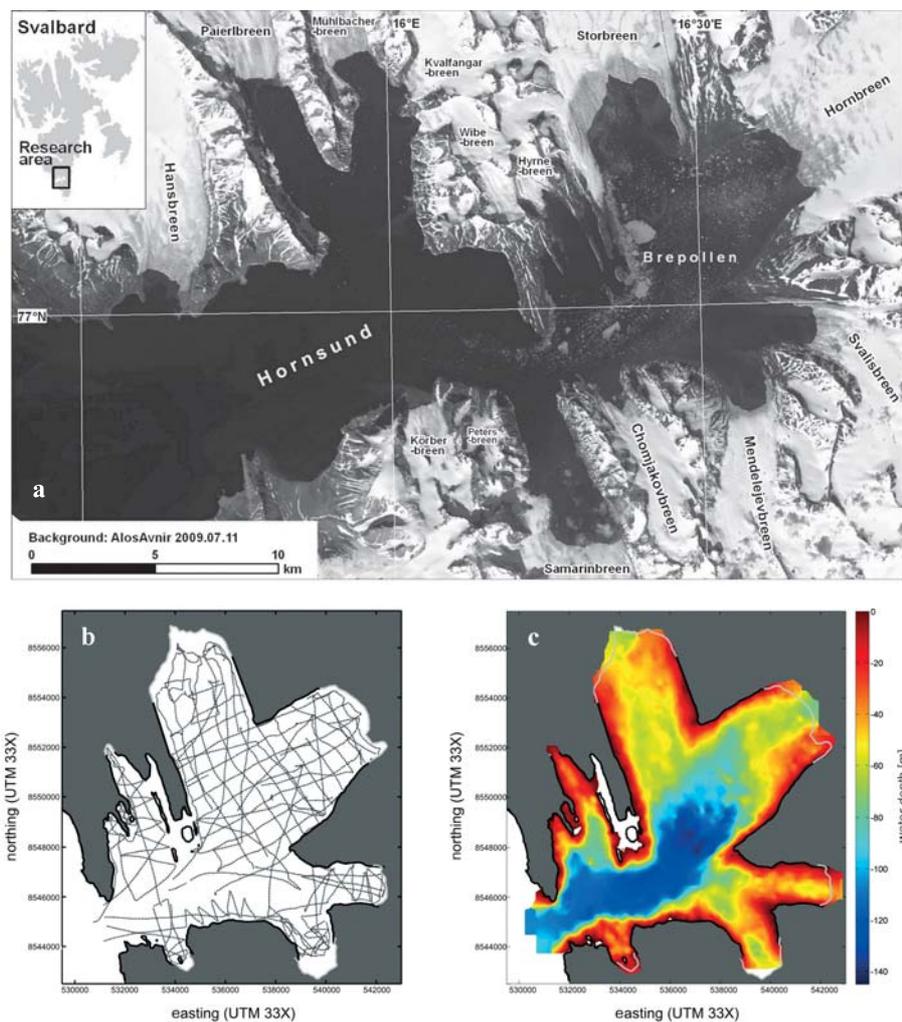

Figure 1. Map of Hornsund Fiord, Svalbard (a), locations of bathymetric profiles at Brepollen – grey dashed lines (b) and interpolated Brepollen bathymetry (c)

western Spitsbergen (Figure 1a). Bathymetric data were collected from a small boat equipped with a low-cost Lowrance LMS-527cDF echosounder during the summers of 2007 and 2008. A total of 120 bathymetric sections with an overall length of 384 km were made (Figure 1c). An interpolated bathymetry map for Brepollen (Figure 1b) was prepared on a 25 m grid (Moskalik et al. 2013a). It was assumed that it showed all forms larger than ten times the size of the grid; forms smaller than 250 m therefore required detailed analysis. A consequence of this methodology was that the bathymetric profiles used in the analysis had a minimum length of 256 m. In order to select sections for analysis, two classifying parameters were implemented. Every measurement on a bathymetric profile could become an Initial Profile Point (IPP) for the analysis on condition that there was an End Profile Point (EPP) on the profile 256 m distant along the measuring route. The first parameter was calculated by finding the average deviation of the records between IPP and EPP from a linear fit between them. The lower the value of this parameter, the closer the location of a depth measurement to the straight segment. The other parameter was the real distance between IPP and EPP; this was used if measurements were being made while sailing haphazardly in the vicinity of a specific point. It was assumed that when the average deviation from the linear fit was more than 2% of its length or the distance between IPP and EPP was less than 98% of its length, the profiles did not fulfil the straightness requirement.

3. Methodology

The following data analysis scheme was employed to characterise morphological seabed differences:

- calculation of mathematical parameters describing bathymetric section diversification;

- the first parameter reduction step, based on the analysis of entire bathymetric profiles; here, chaotic parameters were rejected; in the case of correlated parameters, only one remained;

- standardisation of parameters;

- the second reduction step was based on Principal Component Analysis (PCA);

- determination of the number of clusters;

- assignment of individual profile sections to clusters, based on cluster analysis;

- assignment of morphological feature classes to given clusters.

The paper describes all these steps in detail.

Statistical, spectral and wavelet transformations, as well as fractal and median filtration parameters were used in this work. These parameters were determined not for the depth profiles, but for the deviations from the mean value (MV), linear trend (LT) and square trend (ST) of all straight segments of profiles with a length of 256 m selected by the method described above (Figure 2).
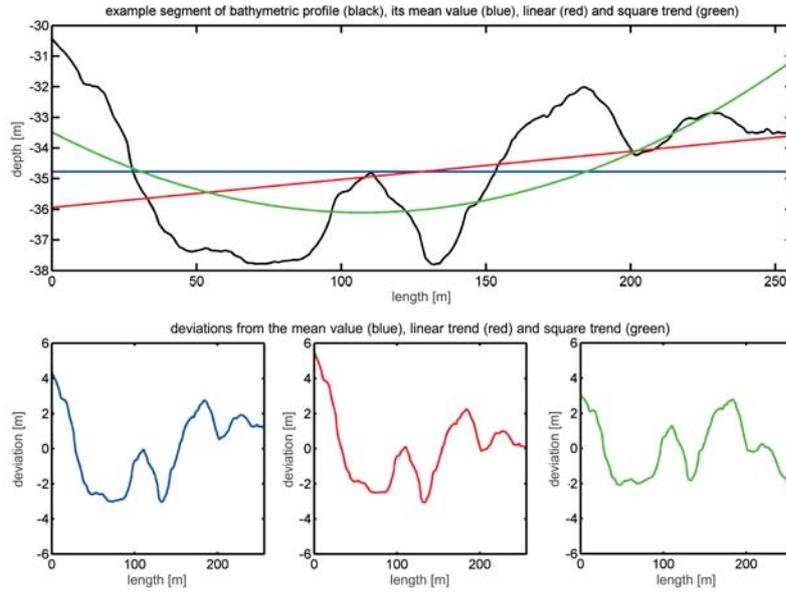

Figure 2. Example segment of bathymetric profile, its mean value, linear, square trend and deviations

The usefulness of statistical parameters for describing morphological diversification was shown in topographical analyses of a whole planet (Aharonson et al. 2001, Nikora & Goring 2004, 2005) but also of smaller regions (Moskalik & Bialik 2011).

The following statistical parameters were determined:

- the average absolute value of deviations (De^{MV}, De^{LT}, De^{ST});

- standard deviation of deviations (σ^{MV}, σ^{LT}, σ^{ST});

- skewness of deviations (Skew^{MV}, Skew^{LT}, SkewST);

- kurtosis of deviations (Kurt^{MV}, Kurt^{LT}, KurtST);

and parameters based on semivariograms of deviations:

- linear regressions (SLR^{MV}, SLR^{LT}, SLR^{ST});

– nugget of semivariogram linear regression (C_0^{MV}, C_0^{LT}, C_0^{ST}).

The range of interaction is the limit of increase in value of semivariograms (ω^{MV}, ω^{LT}, ω^{ST}), with its imposed limit of half of the length of the segments analysed.

The usefulness of spectral analysis for describing morphological features was also demonstrated for planet topography (Nikora & Goring 2006) and also for smaller regions like bathymetric maps (Lefebvre & Lyons 2011) and linear profiles (Goff et al. 1999, Goff 2000, Tęgowski & Łubniewski 2002). The following parameters were determined for the bathymetric profiles collected at Brepollen:

– the total spectral energies in the form of integrals of power spectral density deviations from the bathymetric profile (S_{k1}^{MV}, S_{k1}^{LT}, S_{k1}^{ST}):

$$S_{k_1} = \int\limits_0^{k_{Ny}} C_k dk \,, \tag{1}$$

where k_{Ny} is the Nyquist parameter and C_k is the normalised power spectrum given by Pace & Geo (1988) in the form:

$$C_k = \frac{\log_{10}(10^5 \ S(k) \ S_{\max}(k)^{-1} + 1)}{\log_{10}(10^5 + 1)} \,, \tag{2}$$

where $S(k)$ is the power spectral density of the bathymetric profile;

– relations of the spectral energy to the total spectral energy for each of the deviations take the form:

$$S_{km} = \frac{1}{S_{k_1}} \int\limits_0^{\frac{1}{m}k_{Ny}} C_k dk \tag{3}$$

determined for $m = 2$, 4, 8, 16 (S_{km}^{MV}, S_{km}^{LT}, S_{km}^{ST});

– the eight first spectral moments (M_r^{MV}, M_r^{LT}, M_r^{ST}) of order $r = 0$, ..., 7 defined as

$$M_r = \int\limits_0^\infty k^r \ S(k) dk \,; \tag{4}$$

– average values of wave numbers (k^{MV}, k^{LT}, k^{ST}):

$$k = \frac{M_1}{M_0} \,; \tag{5}$$

– spectral widths ($v^{2\,\text{MV}}$, $v^{2\,\text{LT}}$, $v^{2\,\text{ST}}$, $\epsilon^{2\,\text{MV}}$, $\epsilon^{2\,\text{LT}}$, $\epsilon^{2\,\text{ST}}$) describing the concentration of power spectra around the average wave numbers

$$v^2 = \frac{M_0 \times M_2}{M_1} - 1 \qquad\qquad (6)$$

and

$$\epsilon^2 = \frac{M_0 \times M_4 - M_2^2}{M_0 \times M_4}\,; \qquad\qquad (7)$$

– spectral skewness describing the shape of spectra (γ^{MV}, γ^{LT}, γ^{ST})

$$\gamma = \frac{M_3}{M_2^{3/2}}\,. \qquad\qquad (8)$$

Additional analysis involved the use of wavelet transforms, also used in the analysis of bathymetric measurements (Little et al. 1993, Little 1994, Little & Smith 1996, Wilson et al. 2007). A fundamental problem in wavelet analysis is the selection of the mother wavelet function. For analysing the echo envelope of the acoustic signal, Ostrovsky & Tęgowski (2010) applied six differently defined mother functions. The use of so many different functions did not yield a larger amount of information, however. In the present case, the number of wavelet mother functions was reduced to two: one symmetric and the other asymmetric. The Mexican Hat (mexh) was selected as the symmetric wavelet mother function, while the family of Daubechies wavelets exemplifies the asymmetric mother functions. A wavelet of the order of 7 (db7) was selected from this family. In order to account for wavelet asymmetry, profiles were analysed in both directions, in the same direction as the measurements according to (db7+) and in the opposite direction (db7−). The following parameters were determined for each of the transforms:

– wavelet energies for a given scaling parameter ($E_{j,\,\text{wav}}^{\text{MV}}$, $E_{j,\,\text{wav}}^{\text{LT}}$, $E_{j,\,\text{wav}}^{\text{ST}}$):

$$E_{j,\,\text{wav}} = \int\limits_0^{b_{\max}} C^2(a,\,b)\,db\,, \qquad\qquad (9)$$

where

$$C_{a,\,b} = \int f(x)\Psi(a,\,b,\,x)\,dx \qquad\qquad (10)$$

is the wavelet transform of the bathymetric profile of $f(x)$ and

$$\Psi(a,\,b,\,x) = \frac{1}{\sqrt{a}} \times \Psi\left(\frac{x-b}{a}\right), \qquad a,b \in R, \quad a \neq 0, \qquad (11)$$

where a is the scaling parameter corresponding to the stretching or compression of the mother function, and b is the parameter specifying the wavelet location on the profile. Calculations were performed for wavelet scales $a = 2^j$ for $j = 1, \ldots, 7$. Larger values of parameter a could not be used, because for $j = 7$, the wavelet size is half the length of the bathymetric profile under consideration;

– the entropy of the bathymetric profiles ($h_{\mathrm{wav}}^{\mathrm{MV}}$, $h_{\mathrm{wav}}^{\mathrm{LT}}$, $h_{\mathrm{wav}}^{\mathrm{ST}}$), defined as

$$h_{\mathrm{wav}} = \sum_{j=1}^{j=7} E_{j,\,\mathrm{wav}} \ \times \ \ln(E_{j,\,\mathrm{wav}}). \tag{12}$$

The use of a fractal dimension in the analysis of bottom bathymetry should result from the following assumptions (Herzfeld et al. 1995):

– bathymetry has a non-trivial structure at every scale;

– it cannot be described by simple geometric figures;

– its topological dimension D_T is smaller than the Hausdorff dimension D_H defined as:

$$D_H = \lim_{r \to 0} \ \frac{-\log_{10} N(r)}{\log_{10} r}\,, \tag{13}$$

where $N(r)$ denotes the number of spheres of radius r needed to completely cover the object;

– it is self-similar in the stochastic sense.

It is evident that the bathymetry of a water body formed by numerous geological processes has a non-trivial structure and that it cannot be described by simple geometric figures. The work involving the analysis of bathymetric profiles from the eastern Pacific (Herzfeld et al. 1995) indicates that bathymetry can be treated as a fractal because the assumption that $D_H > D_T$ is fulfilled; however, the assumptions of self-similarity are not satisfied when the image scale is being changed. The fractal dimension is considered to be an appropriate parameter for describing the morphological diversification of bottom surfaces (Wilson et al. 2007). In the case of a flat bottom, the fractal dimension calculated for the bathymetric profile should take a value equal to unity; as irregularities in the seafloor appear and their magnitudes change, its value will rise. In this work, the fractal dimension was determined using indirect methods, such as the box dimension, semivariogram analysis of spectral parameters and wavelet analysis.

For determining the box fractal dimension of the deviations from the bathymetric profile segments ($D_{\text{box}}^{\text{MV}}$, $D_{\text{box}}^{\text{LT}}$, $D_{\text{box}}^{\text{ST}}$), the definition given by Hastings & Sugihara (1994) was used:

$$D_{\text{box}} = \lim_{\Delta s \to 0} \frac{\log_{10} N(\Delta s)}{-\log_{10} \Delta s} \,, \tag{14}$$

where $N(\Delta s)$ determines the number of squares covering a depth profile of a side length Δs. In case of the bathymetric profiles, both the length and depth have the same dimension.

The proposed procedure for determining this parameter consists of four consecutive steps:

- normalisation of the distance, taking the unit profile length to be 256 m;

- normalisation of the depth, considering independently the unit maximum difference in deviation of the depth on all analysed profile segments for each type of deviations;

- determination of the number of squares covering the bathymetric profile for the division of values normalised from 11 to 110 segments;

- determination of the slope coefficient of the curve defined by the dependence of $\log_{10}(N(\Delta s))$ on $-\log_{10}(\Delta s)$, which is equivalent to the box fractal dimension.

Application of a uniform standardisation is valid, taking a standard distance and depth of 256 m, equal to the length of the analysed section. In such analyses, however, the depth differences were often too small in comparison with the length of the segment profile to use the same scale. For this reason, the maximum difference in depth of all segments was used as the depth normalisation.

The other methods used for determining the fractal dimension of bathymetric profile deviations from the mean, linear and quadratic trend were the analyses of (i) the semivariogram ($D_{\text{sem}}^{\text{MV}}$, $D_{\text{sem}}^{\text{LT}}$, $D_{\text{sem}}^{\text{ST}}$), (ii) the power spectral density (D_{FFT}^{MV}, D_{FFT}^{LT}, D_{FFT}^{ST}) and (iii) the wavelet transform ($D_{\text{wav}}^{\text{MV}}$, $D_{\text{wav}}^{\text{LT}}$, $D_{\text{wav}}^{\text{ST}}$). The following relationships can be derived from them:

$$D_{\text{sem}} = 2 - \frac{\alpha}{2} \,, \tag{15}$$

where α is the semivariogram regression coefficient in the log-log scale (Wen & Sinding-Larsen 1997);

$$D_{FFT} = \frac{5 - \beta}{2} \,, \tag{16}$$

where β is the regression coefficient of the spectral density in the log-log scale (Mandelbrot 1982, Wornell & Oppenheim 1992);

$$D_{\mathrm{wav}} = \frac{3}{2} - \gamma\,, \tag{17}$$

where γ is the regression coefficient of the wavelet transform coefficient $C(a, b)$ averaged over the parameter b determining the location depending on the scaling parameter a in the log-log scale (Mandelbrot 1982).

A median filter was also used to analyse the diversity of bottom forms. Operation of the filter resulted in replacement of all the values by the median of the nearest values to each of them (White 2003, White & Hodges 2005). This filter is used to separate different sizes of morphological forms (e.g. Wessel 1998, Adam et al. 2005, Kim 2005, Hiller & Smith 2008, Kim & Wessel 2008). A window of width $2d$ with d increasing in geometric progression was used in the study: $d = 2$ ($\mathrm{MF}_1^{\mathrm{MV}}$, $\mathrm{MF}_1^{\mathrm{LT}}$, $\mathrm{MF}_1^{\mathrm{ST}}$), 4 ($\mathrm{MF}_2^{\mathrm{MV}}$, $\mathrm{MF}_2^{\mathrm{LT}}$, $\mathrm{MF}_2^{\mathrm{ST}}$), 8 ($\mathrm{MF}_3^{\mathrm{MV}}$, $\mathrm{MF}_3^{\mathrm{LT}}$, $\mathrm{MF}_3^{\mathrm{ST}}$), 16 ($\mathrm{MF}_4^{\mathrm{MV}}$, $\mathrm{MF}_4^{\mathrm{LT}}$, $\mathrm{MF}_4^{\mathrm{ST}}$), 32 ($\mathrm{MF}_5^{\mathrm{MV}}$, $\mathrm{MF}_5^{\mathrm{LT}}$, $\mathrm{MF}_5^{\mathrm{ST}}$) and 64 ($\mathrm{MF}_6^{\mathrm{MV}}$, $\mathrm{MF}_6^{\mathrm{LT}}$, $\mathrm{MF}_6^{\mathrm{ST}}$) metres. The next filter, which cuts the size forms up to 128 m, could not be applied to a 256 m long profile segment. This parameter was determined by averaging the absolute values of the residue after filtering.

All the parameters defined above were identified for every profile. Some of them were correlated or their shape was chaotic, providing no information that could define the seabed morphological diversity.

The discussion includes all the parameters used, based on an example bathymetric profile. This profile is characterised by including varied morphology (Figure 3b). The profile's depth varies within the range of 10–120 m. The maximum depth of 120 m was found in the central part of Brepollen, and the profile end is positioned close to the Hyrne glacier calving front.
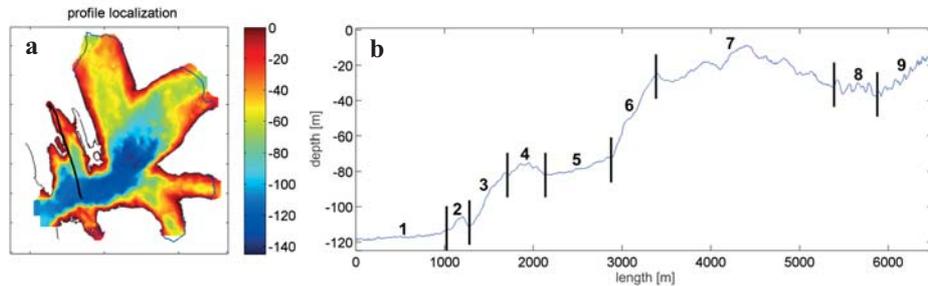

Figure 3. Example profile with sections (b) and its localisation (black line) on the Brepollen bathymetry map (a)

The following profile sections were identified:

– Section 1 – an almost flat seabed 1 km long with depths between 115–120 m.

– Section 2 – a hill at the bottom of a 200 m wide slope more than 10 m high.

– Section 3 – an average slope of 4° to 5° with a slight roughness in its upper parts.

– Section 4 – a hill 10 m high and 200 m wide in the upper part of the slope. The isolation of this section was the result of changes in roughness on the slope. It can be assumed that this is a continuation of section 3; the other slope of the hill is inclined towards section 5.

– Section 5 – a gently inclined sea bottom about 700 m long and descending to a depth of 10 m.

– Section 6 – a hill with a slope steeper than 5°.

– Section 7 – a singular convex form 2 km long with sharp elevations and characteristically increasing of roughness in the direction of the Hyrne glacier.

– Section 8 – part of the profile with forms 50–100 m wide and 10 m high.

– Section 9 – a hill before the glacier front with visibly smaller forms than recorded in section 8.

Analysis of the statistical parameters for the example bathymetric profile indicates that its diversity is reflected by the variability in parameters De, σ, SLR for every type of deviation and C_0^{MV}. Analysis of the other parameters does not reflect this diversity, however: the variations are mostly chaotic. There was a significant correlation of σ with De for every deviation (MV, LT and ST) independently (Figures 4a, 4b, 4c). The slope of the curve is almost the same in every case. The range of values of parameters σ^{LT} and σ^{ST} or De^{LT} and De^{ST} might suggest the erroneous conclusion that they too are correlated, but the evidence for the non-dependence of these parameters is the quantitative distribution of all possible pairs of σ^{LT} and σ^{ST} (Figure 4d). Pairs of these parameters lie within almost the whole area below the linear relation describing the equivalence of σ^{ST} and σ^{LT}. A similar analysis was conducted for the relationship between σ and SLR (Figures 4e, 4f, 4g): this is exponential. A negative linear relationship was also found between C_0^{MV} and SLR^{MV} (Figure 4h). The unequivocal inference from the foregoing is that for every deviation only one of these parameters contributes clearly independent information on the morphological diversity of the seabed.
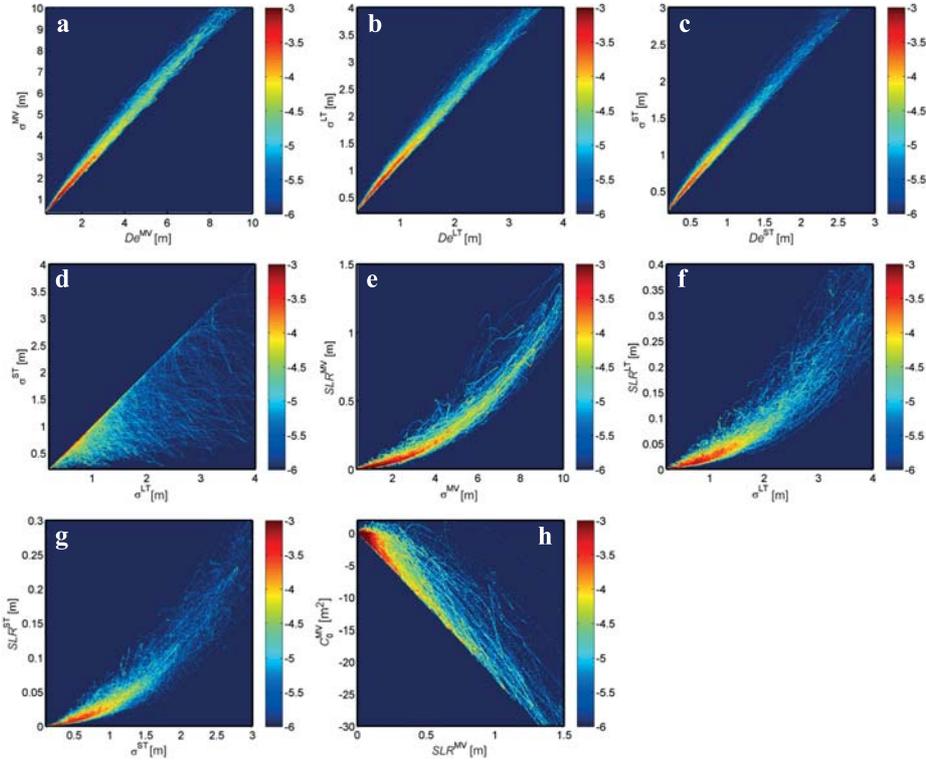
Figure 4. The relationship between: σ and De for MV (a), LT (b), ST deviations (c); SLR and σ for MV (e), LT (f), ST deviations (g); σ^{LT} and σ^{ST} (d); C_0^{MV} and SLR^{MV} (h). The colour scale represents the ratio of the number of pairs of these relations to all pairs on the logarithmic scale

Spectral moments (M_i) and spectral skewness (γ) were found to be the most significant spectral parameters. The higher the order of a spectral moment, the lower the difference between the values. These features are highlighted by the correlation coefficients for M_i and M_j pairs for each deviation (Figure 5). There is also a correlation between the spectral moments for LT and ST (Figure 5); the coefficient of this correlation, of the 2nd order, is close to 1. In view of the above, it was decided that only 0 to 3rd order spectral moments would be used for every deviation.

The similarities between σ and M_0 were also investigated. Detailed analysis showed that for every type of deviation there exists a linear dependence between σ^2 and M_0. It is clear from the above relationship that when spectral analysis was used, the addition of characteristics emerging from statistical parameters did not contribute any new knowledge regarding the sea bottom morphology in Brepollen.
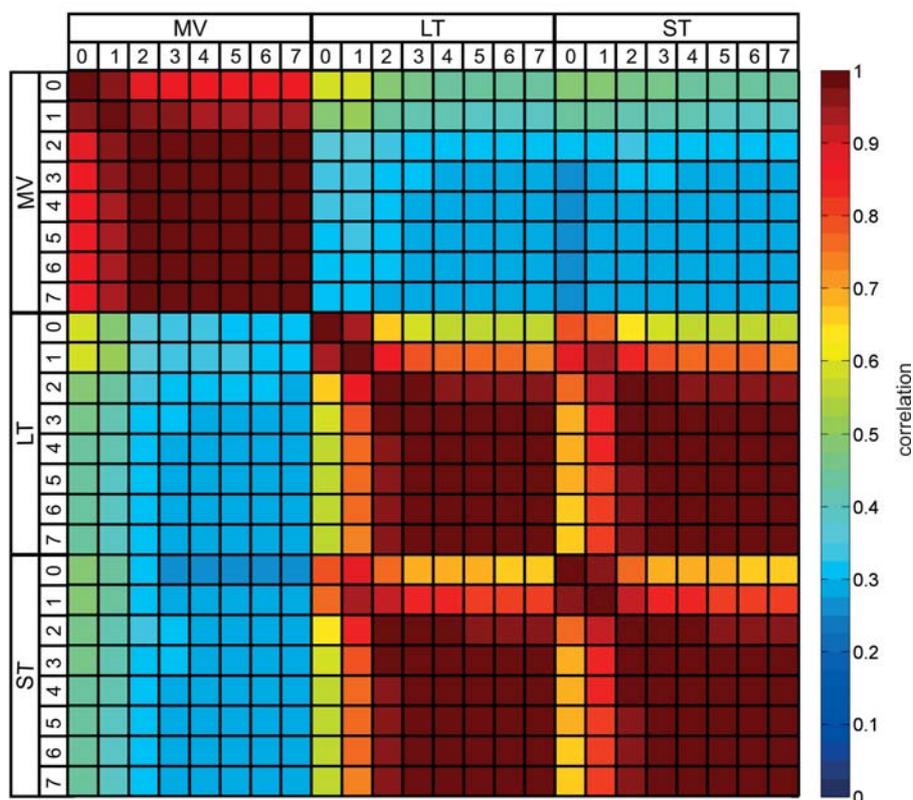
Figure 5. Correlation between M_i and M_j spectral moments of orders 0 to 7 for all types of deviations (MV, LT, ST)

Determination of the wavelet energy for successive scaling parameters is an excellent method for isolating morphological forms on a bathymetric profile, as it takes the magnitude of forms into consideration on the basis of the scaling parameter's size. To verify the applicability of wavelet energies, the correlations between them were calculated (Figure 6). The correlation for every type of wavelet was much less than 1, even in the case of adjacent scaling parameters for the same type of wave. Analysis of the wavelet energies calculated for the example profile showed that the wavelet energy determined using the mexh wavelet for the scaling parameter $a = 2^i$ resembled that of the db7 wavelet for the scaling parameter $a = 2^{i+2}$ when $i = 1, \ldots, 5$. This observation was confirmed by wavelet correlation analysis (Figure 6).

The final point in the discussion of the application of wavelets to bathymetric profile analysis is the possible use of asymmetric wavelets, such as db7. The most effective approach seems to be to investigate the
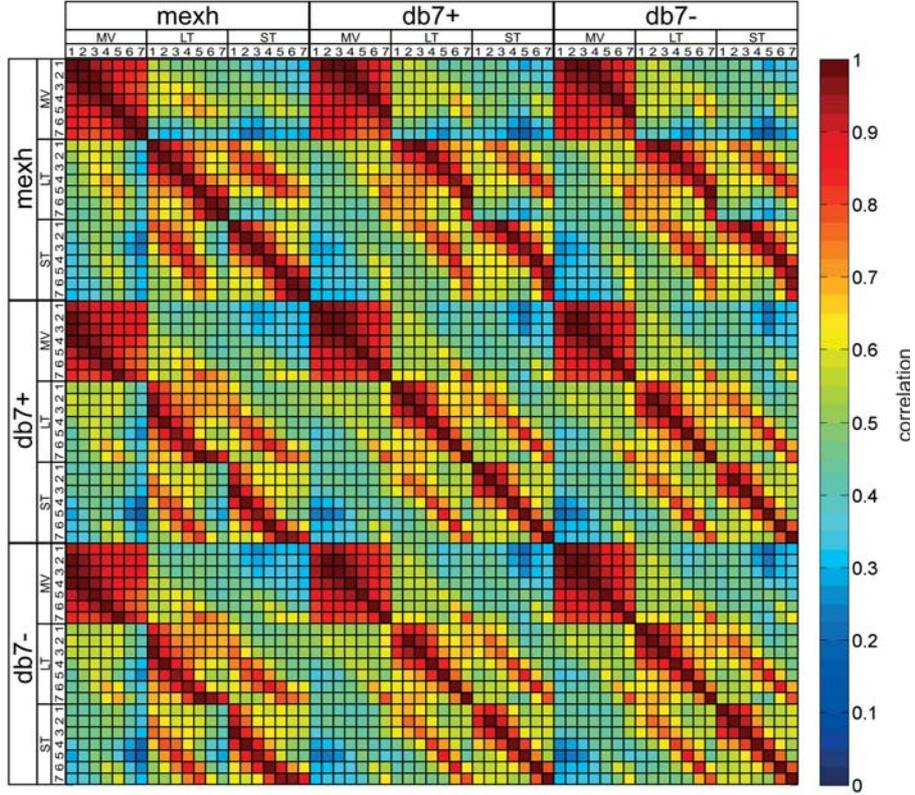
Figure 6. Correlation between wavelet energies for all wavelet mother functions (mexh, db7+, db7-), scaling parameters (2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7) and types of deviations (MV, LT, ST) used

energy correlation between wavelet energies for the same scaling parameter of asymmetric wavelets calculated in both directions of a profile. Only the $(E_{1,\,\mathrm{db7+}}^{(\mathrm{MV,LT,ST})}, E_{1,\,\mathrm{db7-}}^{(\mathrm{MV,LT,ST})})$ correlation was less than 0.9 (Figure 6); in the other cases it was close to 1. It was shown that only the first two energies calculated for db7 wavelets yielded suitable results, because for higher scaling parameters they were correlated with wavelet energies calculated from mexh. It was decided to add three additional parameters, besides the energies for db7, defined as:

$$E_{i,\,\mathrm{db7}\pm} = \frac{E_{i,\,\mathrm{db7+}} + E_{i,\,\mathrm{db7-}}}{2} \quad \text{for} \quad i = 1,\,2$$

$$E_{1,\,|\mathrm{db7}|} = |E_{1,\,\mathrm{db7+}} - E_{1,\,\mathrm{db7-}}|$$

$$(18)$$

for every deviation type MV, LT and ST.

For the fractal dimension, the quality of the results obtained using semivariograms, spectral and wavelet analyses was insufficient. Box size counts were found to be the most efficient methods. The application of a median filter to bathymetric profile segments was also a good way of finding diverse forms on the example profile (Figure 7).

The above analyses demonstrate that to describe the diverse morphology of Brepollen the following parameters have to be taken into account: M_0, M_1, M_2, M_3, γ, $E_{1,\text{mexh}}$, $E_{2,\text{mexh}}$, $E_{3,\text{mexh}}$, $E_{4,\text{mexh}}$, $E_{5,\text{mexh}}$, $E_{6,\text{mexh}}$, $E_{7,\text{mexh}}$, $E_{1,\text{db7}\pm}$, $E_{2,\text{db7}\pm}$, $E_{1,|\text{db7}|}$, D_{box}, MF_1, MF_2, MF_3, MF_4, MF_5, MF_6. As these parameters could still be independent, the input parameters were reduced by Principal Component Analysis (PCA).

Before embarking on PCA, the distributions of the values of each parameter were analysed. Two types of calculated values were identified: (i) with data where quantity is encompassed within one order of magnitude (γ, D_{box}, MF_1, MF_2, MF_3, MF_4, MF_5, MF_6) and (ii) with data whose values range over several orders of magnitude (M_0, M_1, M_2, M_3, $E_{1,\text{mexh}}$, $E_{2,\text{mexh}}$, $E_{3,\text{mexh}}$, $E_{4,\text{mexh}}$, $E_{5,\text{mexh}}$, $E_{6,\text{mexh}}$, $E_{7,\text{mexh}}$, $E_{1,\text{db7}\pm}$, $E_{2,\text{db7}\pm}$, $E_{1,|\text{db7}|}$). For the second case the common logarithm was determined. The next step included data normalisation:

$$x_m = \frac{x - x_{sr}}{\sigma_x}, \tag{19}$$

where x_m – new parameter value, x – its determined value, x_{sr} – mean value of determined parameters, σ_x – standard deviation of determined parameters.

After such parameter transformation, the mean of each one will be equal to zero and the standard deviation equal to one.

Analysis of the variance of Principal Components (PCs) (Figure 8) showed their diminishing influence on the overall value. For the independent analysis of every deviation, the first ten PCs are sufficient for cluster analysis. Together, these correspond to more than 98% of the cumulative variance. In the analysis of deviation MV, this value was exceeded by the first nine PCs, but despite this, it was decided to use the same number as in the other two cases. When all the parameters were included, 98% of the cumulative variance was exceeded for the first 16 PCs, and this number of parameters was utilised in the cluster analysis.

Cluster analysis, a process for combining series of points into groups, enables common features to be assigned to points on the bathymetric profile; every group represents one feature. The k-means method was used to perform the analysis. The algorithm gathers the cluster points in such
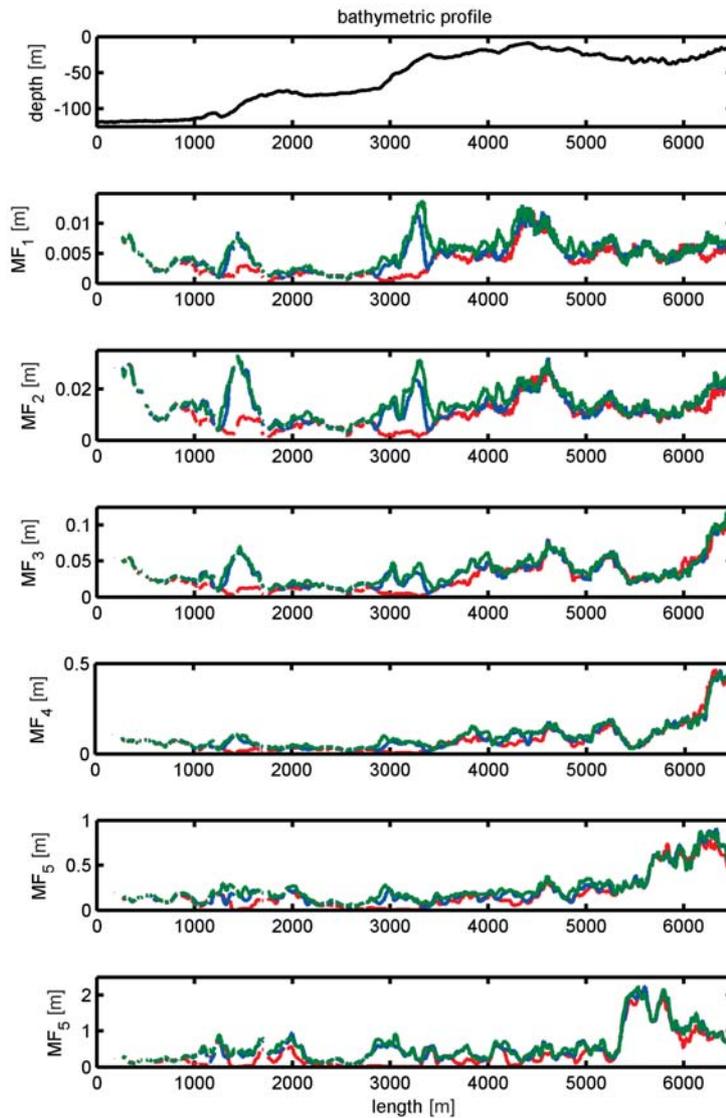
Figure 7. Profile analysis calculated from the mean remainder from median filtration, increasing in geometric progression for isolated forms smaller than $d = 2^1(\mathrm{MF}_1)$, $2^2(\mathrm{MF}_2)$, $2^3(\mathrm{MF}_3)$, $2^4(\mathrm{MF}_4)$, $2^5(\mathrm{MF}_5)$, $2^6(\mathrm{MF}_6)$ metres, for deviations from the mean value (red), linear (blue) and square trend (green)

a way that the cumulative distance between the points and the cluster midpoint, where they are located, is minimal, but that the distance between clusters is a maximum. The square of the Euclidean distance was used as a measure of distance.
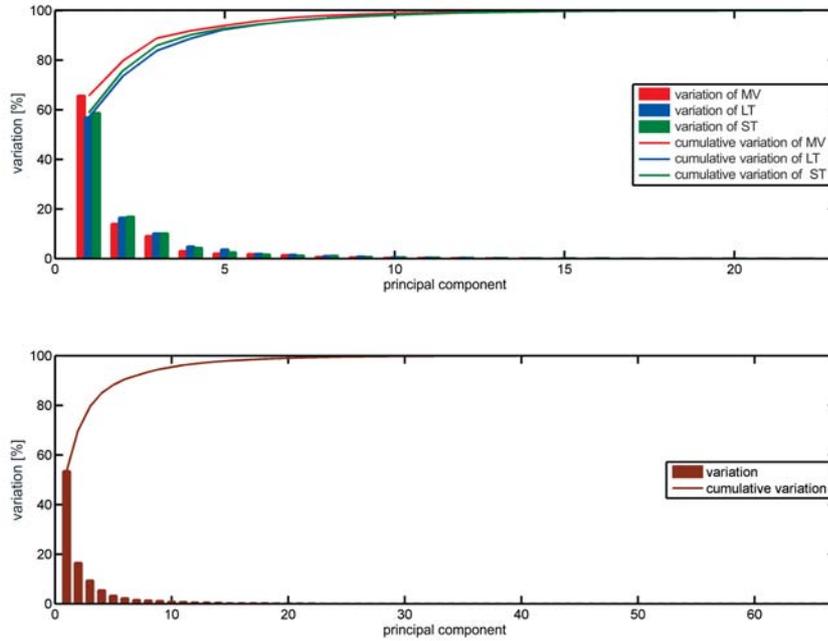
Figure 8. Variation of Principal Components obtained separately for parameters analysed for deviations from the mean value (MV), linear (LT) and square trend (ST) on segments of bathymetric profiles and their cumulative value (upper plot) and for all parameters as one data set (lower plot)

The choice of the number of clusters is a tricky problem. The most convenient situation is when there are environmental pointers to the number of features investigated, as this will then be equal to the number of clusters formed. If such information is unavailable, one can employ automated methods. Of 30 methods of cluster number choice analysed by Milligan & Cooper (1985), the method of Caliński & Harabasz (1974) was identified as one of the most reliable for determining the maximum of the Caliński-Harabasz index CH_{index}. It was defined as

$$CH_{\text{index}} = \frac{B}{K-1} \times \frac{N-K}{W}\,, \tag{20}$$

where N – number of all points, K – number of clusters, B – distance between clusters and W – the distance within clusters.

The magnitudes of B and W are obtained as follows:

$$B = \sum_{k=1}^{K} n_k \|z_k - z\|^2 \,,$$

$$(21)$$

$$W = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \|x_{i \in k} - z_k\|^2 \,,$$

where n_k – number of points in cluster k, z_k – position of the centre of cluster k, z – position of the centre of all points, $x_{i \in k}$ – the i-th point located in cluster k, and $\| \|$ is the distance norm (Maulik & Bandyopadhyay 2002).

Ray & Turi (1999) derived another method of determining cluster numbers. Their index makes direct use of the cluster assumption choice and is defined as follows:

$$II_{\text{index}} = \frac{\text{intra}}{\text{inter}} = \frac{N^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \|x_{i \in k} - z_k\|^2}{\min\left(\|z_i - z_j\|^2\right)} \,,$$
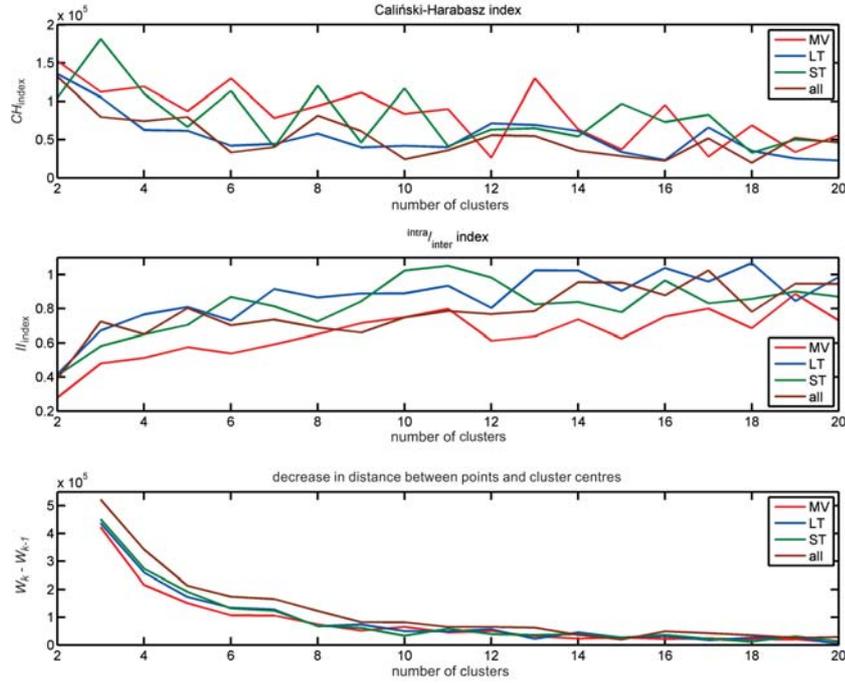
$$(22)$$

Figure 9. Indices for obtaining the number of clusters

where 'intra' is the mean distance between the points and the centre of the cluster containing them, while 'inter' is the minimum distance between the clusters.

In these cases the number of clusters involves finding the maximum of CH_{index} or minimum of II_{index}.

Both indices were determined for numbers of clusters from 2 to 20 in all the cases analysed (Figure 9). In general CH_{index} decreases and II_{index} increases with increasing numbers of clusters. Despite the many deviations from the above trend for both indices it was difficult to define the cluster number. A small number of clusters was found to be the most appropriate.

To identify the maximum number of clusters, the total distance between the points and each cluster centre (where they are located) was defined:

$$W_K = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \|x_{i\in k} - z_k\|^2 \,. \tag{23}$$

By analysing the $W_K - W_{K-1}$ dependence (Figure 9), on the assumption that the value must not be too high, 6 was chosen as the most appropriate value.

4. Results of clustering and discussion

Cluster analysis was performed for two to six clusters for deviation types MV, LT, ST separately and for all the types. In order to assign a specific cluster to a seabed morphological type, the results for the example profile were analysed first. To distinguish individual clusters a marker $C_{n\_s}^{(\text{deviation type})}$ was introduced, where the possible deviation types are MV, LT, ST or all for the whole set of parameters, n denotes the number of clusters and s is the cluster number in a specific set. For example, $C_{4\_3}^{\text{MV}}$ signifies the third of four clusters for the deviation from the mean depth.

For deviation type MV (Figure 10a) the most characteristic differentiation is related to the slope of a hill. Clusters $C_{2\_1}^{\text{MV}}$, $C_{3\_1}^{\text{MV}}$, $C_{4\_4}^{\text{MV}}$, $C_{5\_5}^{\text{MV}}$, $C_{6\_1}^{\text{MV}}$ correspond to the steepest slopes, while $C_{2\_2}^{\text{MV}}$ corresponds to gentle slopes and flat areas. For three clusters the steepness of a hillside decreases in the sequence $C_{3\_1}^{\text{MV}}$–$C_{3\_2}^{\text{MV}}$–$C_{3\_3}^{\text{MV}}$. For a larger number of clusters, however, it is hard to state whether the differentiation continues to indicate variations in the global slope or whether it indicates more diverse sea bottoms. No direct interpretation of a seabed was obtained for the clusters calculated for deviation types LT and ST (Figures 10b,c).

The differentiation distribution of the example profile was the most complete when all the parameters were taken into account (Figure 10d). For two clusters the distribution was almost analogous to that of MV, that is, flat
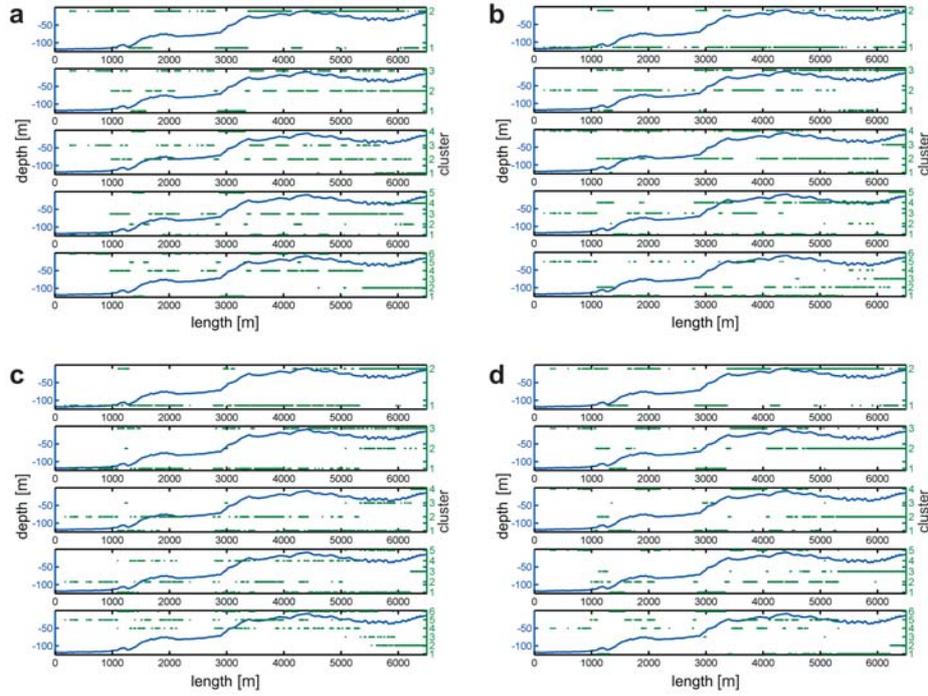
Figure 10. Division of an example bathymetric profile into 2, 3, 4, 5 and 6 clusters based on all profiles for MV (a), LT (b), ST (c) and all types of deviations (d)

or slightly inclined surfaces ($C_{2\_2}^{\mathrm{all}}$) and slopes ($C_{2\_1}^{\mathrm{all}}$). Where three clusters were determined, steep slopes ($C_{3\_1}^{\mathrm{all}}$), a flat seabed, gently sloping hillsides with small morphological forms ($C_{3\_3}^{\mathrm{all}}$) and strongly undulating sections ($C_{3\_2}^{\mathrm{all}}$) were distinguished. Adding a fourth cluster precluded further profile classification. The greatest sea bottom diversity on this profile was found with five clusters. It was classified as follows: (i) a flat seabed ($C_{5\_5}^{\mathrm{all}}$), (ii) sections with gently inclined slopes and small forms ($C_{5\_2}^{\mathrm{all}}$), (iii) areas with diverse morphology and numerous bottom forms ($C_{5\_3}^{\mathrm{all}}$) and (iv) steep slopes ($C_{5\_1}^{\mathrm{all}}$). No forms associated with cluster $C_{5\_4}^{\mathrm{all}}$ were found. With six clusters the results were very difficult to interpret; increasing the number of clusters did not improve the results any further.

In order to draw a map with the morphological form classification on the example profile, it was suggested that a new interpolation procedure should be used. Since the results were quantified, the percentage of all clusters was identified at a distance of 500 m from every location. This was dictated by the distance used for the Brepollen interpolation, as this allows information from the whole research area to be used (Moskalik et al. 2013a). The maximum value cluster was used as the morphological differentiation

class corresponding to the sea bottom. Maps of seabed diversity from the 2nd to the 5th class from the cluster analysis of all parameters were prepared (Figure 11).
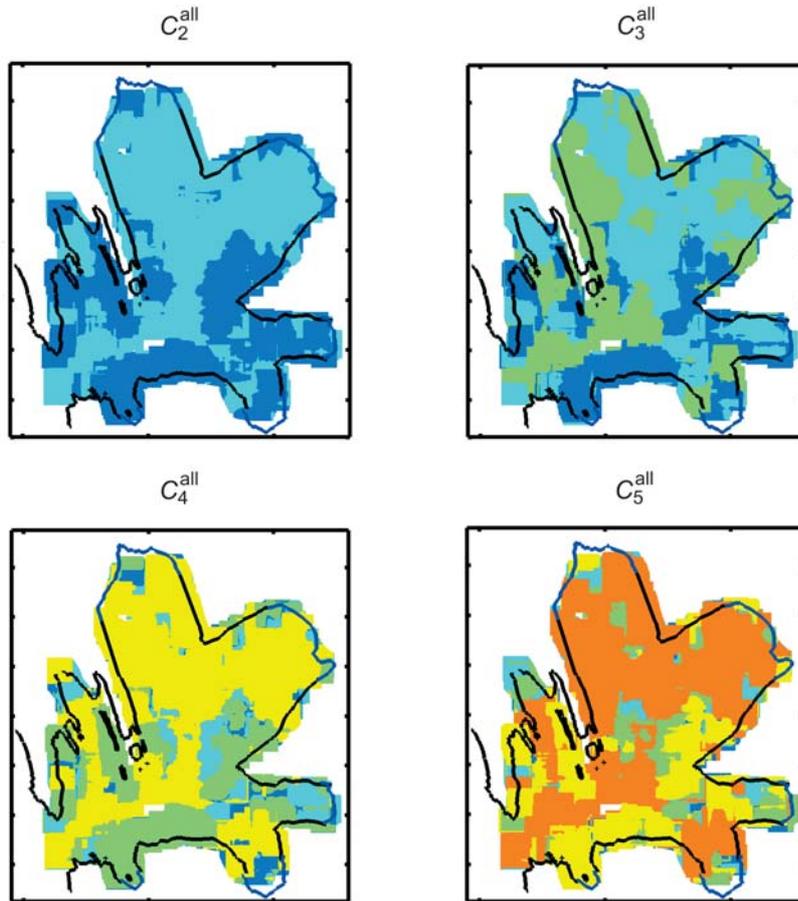

Figure 11. Brepollen region divisions into 2, 3, 4 and 5 morphological differentiation classes, based on cluster analyses of all parameters where the colours are defined as follows: blue – $C^{\text{all}}_{(2,3,4,5)\_1}$, light blue – $C^{\text{all}}_{(2,3,4,5)\_2}$, light green – $C^{\text{all}}_{(3,4,5)\_3}$, yellow – $C^{\text{all}}_{(4,5)\_4}$, orange – $C^{\text{all}}_{5\_5}$

Analysis of the results revealed a rapid increase in information for three clusters than for two. In comparison with the example profile, the results allow one to identify areas, such as: (i) steeply inclined areas ($C^{\text{all}}_{3\_1}$), (ii) almost flat and gently inclined areas ($C^{\text{all}}_{3\_3}$) and (iii) areas characterised by a diverse bottom morphology ($C^{\text{all}}_{3\_2}$). Class $C^{\text{all}}_{3\_1}$ areas are regarded as post-glacial valleys, located in the south-central part of Brepollen. They

are characteristic of the area between central Brepollen and the Hornbreen glacier valley. There are ridges running NE-SW visible on the bathymetric map (Figure 1c). Class $C_{3\_2}^{\mathrm{all}}$ regions are mainly: (i) the Storbreen glacier valley bottom, right down to its extension in central Brepollen, (ii) the northern part of the Hornbreen glacier valley, (iii) the outer part of the Mendelejevbreen glacier valley, (iv) the Svalisbreen valley slopes (v) and the Hyrnebreen glacier front. The final class $C_{3\_3}^{\mathrm{all}}$ is located in (i) the central part of Brepollen, (ii) on the Storebreen glacier valley slopes, (iii) in front of the SE part of the Hornbreen glacier and (iv) in the centre of the Mendelejevbreen glacier valley. The classes in the Mendelejevbreen glacier valley defined the location of the glacier front after its charge in the year 2000 (Głowacki & Jania 2008, Błaszczyk et al. 2009, 2013).

The quality of the information on seabed differentiation obtained from the identification of clusters 4 and 5 was poorer. The central Brepollen bottom and the Store and Horn glacier valleys were assigned to a single class, as when two clusters were determined (Figure 11). These classes highlighted a distinct depression right by the Store glacier front (Figure 1c), at the point where a river flows out from under the glacier.

As can be seen from this example, one should avoid the direct transfer of cluster features from the example profile to the whole of Brepollen. Almost all the easily identified classes are located in (i) the central part of Brepollen, (ii) the Storebreen glacier valley and (iii) the Hornbreen glacier valley. Correct identification of similar classes in the rest of the region is difficult because the distance used during the compilation of maps is nearly half of the width of the glacier valleys. Since every class can occur in these two valleys it can be assumed that similar forms are present in both.

5. Conclusions

Despite the rapid development of acoustic methods and the use of technologically advanced multibeam echosounders during seafloor scanning performed from large vessels in post-glacial regions, it is still necessary to supplement such activities using single beam echosounders from small boats. In this work the bottom morphology of Brepollen (Hornsund, Spitsbergen) was described by analysing 256 m segments of bathymetric profiles. Among the suggested statistical, spectral, wavelet, fractal dimension and median filter parameters, the following were identified as being the most useful: (i) low-order spectral moments, (ii) spectral skewness, (iii) wavelet energies, (iv) box fractal dimension, (v) mean of the remainder from median filtration. The other parameters were either significantly correlated with parameters (i)–(v), or else they could not be used directly to characterise the shape of a bathymetric profile.

Cluster analysis revealed at least three morphological classes in Brepollen: (i) steep slopes (southern Brepollen), (ii) flat sea bottoms (central Brepollen) and gentle slopes (the Store glacier valley and the southern part of the Horn glacier valley), (iii) the most morphologically diverse region (the central Store valley, the northern part of the Horn glacier valley and the NE part of central Brepollen with the adjacent Horn and Store valleys).

Acknowledgements

References

Adam C., Vidal V., Bonneville A., 2005, *MiFil: a method to characterize seafloor swells with application to the south central Pacific*, Geochem. Geophy. Geosy., 6 (1), Q01003, 1–25, http://dx.doi.org/10.1029/2004GC000814.

Aharonson O., Zuber M. T., Rothman D. H., 2001, *Statistics of Mars' topography from the Mars Orbiter Laser Altimeter: slopes, correlations and physical models*, J. Geophys. Res., 106 (E10), 23723–23735, http://dx.doi.org/10.1029/2000JE001403.

Błaszczyk M., Jania J. A., Hagen J. O., 2009, *Tidewater glaciers of Svalbard: recent changes and estimates of calving fluxes*, Pol. Polar Res., 30 (2), 85–142.

Błaszczyk M., Jania J. A., Kolondra L., 2013, *Fluctuations of tidewater glaciers in Hornsund Fjord (Southern Svalbard) since the beginning of the 20th century*, Pol. Polar Res., 34 (4), 327–352, http://dx.doi.org/10.2478/popore-2013-0024.

Caliński T., Harabasz J., 1974, *A dendrite method for cluster analysis*, Commun. Stat., 3, 1–27.

Dowdeswell J. A., Hogan K. A., Evans J., Noormets R., O'Cofaigh C., Ottesen D., 2010, *Past ice-sheet flow east of Svalbard inferred from streamlined subglacial landforms*, Geology, 38 (2), 163–166, http://dx.doi.org/10.1130/G30621.1.

Forwick M., Baeten N. J., Vorren T. O., 2009, *Pockmarks in Spitsbergen fjords*, Norw. J. Geol., 89 (1–2), 65–77.

Głowacki P., Jania J. A., 2008, *Nature of rapid response of glaciers to climate warming in Southern Spitsbergen, Svalbard*, [in:] *The first International Symposium on the Arctic Research (ISAR-1) – Drastic Change under Global Warming*, Nat. Comm. Japan, Tokyo, 257–260.

Goff J. A., 2000, *Simulation of stratigraphic architecture from statistical and geometrical characterizations*, Math. Geol., 32 (7), 765–786, http://dx.doi.org/10.1023/A:1007579922670.

Goff J. A., Orange D. L., Mayer L. A., Hughes Clarke J. E., 1999, *Detailed investigation of continental shelf morphology using a high-resolution swath sonar survey: the Eel margin, northern California*, Mar. Geol., 154 (1–4), 255–269, http://dx.doi.org/10.1016/S0025-3227(98)00117-0.

Hastings H. M. G., Sugihara G., 1994, *Fractals – a user's guide for the natural sciences*, Oxford Univ. Press, Oxford, New York, 7–77.

Herzfeld U. C., Kim I. I., Orcutt J. A., 1995, *Is the ocean floor a fractal?*, Math. Geol., 27 (3), 421–462, http://dx.doi.org/10.1007/BF02084611.

Hiller J. K., Smith M., 2008, *Residual relief separation: digital elevation model enhancement for geomorphological mapping*, Earth Surf. Proc. Land., 33 (14), 2266–2276, http://dx.doi.org/10.1002/esp.1659.

Kim S.-S., 2005, *Separation of regional and residual components of bathymetry using directional median filtering*, M. Sc. thesis, Univ. Hawaii, 49 pp.

Kim S.-S., Wessel P., 2008, *Directional median filtering for regional-residual separation of bathymetry*, Geochem. Geophy. Geosy., 9 (3), Q03005, 11 pp., http://dx.doi.org/10.1029/2007GC001850.

Lefebvre A., Lyons A. P., 2011, *Quantification of roughness for seabed characterisation*, [in:] *Underwater Acoustic Measurements (4th UAM) – Technologies & Results, Kos, Greece*, Proc. Book 4th Int. Conf. Exhibit., 1623 –1630.

Little S. A., 1994, *Wavelet analysis of seafloor bathymetry: an example*, [in:] *Wavelets in geophysics*, E. Foufoula-Georgiou & P. Kumar (eds.), Acad. Press Inc., San Diego, London, 167–182.

Little S. A., Carter P. H., Smith D. K., 1993, *Wavelet analysis of a bathymetric profile reveals anomalous crust*, Geophys. Res. Lett., 20 (18), 1915–1918, http://dx.doi.org/10.1029/93GL01880.

Little S. A., Smith D. K., 1996, *Fault scarp identification in side-scan sonar and bathymetry images from the Mid-Atlantic Ridge using wavelet-based digital filters*, Mar. Geophys. Res., 18 (6), 741–755, http://dx.doi.org/10.1007/BF00313884.

Mandelbrot B. B., 1982, *The fractal geometry of nature*, W. H. Freeman & Co., New York, 468 pp.

Maulik U., Bandyopadhyay S., 2002, *Performance evaluation of some clustering algorithms and validity indices*, IEEE T. Pattern Anal., 24 (12), 1650–1654, http://dx.doi.org/10.1109/TPAMI.2002.1114856.

Milligan G. W., Cooper M. C., 1985, *An examination of procedures for determining the number of clusters in a data set*, Psychometrica, 50 (2), 159–179, http://dx.doi.org/10.1007/BF02294245.

Moskalik M., Bialik R. J., 2011, *Statistical analysis of topography of Isvika Bay, Murchisonfjorden, Svalbard*, [in:] *GeoPlanet: Earth and planetary sciences, experimental methods in hydraulic research*, P. Rowiński (ed.), 1st edn., Springer, Berlin–Heidelberg, 225–233.

Moskalik M., Błaszczyk M., Jania J., 2013b, *Statistical analysis of Brepollen bathymetry as a key to determine average depth on a glacier foreland*, Geomorphology, http://dx.doi.org/10.1016/j.geomorph.2013.09.029, (in press).

Moskalik M., Grabowiecki P., Tęgowski J., Żulichowska M., 2013a, *Bathymetry and geographical regionalization of Brepollen (Hornsund, Spitsbergen) based on bathymetric profiles interpolations*, Pol. Polar Res., 34 (1), 1–22, http://dx.doi.org/10.2478/popore-2013-0001.

Nikora V., Goring D., 2004, *Mars topography: bulk statistics and spectral scaling*, Chaos Solit. Fractals, 19 (2), 427–439, http://dx.doi.org/10.1016/S0960-0779(03)00054-7.

Nikora V., Goring D., 2005, *Martian topography: scaling, craters, and high-order statistics*, Math. Geol., 37 (4), 337–355, http://dx.doi.org/10.1007/s11004-005-5952-4.

Nikora V., Goring D., 2006, *Spectral scaling in Mars topography: effect of craters*, Acta Geophys., 54 (1), 102–112, http://dx.doi.org/10.2478/s11600-006-0009-8.

Ostrovsky I., Tęgowski J., 2010, *Hydroacoustic analysis of spatial and temporal variability of bottom sediment characteristics in Lake Kinneret in relation to water level fluctuation*, Geo-Mar. Lett., 30 (3–4), 261–269, http://dx.doi.org/10.1007/s00367-009-0180-4.

Ottesen D., Dowdeswell J. A., 2006, *Assemblages of submarine landforms produced by tidewater glaciers in Svalbard*, J. Geophys. Res., 111, F01016, http://dx.doi.org/10.1029/2005JF000330.

Ottesen D., Dowdeswell J. A., 2009, *An inter-ice-stream glaciated margin: submarine landforms and a geomorphic model based on marine-geophysical data from Svalbard*, Geol. Soc. Am. Bull., 121 (11–12), 1647–1665, http://dx.doi.org/10.1130/B26467.1.

Ottesen D., Dowdeswell J. A., Benn D. I., Kristensen L., Christiansen H. H., Christensen O., Hansen L., Lebesbye E., Forwick M., Vorren T. O., 2008, *Submarine landforms characteristic of glacier surges in two Spitsbergen fjords*, Quaternary Sci. Rev., 27 (15–16), 1583–1599, http://dx.doi.org/10.1016/j.quascirev.2008.05.007.

Ottesen D., Dowdeswell J. A., Landvik J. Y., Mienert J., 2007, *Dynamics of the Late Weichselian ice sheet on Svalbard inferred from high-resolution sea-floor morphology*, Boreas, 36 (3), 286–306, http://dx.doi.org/10.1111/j.1502-3885.2007.tb01251.x.

Pace N. G., Gao H., 1988, *Swath seabed classification*, IEEE J. Ocean. Eng., 13 (2), 83–90, http://dx.doi.org/10.1109/48.559.

Pastusiak T., 2010, *Issues of non-researched marine regions coverage by electronic maps*, Logistyka, 2, 2069–2086, (in Polish).

Ray S., Turi R. H., 1999, *Determination of number of clusters in K-means clustering and application in colour image segmentation*, [in:] *Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99), Calcutta, India*, Proc.

4th Int. Conf., N. R. Pal, A. K. De & J. Das (eds.), Narosa Publ. House, New Delhi, 137–143.

Statens Kartverk, 2008, Paper chart 526, Hornsund, scale 1:50 000.

Tęgowski J., Łubniewski Z., 2002, *Seabed characterisation using spectral moments of the echo signal*, Acta Acust., 88 (5), 623–626.

The Norwegian Hydrographic Service and Norwegian Polar Research Institute, 1990, *Den Norske Los. Arctic Pilot*, 7, (2nd edn.), U.K. Hydrogr. Office, 2007, NP11 Arctic Pilot Edition 2004, Correction 2007.

Wen R., Sinding-Larsen R., 1997, *Uncertainty in fractal dimension estimated from power spectra and variograms*, Math. Geol., 29 (6), 727–753, http://dx.doi.org/10.1007/BF02768900.

Wessel P., 1998, *An empirical method for optimal robust regional-residual separation of geophysical data*, Math. Geol., 30 (4), 391–408, http://dx.doi.org/10.1023/A:1021744224009.

White L., 2003, *Rivers bathymetry analysis in the presence of submerged large woody debris*, M. Sc. Eng. thesis, Univ. Texas, Austin, 157 pp.

White L., Hodges B. R., 2005, *Filtering the signature of submerged large woody debris from bathymetry data*, J. Hydrol., 309 (1–4), 53–65, http://dx.doi.org/10.1016/j.jhydrol.2004.11.011.

Wilson M. F. J., O'Connell B., Brown C., Guinan J. C., Grehan A. J., 2007, *Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope*, Mar. Geod., 30 (1–2), 3–35, http://dx.doi.org/10.1080/01490410701295962.

Wornell G. W., Oppenheim A. V., 1992, *Estimation of fractal signals from noisy measurements using wavelets*, IEEE T. Signal Proces., 40 (3), 611–623, http://dx.doi.org/10.1109/78.120804.