<table>
<thead>
<tr>
<th>Title:</th>
<th>Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s):</td>
<td>Els Verfaillie (UGent), Vera Van Lancker (UGent) and Marc Van Meirvenne (UGent)</td>
</tr>
<tr>
<td>Document owner:</td>
<td>Els Verfaillie (<a href="mailto:els.verfaillie@ugent.be">els.verfaillie@ugent.be</a>)</td>
</tr>
<tr>
<td>Reviewed by:</td>
<td>Els Verfaillie (<a href="mailto:els.verfaillie@ugent.be">els.verfaillie@ugent.be</a>)</td>
</tr>
<tr>
<td>Workgroup:</td>
<td></td>
</tr>
<tr>
<td>MESH action:</td>
<td>4</td>
</tr>
<tr>
<td>Version:</td>
<td>n/a</td>
</tr>
<tr>
<td>Date published:</td>
<td></td>
</tr>
<tr>
<td>File name:</td>
<td>WE_UGent_Multivariate_geostatistics.pdf</td>
</tr>
<tr>
<td>Language:</td>
<td>English</td>
</tr>
<tr>
<td>Number of pages:</td>
<td>17 (including these pages)</td>
</tr>
<tr>
<td>Summary:</td>
<td>Multivariate geostatistics have been used to obtain a detailed and high-quality map of the median grain-size distribution of the sand fraction at the Belgian Continental Shelf. Sandbanks and swales are the dominant geomorphological features and impose a high-spatial seafloor variability. Interpolation over complex seafloors is difficult and as such various models were investigated. In this paper, linear regression and ordinary kriging (OK) were used and compared with kriging with an external drift (KED) that makes use of secondary information to assist in the interpolation. KED proved to be the best technique since a linear correlation was found between the median grain-size and the bathymetry. The resulting map is more realistic and separates clearly the sediment distribution over the sandbanks from the swales. Both techniques were also compared with a simple linear regression of the median grain-size against the bathymetry. An independent validation showed that the linear regression yielded the largest average prediction error (almost twice as large as with KED). Unlike most static sedimentological maps, our approach allows for defining grain-size classes that can be adapted according to the needs of various applications. These relate mainly to the mapping of soft substrata habitats and of the most suitable aggregates for extraction. This information is highly valuable in a marine spatial planning context.</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Multivariate geostatistics; Median grain-size; Bathymetry; Habitat mapping; Resource maps; Belgian continental shelf</td>
</tr>
</tbody>
</table>
Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas

Els Verfailliea, Vera Van Lanckera, Marc Van Meirvenneb

Abstract

Multivariate geostatistics have been used to obtain a detailed and high-quality map of the median grain-size distribution of the sand fraction at the Belgian Continental Shelf. Sandbanks and swales are the dominant geomorphological features and impose a high-spatial seafloor variability. Interpolation over complex seafloors is difficult and as such various models were investigated. In this paper, linear regression and ordinary kriging (OK) were used and compared with kriging with an external drift (KED) that makes use of secondary information to assist in the interpolation. KED proved to be the best technique since a linear correlation was found between the median grain-size and the bathymetry. The resulting map is more realistic and separates clearly the sediment distribution over the sandbanks from the swales. Both techniques were also compared with a simple linear regression of the median grain-size against the bathymetry. An independent validation showed that the linear regression yielded the largest average prediction error (almost twice as large as with KED).

Unlike most static sedimentological maps, our approach allows for defining grain-size classes that can be adapted according to the needs of various applications. These relate mainly to the mapping of soft substrata habitats and of the most suitable aggregates for extraction. This information is highly valuable in a marine spatial planning context.

Keywords: Multivariate geostatistics; Median grain-size; Bathymetry; Habitat mapping; Resource maps; Belgian continental shelf

1. Introduction

Seabed habitats are subject to increasing pressures from human developments such as fisheries, aggregate extraction, dredging/dumping and windmill farms. In this context, the mapping of habitats and their prediction becomes crucial, both at the level of baseline studies as during the monitoring and decommitment phase. There is a difference between the physical (or abiotic) and the biological (or biotic) part of a seabed habitat (Fig. 1). However, if a full coverage map of the physical habitat is available and if the relations between the physical and the biological habitat are known, it is possible to create a full coverage map of the biological habitat. Nowadays there is an increasing demand for full coverage information. ‘Filling the gaps’ and ‘predictive modelling’ or the prediction of physical and biological information in areas with gaps, is a hot topic (e.g. ICES, 2005) in the
framework of habitat mapping and nature protection. This is one of the aims of the project MESH (Development of a framework for Mapping European Seabed Habitats) and BWZee (Biological Valuation Map of the Belgian Continental Shelf), in which the current research plays an important role.

Data, describing the physical habitat, are available as point information (e.g. sediment samples), as full coverage information (e.g. Digital Elevation Model or DEM) or as full coverage information from a model (e.g. current data, shear stress). The datasets available in this study were sediment samples and a DEM.

For the mapping of soft substrata habitats, it has been shown that the sedimentology (mainly the median grain-size and the silt-clay percentage) is an important parameter to explain and predict the occurrence of (macro)benthos (seabed organisms larger than 1 mm) (e.g. Wu and Shin, 1997; Lee caster, 2003; Van Hoey et al., 2004). Although sediment samples are generally more available than biological samples, it remains difficult to predict (or interpolate) their distribution and this particularly over complex seafloors. As such, a sound methodology for the interpolation of these data is necessary.

The general aim of this paper was to produce a high-quality map of the median grain-size at the Belgian Continental Shelf (BCS), looking for the best interpolation method. This map is a valuable product in the context of aggregate extraction, habitat mapping, ecological valuation, spatial planning and sediment transport.

2. Material and methods

2.1. Data description

Grain-size data was derived from a sedimentological database (‘sedisurf@’) hosted by Renard Centre of Marine Geology, Ghent University. The dataset is a compilation of sample information since 1976 and contains more than 6000 samples.

As a second variable, a high-resolution DEM was compiled based on data from the Ministry of the Flemish Community (Department of Environment and Infrastructure, Waterways and Marine Affairs Administration, Division Coast, Hydrographic Office) and completed with data from the Hydrographic Office of the Netherlands and the United Kingdom. Regarding the Belgian shelf, this is a very valuable source of information as its very large density allowed an interpolation to a resolution of 80 m, using a simple inverse distance algorithm. From the DEM, a slope map was derived. Based on the DEM and the slope map, homogeneous zones at the BCS could be defined (Fig. 2). These zones allow a distinction between sandbanks, swales and foreshore zones. The delimitation of the zones was done by alternatively inspecting the DEM and the slope map and the visual drawing of polygons in a geographical information system (GIS). From each zone the amount of samples, the variation of the grain-size (e.g. mean value, variance, sum, etc.) can be queried in GIS. In this way it is possible to get an impression of the variation of the samples within each zone and to carry out a quality control of the grain-size values. The quality control of the sediment samples was done assuming that samples inside the same zone, are more similar than samples from different zones. Extreme values can be identified and if necessary, removed. For that purpose a sound knowledge of the sedimentological data is needed. On this basis 83 points were removed out of the dataset.

2.2. Linear regression

A well-known approach consists of modelling the relation between the median grain-size and the
depth using a linear function of the type
\[ z(x) = a_0 + a_1 y(x) \]
with \( z(x) \) equal to the measurement of median grain-size at location \( x \), \( a_0 \) the intercept constant value, \( a_1 \) the slope constant value; \( y(x) \) the measurement of depth at location \( x \).

With this relation, each depth value can be converted into a median grain-size value. This type of regression has the major shortcoming that the median grain-size is only derived from the depth at the same location \( x \), regardless of the surrounding values (Goovaerts, 1999).

### 2.3. Geostatistical approach

Geostatistical interpolation techniques (generally known as kriging) have the advantage that they are stochastic in contrast with deterministic techniques like trend surfaces. The latter predict an unknown value in a unique way without an associated measure of uncertainty. Stochastic techniques provide a number of possible values, with a probability of occurrence. A unique solution cannot be expected (Goovaerts, 1997). Moreover, geostatistical techniques have the advantage that they make use of the spatial correlation between neighbouring observations, to predict values at unsampled places (Goovaerts, 1999). These techniques give an indication of the errors and uncertainties associated with the interpolated values, based on a variance surface of the estimated values (Burrough and McDonnell, 1998). Multivariate geostatistics can be used if there is a relation between the predicted variable (e.g. median grain-size) and a secondary variable (e.g. bathymetry). It is possible to include this secondary
information into the interpolation. This additional information results in a more accurate and complete prediction of the variable than without the secondary information. In practice, secondary information is often cheaper or easier to obtain, and as such can complement the sparsely sampled (primary) observations. More details on the applied geostatistical analysis can be found in Goovaerts (1997), Deutsch and Journel (1998) and Wackernagel (1998).

2.3.1. Variogram analysis

The variogram \( \gamma(h) \) represents the average variance between observations separated by a distance \( h \). The value plays an important role in the description and interpretation of the structure of the spatial variability of the investigated regionalized variable. It is often cheaper or easier to obtain, and as such can complement the sparsely sampled (primary) observations. More details on the applied geostatistical analysis can be found in Goovaerts (1997), Deutsch and Journel (1998) and Wackernagel (1998).

\[
\gamma(h) = \frac{1}{2N(h)} \sum_{z_{x_0}} \left\{ z(x) - z(x + h) \right\}^2
\]

with \( z(x_0) \) equal to the measurement at location \( x_0 \), \( z(x_0 + h) \) the measurement at location \( x_0 + h \). \( \gamma(h) \) the variogram for distance vector \( = \text{lag} \ h \) between measurements \( z(x_0) \) and \( z(x_0 + h) \), the \( N(h) \) number of couples of measurements separated by \( h \).

A variogram is presented as a graph (Fig. 3), where the calculated variogram values (dots) represent the experimental variogram. The fitting of a theoretical variogram (curve) is an important step in the variogram analysis. Hereby, the ‘sill’ is the total variance of the variable, the ‘range’ is the maximal spatial extent of spatial correlation between observations of the variable and the ‘nugget’ is the random error.

The theoretical variogram can be composed of nested models or structures. Common models are the nugget model, spherical model, exponential model, Gaussian model and power model. Direction dependant variograms can be set up in the case of anisotropic variability. The formulas of these models can be found in e.g. Journel and Huijbregts, 1978; Wackernagel, 1998.

For the variogram analysis the programme Variowin 2.21 (Pannatier, 1996) was used.

2.3.2. Interpolation with kriging

A univariate and a multivariate variant of kriging were compared:

- **Ordinary kriging** (OK) of median grain-size (d50) using directional variograms;
- **Kriging with an external drift** (KED) of d50 with bathymetrical values as secondary information and with an omnidirectional variogram.

For the geostatistical analysis the software GSLIB 1998 (Deutsch and Journel, 1998) was used. OK is the most frequently used kriging technique. The OK algorithm uses a weighted linear combination of sampled points situated inside a neighbourhood (or interpolation window) around the location \( x_0 \), where the interpolation is conducted. An underlying assumption is that the mean value (\( m \)) is locally stationary (i.e. that it has a constant value inside the interpolation neighbourhood). The algorithm can be written as

\[
Z^*(x_0) = \sum_{z=1}^{n(x_0)} \left\{ \lambda_z [Z(x_z) - m] \right\} + m
\]

\[
= \sum_{z=1}^{n(x_0)} \{ \lambda_z Z(x_z) \} + \left[ 1 - \sum_{z=1}^{n(x_0)} \lambda_z \right] m
\]

with \( \lambda_z \) equal to the weights attributed to the \( n(x_0) \) observations \( z(x_z) \); \( n \) the total number of observations \( z(x_z) \); \( n(x_0) \) the subset of \( n \), lying inside the interpolation window.

The weights \( \lambda_z \) are obtained by solving a set of equations (the kriging system) involving knowledge of the variogram (see e.g. Goovaerts, 1997). These weights are constrained to sum to one, leading to the elimination of the parameter \( m \) from the estimator which is thus written as

\[
Z^*_{OK}(x_0) = \sum_{z=1}^{n(x_0)} \lambda_z Z(x_z) \quad \text{with} \quad \sum_{z=1}^{n(x_0)} \lambda_z = 1.
\]
KED is a multivariate variant of ‘Kriging with a Trend Model’ (KT), formerly called ‘Universal Kriging’. KED and KT are non-stationary methods, meaning that the statistical properties of the variable are not constant in space (i.e. no constant mean within the interpolation neighbourhood). With KT, the trend is modelled as a function of the spatial coordinates, whilst for KED the trend \( m(x_0) \) is derived as a local linear function of the secondary variable \( z_2(x_0) \), which is formulated in each interpolation window (Goovaerts, 1997)

\[
m(x_0) = b_0 + b_1 z_2(x_0)
\]

with \( m(x_0) \) the trend at location \( x_0 \), \( b_0, b_1 \) the unknown parameters of the trend, calculated in each interpolation window from a fit to the observations, \( z_2(x_0) \) the secondary variable at location \( x_0 \).

The KED estimator has the same form as the OK estimator.

For non-stationary geostatistics such as KED, \( Z(x) \) can be decomposed into a deterministic function or a drift \( m(x) \) and a residual random function \( Y(x) \) (Wackernagel, 1998)

\[
Y(x) = Z(x) - m(x).
\]

The underlying variogram associated with \( Y \) is directly accessible, when the drift is not active in a particular direction of space. The variogram in this direction can be extended to the other directions under an assumption of isotropic behaviour of the underlying variogram (Wackernagel, 1998).

KED is a multivariate geostatistical technique, as it makes use of secondary information. However, this secondary data must be available at all primary data locations as well as at all locations being estimated. A more complex multivariate geostatistical technique is cokriging, which does not require this secondary information to be known at all locations being estimated. Cokriging is much more demanding than other kriging techniques because both direct and cross variograms must be inferred and jointly modelled and because a large cokriging system must be solved (Goovaerts, 1997).

2.3.3. Validation

To enable a thorough quality control of the geostatistical analysis, the sedisurf@ database was divided into two subsets: a prediction and an independent validation dataset. The proportion of both datasets is respectively, 70% and 30% of the whole dataset. The validation dataset was selected using a random selection of data points.

Several indices are suitable to evaluate the interpolation. These indices are all a measure of the estimation error that is the difference between the estimated and the observed value: \( z^*(x_a) - z(x_a) \).

1. The mean estimation error (MEE), which has to be about zero to have an unbiased estimator.

\[
\text{MEE} = \frac{1}{n} \sum_{z=1}^{n} (z^*(x_z) - z(x_z)).
\]

2. The mean-square estimation error (MSEE), which has to be as low as possible and which is useful to compare different procedures. The root mean-square estimation error (RMSEE) is used to obtain the same units as the variable. This parameter has to be compared to the variance or the standard deviation of the dataset.

\[
\text{MSEE} = \frac{1}{n} \sum_{z=1}^{n} (z^*(x_z) - z(x_z))^2.
\]

3. The mean absolute estimation error (MAEE), which is analogous to the MSEE, but less sensitive to extreme deviations.

\[
\text{MAEE} = \frac{1}{n} \sum_{z=1}^{n} |z^*(x_z) - z(x_z)|.
\]

4. The Pearson correlation coefficient between \( z^*(x_a) \) and \( z(x_a) \), which indicates the degree of linear correlation between observed and estimated values. This value always has to be considered in combination with the MEE. The correlation coefficient is itself a measure of the proportion of variance explained, hence is related to MSEE.

3. Results

3.1. Linear regression

The relation between median grain-size and depth was modelled as

\[
d50 = 179.84 + 5.94 \times \text{depth},
\]

resulting in the map of the median grain-size shown at Fig. 4. This map is a simple rescaling of the DEM, converted into grain-size values between 179 and 508. Linear regression is not an exact interpolator, meaning that the interpolated map does not honour observations; the measurements are only
used to calculate a linear regression function. As the map is a transformation of the DEM, it shows very clearly the anisotropy, but the typical, more patchy pattern of the grain-size is completely lost (compare with Fig. 10a and b).

3.2. Geostatistical approach

3.2.1. Exploratory data analysis

The histogram of the grain-size data (Fig. 5) shows a symmetric distribution.

At every location where the median grain-size $d_{50}$ is known the depth is also known from the DEM (Fig. 6). The Pearson correlation coefficient $r_{ij}$ between both variables is 0.46, indicating a moderately strong correlation. The Spearman rank correlation is slightly larger (0.52) indicating the presence of some outliers (as can be seen at Fig. 6) reducing the Pearson correlation coefficient.

The scatterplot suggests the existence of two populations (one parallel with and one perpendicular to the $X$ axis). However after splitting the two populations, the correlations did not improve. To preserve the added value of the secondary variable in the geostatistical analysis, the decision was made to keep the dataset as a single entity.

3.2.2. Variogram analysis

The maximal diagonal distance at the BCS is about 90 km. Following a rule of thumb, the product of the lag interval distance and the number of lags should not exceed half of this largest dimension: i.e. between 30 and 45 km. Consequently, the variogram surface was calculated using

![Fig. 4. Map of median grain-size, on the basis of linear regression.](image-url)
11 lags of 3000 m. This variogram surface (Fig. 7) shows a clear anisotropy. The direction of the largest continuity is about 50° (expressed as a trigonometric angle), corresponding to the direction of the sandbanks at the BCS and to the smallest variogram values. This indicates that the sandbanks have a strong influence on the spatial variability of the data. This is the case for the median grain-size (Fig. 7, left), but it is stronger with the depth values (Fig. 7, right). The direction of the largest discontinuity is about 130°, corresponding to the direction perpendicular to the sandbanks. To characterize the spatial variability in different directions directional variograms were calculated in the directions: 40°, 85°, 130° and 175°.

For the directional variograms computed over a large distance (i.e. over a distance of 33 km or 11 lags of 3000 m.), a sill was reached in the direction of the largest continuity (40°). In the directions perpendicular to this direction, no sill was reached. This indicates a spatial trend, i.e. an increasing variability with increasing distance, which is caused by a non-stationary mean (i.e. a non-constant mean median grain-size over the BCS). Therefore, for OK the variogram (Fig. 8) was restricted to a distance of 10 km (with 20 lags of 500 m), which is large enough to cover the interpolation window.

For KED the experimental variogram (Fig. 9) was estimated using increasing lag spacings between 500 and 1000 m. In this way it was possible to model accurately both the short and long distance patterns during the variogram analysis. The short distance variability is important for the fitting of the nugget and initial behaviour of the variogram, while the long distance variability is important for the fitting of the range and eventually compound or ‘nested’ models. Only the variogram in the direction of the
largest continuity (50°) was calculated, because we consider it as representative for stationary conditions without a trend. For KED a linear trend with the depth (causing the anisotropy) was calculated within each interpolation window. The variogram is also shown at a distance of 10 km, to make it comparable with the directional variograms of OK, although the sill would not change anymore over a larger distance (as 50° is the direction of the largest continuity).

To the variogram of OK an exponential model was fit with a nugget of 1240 μm², a range of 2200 m in the direction of the largest continuity and a range of 880 m in the direction of the lowest continuity. This represents a geometrical anisotropy, meaning that there are different ranges in different directions. This anisotropy is modelled using an ellipse, with the largest and the smallest range as, respectively the main axis and the side axis. The ratio between the largest and the smallest ranges, is the anisotropy ratio. The anisotropy ratio is 0.40. The sill of the structure has a value of 7740 μm².

The theoretical variogram of KED was best modelled as a nested structure. The nugget is 1560 μm², the first structure is an exponential model with a range of 2400 m and the second structure is a spherical model with a range of 9000 m, the sill is 7410 μm².
3.2.3. Interpolation with kriging

For the calculation of the final OK map, the fitted variogram parameters (nugget, range, sill, anisotropy ratio) were used. Minimum two and maximum 16 observations were required for the interpolation. Quadrants (i.e. circles divided in four equal parts) were used, with a maximal amount of observations of four per quadrant. The search radius was 5000 m. So, points further than 2200 m (i.e. maximal range) were also involved in the interpolation. This is advantageous for locations with a low density of data (e.g. in the northern part of the BCS). These observations obtain very low weights, because they are located outside of the distance of the maximal range, but still carry some information.

The result of the OK (Fig. 10, top) is an almost full coverage map. A strip in the northeast of the BCS is not covered, because data for interpolation are lacking. This map appears quite continuous (excepted the three spots in the northern part of the BCS), without the ‘bull’s eyes’ or concentric patterns around data points, typical for ‘inverse distance’ interpolations. However, the map shows grain-sizes with continuous values across the sandbanks. As no secondary bathymetry information was used for this map, the topography of the seabed cannot be observed inside of the pattern of the median grain-size values.

For the calculation of the KED map, the parameters from the variogram analysis were also used. Besides minimum two and maximum 16 observations and quadrants with maximum four observations are used. The maximal search radius is 9000 m, corresponding to the maximal range.

The result of KED (Fig. 10, bottom) looks much more realistic than the result of OK. Moreover the median grain-size varies in proportion to the depth. This is very clear in the Hinderbanks region (northern part of BCS). Values between 400 and 500 μm are mainly found in the swales, while the values between 350 and 400 μm are dominantly found at the sandbanks. This pattern is also clear closer to the coast. Unlike OK, KED made use of the secondary information of the bathymetry. The topography pattern of the seabed can be clearly seen inside of the median grain-size map.

3.2.4. Validation

The scatter plots (Fig. 11) of the observed versus the estimated values give a first indication in the validation of different techniques. For linear regression the correlation coefficient is much lower than the values for both OK and KED, which demonstrates the inefficiency of this technique compared with both kriging techniques. The correlation coefficient between both values is slightly larger for KED than for OK, indicating that KED gives better results.

However, scatter plots have to be considered in combination with validation indices (Table 1). Linear regression yields the largest error for each validation index. KED provides a better result compared to OK, next to a visually more realistic map.

The estimation variance of the kriging analysis gives an indication of the overall reliability of the kriging. This is not an absolute measure of reliability of the kriging estimate (Journel, 1993; Armstrong, 1994; Goovaerts, 1997), but it gives more an indication of the sampling density (a high sampling density means logically a high quality). This is valuable information as it can be used to guide future sampling campaigns. Where the variance reaches high values, new samples are preferably taken. This allows filling gaps and monitoring on a purposive and efficient manner. Fig. 12 shows the estimation variance of KED. As the KED and the OK map are based on the same samples, only the result of KED is given in Fig. 12. For the interpolation the extreme minimum of two observations was used, to obtain a map that approaches a full coverage map. Fig. 12 indicates clearly where this minimum of two observations is too low to give a reliable grain-size value.

4. Discussion

The result of KED is a high-quality and high-resolution map (250 × 250 m) of the median grain-size at the BCS, using the bathymetry as secondary information (Fig. 10 bottom). Leecaster (2003) also used a multivariate kriging technique in combination with bathymetry for the mapping of the grain-size in Santa Monica Bay, California. She showed that the inclusion of depth in the model improved the prediction in the depth-defined areas like canyons, canyon lips and shortbanks. Most applications of multivariate geostatistics using a DEM as secondary information are, however, found in the soil science (Bourennane et al., 2000; Bourennane and King, 2003, Hengl et al., 2004) and climatology (Goovaerts, 1999; Hudson and Wackernagel, 1994 and Martinez-Cob and Cuenca, 1992). As such, it was a challenge to apply and test these techniques in
Fig. 10. Maps of median grain-size, on the basis of ordinary kriging (top) and kriging with an external drift (bottom). The topography of the seabed can be recognized inside of the map below, because this methodology uses the bathymetry to assist with the interpolation.
a complex marine environment, dominated by a high spatial variability imposed by sandbanks. Moreover, the technique was used over a large area (3600 km²) comprising a nearshore, coastal and offshore zone, each with different morphologies. Although the data availability drastically decreased in an offshore direction, the results were very satisfactory.

By comparing and validating linear regression and the two-kriging techniques (OK and KED), it is obvious that kriging is a better interpolation method. Moreover, for data which are unevenly

---

**Table 1**

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>OK</th>
<th>KED</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEE (mean est err)</td>
<td>-9.17</td>
<td>-8.09</td>
<td>-5.71</td>
</tr>
<tr>
<td>MSEE (mean sq err)</td>
<td>12469.29</td>
<td>7409.95</td>
<td>6745.60</td>
</tr>
<tr>
<td>RMSEE (root mean sq err)</td>
<td>111.67</td>
<td>86.09</td>
<td>82.13</td>
</tr>
<tr>
<td>MAEE (mean abs est err)</td>
<td>74.89</td>
<td>54.97</td>
<td>50.29</td>
</tr>
<tr>
<td>Pearson correlation coefficient r</td>
<td>0.42</td>
<td>0.72</td>
<td>0.80</td>
</tr>
</tbody>
</table>

The different parameters are explained in Section 2.
distributed (such as the samples of the median grain-size), kriging has a declustering effect because of its well-known ‘screening effect’ (see e.g. Govaerts, 1997). This results from the fact that kriging considers both the distance to the interpolation point as the sampling configuration (i.e. distance between observations). Consequently, kriging is preferable to non-declustering techniques (such as linear regression or inverse distance interpolation) for situations with unevenly distributed data.

In cases of a general anisotropy or trend (drift), one solution is to use a small search neighbourhood so that one can assume local stationary conditions within it. This is clearly not a solution in this situation where quite abrupt local changes of the bathymetry occur which needs to be modelled as a local trend. So even locally stationary conditions cannot be assumed over the entire study area. Therefore a non-stationary method like KED should be used. A discussion on this topic is given in Meul and Van Meirvenne (2003).

For applications outside the Belgian shelf, some precautions are needed as the correlation between the bathymetry and the grain-size will depend on the morphology, topography and on the substrate type. However, it is likely that some level of correlation will exist (e.g. Leecaster, 2003). The study area of this paper has a very definite presence of the sandbanks dominating the topography of the seabed. The grain-size is expected to vary following the alternation of sandbanks and swales. However, it remains open to discussion whether the

---

Fig. 12. KED estimation variance of the median grain size.
environmental setting of the study area defines the benefit of KED and whether this should be investigated first. Where the bathymetry is not a dominating characteristic of the study area, other secondary information (e.g. current and wave parameters) might control the pattern of grain-sizes. At a local scale, it can be expected that the grain-size is more related with the geomorphology than with the bathymetry as such. However, this kind of data is more difficult to obtain and is more vulnerable to subjectivity than bathymetry data. However, future research will test the potential correlation as, nowadays, several algorithms exist to estimate the morphological variance from bathymetric-derived features (such as depressions, crests, flats and slopes, both on a large- and small-scale).

The calculation of the bathymetric position index is the most widely available technique and is a measure of where a location, with a defined elevation, is relative to the overall landscape (Weiss, 2001; Iampietro and Kvitek, 2002; Lundblad et al., 2006). Further research will focus on the relation between bathymetric-derived features and physical datasets such as the median grain-size. Furthermore the geomorphology will be analysed in the context of marine habitat mapping, as topographic features are assumed to be important possible habitats for marine organisms.

If a good correlation can be found between the grain-size distribution and the bathymetry (or other environmental variables such as bathymetric-derived features), detailed grain-size maps can be produced. These have considerable advantages compared to the traditional static sedimentological mapping. Most sedimentological maps are based on the Folk classification (Folk, 1974) giving a percentage of gravel, sand and mud. These maps remain highly valuable, but are rather difficult to use for detailed purposes. Nowadays, there is however a need for detailed maps that give a direct reflection of the grain-size itself. Median grain-size values become more widespread available as also bathymetry data that can assist the interpolation. This combination of information is crucial to define the most suitable areas for aggregate extraction and to reserve these areas in a spatial planning context. Moreover, numerical sediment maps are needed to serve as an input layer for various modelling initiatives. These relate to sediment transport modelling or to the predictive modelling of the distribution of soft substrata habitats. In literature, it has been shown that macrobenthic communities in sandy shelf environments have a clear relationship with well-defined ranges of median grain-size and silt-clay percentage (Van Hoey et al., 2004; Lu, 2005; Willems et al., in press). As such a mapping of these variables, and their querying, enables direct predictions that are biologically relevant. This calls however for detailed sedimentological maps and these are rarely available. In an international context, there is also a growing interest in sedimentological maps and this related to the concept of ‘Marine Landscapes’ (Roff and Taylor, 2000; Roff et al., 2003; Golding et al., 2004). Generally, marine landscape modelling is an approach that uses geophysical data as a surrogate for biological mapping. Biological data are only used for the validation of the marine landscapes in terms of their biological relevance. In most cases, the approach remains rather broad-scale, mostly because of the limited detail of the sedimentological maps that are used in the analysis. Schelfaut (2005) used however the detailed grain-size map, described in this paper, and was able to obtain very detailed marine landscapes with high relevance towards the biological value. Future research will focus on the mapping of other target variables, such as the silt-clay percentage, because of its importance for the mapping of the occurrence of the macrobenthos in soft substrata (Van Hoey et al., 2004; Lu, 2005; Willems et al., in press).

5. Conclusions

There is a growing need for a detailed mapping of the seafloor and this is required at a full coverage basis. Apart from the bathymetry, the most crucial variable is sedimentology, as it rules sediment transport processes and it is often the missing link for the prediction of the occurrence of soft substrata habitats or macrobenthic communities/species. The median grain-size was chosen as environmental parameter, as this parameter is the most calculated by a wide variety of scientists and the most frequently used in modelling studies. Hence, a sound interpolation of these data is highly valuable for a wide range of disciplines.

Kriging techniques proved to be the most promising tools to obtain a detailed and high-quality map of the median grain-size distribution. These techniques differ from other linear estimation techniques in their aim to minimize the error variance. In addition, kriging with an external drift allowed using correlated secondary information
such as bathymetry to assist in the interpolation. Linear regression, ordinary kriging (OK) and kriging with an external drift were compared and validated using an independent dataset. Several validation indices were involved. The independent validation showed that the KED map of the median grain-size is much better than the results obtained using linear regression and better than using OK.

KED enabled to obtain a high-quality and high-resolution map (250 × 250 m) of the median grain-size at the BCS, using the bathymetry as secondary information. The resulting map is more realistic and separates clearly the sediment distribution over a complex of sandbanks and swales.

Acknowledgements

This paper is part of the Ph.D. research of the first author, financed by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). The study frames into the research objectives of the project MESH (“Development of a framework for Mapping European Seabed Habitats”), financed by the INTERREG IIIB North West Europe Programme. It also frames into the research objectives of the projects MAREBASSE (“Management, Research and Budgeting of Aggregates in Shelf Seas related to End-users”, EV/02/18A) and BWZee (“A Biological Valuation Map for the Belgian Continental Shelf”, EV/88/37B), financed by the Belgian Science Policy.

The bathymetric data were delivered by the Ministry of the Flemish Community, Department of Environment and Infrastructure, Waterways and Marine Affairs Administration, Division Coast, Hydrographic Office. The sedisurf@ database has been compiled by the Renard Centre of Marine Geology and consists out of data from various institutes.

Finally, the authors would like to thank the reviewers for their very useful comments and suggestions.

References


