Chapter 3

Geostatistical modelling of sedimentological parameters using multi-scale terrain variables: application along the Belgian Part of the North Sea

ACCEPTED FOR PUBLICATION AS:

Verfaillie, E., Du Four, I., Van Meirvenne, M. and Van Lancker, V., accepted. Geostatistical modelling of sedimentological parameters using multiscale terrain variables: application along the Belgian Part of the North Sea. International Journal of Geographical Information Science.

3 Geostatistical modelling of sedimentological parameters using multi-scale terrain variables: application along the Belgian Part of the North Sea

Abstract

The sediment nature and processes are the key to the understanding of the marine ecosystem, and can explain particularly the presence of soft-substrata habitats. For predictions of the occurrence of species and habitats, detailed sedimentological information is often crucial.
This paper presents a methodology to create high quality sedimentological data grids of grain-size fractions and the percentage of silt-clay. Based on a multibeam bathymetry terrain model, multiple sources of secondary information (multi-scale terrain variables) were derived. Through the use of the geostatistical technique, Kriging with an external drift (KED), this secondary information was used to assist in the interpolation of the sedimentological data. For comparison purposes, the more commonly used Ordinary Kriging technique, was also applied. Validation indices indicated that KED gave better results for all of the maps.

Keywords: Multivariate geostatistics; sedimentology; topography; ecogeographical variables; Belgian part of the North Sea

3.1 Introduction

For marine habitat mapping and spatial planning purposes, high quality maps of ecogeographical variables (EGVs), that assist in the prediction of the occurrence of biological species or communities are invaluable (Derous et al. 2007 and Degraer et al. 2008). For soft substrata habitats, the grain-size and the silt-clay percentage are often the most determining EGVs for the modelling of macrobenthic species (Wu and Shin 1997; Van Hoey et al. 2004; Willems et al. 2008). As such, interpolated data of these sedimentological variables are required, if full-coverage maps of macrobenthos are needed for scientific or management purposes. However, the occurrence of macrobenthic species or communities is known to be patchy or bound to topographic variation (Rabaut et al. 2007); as such, more detailed sedimentological information is required if targeted predictions of macrobenthos are to be made (e.g. impact assessments). Consequently, (multi-scale) terrain characteristics are believed to be important EGVs also (Guisan and Thuiller 2005; Baptist et al. 2006 and Wilson et al. 2007).

EGVs that cover entire parts of the seafloor (e.g. derived from high-resolution multibeam bathymetry), represent well the topographical and morphological variation; however, this is seldom the case when the sedimentological variability is considered. Mostly, sedimentological data are interpolated from poorly distributed sediment sampling points and most often inadequate techniques are being used for the interpolation. Verfaillie et al. (2006) and Pesch et al. (2008) argumented already that the quality of the sedimentological maps can be improved significantly, if complex geostatistical interpolation methods are applied.

Multivariate geostatistics can be considered when there is a linear correlation between the variable and a secondary dataset. In Verfaillie et al. (2006), one secondary dataset (Digital Terrain Model or DTM) was used to create a high quality map of the median grain-size of the sand fraction (fraction between 63 and 2000 µm), based on Kriging with an External Drift (KED). However, if more than one secondary dataset is available, that correlates with the sedimentological variable, improved results can be obtained (e.g. Kyriakidis et al. 2001; Bourennane and King 2003; Reinstorf et al. 2005; Hengl et al. 2007a and Miras-Avalos et al. 2007). Furthermore, Verfaillie et al. (2006) demonstrated that interpolations based on linear regression and Ordinary Kriging (OK), resulted in respectively bad and relatively good results, compared to KED.

Our aim was to produce high quality maps of d_s10 (10th percentile of the sand fraction), d_s50 (median grain-size of the sand fraction), d_s90 (90th percentile of the sand fraction) and silt-clay% (fraction below 63 µm) using KED (Goovaerts 1997) with multiple secondary datasets, derived from multibeam bathymetry. For unimodal sandy sediments, maps of the d_s10 and d_s90 are in principle very similar to those of d_s50. Still, for skewed grain-size distributions, with extreme fine or coarse fractions, those variables can be important to explain presences of certain species or communities.

This paper will demonstrate particularly the strength of advanced geostatistical techniques to model a suite of sedimentological variables, using multiple secondary EGVs.

3.2 Materials and methods

3.2.1 Study area and datasets

The study area (Figure 3.1) was situated on the Belgian part of the North Sea (BPNS), at about 16 km away from the harbour of Zeebrugge and very close to the Belgian-Dutch border. Depths were between 15 and 24 m MLLWS (Mean Lowest Low Water at Spring tide). Important geomorphological and ecological values characterise this area. Large- to very large sand dunes (sensu Ashley 1990) were present in the area, reaching heights of 2.5 m, with wavelengths of a few hundred meters.

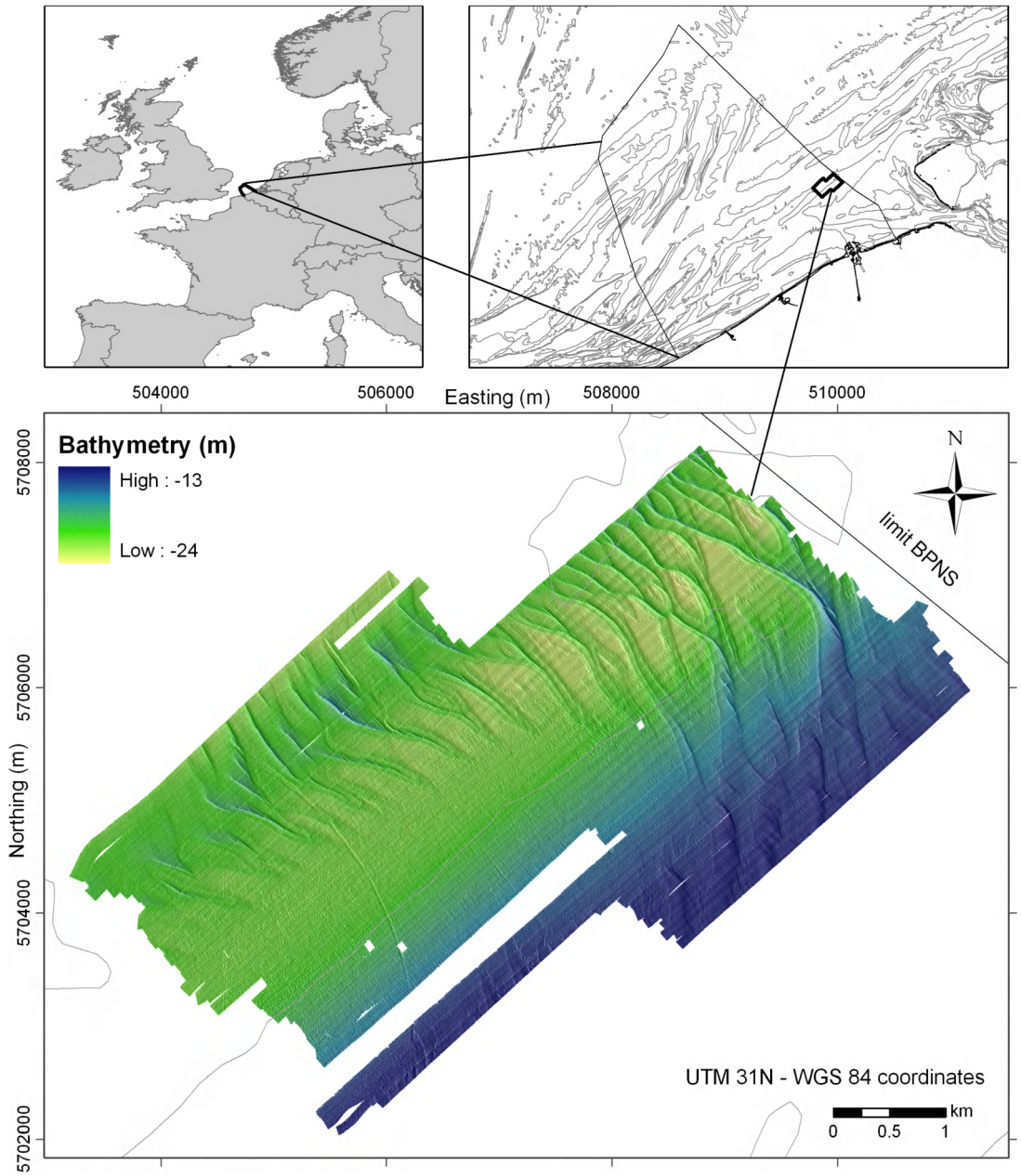

Figure 3.1: Study area (bottom), located in Europe (top left) and the Belgian part of the North Sea (BPNS) (top right).
Large- to very large sand dunes (sensu Ashley 1990) are present in the area.

The sedimentological dataset consisted out of 97 samples, collected during 2 campaigns (RV/*Belgica* 2006/11/20-24 and 2007/11/26-30). A stratified random sampling approach was chosen, based on previously acquired multibeam bathymetry. Sedimentological samples were analyzed with a Malvern Mastersizer 2000 laser particle size analyzer (Malvern Instruments 2008). New multibeam bathymetry (Kongsberg Simrad EM1002S) data were acquired also during the 2 sampling campaigns. For this study, the bathymetry datasets were processed at a resolution of 5 m.
Software used was Variowin 2.21 (Pannatier 1996) for the variogram analysis of the sedimentological datasets; gstat 0.9-42 (Pebesma 2004), implemented in R 2.6.1 (R version 2.6.1 2007) for the geostatistical analysis; ArcGIS 9.2 for GIS analyses and modelling; Biomapper 3.2 (Hirzel et al. 2002b; Hirzel et al. 2006) for the Principal Component Analyses (PCA); and SPSS 15.0 for the correlation analysis of the sedimentological data with the EGVs.

3.2.2 Research strategy

The research strategy consisted out of three steps (Figure 3.2): (1) the selection of relevant EGVs as secondary variables for KED; (2) geostatistical interpolation, based on KED and OK; and (3) comparison of the results.

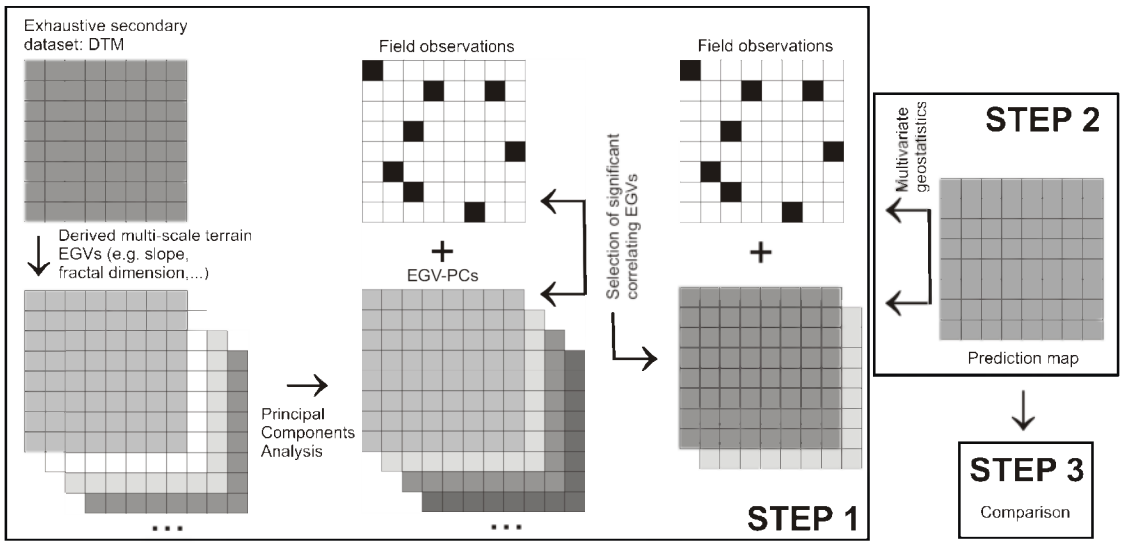

Figure 3.2: Research strategy:
Step 1: The full coverage Digital Terrain Model (DTM) was subjected to a multi-scale terrain analysis, resulting in a set of derived Ecogeographical Variables (EGVs). After a Principal Components Analysis, a Pearson correlation between the field observations and the secondary datasets was calculated. Only significantly ($p \leq 0.05$) correlating Principal Components (PCs or EGV-PCs) were retained as secondary variables for Kriging with an external drift (KED); Step 2: Field observations were interpolated using KED with the selected EGV-PCs as secondary information. Ordinary Kriging (OK) was also applied on the field observations without secondary information (not shown in the scheme); and Step 3: Results of KED and OK are compared and evaluated.

3.2.3 Selection of EGVs as secondary variables for KED

Based on the DTM, a range of multi-scale characteristics were derived that could be used as secondary datasets for KED (slope, eastness, northness, profile curvature, plan curvature, mean curvature and fractal dimension; cfr. Wilson et al. 2007, for an overview and description). Each variable was calculated on 5 different spatial scales, ranging from fine- (15 m) to large-scale (155 m). Window sizes of 3, 7, 13, 21 and 31 cells were applied (with a resolution of 5 m, this corresponded respectively to lengths of 15, 35, 65, 105 and 155 m). In this paper, the dataset of multi-scale characteristics were called 'terrain EGVs'.
To avoid multicollinearity (i.e. high degree of linear correlation) of the terrain EGVs, a PCA was applied. The PCA is based on a correlation matrix, implying that the Kaiser-Guttman criterion can be applied (Legendre and Legendre 1998). This means that Principal Components (PCs) with eigenvalues larger than 1 were preserved as meaningful components for the analysis.
A Pearson correlation coefficient was calculated between the PCs (or EGV-PCs) and the sedimentological point data (d_s10, d_s50, d_s90 and silt-clay%). The selection of EGV-PCs as secondary datasets for the geostatistical modelling was based on statistically significant correlations ($p \leq 0.05$) and the visual inspection of linearity on a scatter plot.

3.2.4 Interpolation with OK and KED

Kriging requires a variogram analysis. The variogram $\gamma(\mathbf{h})$ represents the average variance between observations, separated by a distance $\mathbf{h}$. This value is important in the description and interpretation of the structure of the spatial variability of the investigated regionalized variable (Journel and Huijbregts 1978). The 'sill' is the total variance s^2 of the variable, the 'range' is the maximal spatial extent of spatial correlation between observations of the variable and the 'nugget variance' represents random error or small-distance variability.
Geostatistics is based on the concept of Random Functions, whereby the set of attribute values $z(\mathbf{x})$ at all locations $\mathbf{x}$ are considered as a particular realization of a set of spatially dependent Random Variables $Z(\mathbf{x})$ (Meul and Van Meirvenne 2003).
To compare the resulting maps of predictions of the sedimentological data, the datasets were interpolated, both with OK and KED.
OK is the most frequently used kriging technique. The OK algorithm uses a weighted linear combination of sampled points, situated inside of a neighbourhood (or interpolation window) around the location $\mathbf{x}_0$ 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:

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \{\lambda_\alpha \cdot [Z(\mathbf{x}_\alpha) - m]\} + m = \sum_{\alpha=1}^{n(x_0)} \{\lambda_\alpha Z(\mathbf{x}_\alpha)\} + \left[1 - \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha\right] \cdot m \qquad (3.1)$$

with λ_α equal to the weights attributed to the $n(\mathbf{x}_0)$ observations $z(\mathbf{x}_\alpha)$; n the total number of observations $z(\mathbf{x}_\alpha)$; $n(\mathbf{x}_0)$ the subset of n, lying inside the interpolation window. The weights λ_α 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:

$$Z^*_{OK}(\mathbf{x}_0) = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha Z(\mathbf{x}_\alpha) \qquad \text{with} \qquad \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha = 1 \qquad (3.2)$$

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 $m(\mathbf{x}_0)$ is derived from a local linear function of the secondary variable, which is formulated in each interpolation window (Goovaerts 1997):

$$m(\mathbf{x}_0) = b_0 + b_1 u_2(\mathbf{x}_0) \quad (3.3)$$

with $m(\mathbf{x}_0)$ the trend on location $\mathbf{x}_0$; b_0, b_1 the unknown parameters of the trend, calculated in each interpolation window from a fit to observations; $u_2(\mathbf{x}_0)$ the secondary variable on location $\mathbf{x}_0$.
In the case of more than one secondary variable $u_i(\mathbf{x}_0)$, this formula can be extended to:

$$m(\mathbf{x}_0) = b_0 + b_1 u_2(\mathbf{x}_0) + b_2 u_3(\mathbf{x}_0) + \ldots + b_{i-1} u_i(\mathbf{x}_0) \qquad (3.4)$$

with $m(\mathbf{x}_0)$ the trend at location $\mathbf{x}_0$; b_0, b_1, b_2, b_{i-1} the unknown parameters of the trend, calculated in each interpolation window from a fit to the observations ; $u_2(\mathbf{x}_0)$, $u_3(\mathbf{x}_0)$, …, $u_i(\mathbf{x}_0)$ the secondary variables at location $\mathbf{x}_0$, depending on the number of secondary variables i-1.
The KED estimator has the same form as the OK estimator.
At each location where the primary sedimentological variable $z(\mathbf{x}_\alpha)$ was observed, the residual $r(\mathbf{x}_\alpha)$ was computed:

$$r(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha) - m(\mathbf{x}_\alpha) \quad (3.5)$$

A major problem concerning KED is that the underlying (trend-free) variogram is assumed to be known. This means that the variogram, estimated from the raw data, is biased if the mean changes from place to place. As such, it is necessary to remove the local mean and estimate the residual variogram (Lloyd 2005). A solution to estimate the underlying variogram, associated with $r(\mathbf{x}_\alpha)$, is to use the variogram in a direction where the drift is not active (Goovaerts 1997; Wackernagel 1998; Hudson and Wackernagel 1994; Lloyd 2005 and Verfaillie et al. 2006). The variogram in this direction can be extended to other directions under the assumption of isotropic behavior of the underlying variogram.
For KED, the 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).
The selected EGV-PCs were used as secondary datasets for KED, resulting into sedimentological data grids of d_s10, d_s50, d_s90 and silt-clay%.

KED was computed in R, based on Hengl (2007b) and Hengl (pers. comm.).

3.2.5 Comparison of OK and KED

To enable a thorough quality control of the geostatistical analysis, based on both OK and KED, a 5-fold cross validation was performed (Fielding and Bell 1997), meaning that the sedimentological dataset was split into 5 partitions and that each partition was withheld one after the other. Several indices are suitable to evaluate the interpolation. These indices are all a measure of the estimation error, which is the difference between the estimated and the observed value:

$$z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)$$

(a) The *mean estimation error* (MEE), which has to be around zero to have an unbiased estimator.

$$\text{MEE} = \frac{1}{n}\sum_{\alpha=1}^{n}\left(z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)\right) \qquad (3.6)$$

(b) The *mean square estimation error* (MSEE), which has to be as low as possible and 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.

$$\text{MSEE} = \frac{1}{n}\sum_{\alpha=1}^{n}\left(z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)\right)^2 \qquad (3.7)$$

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

$$\text{MAEE} = \frac{1}{n}\sum_{\alpha=1}^{n}\left|z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)\right| \qquad (3.8)$$

(d) The *Pearson correlation coefficient* between $z^*(\mathbf{x}_\alpha)$ and $z(\mathbf{x}_\alpha)$, indicates the degree of linear correlation between observed and estimated values. This value has to be considered in combination with the MEE. The correlation coefficient is, in itself, a measure of the proportion of variance explained, hence is related to MSEE.
The validation indices permit comparing the results of OK and KED.

3.3 Results

3.3.1 Selection of EGVs as secondary variables for KED

PCA resulted in 9 PCs, explaining 81.4 % of the total variance. Table 3.1 gives an overview of the selected PCs with the corresponding EGVs with high factor loads ($-0.5 < r$ and $r > 0.5$). The Pearson correlation coefficients of all 9 PCs with the values of d_s10, d_s50, d_s90 and silt-clay% and the significant linear correlations are presented in Table 3.2. All of the sedimentological variables showed a significant correlation with PC2 and PC6. A selection of scatter plots is presented in Figure 3.3. As the scatter plots of d_s10, d_s50 and d_s90 are very similar for PC2 and PC6, only the scatter plots of d_s90 are given. The correlation coefficient between the silt-clay% and PC2 and PC6 is very weak and only significant at the 0.05 level (Table 3.2). As such, these

scatter plots are not presented in Figure 3.3 and it is expected that the secondary variables PC2 and PC6 will not contribute significantly to the KED interpolation of the silt-clay%. PC2 was mainly explained by multi-scale slope and fractal dimension, while PC6 by multi-scale plan curvature (Table 3.1). Those PCs were the major contributors for the KED analysis. Moreover, d_s90 correlated weakly with PC1 as well, mainly explained by multi-scale mean and profile curvature. This means that the sediment variation was mainly correlated with the combined pattern of slope, fractal dimension and plan curvature and this on different spatial scales.
The correlation coefficient between the sedimentological variables and the other 6 PCs (PC3, PC4, PC5, PC7, PC8 and PC9) were not given, as they were not statistically significant and thus not having a linear relation.

Table 3.1: Principal Components (PCs) showing significant correlations with the sedimentological variables (cfr. Table 3.2), with their corresponding ecogeographical variables (EGVs) and factor loads (between brackets). Only those EGVs are given with factor loads < -0.5 or > 0.5, being the EGVs that are most explaining the PCs.

PC1	PC2	PC6
mcurv_13 (-0.89)	slp_13 (-0.89)	plcurv_21 (-0.67)
mcurv_21 (-0.88)	slp_21 (-0.87)	plcurv_13 (-0.56)
prcurv_13 (-0.83)	slp_7 (-0.79)	plcurv_31 (-0.55)
prcurv_21 (-0.82)	slp_31 (-0.76)	
mcurv_7 (-0.74)	fd_13 (0.65)	
mcurv_31 (-0.72)	slp_3 (-0.62)	
prcurv_31 (-0.67)	fd_7 (0.56)	
prcurv_7 (-0.67)	fd_21 (0.54)	

(mcurv = mean curvature, prcurv = profile curvature, slp = slope, plcurv = plan curvature, fd = fractal dimension, 3, 7, 13, 21 and 33 are multi-scale indices).

Table 3.2: Pearson correlation coefficients between the sedimentological variables and the Principal Components (PCs) and their statistical significance values (p). Only those PCs and correlation coefficients are given that have a statistical significant correlation. Those PCs were used as secondary variables for the Kriging with an external drift analysis.

		PC1	PC2	PC6
d_s10	Pearson correlation		-.537**	.355**
	p		.000	.001
d_s50	Pearson correlation		-.524**	.377**
	p		.000	.000
d_s90	Pearson correlation	-.284**	-.537**	.387**
	p	.008	.000	.000
Silt-clay%	Pearson correlation		.260*	-.263*
	p		.012	.011

**** Correlation is significant at the 0.01 level, * Correlation is significant at the 0.05 level.**

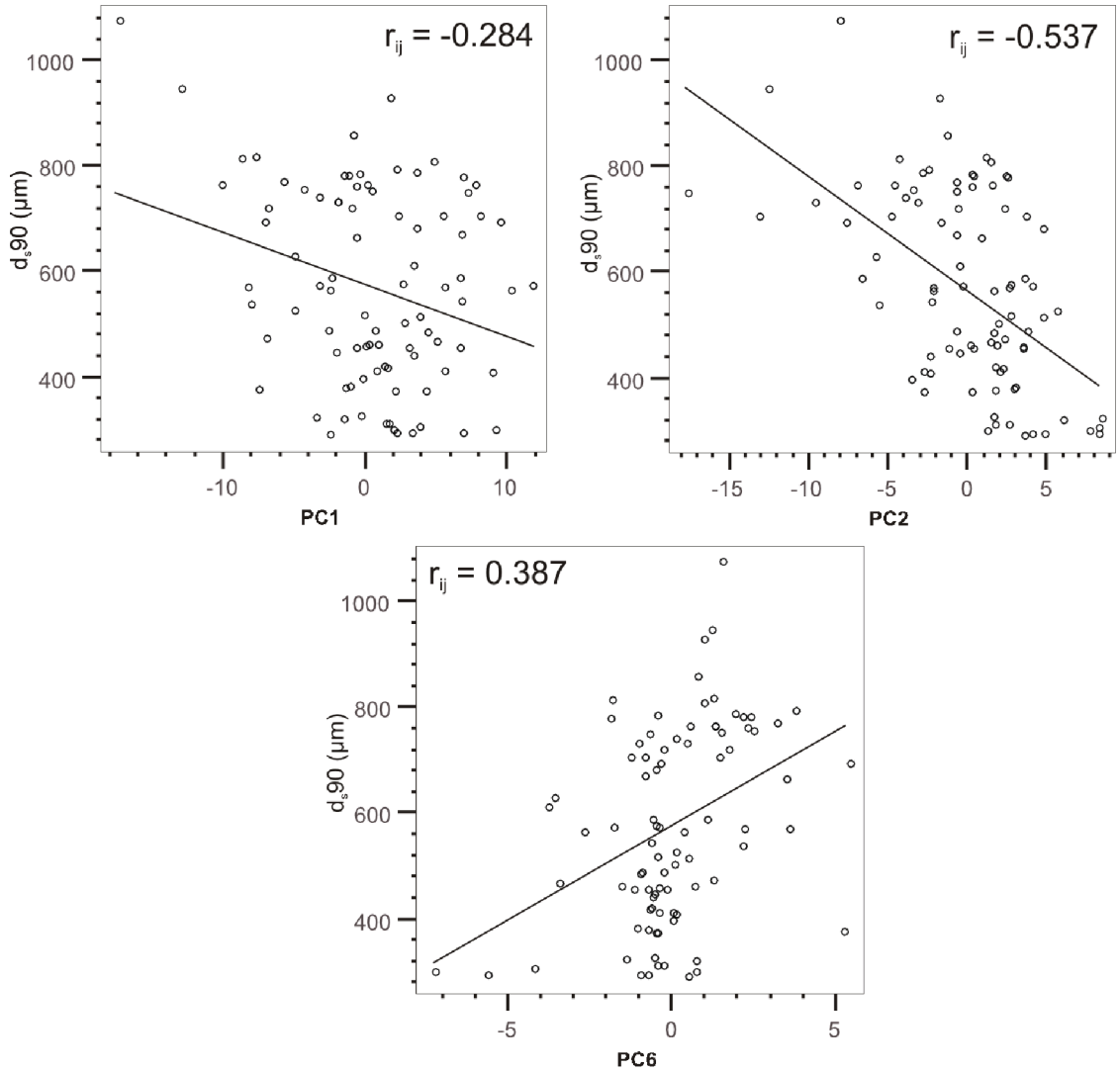

Figure 3.3: Scatter plots showing the Pearson correlation coefficients (r_{ij}) of Table 3.2 between d_s90 and the Principal Components (PCs). Correlation coefficients and scatter plots between d_s10, d_s50 and PC2 and PC6 are very similar; as such scatter plots are not presented. Correlation coefficients between the silt-clay% and PC2 and PC6 are very weak. As such, those scatter plots are not presented.

3.3.2 Interpolation with OK and KED

The variograms for OK and KED are presented in respectively Figure 3.4 and 3.5. All variograms of the sedimentological variables could be fit in a relatively straightforward way, except that of the silt-clay%, which behaved more unstable, due to the relative small values of this variable and the impact of a larger-scale trend.
The variogram surface for each sedimentological variable did not show any obvious anisotropy, still the direction of the strike of the sand dunes (120°, expressed as a trigonometric angle) was considered as the direction of the highest continuity. This means that, in this direction, it was expected that the sedimentological variables were more continuous than in other directions. It is logical that in the direction of the strike of a sand dune, similar sedimentological characteristics are found, while those characteristics are different in a perpendicular direction. Two OK variograms and data

grids per sedimentological variable were created, with an omnidirectional and a directional variogram (being the direction of the strike of the sand dunes). The two results were compared, based on their validation indices: for d_s10 and silt-clay%, a directional variogram gave the best result, whilst for d_s50 and d_s90, an omnidirectional variogram scored best.
For KED, the direction of the strike of the sand dunes, was considered as a drift-free direction. As such, the variogram of this direction was considered as omnidirectional and was used for the analysis.

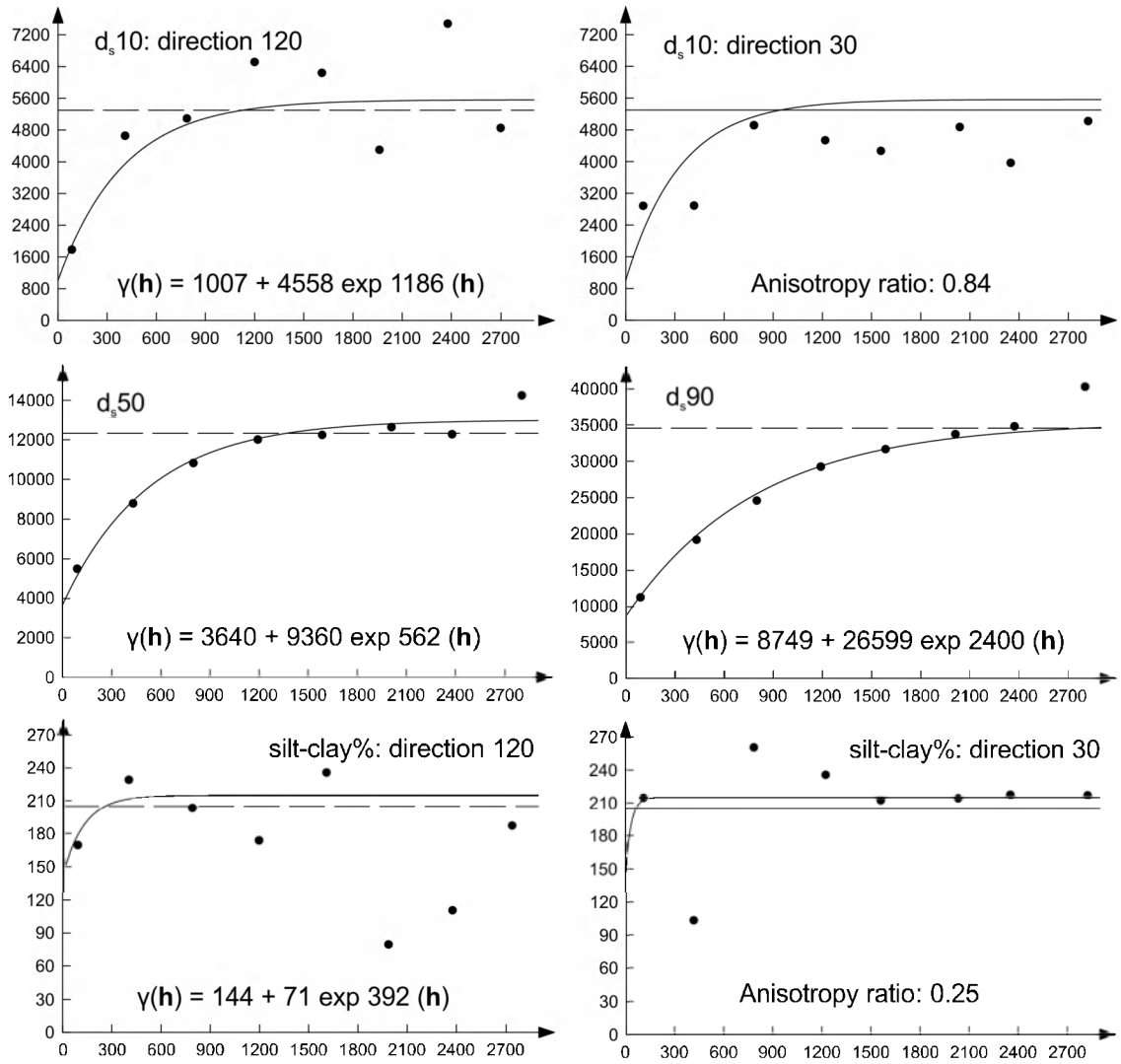

Figure 3.4: Experimental and fitted variograms for Ordinary Kriging (OK): X-axis represents lag distance (m) and the Y-axis is the semivariance (units are µm² for d_s10, d_s50, d_s90 and %² for silt-clay%). Variogram models are expressed as $\gamma(h) = C_0 + C_1 \exp a(h)$, with C_0 = nugget effect, C_1 = sill, exp = exponential model and a(h) = practical range. Practical ranges are equal to the distance at which 95% of the sill has been reached. Directions are expressed as trigonometric angles (zero degrees = east increasing counter clock wise).

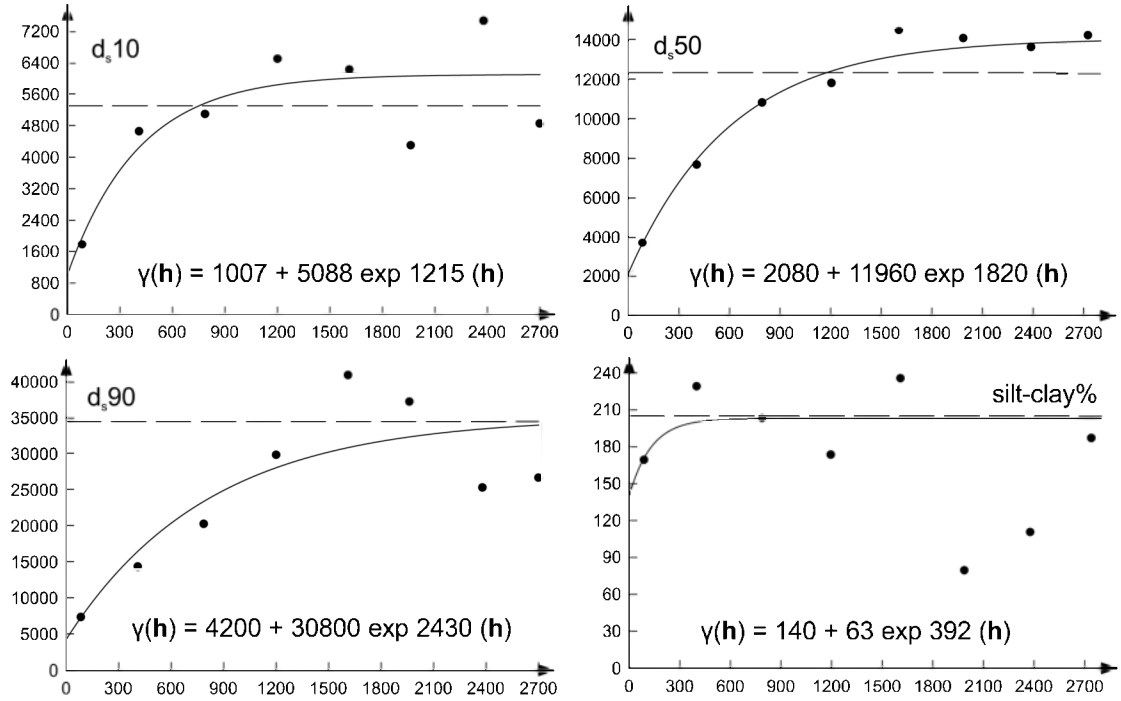

Figure 3.5: Experimental and fitted variograms for Kriging with an external drift (KED), in the direction of the strike of the sand dunes (120° expressed as a trigonometric angle; zero degrees = east increasing counter clock wise); they are considered omnidirectional, because of the assumption that this direction is drift-free. The X-axis represents the lag distance (m) and the Y-axis is the semi-variance (units are µm² for d_s10, d_s50, d_s90 and %² for silt-clay%). Variogram models are expressed as $\gamma(h) = C_0 + C_1 \exp a(h)$, with C_0 = nugget effect, C_1 = sill, exp = exponential model and a(h) = practical range. Practical ranges are equal to the distance at which 95% of the sill has been reached.

Figure 3.5 shows the maps of the resulting sedimentological data grids, modelled with OK and KED. The blanked zones are due to missing data; their surface area has been enlarged due to the multi-scale analysis (with window sizes of maximum 31 cells).
The results of d_s10, d_s50 and d_s90 are very similar. As such, no outliers of extreme fine or coarse fractions are present; the sediment is very homogeneous and well sorted. The OK maps are smooth and rather unnatural, in the sense that they show concentric patterns around the data points, whilst the KED maps reflect well the variation of the natural environment. Still, the two methodologies showed the same trend: coarser grain-sizes on the sand dunes and finer grain-sizes between and away from the sand dunes. The influence of the underlying topography was very clear in the results from KED. The same trend, showing a difference between the sand dunes (low silt-clay%) and the area away from the dunes (higher silt-clay%), holded true for the silt-clay%. The rough, mottled pattern away from the dunes, and visible on all of the KED maps, was due to the presence of dense colonies of tube worms; their existence was validated with extensive terrain verification.

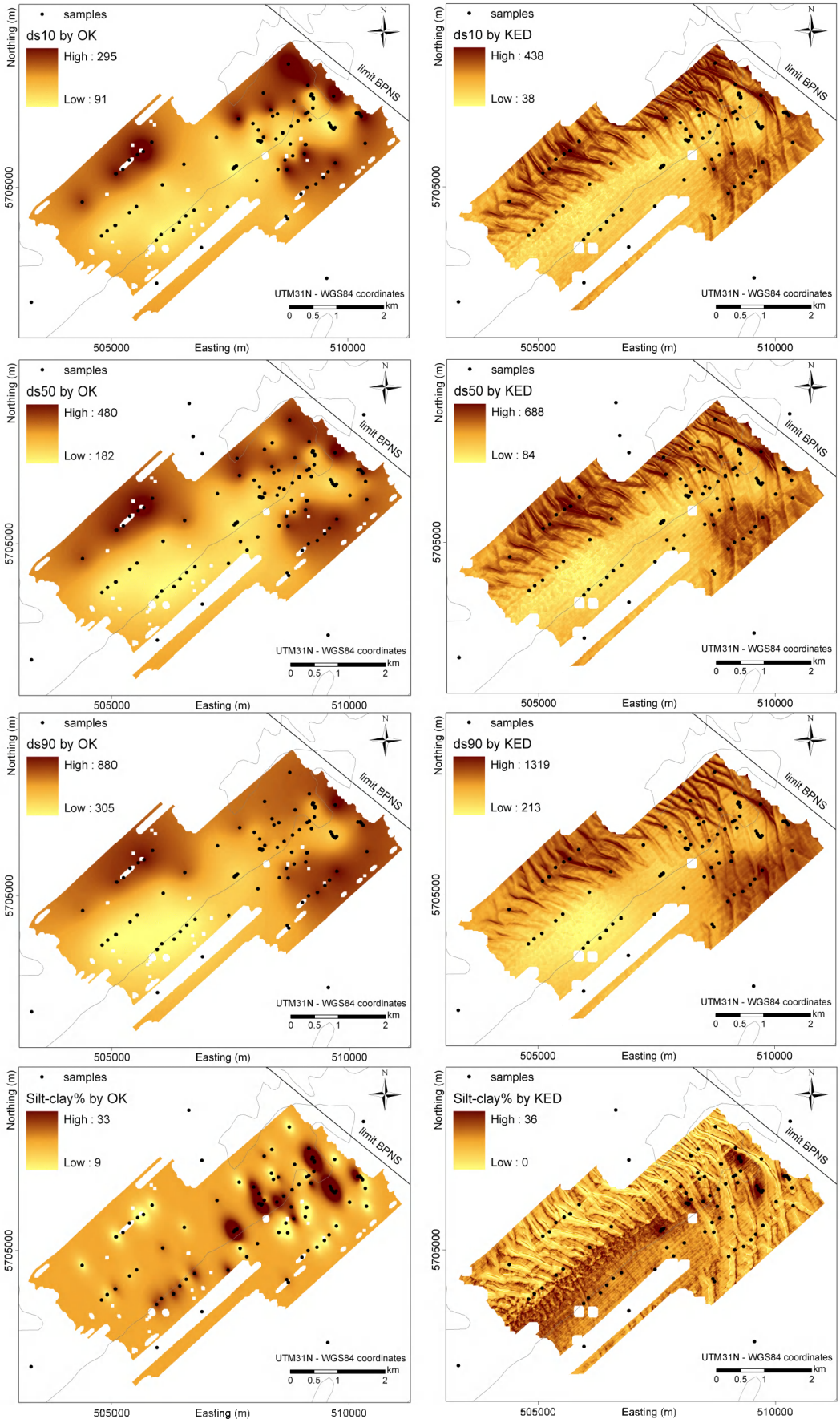

Figure 3.6: Sedimentological maps, based on Ordinary Kriging (OK) (left) and Kriging with an external drift (KED) (right).

3.3.3 Comparison of OK and KED

The validation indices are given in Table 3.3. KED provided a better result, compared to OK for all of the indices of d_s10, d_s50 and d_s90. From this, the KED results of d_s10, d_s50 and d_s90 could be considered better than those of OK.
For the silt-clay%, the result of OK was highly comparable to the result of KED. The MEE and Pearson correlation coefficient between the observed and the estimated values were better for OK compared to KED. The other validation indices were slightly better for KED compared to OK. This was due to the low correlation coefficient between silt-clay% and PC2 and PC6 (Table 3.2), meaning that the contribution of the secondary variables for KED was limited. The significant correlation coefficients between d_s10, d_s50, d_s90 and the PCs were all significant at the 0.01 level, while for silt-clay%, the correlation was significant at the 0.05 level (the lower the significance level, the stronger the evidence) (Table 3.2).
Next to the better validation indices, KED gave visually more natural maps.

Table 3.3: Validation indices (cfr. Materials and Methods) of different sedimentological data grids.
Except for the MEE and the Pearson correlation coefficient of the silt-clay%, all validation indices give better results for Kriging with an external drift (KED) compared to Ordinary Kriging (OK).

	d_s10_{OK}	d_s10_{KED}	d_s50_{OK}	d_s50_{KED}	d_s90_{OK}	d_s90_{KED}	$Sc\%_{OK}$	$Sc\%_{KED}$
MEE	2.44	**-0.55**	6.85	**-1.22**	3.09	**2.48**	**-0.42**	-0.51
RMSEE	63.01	**56.50**	93.51	**82.78**	134.78	**121.68**	13.09	**13.04**
MAEE	46.32	**40.47**	71.99	**64.69**	104.04	**93.82**	9.90	**9.82**
r	0.52	**0.64**	0.55	**0.67**	0.68	**0.75**	**0.50**	0.46

Sc% = silt-clay%, in bold are the best results.

3.4 Discussion

The aim of this paper was to create high quality sedimentological data grids, using multiple sources of secondary information. Next, the following items will be discussed: the secondary variables for KED and the comparison between OK and KED.

3.4.1 Secondary variables for KED

The proposed methodology allowed using a whole set of secondary variables. Here, 34 multi-scale terrain EGVs were derived from the DTM (slope, eastness, northness, profile curvature, plan curvature, mean curvature and fractal dimension). All of them were calculated on 5 different spatial scales, ranging from fine- to large-scale. A PCA reduced the large number of secondary variables to 9 PCs. Three of these PCs correlated significantly with the sedimentological variables. The PCA allowed maintaining a maximum of information, but avoided redundancy of correlating data.
For all of the sedimentological variables, there was a similar subset of PCs and EGVs, correlating significantly with the sedimentology (Table 3.1): mean, profile and plan curvature; slope and fractal dimension, on all different spatial scales. This means that a combination of different spatial scales was important in explaining the

sedimentological variation. Mainly the larger window sizes of 13, 21 and 31 (or 65, 105 and 155 m) were well represented, but also the smaller window sizes of 3 and 7 cells (or 15 and 35 m) were important. Mainly the larger distances were well suited to explain the sedimentological variability imposed by bedforms having wavelengths of around 100 m (very large dunes sensu Ashley 1990), but the smaller distances corresponded more with the smaller dunes (large dunes sensu Ashley 1990). Mainly the EGVs, associated with PC2 and PC6 (multi-scale slope, fractal dimension and plan curvature), were responsible for the overall sedimentological variation, as all of the sedimentological variables were correlated with those PCs. Such a slope – grain-size correlation has also been detected on sandy beaches (McLachlan 1996), while Azovsky et al. (2000) detected a correlation between grain-size and fractal dimension. Fractal dimension (Mandelbrot 1983) is often referred to as a measure of the surface complexity; as such it can be linked to habitat complexity of macrofauna (Kostylev et al. 2005).
Besides topography, possibly other EGVs correlate with the sedimentology and could be valuable secondary datasets for a multivariate geostatistical interpolation: e.g. the correlation between silt and nutrient richness (Greulich et al. 2000); between sand and organic matter content (Mantelatto and Fransozo 1999); and between grain-size and bottom current strength (Revel et al. 1996). Still, no high resolution datasets, other than the DTM, were available for this study area.
Categorical EGVs could be valuable secondary datasets as well (Hengl et al. 2007c). An example of such a dataset could be acoustic seabed classes of the sediment, derived from the classification of multibeam backscatter strength (Van Lancker et al. 2007) or side-scan sonar classes. Still, this information was not available for this study area.

3.4.2 Comparison of KED and OK

Validation indices, as presented in Table 3.3, are a valuable tool, though they permit only a comparison of different interpolation methods, applied on the same dataset. A d_s50_{OK} and a d_s50_{KED} map can be compared and the best result can be evaluated. It is more difficult to compare results from e.g. the d_s10_{KED}, d_s50_{KED}, d_s90_{KED} and silt-$clay_{KED}$ data grids. To overcome this issue, the correlation coefficients of the observed versus the estimated values can be compared. For this study, the coefficient indicates that d_s90_{KED} map is the most reliable.
The validation indices can be compared with the accuracy of the sedimentological variables. The accuracy of the sedimentological analyses is in the range of 1 % (Malvern Instruments 2008). The differences between OK and KED were well above this analytical accuracy. For example, the RMSEE of d_s50 reduced with 10.73 µm (Table 3.3), which represents a relative gain of 11.45 %. For the silt-clay%, where the RMSEE only reduced with 0.05 %, the difference in accuracy between OK and KED was negligible. The interpolation of the silt-clay% was less straightforward than the interpolation of the d_s10, d_s50 and d_s90. This poor increase in accuracy between both interpolation methods was mainly due to the small correlation coefficients between the silt-clay% and the PCs.

3.5 Conclusion

This paper proposed a multivariate geostatistical approach to obtain high quality sedimentological data grids of d_s10, d_s50, d_s90 and silt-clay%. KED was used with multiple secondary variables on different spatial scales, all derived from a DTM of the bathymetry. The sedimentological data were interpolated also with OK, and validation indices enabled to compare both results. For all of the sedimentological variables, KED gave the best result, although the results for the silt-clay% for both OK and KED, were very similar. The maps, based on KED, showed a different pattern on the sand dunes and away from and between the sand dunes. The sand dunes are composed of coarser sand, whilst the zones away from them have finer grain-sizes. The same difference can be observed for the silt-clay%: a high silt-clay% away from the dunes is observed and a low silt-clay% on the sand dunes. This pattern is not at all clear when the results, obtained with OK, were evaluated.
These highly detailed sedimentological data grids are the key for the adequate prediction of biological species, communities or habitats. This is especially the case for the predictive modelling of soft-substrata macrobenthos, of which the occurrence relates highly with sedimentological gradients (e.g. Degraer et al. 2008).

3.6 Acknowledgements

This paper is part of the PhD research of the first author, financed by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). It contributes also to the EU INTERREGIIIB project MESH (''Development of a framework for Mapping European Seabed Habitats''). In addition, it frames into the research objectives of the project MAREBASSE (''Management, Research and Budgeting of Aggregates in Shelf Seas related to End-users'', Belgian Science Policy, SPSDII, contract EV/02/18A). The sedisurf@ database has been compiled at the Renard Centre of Marine Geology and consists of data from various institutes.
Tomislav Hengl is acknowledged for his contribution to the R-gstat computation.