

## **Chapter 6**

# **The relevance of ecogeographical variables for marine habitat suitability modelling of *Owenia fusiformis***

SUBMITTED FOR PUBLICATION AS:

Verfaillie, E., Degraer, S., Du Four, I., Rabaut, M., Willems, W. and Van Lancker, V. (submitted to Estuarine, Coastal and Shelf Science) The relevance of ecogeographical variables for marine habitat suitability modelling of *Owenia fusiformis*.

## 6 The relevance of ecogeographical variables for marine habitat suitability modelling of *Owenia fusiformis*

---

### **Abstract**

Predictive modelling of ecologically relevant marine habitats requires predictor or ecogeographical variables (EGVs) that govern the potential distribution of a species or community.

This paper shows for the Belgian part of the North Sea how different combinations of EGV subsets result in different habitat suitability models of the macrobenthic species *Owenia fusiformis*. This tube-building polychaete belongs to the ecologically rich *Abra alba* community, is living in a well-defined habitat and is strongly linked to the sediment composition and topography of the seabed. Therefore, subsets of sedimentary, multi-scale topographical and other EGVs (e.g. hydrodynamics) were used to predict the distribution of the species. In earlier studies, only sedimentological and bathymetrical EGVs were considered for this species.

The habitat suitability models were derived from ecological niche factor analysis (ENFA) and the best model was selected using cross-validation. The validation showed that topographical EGVs on the smallest spatial scale were crucial for a good habitat suitability model. Surprisingly, a model based exclusively on these topographical EGVs is better than a model that also includes sedimentological and hydrodynamical EGVs.

*Keywords:* Habitat suitability modelling, ecogeographical variables, *Owenia fusiformis*, ecological niche factor analysis, Belgian part of the North Sea

---

## 6.1 Introduction

*Owenia fusiformis* is a macrobenthic species that is common on the Belgian part of the North Sea (BPNS). The species belongs to the ecologically important *Abra alba* community (Van Hoey et al. 2004; Degraer et al. 2008), characterized by high densities and diversity. *Owenia fusiformis* can be considered as an ecosystem engineer influencing the benthic community locally: there is a positive correlation between the occurrence of *O. fusiformis* and the species abundance/richness of the community (Somaschini 1993). During the last decades, this tube-building polychaete has shown a considerable increase in density. Between 1976 and 1986, the species was found in low densities of maximum 15 ind./m<sup>2</sup>, whereas in the period 1994-2001, densities increased to 500 ind./m<sup>2</sup> (Degraer et al. 2006). In recent years, the species has reached average densities of 165 ind./m<sup>2</sup>, with dense aggregations of more than 4000 tubes per m<sup>2</sup> locally (Marine Biology Section, Ugent – Belgium, 2008).

The species lives in flat, soft-sediment environments. It prefers fine-to-medium sands with a grain-size between 100 and 500 µm and is characteristic of sheltered areas with high percentages of organic matter (Fager, 1964). *O. fusiformis* forms dense aggregations or patches. This clustering behaviour is comparable to that of the tube-building, habitat-engineering polychaete *Lanice conchilega*, (Rabaut et al. 2007). The tubes of *O. fusiformis* are shorter (i.e. 12-13 cm; Fager 1964) and the organisms have a longer lifespan (i.e. on average three years; Ménard et al. 1989).

The flexible tube of cemented sand grains and shell fragments is longer than the worm itself (Hartmann-Schröder 1996) and protrudes from the surface. When the species occurs in dense aggregations, the biogenic structures of protruding tubes can be detected by acoustic methods. The patches described in Van Lancker et al. (2007) protruded 18 to 40 cm above the surrounding sediment.

Classification and modelling techniques used in habitat mapping are all based on the assumption that the suitability of an area for a certain species is related to its predictor or ecogeographical variables (EGVs), corresponding to the environmental variables relating to factors of potential relevance to the focal species (e.g. substrate type, topographical position, hydrodynamical regime or presence/absence of another species). The ecogeographical information is generally more widely available than ecological sample information of the focal species. As such, it becomes possible to predict how suitable a habitat is for a certain species, using specific combinations of these EGVs in a habitat suitability model (HSM).

For soft-substrate habitats, grain-size and silt-clay percentage are commonly thought to be the most influential EGVs for the modelling of macrobenthic species (Van Hoey et al. 2004; Degraer et al. 2008; Willems et al. 2008). However, the spatial distribution of some macrobenthic species or communities is known to be patchy or bound to topographical variation (Rabaut et al. 2007). Multi-scale topographical characteristics are therefore believed to be important EGVs too (Guisan and Thuiller 2005; Baptist et al. 2006; Wilson et al. 2007).

This paper demonstrates the influence of different combinations of EGVs on the result of the habitat suitability (HS) modeling of *O. fusiformis*. The different combinations of EGVs include or exclude specific (multi-scale) topographical, sedimentological or other EGVs.

## **6.2 Material and methods**

### **6.2.1 Study area and datasets**

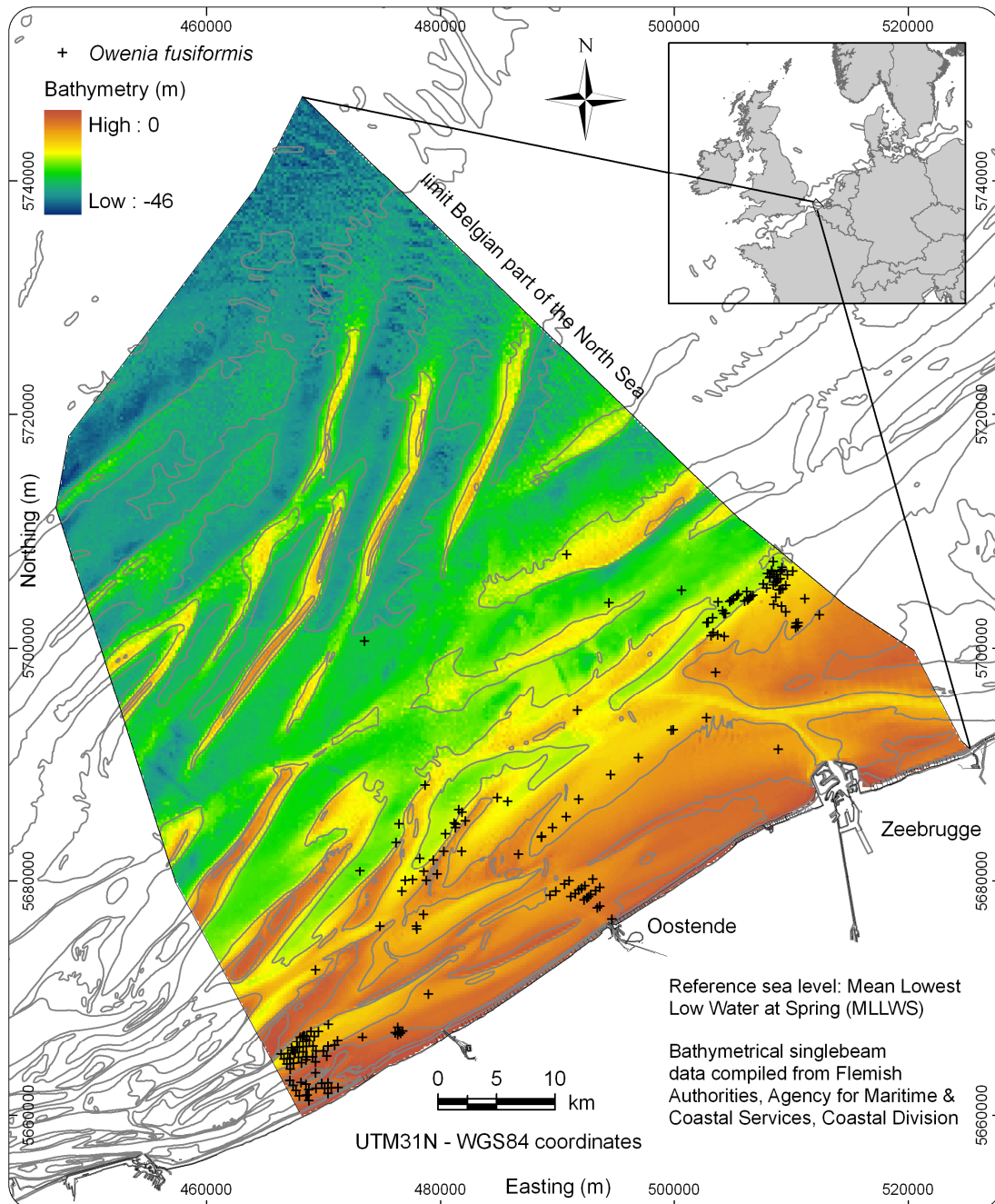
The study area covers the entire BPNS with a surface of 3600 km<sup>2</sup>, situated on the North-West European Continental Shelf (Figure 6.1). The BPNS is relatively shallow and dips gently from 0 to -50 m MLLWS (Mean Lower Low Water Springs). A highly variable topography, dominated by a series of sandbanks and swales, characterizes the seabed surface.

Thirty-seven full-coverage topographical, sedimentological and hydrodynamical EGVs were used, all available as raster maps of 250 x 250 m resolution (Table 6.1). The variables were made more symmetrical by the Box-Cox standardizing algorithm (Sokal and Rohlf 1981), a procedure that produces a distribution as close to a Gaussian as possible.

*Owenia fusiformis* was found in 193 stations at the BPNS (Figure 6.1) (Marine Biology Section, Ugent – Belgium, 2008). For each biological sample, the density of *O. fusiformis* was calculated. Densities were expressed as individuals/m<sup>2</sup> and were divided into four classes, giving different weights to the presence maps: 1) 1-10 specimens; 2) 11-100 specimens; 3) 101-1000 specimens, 4) 1001-6000 specimens.

### **6.2.2 Research strategy**

The research strategy comprised four steps: (1) selection and production of EGVs; (2) Ecological Niche Factor Analysis (ENFA), identifying ecologically relevant factors; (3) production of HS models of *O. fusiformis*, using the geometric mean distance algorithm and different combinations of EGVs; (4) cross-validation of different HS models, enabling the identification of the best explanatory EGVs for each study area. Software used was ArcGIS 9.2 for GIS analyses and mapping; Landserf 2.2.0 (Wood 2005) for multi-scale topographical analyses; Biomapper 3.2 (Hirzel et al. 2002b; and Hirzel et al. 2006) for ENFA, HS modelling and validation.



**Figure 6.1: Bathymetrical map of the Belgian part of the North Sea, located in northwestern Europe, with indication of presence samples of *Owenia fusiformis*.**

### 6.2.3 Step 1: Selection and production of EGVs

The median grain-size of the sand fraction and the silt-clay percentage were used as sedimentological EGVs (Table 6.1).

Following Wilson et al. (2007), terrain-analysis variables can be grouped into four classes, all of them being derivatives from a digital terrain model (DTM):

- slope;
- orientation (aspect);
- curvature and relative position of features;
- terrain variability.

All of these variables can be computed in GIS software or specialized terrain-analysis software (e.g. LandSerf from Wood (2005)).

Terrain analysis may be performed on a DTM represented as a raster grid, or on a continuous representation of a DTM as a double-differentiable surface (e.g. Wood 1996). The latter approach offers great flexibility in the choice of algorithms of terrain analysis and in the scales at which the analyses are performed. Following Evans (1980), a DTM is approximated by a bivariate quadratic equation:

$$Z = aX^2 + bY^2 + cXY + dX + eY + f \quad (6.1)$$

with  $Z$  = height of the DTM surface;  
 $X$  and  $Y$  = horizontal coordinates;

The coefficients  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $e$  and  $f$  in equation 6.1 can be solved within a window using simple combinations of neighboring cells, which is the basis for terrain analysis in most GIS software, regardless of their using grid-based methods or a mathematical representation of the DTM.

For a terrain analysis across a variety of spatial scales, Wood (1996) solves this equation for an  $n$  by  $n$  matrix with a local coordinate system ( $x$ ,  $y$ ,  $z$ ) defined with the origin at the central pixel. The user may specify any odd number ( $n$ ) for the size of the square analysis window defining the part of the raster DTM to be analyzed in relation to each central pixel in turn. To compute the terrain parameters in LandSerf (Wood 2005), an analysis window is effectively moved across the raster DTM surface such that each pixel in turn becomes the central pixel on which calculations are based.

For the BPNS, the multi-scale terrain analysis resulted in a range of EGVs such as slope and other bathymetrical derivatives, measured over four spatial scales (see Table 6.1 for an overview). Window sizes of 3, 9, 17 and 33 pixels, corresponding to 750, 2250, 4250 and 8250 m, respectively, were used for this analysis. The window sizes of 3, 9, 17 and 33 pixels, were chosen because they provide an adequate cover of different spatial scales, following Wilson et al. 2007.

Hydrodynamical EGVs were the maximum bottom shear stress and the maximum current velocity (Table 6.1). Furthermore, the MERIS satellite-derived EGVs maximum Chlorophyll *a* concentration and maximum Total Suspended Matter were used (Table 6.1). Finally, the distance to the coastline was computed (Table 6.1). The resolution of all of the EGVs was 250 x 250 m, except for maximum Total Suspended Matter and maximum Chlorophyll *a* concentration; these had original resolutions of 1 x 1 km, but were resampled to 250 x 250 m.

**Table 6.1: Full-coverage ecogeographical variables (EGVs) as input for the Ecological Niche Factor Analysis and habitat suitability modelling. The multi-scale topographical EGVs have window sizes of 3, 9, 17 and 33 pixels. Fractal dimension EGVs have only 3 window sizes of 9, 17 and 33 pixels. (/ = no unit)**

EGV	Unit	Reference or procedure
<p><i>Sedimentology</i></p> <p><math>d_x</math> = diameter for which x% of the sand fraction (63-2000 <math>\mu\text{m}</math>) in the sample has a smaller diameter</p> <ul style="list-style-type: none"> <li><math>d_{50}</math> = median grain-size of sand fraction</li> <li>Silt-clay percentage (silt-clay %) = 0-63 <math>\mu\text{m}</math></li> </ul>	<p><math>\mu\text{m}</math></p> <p>%</p>	<p>Reference: sedisurf@database hosted at Ghent University, Renard Centre of Marine Geology.</p> <p>Reference: Verfaillie et al. 2006</p> <p>Reference: Van Lancker et al. 2007</p>
<p><i>Topography</i></p> <ul style="list-style-type: none"> <li>Digital terrain model (DTM) of multibeam bathymetry</li> </ul>	m	<p>Reference: DTM from Flemish Authorities, Agency for Maritime and Coastal Services, Flemish Hydrography</p>
<ul style="list-style-type: none"> <li>Slope (slp) = a first derivative of the DTM</li> </ul>	$^{\circ}$	<p>Procedure: Evans (1980); Wilson et al. (2007)</p>
<p>Aspect = a first derivative of the DTM</p> <p>Indices of northness and eastness provide continuous measures (-1 to +1), describing the orientation of the slopes.</p> <ul style="list-style-type: none"> <li>Eastness (eastn) = sin (aspect)</li> <li>Northness (northn) = cos (aspect)</li> </ul>	/	<p>Procedure: Wilson et al. (2007); Hirzel et al. (2002a)</p>
<p>Surface curvature = a second derivative of the DTM</p> <ul style="list-style-type: none"> <li>Profile curvature (prcurv) = rate of change of slope along a profile in the surface; useful to highlight convex and concave slopes</li> <li>Plan curvature (plcurv) = rate of change of aspect in plan across the surface; useful for defining ridges, valleys and slopes</li> <li>Mean curvature (mcurv) = average value obtained from maximum and minimum profile curvature, providing an indication of the relative position of features</li> </ul>	/	<p>Procedure: Evans (1980); Wilson et al. (2007)</p>
<ul style="list-style-type: none"> <li>Rugosity (rug) = ratio of the surface area to the planar area, across the neighbourhood of the central pixel</li> </ul>	/	<p>Procedure: Jenness (2002); Lundblad et al. (2006); Wilson et al. (2007)</p>
<p>Bathymetric Position Index (BPI) = measure of where a location, with a defined elevation, is relative to the overall landscape</p> <ul style="list-style-type: none"> <li>BPI with a window size of 6 pixels (BPI_1500m)</li> </ul>	/	<p>Procedure: Lundblad et al. (2006); Wilson et al. (2007)</p>
<ul style="list-style-type: none"> <li>Fractal dimension (fd) = a measure of the surface complexity</li> </ul>	/	<p>Procedure: Mandelbrot (1983); Wilson et al. (2007)</p>
<p><i>Hydrodynamics</i></p> <ul style="list-style-type: none"> <li>Maximum bottom shear stress (bstrx) = frictional force, exerted by the flow per unit area of the seabed</li> <li>Maximum current velocity (mmax)</li> </ul>	<p><math>\text{N/m}^2</math></p> <p>m/s</p>	<p>Reference: Management Unit of the North Sea Mathematical Models and the Scheldt estuary (MUMM)</p>
<p><i>Satellite-derived variables</i></p> <ul style="list-style-type: none"> <li>Maximum Chlophyl a concentration (max Chl a) over a 2-year period (2003-2004)</li> <li>Maximum Total Suspended Matter (max TSM): measure for turbidity over a 2-year period (2003-2004)</li> </ul>	<p>mg/<math>\text{m}^3</math></p> <p>mg/l</p>	<p>Reference: MERIS data processed by MUMM in the framework of the BELCOLOUR-2 project (ESA ENVISAT AOID3443)</p>
<ul style="list-style-type: none"> <li>Distance to coast (distcst)</li> </ul>	km	Computed in GIS

#### 6.2.4 Step 2: Ecological Niche Factor Analysis

ENFA (Hirzel et al. 2002) is a statistical technique, based on Hutchinson's (1957) concept of the ecological niche, that computes suitability functions for species by comparing the EGVs of the species with those of the whole set of cells. Unlike other HS modelling techniques, ENFA needs only presence data of species. This was appropriate for the current dataset, since some samples were taken only where the presence of *O. fusiformis* was likely, as determined from highly detailed multibeam observations.

Contrary to PCA, where axes are chosen to maximize the variance of the distribution, ENFA computes ecologically relevant factors. Still, the output of ENFA is similar to that of PCA, with the results being sets of new, linearly independent variables, combining the original EGVs.

Species are generally expected to show non-random distributions with respect to EGVs, meaning that a species with e.g. an optimum depth is expected to occur within this optimal range. As such, the depth distributions of the cells in which species are observed, in comparison with the whole set of cells, may be quantified. These distributions may be different for different species, regarding their mean and standard deviations.

For one single EGV, the species' marginality ( $M$ ) can be defined as the absolute difference between the global mean ( $m_G$ ) and the species mean ( $m_S$ ), divided by 1.96 standard deviations ( $\sigma_G$ ) of the global distribution. A large  $M$  value close to 1, means that the species lives in a very particular habitat relative to the reference set. The operational definition of marginality as it is implemented in the Biomapper software (Hirzel et al. 2002b), is a multivariate extension of the species' marginality (i.e. the *global* marginality). This is an overall marginality  $M$  computed over all EGVs, allowing the comparison of the marginalities of different species within a given area. Similarly, the specialization  $S$  for one single EGV can be defined as the ratio of the standard deviation of the global distribution ( $\sigma_G$ ) to that of the focal species ( $\sigma_S$ ). Any  $S$  value exceeding 1, indicates some form of specialization (i.e. a narrow niche breadth in comparison with the available conditions). Again, a global specialization index for all of the EGVs, can be computed. This value ranges from 1 to infinity. For ease of interpretation, the global tolerance coefficient, defined as the inverse of the specialization, is usually preferred as it ranges from 0 to 1. It is an indicator of the species' niche breadth.

The multivariate niche (i.e. for all of the EGVs) can be quantified on any of its axes by an index of marginality and specialization. ENFA chooses its first axis to account for all the marginality of the species, and the following axes to maximize specialization.

The broken-stick method (MacArthur 1960; Frontier 1976; and Legendre and Legendre 1998) was used to decide on the number of factors to retain for the HS modelling.

To test and compare the performance of the models, eight combinations of EGVs were used to compute the factors (Table 6.2): (1) all EGVs; (2) all topographical EGVs on all of the spatial scales; (3) all topographical EGVs computed with window size 3 (including DTM and BPI); (4) all topographical EGVs computed with window size 9; (5) all topographical EGVs computed with window size 17; (6) all topographical EGVs computed with window size 33; (7) sedimentological EGVs; (8) sedimentological EGVs and all topographical EGVs computed with window size 3 (including DTM and BPI).



**Table 6.2: Different combinations of EGVs (/ = not used for analysis; X = used for analysis; M = used at multi-scale, with window sizes of 3, 9, 17 and 33 pixels; 3; 9; 17 and 33: used with window sizes of 3, 9, 17 and 33 pixels, respectively).**

	Combinations							
	1	2	3	4	5	6	7	8
d <sub>s</sub> 50	X	/	/	/	/	/	X	X
silt-clay %	X	/	/	/	/	/	X	X
DTM	X	X	X	X	X	X	/	X
slp	M	M	3	9	17	33	/	3
eastn	M	M	3	9	17	33	/	3
northn	M	M	3	9	17	33	/	3
prcurv	M	M	3	9	17	33	/	3
plcurv	M	M	3	9	17	33	/	3
mcurv	M	M	3	9	17	33	/	3
fd	M	M	/	9	17	33	/	/
rug	3	3	3	/	/	/	/	3
BPI_1500m	X	X	X	/	/	/	/	X
bstrx	X	/	/	/	/	/	/	/
mmax	X	/	/	/	/	/	/	/
max Chl a	X	/	/	/	/	/	/	/
max TSM	X	/	/	/	/	/	/	/
distcst	X	/	/	/	/	/	/	/

### 6.2.5 Step 3: Habitat suitability modelling

Several algorithms are implemented in Biomapper to compute, for each grid cell, the suitability for *O. fusiformis*: median (of the species distribution on all selected niche factors), distance harmonic mean, distance geometric mean (GM) and minimal distance algorithm (Hirzel et al. 2002; Hirzel and Arlettaz 2003). The GM algorithm (Hirzel and Arlettaz 2003) was selected for the present study, as validation (Step 4) showed that the other algorithms gave systematically worse results than the GM algorithm. For this algorithm, no assumptions are made on species distribution. The suitability of any point  $\mathbf{P}$  in the environmental factor space is the geometric mean  $H_G$  of  $N$  species-observation points  $\mathbf{O}_i$ , which is computed from the distances to all observations:

$$H_G(\mathbf{P}) = \sqrt[N]{\prod_{i=1}^N \delta(\mathbf{P}, \mathbf{O}_i)} \quad (6.2)$$

For the observations of *O. fusiformis*, a frequency distribution and a GM algorithm value are computed for each ecological niche factor. The farther away a grid cell is from this value, the less suitable it is. The suitability index ranges from 0 to 1.

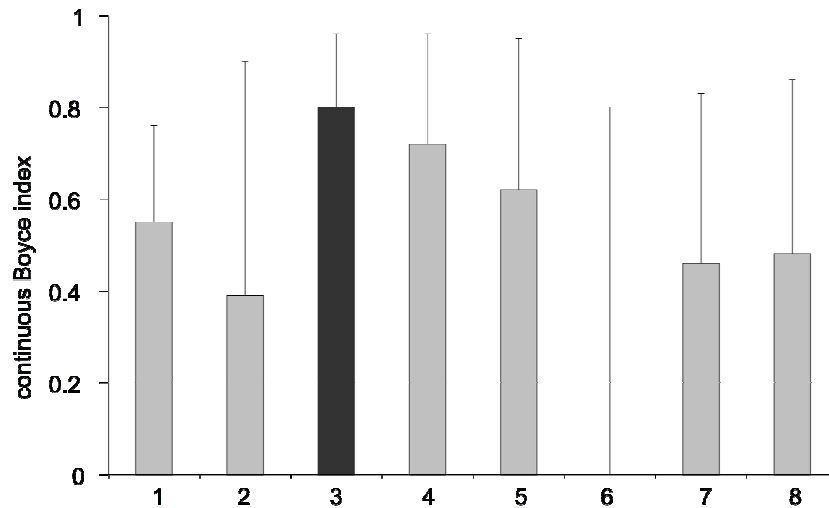
### 6.2.6 Step 4: Validation

A k-fold cross-validation splits the species data into k sets. Then, k-1 sets are used to compute a HS model and the remaining set is used to validate this model. This strategy is repeated k times. For each of the models, the cross-validation results in k different HS models. By determining how the results vary, their predictive power is assessed. For this study, use was made of the continuous Boyce index (Hirzel et al. 2006), following the approach of Boyce et al. (2002). This method partitions the habitat suitability range into  $b$  classes. For each class  $i$ , it calculates 2 frequencies: 1)  $P_i$ , the predicted frequency of evaluation points; and 2)  $E_i$ , the expected frequency of evaluation points or the frequency expected from a random distribution across the study area, given by the relative area covered by each class. For each class  $i$ , the predicted to expected ( $P/E$ ) ratio is  $F_i$ . A low-suitability class should contain fewer evaluation presences than expected by chance, resulting in  $F_i < 1$ , whereas high-suitability classes are expected to have  $F_i$  values higher than 1. A good model is thus expected to show a monotonously increasing curve (increase of  $F_i$  and increase of habitat suitability). Boyce et al. (2002) proposed to measure this monotonous increase by the Spearman rank correlation coefficient between  $F_i$  and  $i$  (i.e. “Boyce index” varying between -1 and 1, corresponding to a bad and a good model, respectively). Because of the sensitivity to the number of suitability classes  $b$  and to their boundaries, Hirzel et al. (2006) proposed a new index (“continuous Boyce index”), determined by using a moving window of width  $W$  (e.g. 0.2), instead of fixed classes. The computation of this index starts with a first class covering the suitability range  $[0, W]$  whose  $P/E$  ratio is plotted against the average suitability values of the class,  $W/2$ . In the next step, the moving window is shifted a short distance upwards and  $P/E$  is plotted again. This is repeated until the window reaches the last possible range  $[1-W, 1]$ . This results generally in a smooth  $P/E$  curve, from which a continuous Boyce index is computed.

For this study, the continuous Boyce index was calculated with a moving window size of 0.2. The eight HS models resulting from the eight combinations of EGVs (Table 6.2) were validated with a k value of 10.

## 6.3 Results

The validation using combination set 3 (all topographical EGVs computed with window size 3 (including DTM and BPI); Table 6.2) gave the best result (Figure 6.2). This model was thus selected as the final model for the BPNS. Combination sets 4, 5 and 1 (i.e. topographical EGVs with window sizes of 9 and 17; and all EGVs, respectively) also gave rather good results (indices  $> 0.5$ ). The model obtained by combination set 6 (i.e. topographical EGVs with a window size of 33, i.e. computed on a large spatial scale of more than 8 km), resulted in a very bad model (with a validation index equal to zero). Combination sets 2, 7 and 8 gave intermediate results.



**Figure 6.2: Continuous Boyce indices for the Belgian part of the North Sea for different combinations of EGVs (X-axis) (Table 6.2). The best combination is marked in black (combination 3; all topographical EGVs computed with window size 3 (including DTM and BPI) Error bars correspond to 1 standard deviation.**

ENFA, based on the final model, results in a global marginality coefficient of 0.63; a global specialization coefficient of 1.92 (i.e. a global tolerance coefficient of 0.52), indicating that *O. fusiformis* lives in conditions rather uncommon for the BPNS and that its niche breadth is generally slightly narrow. After comparing the ENFA eigenvalues with the broken-stick distribution, the first 7 factors were kept as significant for the analyses (Table 6.3). They explain 97% of the information (i.e. 100% of the marginality and 95% of the specialization).

*Owenia fusiformis* shows the highest marginality score for the bathymetry (DTM), meaning that the species prefers higher-than-average values (i.e. a preference for shallow locations). However, the specialization for this predictor is rather low, meaning that the species is quite tolerant regarding the depth. Regarding the BPI; the species prefers flatter-and-more-depressed-than-average conditions (i.e. preference for depressions and flat areas). Regarding eastness and northness, the species has a preference for more western and more northern orientations than average, although the species is again very tolerant for these predictors.

The HS map of *O. fusiformis* in Figure 6.3 shows that high suitabilities are mainly expected in the coastal zone extending no more than 30 km offshore. Highest suitabilities are found at the N side of the Vlakte van de Raan, extending towards the Netherlands; at the S and N side of the Oostendebank; at the S and N side of the Nieuwpoortbank; and between the Middelkerkebank and Kwintebank. To a lesser extent, high suitabilities are found at the N side of the Thorntonbank; at the S and N side of the Akkaertbank; and at the S side of the Wenduinebank.

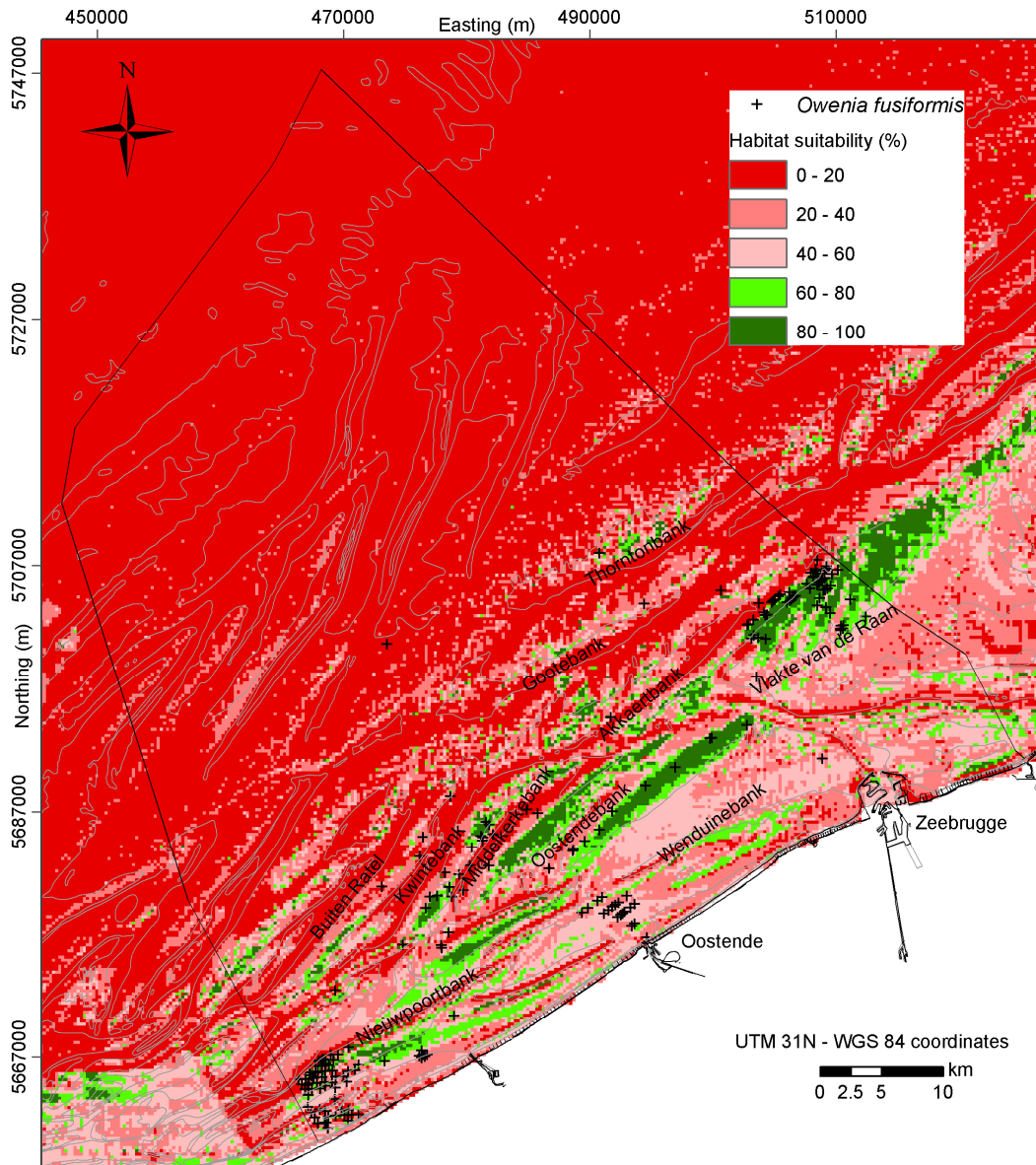
**Table 6.3: Correlation between ENFA factors and the EGVs for the final model. The percentages indicate the amount of specialization accounted for by the factor (factor 1 explains 100% of the marginality).**

	Factor 1 (20%)	Factor 2 (33%)	Factor 3 (13%)	Factor 4 (10%)	Factor 5 (9%)	Factor 6 (7%)	Factor 7 (3%)
BPI_1500m	---	0	*	**	**	*****	*
DTM	+++++++	0	***	0	*	***	0
eastn3	--	0	**	*	*****	**	*
mcurv3	--	*****	**	****	*****	0	*****
northn3	++	0	**	*	****	**	0
plcurv3	+	*	0	**	***	***	*****
pcurv3	--	*****	0	*****	**	***	*****
rug	---	*	*****	*	***	**	0
slp3	--	0	***	**	*	**	**

Factor 1 is the Marginality factor. Positive values indicate higher-than-average values. Negative values mean the reverse. The greater the number of symbols, the higher the correlation. 0 indicates a very weak correlation.

Factors 2 to 7 are Specialization factors. The symbol \* means *O. fusiformis* is found occupying a narrower range of values than available. The greater the number of \*, the narrower the range. 0 indicates a very low specialization.

Considering HS values  $\geq 60\%$  as highly suitable for the species, the following ranges of EGVs correspond to optimal environmental conditions (EGV ranges are obtained by selecting pixels with HS values higher than 60 %): (1) shallow-water environments (8 - 20 m); (2) fine sandy sediments (median grain-size between 145 and 285  $\mu\text{m}$ ); (3) moderately high silt-clay % (0.5 - 20 %); (4) high amounts of Total Suspended Matter ( $> 25 \text{ mg/l}$ ); (5) maximum Chlorophyll a concentration of 30 to 40  $\text{mg/m}^3$ ; (6) moderate maximum bottom shear stresses (1.2 - 2.3  $\text{N/m}^2$ ); (7) maximum current velocities of 0.8 to 1.0 m/s; (8) topographies with northwest orientations (northness  $> 0.5$  and eastness  $< 0$ ); and (9) flat topographies (slope around  $0^\circ$ ). Table 6.4 gives a summary of the mean and standard deviations of the EGVs. As multi-scale EGVs give similar values for different window sizes, only values for window size 3 are given. Although only topographical EGVs with a window size of 3 are used for the final model, other EGV ranges are also given in Table 6.4.



**Figure 6.3: Final habitat suitability model of the Belgian part of the North Sea for *Owenia fusiformis*. The optimal niche of *Owenia fusiformis* lies in the coastal zone (maximum 30 km offshore), in between the major Coastal Banks.**

**Table 6.4: Mean and standard deviations (SD) for the EGVs on the BPNS, corresponding to HS values  $\geq 60\%$ . The EGVs and their units are described in Table 6.1.**

EGV	Mean $\pm$ SD
d <sub>s</sub> 50	214.91 $\pm$ 68.43
silt-clay %	10.89 $\pm$ 10.73
DTM	-13.22 $\pm$ 5.49
slp	0.19 $\pm$ 0.14
eastn	-0.44 $\pm$ 0.38
northn	0.57 $\pm$ 0.58
prcurv	0.00 $\pm$ 0.01
plcurv	0.16 $\pm$ 2.47
mcurv	0.00 $\pm$ 0.01
rug	1.00 $\pm$ 0.00
BPI_1500m	0.07 $\pm$ 0.91
bstrx	1.79 $\pm$ 0.49
mmax	0.91 $\pm$ 0.11
max Chl a	35.25 $\pm$ 3.43
max TSM	34.21 $\pm$ 7.84
distcst	10.73 $\pm$ 5.99

## 6.4 Discussion

The aim of this paper is to examine which EGVs are the most important predictors of the species *O. fusiformis*.

### 6.4.1 Comparison of EGV conditions with values from the literature

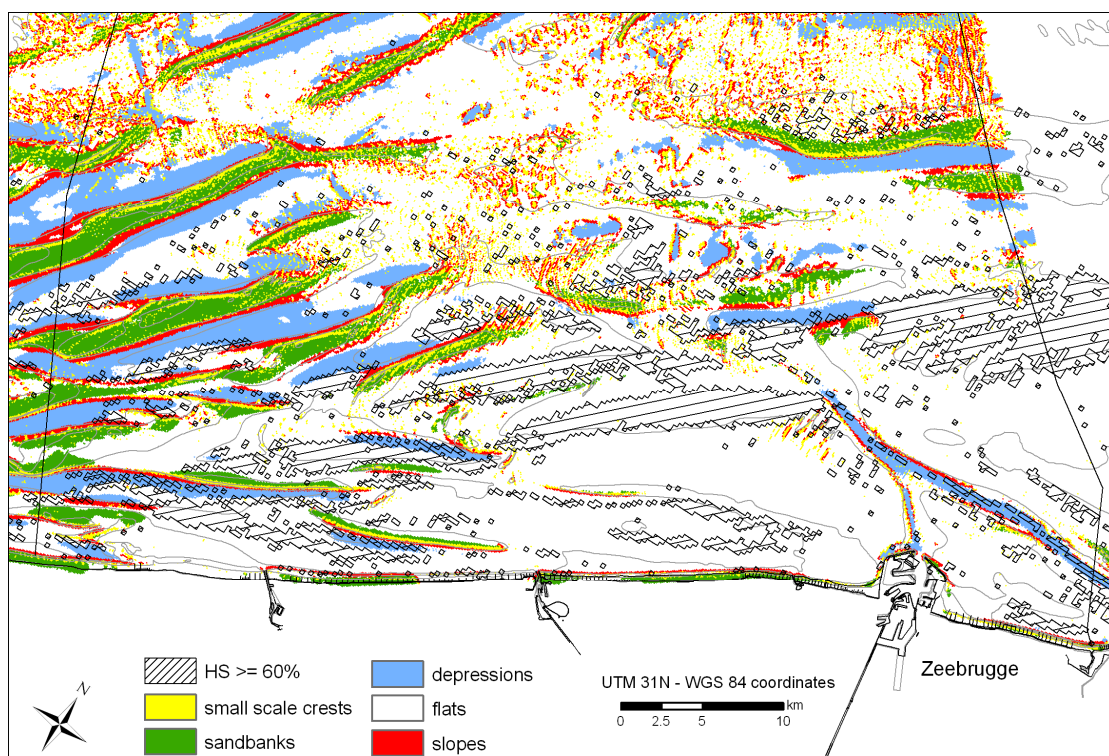
For the Baie de Seine (France), *O. fusiformis* has been described as one of the ten most abundant species (ind/m<sup>2</sup>) of the *Abra alba* community (Fromentin et al. 1997; Van Hoey et al. 2004). However, d<sub>s</sub>50 of that community varied between 80-120  $\mu$ m (being much finer sediment than what has been encountered in this study), while the mean depth is around 10.5 m (somewhat shallower than the value determined in this study). In Van Hoey et al. (2005), *Owenia fusiformis* was also identified as one of the ten most abundant species of the *A. alba* community on the BPNS. According to those authors, the mean d<sub>s</sub>50 of that community was 222  $\pm$  45  $\mu$ m, the mean silt-clay% was 14  $\pm$  11 % and the mean depth was 10.8 m. These results are highly comparable with the values found for *O. fusiformis* in this study. This means that regarding the sedimentological and the depth conditions, the *A. alba* community and *O. fusiformis* are very similar. Dauvin et al. (2004) described that *O. fusiformis* has a preference for sand with high percentages of silt-clay, which also corresponds to the results of this study.

To our knowledge, only sedimentological and bathymetrical EGVs have ever been considered for this species. This study demonstrates that other conditions (such as bathymetric derivatives) are also important predictors. Although a good model is obtained using all of the EGVs (combination set 1 in Figure 6.2) and thus including depth and sedimentology, the best model is obtained from topographical variables

alone, on a spatial scale of 750 m (including depth and BPI on a spatial scale of 1500 m; combination set 3 in Figure 6.2). This is a surprising result, as those EGVs do not contain any information other than topography (e.g. sedimentology or hydrodynamics).

#### 6.4.2 Spatial scale and spatial structure

Figure 6.2 shows that combinations 3, 4 and 5 give similar good results for modelling the occurrence of *O. fusiformis*. This indicates that the terrain variables with window sizes of 3, 9 and 17 pixels (or spatial scales of 750, 2250 and 4250 m, respectively; similar to the spatial scale of a sandbank) contain overlapping and no complementary information. The fact that the model based on a window size of 33 pixels (or 8250 m) has a zero validation index, means that terrain variables calculated on this spatial scale are too broad or too general to be good predictors for *O. fusiformis*.

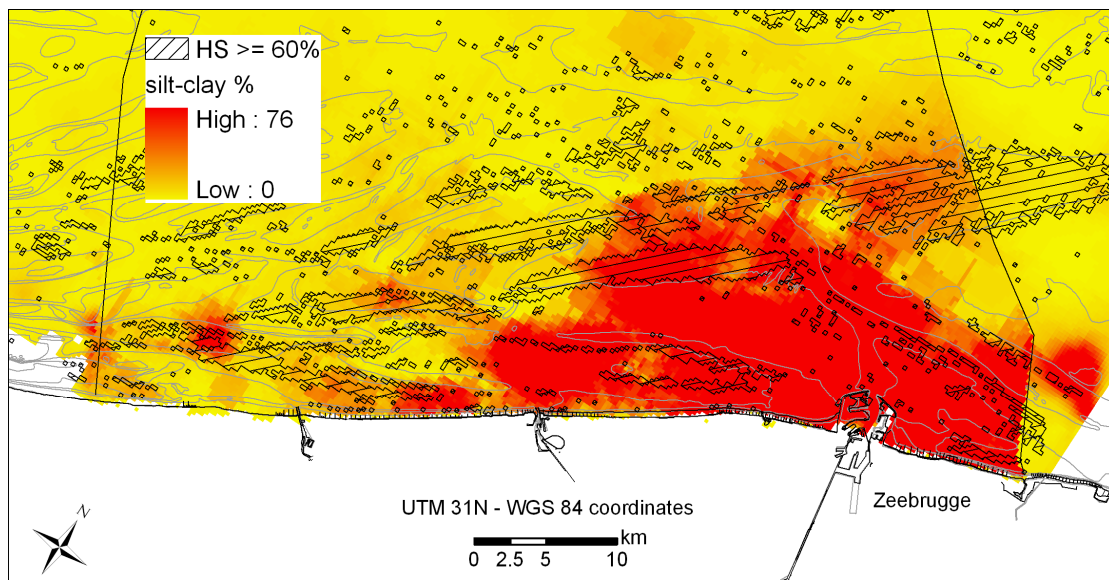


**Figure 6.4: Habitat suitability (HS) zones of  $\geq 60\%$  plotted on BPI classification: small scale crests (bedforms), sandbanks, depressions (or swales), flats and slopes. The high-suitability zones exclude the large sandbanks. The area shown is the southern part of the Belgian part of the North Sea.**

It is remarkable that high-suitability zones are located in areas where no sandbanks are present (Figure 6.4; representing a classification of topographical features, based on BPI and slope). The zones with higher HS are located mainly in shallow, flat zones. Still, many other shallow, flat zones have low habitat suitabilities (mainly in the southeastern coastal zone and farther away from the coastline). This is due to a combination of non-suitable EGVs, but it is difficult to determine which conditions are most constraining and which are not. At first sight, the low suitabilities around the harbour of Zeebrugge (SW part of the BPNS) are due to high silt-clay %. Still, as this predictor is not an input EGV for the final model, the low HS cannot be predicted by

this variable. Two EGVs that are input variables for the final model are rugosity and slope, and in this area of low suitability, the rugosity and the slope are generally very low compared to the rest of the BPNS (i.e. a completely flat area, without any small-scale topography that can be computed on a resolution of 250 x 250 m). This area of very low rugosity (equal to 1.00) and slope (quasi equal to 0.00), corresponds to a large extent with the zone of high silt-clay % around the harbour of Zeebrugge (Figure 6.5). Where, in this zone of high silt-clay %, slightly higher rugosity and slope values are observed and where small scale crests and the anthropogenic navigation channel are located (Figure 6.4), the model predicts higher suitabilities (although this is probably not correct in the case of the navigation channel). This suggests that *O. fusiformis* has a preference for very flat areas, that are in a very minor extent affected by some rugosity and some slope (order of magnitude between 0.1 and 0.2°). An interesting question that rises from these observations, is whether the dense aggregations of *O. fusiformis*, that protrude above the surface are reflected by these slightly higher rugosity and slope values. Further research regarding this issue, is needed.

These observations indicate as well that sedimentological and hydrodynamical EGVs, considered as crucial for predicting the species until now, can be replaced to a certain extent by topographical EGVs. Further research is necessary to examine why sedimentological and hydrodynamical EGVs are not selected by the best model. A possible cause is the fact that these EGVs are correlated to the topographical EGVs and as such contain redundant information, that is covered sufficiently by the information contained in the topographical EGVs.



**Figure 6.5: Habitat suitability (HS) zones of  $\geq 60\%$  plotted on silt-clay %. The high-suitability zones around the harbour of Zeebrugge are located mainly in the zone of higher silt-clay %, although topographical EGVs (particularly slope and rugosity) make a subtle difference between lower (extreme low slope and rugosity) and higher (slightly higher slope and rugosity) suitabilities. The area shown is the southern part of the Belgian part of the North Sea.**

Of course, as explained by Legendre (1993), the occurrence of species is not only predicted by non-spatially and spatially structured environmental variance (corresponding to the variance of the EGVs). Species distribution and density can also



be associated with a spatial structure that cannot be explained by environmental conditions, such as biotic processes within the population or community (e.g. competition, predation, recruitment processes, etc.). An example of such a recruitment process was demonstrated by Callaway (2003), where juveniles of the tube-building polychaete *L. conchilega* attached to adults, stimulating aggregation of the species.

## **6.5 Conclusion**

Past studies demonstrated that the occurrence of *O. fusiformis* can be predicted by the sedimentology and the depth. The present study shows that multi-scale topographical EGVs are also important predictors. Remarkably, the selection of one single spatial scale comparable with the spatial scale of a sandbank (i.e. between 750 and 4250 m) is sufficient to obtain a successful model of the spatial distribution of *O. fusiformis*.

The main predictors selected by the best model are all terrain EGVs computed with a window size of 3 pixels or 750 m. This indicates that sedimentological EGVs, considered as crucial for predicting the species until now, contain overlapping information with the topographical EGVs.

The shallow coastal zone, away from the large sandbanks, has the highest habitat suitability for *O. fusiformis* on the BPNS. The species prefers mainly flat areas with some minimal rugosity and slope.

## **6.6 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 also contributes to the EU INTERREGIII B 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”, EV/02/18A), financed by the Belgian Science Policy. The sedisurf@database has been compiled by the Renard Centre of Marine Geology (Ghent University) and consists of data from various institutes.