niet in kp

DI : 74852

FINAL

Prepared for:

ECOFLAT

EU ENVIRONMENT & CLIMATE PROGRAMME

Modelling habitat characteristics in an estuary

Hydrodynamic modelling, sediment transport modelling and upscaling in the Westerschelde estuary

Research Report

December 1999

T1542 / Z2152

WL | delft hydraulics

FINAL

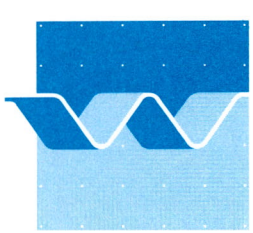
Modelling habitat characteristics in an estuary

Hydrodynamic modelling, sediment transport modelling and upscaling in the Westerschelde estuary

M. Baptist
A. Blauw
J. Boon
M. Marchand
P. Thoolen

Research Report

December 1999



WL | delft hydraulics

WL | delft hydraulics

CLIENT: EU Environment & Climate Programme

TITLE: Modelling habitat characteristics in an estuary

ABSTRACT:

The aim of ECOFLAT is to understand the role of intertidal flats in the ecology of estuaries, and to upscale from process understanding at the small scale to predictions relevant for management at the estuarine scale. The objectives of the study of WL|Delft Hydraulics within the ECOFLAT project are threefold:

- to develop a 2 D hydrodynamic mathematical model for the tidal flat, with a spatial resolution in the order of tens of metres.
- to model the transport and sedimentation/erosion of sediments on the tidal flat.
- to predict macrobenthic occurrence and biomass on an estuary-wide scale.

A hydrodynamic model was developed which is able to reproduce the local water levels and current velocities within the (measuring) accuracy margins. The model is considered applicable for the calculations of transport and sediment/erosion processes. The transport model uses physical processes only and proved to be capable of reproducing the concentrations in the water column and the sedimentation areas adequately. However, the yearly sedimentation rates (extrapolated from the net sedimentation rate for the month of June) proved to be somewhat lower than the estimate derived from the depth profiles of radionuclides (8mm and 1.3 cm, respectively). Furthermore, the spatial distribution of sedimentation as calculated from the model did not compare very well with the actual distribution of areas on the Molenplaat with high percentages of silt. One of the possible explanations to this could be that the dynamics of fine silt cannot be modelled adequately by physical processes only, biological processes may have a large impact on sediment dynamics.

Different methods have been used to predict macrobenthic distribution along the entire Westerschelde estuary. Based on an extensive database of measurements, whereby macrobenthic biomass of various species has been correlated with environmental parameters (depth, median grainsize, salinity and maximum current velocity), models have been developed for predicting macrobenthic biomass. The conclusion was that this extensive set of data for the Westerschelde does not give a clear correlation between local abiotic conditions and the abundance and distribution of macrozoobenthos. Apparently, the available environmental parameters are not explaining all the variety and heterogeneity in biomass in the Westerschelde.

REFERENCES: EU Environment and Climate Programme contract number ENV4-CT96-0216

VER.	ORIGINATOR		DATE	REMARKS	REVIEW		APPROVED BY	
1.0	Marchand et al.		31-8-1999					
2.0	Marchand et al.		27-10-1999		M. de Vries		T. Schilperoort	

PROJECT IDENTIFICATION:		T1542/Z2152						
KEYWORDS:		estuary, hydraulics, sedimentation, erosion, modelling, macrozoobenthos, habitat suitability analysis						
CONTENTS:	TEXT PAGES	iv+54	TABLES	6	FIGURES	55	APPENDICES	2
STATUS:	☐ PRELIMINARY		☐ DRAFT		☒ FINAL			

Contents

Summary

The aim of ECOFLAT is to understand the role of intertidal flats in the ecology of estuaries, and to upscale from process understanding at the small scale to predictions relevant for management at the estuarine scale. The objectives of the study of WL | Delft Hydraulics within the ECOFLAT project are threefold:

- to develop a 2 D hydrodynamic mathematical model for the tidal flat, with a spatial resolution in the order of tens of metres.
- to model the transport and sedimentation/erosion of sediments on the tidal flat.
- to predict macrobenthic occurrence and biomass on an estuary-wide scale.

Hydrodynamic modelling

The objective of the hydrodynamic model was to develop a model of the water levels and currents on the Molenplaat area in the Westerschelde with a spatial resolution in the order of tens of metres. The resulting Molenplaat-model is a two-dimensional model which is nested in a larger hydrodynamic model of the Westerschelde (Scaldis model). Activities included the set-up of the model and calibration and verification with existing field data. It was concluded that the model is able to reproduce the local water levels sufficiently accurate. Also the current velocities at the Molenplaat are being reproduced within the (measuring) accuracy margins. Hence, the model was considered applicable for the calculations of transport and sediment/erosion processes.

Sediment transport modelling

The main goal for the sediment transport model is to create a similar spatial sedimentation and resuspension pattern as present in the available data. As a further check on the functioning of the model also the dynamics of suspended matter in the water column are calibrated to measured concentration patterns. The transport model uses physical processes only and proved to be capable of reproducing the concentrations in the water column and the sedimentation areas adequately. However, the yearly sedimentation rates (extrapolated from the net sedimentation rate for the month of June) proved to be somewhat lower than the estimate derived from the depth profiles of radionuclides (8mm and 1.3 cm, respectively). Furthermore, the spatial distribution of sedimentation as calculated from the model did not compare very well with the actual distribution of areas on the Molenplaat with high percentages of silt. One of the possible explanations to this could be that the dynamics of fine silt cannot be modelled adequately by physical processes only, biological processes may have a large impact on sediment dynamics.

Upscaling

The upscaling activities of ECOFLAT are essential in order to generate relevant information for the management of an entire estuary. Macrobenthic animals are an important ecological target variable for environmental management in themselves. Moreover, they are a key variable in the assessment of the potential of the estuary for wintering birds. Not only biomass values of the macrofauna, but also the composition of the fauna, its distribution with respect to depth in the sediment and with respect to exposure time of the flat area are important variables for the birds. Different methods exist to predict macrobenthic distribution along the entire Westerschelde estuary. Based on an extensive

database of measurements, whereby macrobenthic biomass of various species has been correlated with environmental parameters (depth, median grainsize, salinity and maximum current velocity), models have been developed for predicting macrobenthic biomass.

The conclusion was that this extensive set of data for the Westerschelde does not give a clear correlation between local abiotic conditions and the abundance and distribution of macrozoobenthos. Apparently, the available environmental parameters are not explaining all the variety and heterogeneity in biomass in the Westerschelde.

1 Introduction

1.1 The ECOFLAT project

ECOFLAT is a research project funded by the European Commission in the framework of ENVIRONMENT & CLIMATE Programme (contract number ENV4-CT96-0216) and is part of ELOISE (European Land Ocean Interactions Studies). The aim of ECOFLAT is to understand the role of intertidal flats in the ecology of estuaries, and to upscale from process understanding at the small scale to predictions relevant for management at the estuarine scale. The ECOFLAT project is a co-operation between the Netherlands Institute of Ecology (NIOO-CEMO), Centre des faibles radioactivités (CFR), Plymouth Marine Laboratory (PML), WL | Delft Hydraulics (WL), University of Gent (UG), the Dutch National Institute for Coastal and Marine Management (RIKZ), Institute for Nature Conservation (IN) and Southampton University - Dept. of Oceanography (SUDO).

Tidal flats are important components of macro- and mesotidal estuaries and coastal systems all over the world. They are complex macromorphological structures, shaped in close relation to hydrodynamical factors in the estuary. A single tidal flat can consist of a series of habitats ranging from small sanddunes and megaripples to extensive flat muddy areas and may be inhabited by very different biological communities. Since they can occupy more than half of the total estuarine surface, tidal flats have a significant role in the metabolic processes and material fluxes within the estuary, especially in the degradation and transformation of organic material either advected or produced in situ. Understanding the ecological role of tidal flats in the functioning of the estuarine ecosystem is therefore a central question in the study of land-sea interactions.

Previous research on tidal flats has largely been descriptive; it has generally not adopted a multidisciplinary approach which is necessary to examine the complex interactions between physical, chemical and biological processes that are characteristic of the behaviour of tidal flats as geomorphological and metabolic systems. Thus there exists a large literature on purely sedimentological and purely biological and ecological problems, whereas very few multidisciplinary studies have been undertaken. In current ecological models of estuaries tidal flats are either absent or only present in mass balances on a gross temporal and spatial scale not corresponding to the scales at which the ecological processes actually take place. The lack of mechanistic understanding and representation of the processes limits the predictive capabilities of the models under scenarios relevant to the many ecological problems facing estuarine ecology: eutrophication, human interference with the hydrography, global change and loss of biodiversity.

Many important processes regulating eco-metabolism and material fluxes in estuarine environments have characteristic time scales of only a few hours. The tidal cycle causes current velocities and water levels to change from zero to their maximum over about six hours. As a consequence, the intertidal sediment surface changes two times a day from a

sediment-water interface to a sediment-atmosphere interface with associated shifts in the biogeochemical processes. Of interest are also the lags that exist between the diurnal and the tidal cycles, causing a fourteen-day cyclicity in the light received by benthic diatoms during ebb tide and therefore in benthic photosynthesis. Spatially, the scales of interest to the processes of carbon and nutrient exchanges range from a few mm (vertically), over a few cm (the scale of stream ripples, important for microdistribution of algae, microbial organisms, detritus, meiofauna) to a few tens of m (the scale of gradients in sediment composition and texture relevant to macrobenthos).

The ECOFLAT project aims at bridging the gap between the small scales at which the processes operate and are measured, and the large scales (estuarine scale, regional scale) at which the environmental problems are posed. This aim will be achieved by explicitly studying the small scale processes and by integrating them in mathematical process models. From these models the relationships between estuarine characteristics (typology) on the one hand, and ecological variables relevant to estuarine management and to extrapolation on a regional scale on the other hand, should become easier to predict, because based on a better mechanistic understanding. This upscaling to the estuarine scale will be tested against existing knowledge and databases available from monitoring authorities. In the process studies, emphasis will be placed on the study of the carbon and nutrient cycling within the flats and to the dynamics of carbon and nutrient exchange between the flat and the surrounding estuarine ecosystem. In this project, we will start on a single tidal flat system in order to integrate the different disciplines, but explicitly include the study of a validation site at the end of the project.

1.2 Objectives of the study

The objectives of the study of WL|Delft Hydraulics within the ECOFLAT project are threefold:

- to develop a 2 D hydrodynamic mathematical model for the tidal flat, with a spatial resolution in the order of tens of metres.
- to model the transport and sedimentation/erosion of sediments on the tidal flat.
- to predict macrobenthic occurrence and biomass on an estuary-wide scale.

In this final report the methods, results and conclusions of these three (related) activities is documented.

2 Hydrodynamic modelling

2.1 Objectives

The objective of the hydrodynamic model is to simulate the water levels and currents on the Molenplaat area in the Westerschelde, with a spatial resolution in the order of tens of metres. Ultimately, this model will be used to describe the transport and sedimentation/erosion processes on the Molenplaat (see chapter 3). The Molenplaat-model is a two-dimensional model which is nested in a larger hydrodynamic model of the Westerschelde (Scaldis model). Activities during the hydrodynamic modelling included the set-up of the model and calibration and verification with existing field data.

2.2 Description of the Scaldis model

2.2.1 Introduction

The SCALDIS model is a two-dimensional hydrodynamic model developed by the National Institute for Coastal and Marine Management (RIKZ) with the purpose to calculate water levels and depth averaged currents in the Westerschelde. It is based on the computer package WAQUA (see Dekker, 1993 and Dekker 1994). For the present project, SCALDIS has been used to generate boundary conditions for the detailed Molenplaat model.

2.2.2 Model set-up

Model boundaries

The seaward boundary of SCALDIS is situated along the line Westkapelle - Vlakte van de Raan - Zeebrugge. The upstream boundary is Gent (Belgium).

Grid schematisation

The coastal area is defined according to a rectangular grid with grid cells varying in size between 100m and 400m. At the Westerschelde the grid size is uniform 100x100 m^2, while at the Zeeschelde a curvilineair grid has been defined with a gridsize varying between 30 and 100 m. The area between Rupelmonde and the weir near Gent has been modelled very schematically. See figure 2.1.

Bathymetry

The original bathymetry of SCALDIS has been based on measurements from 1988. Later the bathymetric data of 1995 became available. This data was used to create a new, more recent bottom schematisation for both the SCALDIS model and the Molenplaat model. The model was validated applying this new bathymetry.

Boundary conditions

The seaward part of the model has been regarded as an open boundary where the water levels as a function of tidal components have been defined. At several locations along the Westerschelde sources have been defined in order to simulate the discharge of fresh water, i.c. the discharge of the Schelde at Gent and the Rupeltak, the Antwerp Docks, the Zoommeer and the canal of Terneuzen to Gent. For the upstream discharge of the Schelde the yearly average of 1992 has been chosen.

Salinity

At the start of a simulation run, SCALDIS uses a non-uniform horizontal salinity profile, calculated with the E-WestII model (Dekker et al., 1993), and based on a high or low upstream discharge of the Schelde, respectively (depending on the chosen simulation period). The marginal consequences of the horizontal density differences on the tidal movement can thus be taken into account to a certain extend.

Numerical and physical parameters

time step for the calculation	0.5 minutes
bottom roughness	Manning coefficient = 0.022
gravitation constant	9.813 m/s^2
Viscosity coefficient	10 m^2/s
density of water	1023 kg/m^3
density of air	1205 kg/m^3

2.2.3 Generation of boundary conditions for the Molenplaat model

Several runs have been performed in order to derive the boundary conditions for the Molenplaat model. Verification of the results has been performed by using the high and low water levels at Hansweert, Vlissingen, Terneuzen and Bath, derived from the tide tables. A detailed description of the verification can be found in the reports of Herman et al., 1995 and Thoolen et al. 1997. Model validation has been performed by Rijkswaterstaat. It has been reported that for some conditions the model does not always generate accurate predictions of the tide. Differences between calculated and measured water levels are in the order of 0.05 to 0.13 cm. Calculated currents can deviate from the measured data in the order of 10 to 20 % (see Dekker et al., 1994).

2.3 Description of the Molenplaat model

2.3.1 Introduction

The Molenplaat model is a two-dimensional hydrodynamic model based on the software package Delft2D-Flow, and is nested in the SCALDIS model. It has been developed to predict water levels and stream currents at the Molenplaat. As the Molenplaat is situated near other intertidal areas (Schaar van Ossenisse, Rug van Baarland), it was necessary to include these intertidal flats in the Molenplaat model as well (see figure 2.2). The reason for this is that an open boundary which is defined at a tidal flat which is drying at low tide generates more model disturbances than a boundary which is continuously wet. A second reason for inclusion is that the surrounding tidal flats act as a source for suspended sediment which can be deposited on the Molenplaat.

2.3.2 Model set up

Boundaries

In the southwest the model is limited by an open current boundary perpendicular to the banks, just beyond the Rug van Baarland. In the southeast the model is limited by an open water level boundary almost perpendicular to the banks, just beyond the Plaat van Ossenisse (see figure 2.2).

Grid schematisation

The model has been defined on a rectangular grid of which the x-axis is parallel to the gully between the Molenplaat and the Rug van Baarland. The grid size varies between 40x40m^2 on the Molenplaat itself and 100x100m^2 on the open edges of the model (see figure 2.3).

Bathymetry

The bottom topography has been primarily based on soundings (on a grid of 20x20m^2) from 1994 and 1995, supplemented with depth data from the SCALDIS model for the southeastern and southwestern edges.

Salinity

In the surrounding area of the Molenplaat the horizontal salinity is quite homogenuously distributed (approx. 25 ppt). Therefore, for the Molenplaat model a uniform density distribution is assumed.

Numerical and physical parameters

Time step for the calculation	0.25 minutes
bottom roughness	Manning coefficient = 0.013, 0.022, 0.033 and 0.045
gravitation constant	9.813 m/s^2
Viscosity coefficient	10 m^2/s
density of water	1017 kg/m^3
density of air	1205 kg/m^3

2.3.3 Calibration and verification

Measuring campaign of the 'Zeekat'

From thursday 16 june till and including monday 26 june 1995 field measurements have been performed at the Molenplaat with the pontoon 'Zeekat' of Rijkswaterstaat. Alongside this pontoon a measuring transect was constructed perpendicular to the edge of the flat. At 9 points along this transect current velocity and water level gauges were positioned. Current velocities have been measured at several cm above the bottom. Transformation to depth averaged velocities (in order to compare them with model results) has been performed assuming a logarithmic distribution of velocities with depth. A detailed description of the measurements and analysis can be found in Herman et al., 1995 and Thoolen et al., 1997. With the resulting data the Molenplaat model has been calibrated. A total of 16 calibration and verification runs have been made with the Molenplaat model, the results of which have been reported in Herman et al.,1995. Hereafter the main findings have been summarised.

Influence of the bottom roughness parameter

The value of the bottom roughness parameter has a large influence on the calculated current velocities on the Molenplaat area. The central part of the Molenplaat is silty (at least during the measuring campaign) and smooth, whereas the western part consists of sand and megaripples. It appeared that a uniform bottom roughness for the entire Molenplaat (Manning coefficient n = 0.022) did not gave satisfactory results. Instead, a spatial differentiation of the Manning coefficient was applied. On the central part of the Molenplaat a lower coefficient was used, based on the fact that at that location the bottom roughness is largely determined by the median grain size of approx. 290 μm (in absence of dunes and ripples). An estimation of the roughness parameter k_s is: 3 * median grain size = 0.00087m, which is more or less equivalent to n = 0.013 (Van Rijn, 1990).

At the western side of the Molenplaat megaripples occur, which predominantly determine the bottom roughness. Assuming an average length of the megaripples of 10m and a height of 1 m, this leads to an estimated k_s of 0.706m, which more or less is equivalent to n = 0.045. At the eastern side of the Molenplaat also a higher bottom roughness (n = 0.033) has been used as the sediments are more sandy there than on the centre of the Molenplaat. For the other parts of the Molenplaat and for the rest of the model n = 0.022 has been applied.

In addition to this spatial differentiation also a difference in bottom roughness between ebb and flood situations might be assumed. Due to the specific orientation of the megaripples and dunes at the bottom and the difference in current directions between ebb and flood, it is possible that the bottom roughness during ebb tide is higher than during flood tide. This aspect has not been taken in to account in the model, however.

Influence of the wind

A wind velocity of 5 m/s is sufficient in order to have a significant effect on the value and direction of the current velocities on the tidal flat, for the calibration and validation period. In this study calculations have been performed with a constant wind and without taking into account a possible wind set-up of the waterlevels at the boundaries of the model.

2.4 Conclusions and discussion

During the calibration- and verification period the values and phasing of the current velocities are modelled sufficiently accurate by the Molenplaat model. The same applies for the water levels during the calibration period. However, during the verification period the simulations of the water levels are less sound, but - besides one exception- still within an accuracy margin of 10%.

Bottom roughness is the major calibration parameter and plays an important role in the value of the velocities on shallow parts of the Molenplaat. Apart from the spatial differences in bottom composition, the bottom roughness may also be partly tide-dependent. This aspect could not be incorporated in the model.

It can be concluded that the model is able to reproduce the local water levels sufficiently accurate. Also the current velocities at the Molenplaat are being reproduced within the (measuring) accuracy margins. This implies that the model is applicable for the calculations of transport and sediment/erosion processes (see next chapter).

3 Sediment transport modelling

3.1 Introduction

One of the objectives of the Ecoflat project is the "Development of predictive mathematical models describing the processes acting in and on tidal flats and relating them to the main forcing factors for the tidal flat system". The transport of suspended solids is expected to be an important forcing factor for the distribution of macrobenthos and microphytobenthos and for the biochemistry of tidal flats. The link between these phenomena and the transport of suspended solids is on the one hand the supply of particulate organic material and on the other hand the stability of the substrate for biota. Particulate organic matter serves as food for macrobenthos. Deposited organic matter is mineralised in the bottom by chemical and bacterial processes. The mineralisation of organic material is in turn an important steering factor for the biochemistry of tidal flats and it supplies the nutrients for microphytobenthos growth. A mathematical model for sediment transport on the Molenplaat could therefore be a useful tool in the upscaling of processes to the whole Westerschelde estuary and tidal flats in general.

The main goal for the sediment transport model is to create a similar spatial sedimentation and resuspension pattern as observed in the available data. As a further check on the functioning of the model also the dynamics of suspended matter in the water column are calibrated to measured concentration patterns.

3.2 Analysis of data

Two sets of measured data were available for calibration of the model:

- spatial information on the grain size distribution of sediment on the Molenplaat;
- time series of suspended solids in the water column on the Molenplaat.

As part of the Ecoflat project, samples have been taken to assess the grain size distribution of the Molenplaat sediment. This monitoring campaign was performed by the NIOO-CEMO. The results for the percentage fine silt (< 63 μm) in June 1995 are shown in Figure 3.1. It shows that fine silt percentages are highest in the southern central part of the Molenplaat. Coarse matter settles faster than fine matter. If fine silt can settle, this is an indication that the area is suitable for net sedimentation. We assume that the southern central part of the Molenplaat is the main sedimentation area. Besides the suitability for sedimentation, the transport and supply of suspended solids is also a determining factor for the amount of net sedimentation.

Measurements of suspended solids concentrations in the water column of the Molenplaat are available for 6 monitoring stations. The monitoring was performed by RIKZ-Middelburg in 1996 and 1997 as part of the Ecoflat project. The locations of the monitoring points situated on the Molenplaat are presented in Figure 3.2. The monitoring points are plotted together

with the maximum water level (from the hydrodynamic model, see chapter 2), so the shape of the tidal flat and the relative depth of the monitoring points can be seen.

The measured concentrations of suspended solids at the monitoring stations are presented in Figures 3.3 to 3.5. For those stations where measurements of more than one tidal cycle are available, the results have been plotted in such a way, that the water levels at the monitoring times are synchronised. The synchronised water level measurements are also shown in the plots. The tidal movement can be clearly recognised in the measured data. Concentrations increase fast with the rising and falling of the tide. At the turning of the tide (flood slack as well as ebb slack) the water movement is at its lowest and the concentrations of suspended solids drop. Measurements have only been performed for water depths larger than 5 centimetres. The water levels that are presented in the plots are measured at a station in deeper water (Hansweert).

Only for two stations more than one tidal cycle has been monitored. Considering the large variance of the measurements in the several tidal cycles, conclusions can not be drawn on measurements of one tidal cycle. For this reason a sound comparison can only be made for station 'Molenplaat 3' and station 'Zeekat spring'. The range of concentrations in station 'Molenplaat 3' appears to be somewhat larger than in station 'Zeekat spring'. Also, in station 'Zeekat spring' the first concentration peak appears to be higher than the second whereas in station 'Molenplaat 3' both peaks are equally high. The first peaks represent a water flow from sea and the second peaks represent water flowing from inland.

3.3 Model set up

Transport model

Sediment transport is to a large extent determined by the hydrodynamics of the area. A hydrodynamic model is therefore needed as the basis of any sediment transport model. In Figure 3.6 (below) the relation between the hydrodynamic model, and the input and output of both the hydrodynamic and the sediment transport model are presented in a scheme. WL|Delft Hydraulics has developed a detailed hydrodynamic model of the tidal flat 'Molenplaat' (see chapter 2). This hydrodynamic model is the basis of the suspended solids model. The hydrodynamic model is calibrated for June 1995. The sedimentation model is run and calibrated for June 1995 as well. A time step of 10 seconds is used for the calculations.

The sediment model is a two-dimensional model since stratification is not important on tidal flats. Only one fraction of suspended sediment is modelled. In reality the measured suspended solids concentration consists of a mixture of different types of sediment. Differences in sediment types can be due to the origin of the material (organic matter, silt, sand, riverine material, marine material) and the shape (filaments, grain size, flocs). The model aims at modelling fine material, with a grain size smaller than 63 μm, including organic matter. This is the most important fraction for biota, as it serves as food for filterfeeders and bacteria.

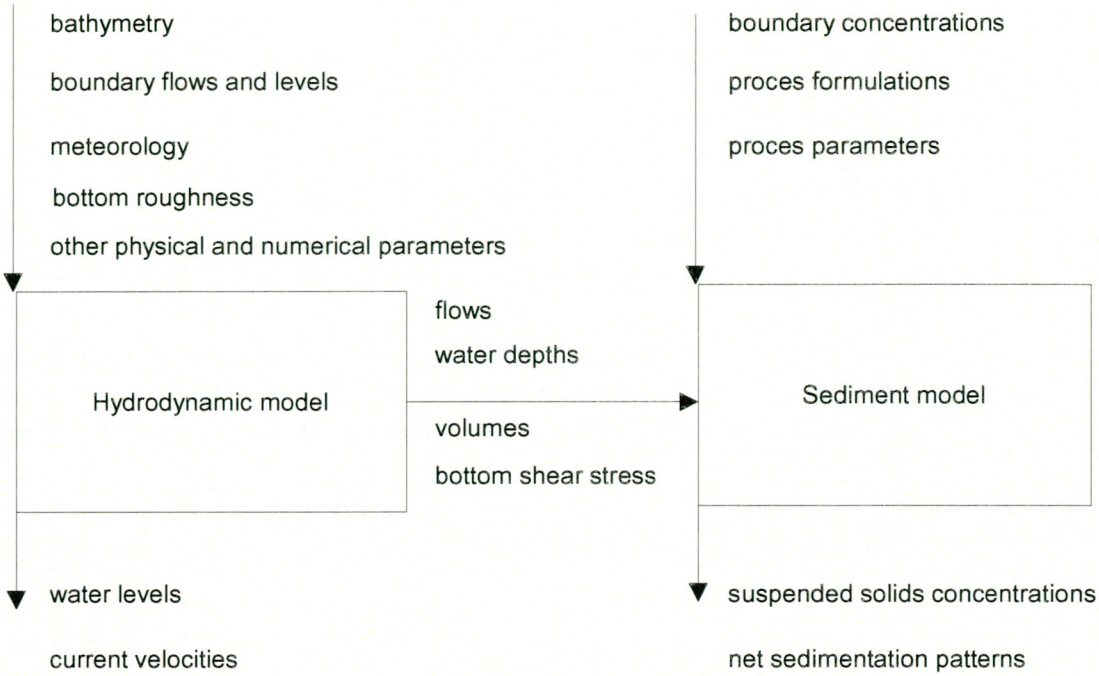


Figure 3.6: Schematic representation of the relation between the hydrodynamic model and the sediment model

Figure 3.7 shows the maximum bottom shear stress over the tide in the hydrodynamic model. The lowest shear stresses are found in the central area of the Molenplaat. This is the area where hydrodynamic conditions are most suitable for net sedimentation and least suitable for resuspension. The high percentages of very fine silt in the monitoring campaign by the NIOO-CEMO in June 1995 were also located in the central area but more to the south. The spatially varying pattern of the bottom roughness, as used in the hydrodynamic model, is likely to play an important role in the calculated pattern of the bottom shear stress. In reality the bottom shear stress is expected to be influenced by the bottom roughness in the same way.

Boundary conditions

The boundary concentrations were based on measurements of suspended solids concentrations from the WAKWAL-database. Average concentrations were calculated for June from data of 1982 to 1993 for the stations 'Hansweert' (upstream boundary) and 'Terneuzen' (downstream boundary). The average concentrations were 37 mg/l at 'Hansweert' and 23 mg/l for 'Terneuzen'.

Van Maldegem (1992) has also calculated averaged suspended solids concentrations for the Westerschelde (for the period 1970 - 1990). He finds somewhat higher concentrations than the averaged concentrations in the WAKWAL database. At Terneuzen he calculates a summer average of about 45 mg/l and a winter average of about 100 mg/l. At Hansweert he calculates a summer average of about 35 mg/l and a winter average of about 90 mg/l. From this can be concluded that the model boundary at Terneuzen (23 mg/l) might be too low.

Process formulations

For the calculation of sedimentation and erosion fluxes the formulations of Krone and Partheniades have been used. In these formulations the bottom shear stress is the steering factor for sedimentation and erosion processes. The bottom shear stress is a measure for the turbulence in the water created by the shear stress between the bottom and the moving water above. Three causes for water movement and turbulence can be distinguished: current velocity, wave energy and ship movements. Of these only the current velocity has been taken into account. At high wind speeds the influence of wind induced waves cannot be neglected, but a test run has shown that the incorporation of wind induced waves does not improve the results of the sediment transport model (Thoolen et al., 1997). The sediment model uses the bottom shear stresses as calculated by the hydrodynamic model: without the bottom shear stress by wind induced waves.

The bottom shear stress is calculated in the hydrodynamic model using the following formulation:

$$\tau = \frac{\rho_l \,.\, g \,.\, Velocity^2}{Chezy^2} \qquad (3.1)$$

Where:

τ	=	bottom shear stress causes by current	[Pa]
ρ_l	=	density of water	[g/m^3]
g	=	gravitational acceleration	[m^2/s]
Velocity	=	current velocity	[m/s]

and:

$$Chezy = \frac{\sqrt[6]{H}}{n} \qquad [m^{1/2}/s] \qquad (3.2)$$

where:

H	=	water depth	[m]
n	=	manning coefficient for bottom roughness	[m$^{-1/3}$/s]

From the formulations it follows that the bottom shear stress increases with higher current velocities and decreases in deeper water. Resuspension occurs when the bottom shear stress exceeds a critical value (τ_{Res}). Sedimentation occurs when the bottom shear stress drops below a critical value (τ_{Sed}). When the bottom shear stress is between τ_{Sed} and τ_{Res} no sedimentation or erosion takes place. In Figure 3.8 the relation between bottom shear stress and sedimentation and erosion is shown in a scheme.

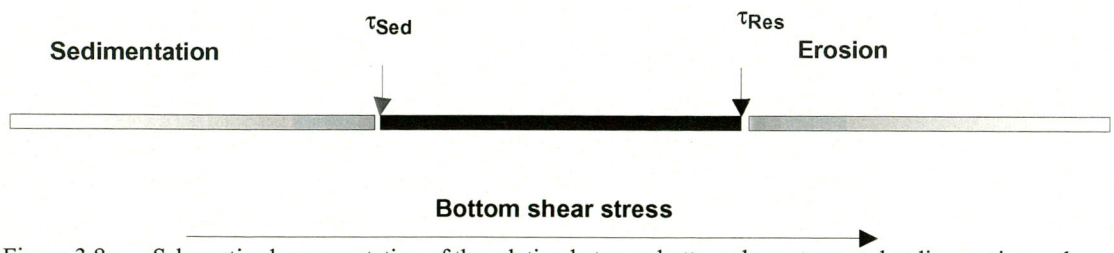


Figure 3.8: Schematised representation of the relation between bottom shear stress and sedimentation and erosion.

The resulting sedimentation or erosion flux per segment is calculated using the formulations of Krone and Partheniades. The sedimentation flux depends on the ratio between the actual bottom shear stress and the critical shear stress for sedimentation and on the settling velocity of suspended matter. The erosion flux depends on the ratio between the actual bottom shear stress and the critical bottom shear stress for erosion and on the maximum resuspension flux (=erosion parameter).

Sedimentation according to Krone:

$$f_{Sed} = P_{Sed} \cdot (V_{Sed} \cdot (SS)) \tag{3.3}$$

$$P_{Sed} = \max\left(0, 1 - \left(\frac{\tau}{\tau_{Sed}}\right)\right) \tag{3.4}$$

Erosion according to Partheniades:

$$f_{Res} = P_{Res} \, x \, (Z_{Res}) \tag{3.5}$$

$$P_{Res} = \max\left(0, \left(\frac{\tau}{\tau_{Res}}\right) - 1\right) \tag{3.6}$$

Where:

f_{Sed}	=	sedimentation flux	[$g.m^{-2}.d^{-1}$]
P_{Sed}	=	sedimentation probability	[-]
V_{Sed}	=	sedimentation velocity	[m/d]
SS	=	concentration suspended solids	[g/m^3]
f_{Res}	=	erosion flux	[$g.m^{-2}.d^{-1}$]
P_{Res}	=	erosion probability (fraction of maximum resuspension flux resuspended at current bottom shear stress)	[-]
Z_{Res}	=	maximum resuspension flux (= erosion parameter)	[$g.m^{-2}.d^{-1}$]

Process parameters

The bottom shear stress is calculated in the hydrodynamic model (see chapter 2). That leaves four process parameters that can be used for calibration of the sediment transport model:

- critical bottom shear stress for erosion: τ_{Res}
- critical bottom shear stress for sedimentation: τ_{Sed}
- settling velocity: V_{Sed}
- maximum resuspension flux: Z_{Res}

Besides these process parameters also the horizontal diffusion coefficient could be used for calibration. The horizontal diffusion coefficient however was kept constant during the calibration at 10 m^2/s.

Initial conditions

As a start for the model calculations an initial condition needs to be defined for the concentration of suspended solids in the water column and on the bottom. Theoretically the initial condition for a model run for June 1995 should be equal to the real situation at the end of May 1995. For the concentration in the water column the effect of the initial condition does not last long as transport and sedimentation and erosion processes soon result in a new equilibrium situation.

For the amount of sediment on the bottom the situation is different. The amount of sediment on the bottom has a major influence on the amount of sedimentation. In theory there is an infinite amount of sediment in the Westerschelde. However, only part of this is susceptible for erosion. Where the sediment is too coarse to resuspend, the model assumes there is no sediment. As the model focusses on fine material, the distribution of fine silt throughout the whole model area at the end of May 1995 should be offered as input to the model. This information is however not available. There are several options to approximate the distribution of fine silt at the start of the calculation. One of them would be to start with a clear bottom and let the model run until a more or less stable sediment distribution pattern is reached. Another option is to start with a layer of silt of equal depth over the whole model and let the model run until a suitable sediment distribution pattern is reached. As the starting situation depends on the process parameters in the model itself, it should be renewed for every calibration run.

In the test runs the model was run 19 days to find the initial condition, starting with a layer of 1 cm sediment over the whole model area. The advantage of this option is that erosion areas can also be distinguished. The results of the 20th day were used for comparison to measured data.

3.4 Model testing

Calibration

The model results were calibrated to the measured data described in paragraph 3.2. The parameter settings were optimised to create a net sedimentation area in the model corresponding to the area with a high percentage of fine silt and to calculate a time series of suspended solids in the water column, similar to the measured time series.

Only two of the four available process parameters were used for calibration: the critical bottom shear stress for erosion and the settling velocity. These parameters had most effect on the model results. In table 3.1 an overview is shown of calibration runs. The best results were obtained with τ_{Res} at 0.3 Pa and V_{Sed} at 5 m/day. The resulting sedimentation and erosion pattern is shown in Figure 3.9. The main sedimentation area is located in the central region of the Molenplaat. This is similar to the area with a high percentage of fine silt in the measurements in Figure 3.2. In the measurements however the area with fine silt is located more to the south than in the model.

Table 3.1: Overview of calibration runs (the final run is coloured grey)

Run name	τ_{Sed}	V_{Sed}	τ_{Res}	Z_{Res}
9a	0.1	8.64	0.4	4320
9b	0.1	8.64	0.35	4320
9c	0.1	8.64	0.3	4320
9d	0.1	8.64	0.5	4320
9e	0.1	8.64	0.5	8640
9f	0.1	6.0	0.3	4320
9g	0.1	5.0	0.3	4320
9h	0.1	10	0.3	4320

Figures 3.10 and 3.11 show the time series of suspended solids in the water column in the measurements and in the model at two monitoring points. The dynamics and the range are reproduced pretty well, although the first concentration peak is missing in station 'Zeekat Spring'.

Net sedimentation rates

Using this calibrated model, the net sedimentation flux is estimated on the central part of the Molenplaat. The concentrations at the end of the 20 days calculation were used as the initial conditions for a new 10 days calculation. In the 20 days calculation also a net sedimentation flux was calculated, but this estimate is probably too high since at the beginning of the calculation the 1 cm layer of sediment in the gullies is flushed away. This results in higher concentrations suspended matter and higher sedimentation rates at the beginning of the calculation. The result of the 10 days calculation is shown in Figure 3.12. The sedimentation area is smaller than in the 20 days calculation. This is because in the 10 days calculation, there is less sediment available for sedimentation. In the central area sedimentation rates are about 150 g/m^2 per ten days, which equals 15 g.m^{-2}.day^{-1}. Assuming a dry density of 2650 kg/m^3 and a wet density of 1325 kg/m^3 a net sedimentation flux of 0.9 mm fine silt per month can be calculated, for June conditions (see for details on the calculation appendix B). If the conditions would remain constant over the year this would result in a accretion rate of 1.1 cm fine silt per year.

Sensitivity analysis

It appears that the sedimentation flux is quite sensitive to the availability of sediment. As a final evaluation of the results a sensitivity test has been performed. For this sensitivity analysis the initial condition is the same as for the estimate of the net sedimentation flux. At the start of the calculation the amount of sediment at the bottom is the result of the 20 days calibration run. Similarly to the run to assess the net sedimentation flux, the model has been run for 10 days.

In Table 3.2 an overview is given of the runs performed for the sensitivity analysis. First the impact of the process parameters is tested. In Figure 3.13 the effects of an increased τ_{Res} and an increased V_{Sed} are shown. An increase of the τ_{Res} from 0.3 to 0.5 leads to a larger sedimentation area. The highest sedimentation fluxes are found at the edges of the sedimentation area. An increase of the settling velocity (V_{Sed}) from 5 to 7 m/day results in

somewhat more sedimentation, in the same small sedimentation area. In Figure 3.14 the effect of the boundaries is shown. Doubling the suspended solids concentration at the sea boundary has much more effect than doubling the river boundary. The sensitivity analysis has been performed using a larger time step and a faster numerical calculation method than the calibration runs. The effect is a reduction of the runtime and a little less accurate results. Figure 3.15 shows the difference in results due to the reduction of the runtime.

The sensitivity tests show that the sedimentation pattern remains similar with different changes to the model. Only changes in the critical shear stress for erosion result in major changes in the sedimentation pattern. By varying the proces parameters and boundaries the net sedimentation flux can be doubled compared to the nominal run.

Table 3.2: Overview of the model runs for the sensitivity analysis

description	Figure	τ_{Res}	V_{Sed}	Terneuzen boundary (mg/l)	Hansweert boundary (mg/l)
nominal run	3.15	0.3	5.0	23	37
more accurate calculation method	3.15	0.3	5.0	23	37
higher τ_{Res}	3.13	0.5	5.0	23	37
higher V_{sed}	3.13	0.3	7.0	23	37
double sea boundary	3.14	0.3	5.0	46	37
double river boundary	3.14	0.3	5.0	23	72

3.5 Discussion and conclusions

Sedimentation pattern

The sedimentation area is calibrated to data of the grain size distribution of the Molenplaat sediment. The sedimentation area in the model is located in the central area of the Molenplaat. The areas with high percentages of fine silt in the measurements are also located in the central area of the tidal flat. But unlike the model, the highest percentages of fine silt are found at the south of the central area. Also, the measurements show higher percentages of fine silt east of the central area, than to the west of the centre of the tidal flat. In the model there is no net sedimentation either west nor east of the central area. There are several possible explanations for the differences between the sedimentation pattern in the model and the distribution pattern of fine silt on the Molenplaat.

The sedimentation and erosion pattern in the model is mainly determined by the bottom shear stress pattern, that is calculated by the hydrodynamic model. For this reason the sedimentation area cannot be shifted to the south (which would make it more comparable to the measurements) using this hydrodynamic model. More detailed information on the bottom roughness could improve the modelling of the bottom shear stress by the hydrodynamic model. Using the present hydrodynamic model the sedimentation area can be enlarged by

increasing the critical bottom shear stress for erosion (τ_{Res}). Increasing the settling velocity V_{Sed} results in more sedimentation but has little effect on the sedimentation pattern.

In reality not only the hydrodynamics but also biological factors determine the sedimentation and erosion pattern, like microphytobenthos mats and bioturbation. In June 1995 the chlorophyll content of the Molenplaat sediment has been measured at several locations. The results are presented in figure 3.16. It is striking that the locations with high chlorophyll contents are the southern central area of the Molenplaat and the area east of the centre. The high chlorophyll contents indicate the presence of microphytobenthos mats that limit the resuspension, resulting in higher silt percentages. The model does not incorporate the effect of microphytobenthos mats, so that could explain the difference between the model and the measurements.

Another explanation could be that the bottom shear stresses at the south and the east side of the flat are in reality lower than calculated in the hydrodynamic model. This could explain why microphytobenthos mats develop in these areas. Microphytobenthos growth and fine silt are correlated, partly because both are resuspended at higher bottom shear stresses (Stapel & de Jong, 1998) and partly because microphytobenthos mats limit resuspension.

Another possible explanation for the difference between the sedimentation area in the model and the distribution pattern of fine silt in the measurements is that the grain size distribution is not only determined by the sedimentation of fine silt but also by the sedimentation of other sediment fractions. So the fact that the highest percentages of fine silt are found in the southern central area could mean that most fine silt settles here. But it could also mean that little fine sand settles here. In order to make the model results more comparable to the grain size distribution more fractions would need to be modelled.

Time series of suspended solids

The suspended solids in the water column are composed of both coarse and fine sediments that have different sedimentation and erosion properties. The model uses only one sediment fraction to model the dynamics of the diverse sediment types. Still, the concentrations in the water column and the sedimentation areas are reproduced by the existing model adequately.

Net sedimentation rates

The net sedimentation rate for the central area of the Molenplaat can be estimated using this model as about 0.9 mm in June. If this would be extrapolated to one year this would be about 1.1 cm per year. Other estimates, based on measured depth profiles of radionuclides, vary from 1.1 cm per year at the edge of the flat to 2.5 cm per year at the centre of the flat (Schmidt et al., 1999a). Estimates based on volume differences, corrected for the subsidence due to compaction, tectonics and isostasy indicate a accretion rate of 1.3 ± 0.1 cm per year on average for the whole Molenplaat (Schmidt et al., 1999b).

The calculated accretion rate is comparable to the other estimates of the accretion rate. Still, the model does only cover part of the aspects involved in the sediment dynamics of the

Molenplaat. Aspects that can have a large impact on sediment dynamics and that are not covered by the model are for instance:

- the role of coarser sediment fractions (> 63 μm);
- biological processes, like formation of microphytobenthos mats and bioturbation;
- tidal variations in concentrations SPM at the boundaries (circa 20 to 150 mg/l);
- seasonal variations in concentrations at the boundaries;
- the effect of wind;
- the effect of spring-neap cycles on tidal currents
- vertical gradients of suspended matter in the water column;
- flocculation effects..

The availability of suspended solids concentrations in the present model is probably an underestimate of the conditions for June. The concentrations at the boundaries are kept constant in this model, whereas in reality the concentrations vary from about 20 mg/l to more than 150 mg/l. In the model the concentrations at the boundary near Terneuzen are 23 mg/l, whereas Van Maldegem (1992) finds an average concentration of 45 mg/l.

3.6 Recommendations

Interesting subjects for further research and modelling efforts are:

- the value of τ_{Res} could be made spatially variable, so the effect of coarser sediment, bioturbation by macrobenthos or microphytobenthos mats can be modelled;
- using time series for the boundaries ;
- using a hydrodynamic model, including a spring-neap cycle;
- modelling 2 fractions of sediment: very fine silt and somewhat coarser silt, so the model produces a sediment composition, that can better be compared to measurements;
- modelling riverine silt and marine silt separately, together with measurements of the composition of silt on the Molenplaat;
- extended validation of velocities on the tidal flat in the hydrodynamic model to recent monitoring data, and testing the effect on the sedimentation and erosion pattern.

4 Upscaling

4.1 Objectives

The upscaling activities of ECOFLAT are essential in order to generate relevant information for the management of an entire estuary, based on the knowledge derived from the modelling exercises and field observations of a single tidal flat. The approach which has been followed within the project is to combine physical and biological data at the scale of the estuary with the knowledge from the Molenplaat. Or, in other words, the increased insight in the functioning of the ecosystem at a local scale may show us the way in which to interpret the correlative data on an estuary wide scale.

From the ecological hypotheses regarding forcing with respect to organic material and nutrients, and the hydrodynamical and transport model, providing forcing with respect to sediment composition, stability, current regime and erosional/sedimentation processes, predictions can be made on the spatial and temporal distribution of macrobenthic animals in the estuary. Macrobenthic animals are an important ecological target variable for environmental management in themselves. Moreover, they are a key variable in the assessment of the potential of the estuary for overwintering birds. Not only biomass values of the macrofauna, but also the composition of the fauna, its distribution with respect to depth in the sediment and with respect to exposure time of the flat area are important variables for the birds.

The objective of the upscaling workpackage is to predict macrobenthic biomass and structure at an estuary level, based on available existing data and on the observations made on the Molenplaat. Two methods have been tested and evaluated with respect to their accuracy and validity:

- *habitat suitability evaluation*

  The abiotic environmental conditions (which can be modelled or measured) determine the suitability of a location as a habitat for macrozoobenthic species. Based on suitability curves for species, each location is given a habitat suitability index. Through the use of the EcoMorphological Model (described later) and GIS tools this procedure has been automated.
- *Artificial Neural Networks (ANN) technique*

  ANN is a generic regression method, capable of finding the best multivariate non-linear fit between the abiotic conditions and the abundance of benthic species.

For a more elaborate description on the application of both methods, reference is made to two separate reports, Baptist (1999a) & Baptist (1999b).

4.2 Available data

For Ecoflat an extensive data set on the occurrence of macrobenthos is available. The database that is used is provided by T. Ysebaert of the Institute of Nature Conservation in Brussels. This database consists of 3112 samples taken in the Westerschelde. A comprehensive analysis of this data set is documented in Ysebaert & Meire (1999).

4.2.1 Contents of the database

The database used for this study consists of the following parameters:

- Sample ID
- Date

Abiotic parameters:

- Yearly mean salinity (‰)
- Depth relative to NAP (m)
- Maximum ebb-velocity (m/s)
- Median grainsize (μm)

Biotic parameters:

- density *Cerastoderma edule* (n/m2)
- density *Macoma balthica* (n/m2)
- density *Heteromastus filiformis* (n/m2)
- density *Corophium volutator* (n/m2)
- density *Nephtys cirrosa* (n/m2)
- biomass *Cerastoderma edule* (g AFDW/m2)
- biomass *Macoma balthica* (g AFDW/m2)
- biomass *Heteromastus filiformis* (g AFDW/m2)
- biomass *Corophium volutator* (g AFDW/m2)
- biomass *Nephtys cirrosa* (g AFDW/m2)

Not all 3112 samples contain data for every parameter. The median grainsize in particular is not available in about half of the samples.

The complete database at the Institute of Nature Conservation contains data on more species (Ysebaert & Meire, 1999).

4.2.2 Preliminary data analysis

Before the data is analysed by statistical techniques or by the use of neural networks, a preliminary analysis is performed by using scatter plots of various parameters and Principal Component Analysis. These techniques provide meaningful insight into some of the relationships that are present and their correlation. Ysebaert & Meire (1999) have also analysed these data and came to similar conclusions.

Abiotic parameters

When correlation between abiotic and biotic parameters are looked for it is important to know what dependencies there are between abiotic parameters.

Beforehand, it can be assumed that there is a correlation between three abiotic parameters, i.e., depth, ebb-velocity and median grainsize. These parameters are not independent. It is to be expected that in areas with larger depths, the current velocity may reach higher values, resulting in larger bottom shear stresses that will erode sediment from the bottom, leaving particles with a larger median grainsize.

In stratified systems a relationship may be found between depth and salinity; the salinity is higher at larger depths. This data set contains data taken in the river Scheldt with salinity's between 1‰ and 2‰, via the brackish eastern part, to the saline western part with salinity's up to 32‰. It is not expected that a correlation is found between the wide range of salinity's and depths.

Figures 4.1 to 4.4 show scatter plots between various combinations of abiotic parameters. Figure 4.1 shows a correlation between depth and maximum ebb-velocity as expected. However, this correlation should be interpreted with care. The parameter maximum ebb-velocity is a model-parameter, i.e. this is based on model-results of a 2DH-hydrodynamic model for the Westerschelde. This is a different model than the one described in this report. This model contains a detailed schematisation of the bathymetry of the estuary and uses this as an important parameter to predict tidal velocities. Because the model is a 2-dimensional model, the current velocity does not present the velocity near the bottom, but the depth-averaged velocity.

The scatter plot of maximum ebb-velocity versus median grainsize (Figure 4.2) shows a (rather weak) correlation. Median grainsize increases with higher current velocity. Again, a computed parameter is compared to an observed value. In this particular case, matters of scale can be a problem. The ebb-velocity is computed on a computational grid and represents the mean velocity in a grid cell. The median grainsize is sampled in a much smaller spot. Considering the heterogeneity of surface sediments, comparing these parameters may not lead to a good fit.

Following the relationships between depth and ebb-velocity, and ebb-velocity between grainsize, a correlation between depth and median grainsize can be expected. Figure 4.3 shows the scatterplot of depth versus grainsize. Most samples of grainsize were taken in the intertidal area, where a wide range of grainsize was found. Going to deeper waters, median grainsize does not show a relationship with depth.

Figure 4.4 presents the scatter plot of salinity versus depth. No correlation is found, as expected.

The cross-correlation matrix of abiotic parameters (Table 4.1) shows weak correlation coefficients. This matrix uses the complete data set in which missing values are omitted, leaving 1400 samples. The strongest correlation is between depth and ebb-velocity.

Table 4.1 Cross-correlation matrix for abiotic parameters

	Salinity (‰)	Depth (m)	Ebb-velocity (m/2)	Median grainsize (µm)
Salinity (‰)	1.000	0.2160	0.1321	0.2754
Depth (m)	0.2160	1.000	0.7138	0.3725
Ebb-velocity (m/2)	0.1321	0.7138	1.000	0.4280
Median grainsize (µm)	0.2754	0.3725	0.4280	1.000

Biotic parameters

The purpose of this study is to investigate whether a multivariate correlation between the abiotic parameters and the density or biomass of macrobenthic species exists. Scatter plots of abiotic parameters versus biomass or density of a species are a good start to gain insight in the response of species to environmental conditions.

None of the scatter plots show a good fit on any linear, exponential, logarithmic, power or polynomial function. One of the reasons for this is the large quantity of samples with a density or biomass of zero n/m^2 or zero g AFDW/m^2 respectively. Another reason is the large variance in the density- or biomass-values. A solution for this is to plot the density or biomass on a logarithmic scale. This way the zero-values are omitted and the variance is decreased. Although the scatter plots give a better presentation, correlation between parameters does not improve.

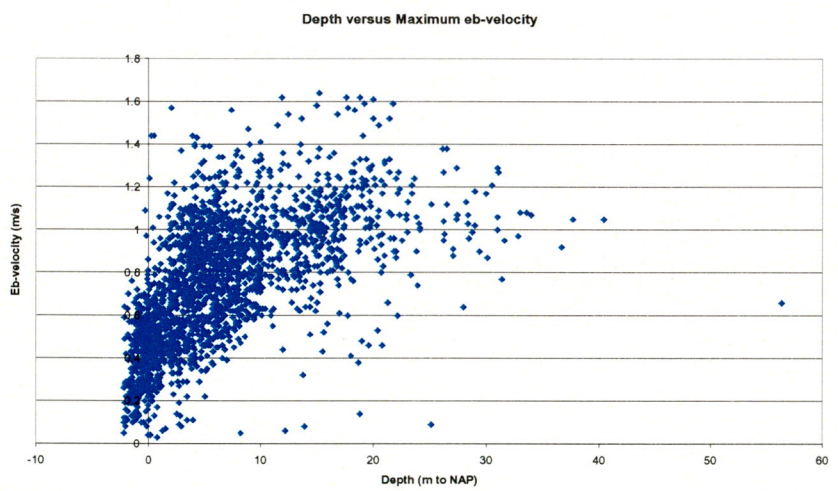


Figure 4.1. Scatter plot of depth versus maximum ebb-velocity

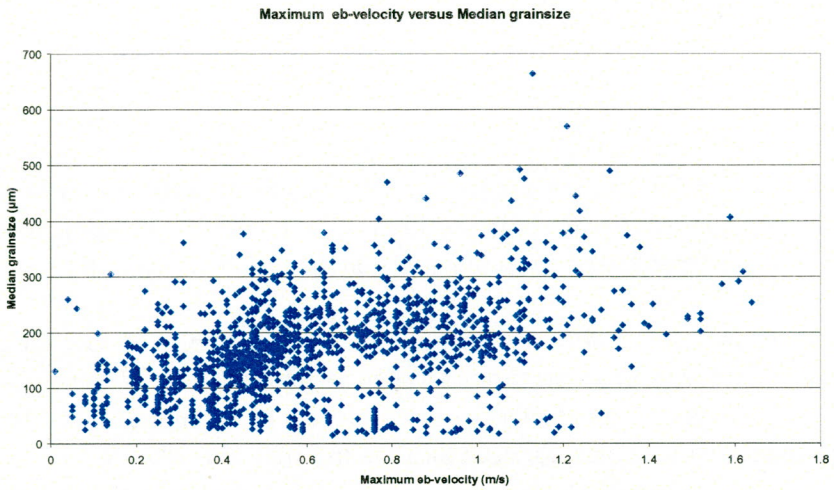


Figure 4.2. Scatter plot of maximum ebb-velocity versus median grainsize

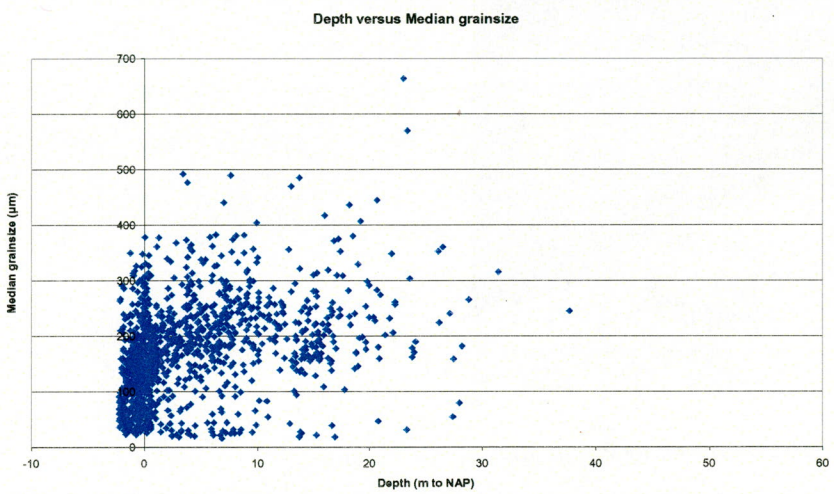


Figure 4.3. Scatter plot of depth versus median grainsize

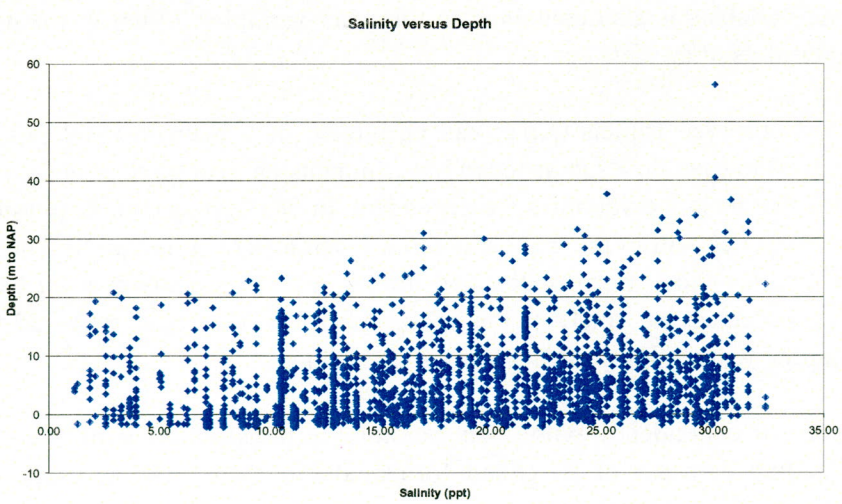


Figure 4.4. Scatter plot of salinity versus depth

4.2.3 Principal Component Analysis

Introduction

Principal Component Analysis (PCA) is a multivariate data analysis technique to assess dependencies in, or the structure of, a data set in the form of N samples in a P-dimensional space. The technique and its interpretation are summarised in the following section.

The idea of principal components is as follows. Suppose that for P variables N observations are available. Such a data set may originate from a water system where at N locations, P (physical, chemical or biological) parameters are observed. In a figure with N=10 observations, and P=3 measured variables, for each location this may look like:

Figure 4.5 Example of N=10 sample points for P=3 parameters

The three variables that are measured in each location may be correlated, i.e. information on one or a subset of the variables is also contained in the other variables. Often this can be recognised from scatter plots of the data-set.

In this example all ten observed triplets (a,b,c) can be plotted in a 3-dimensional (x,y,z)-frame. If the samples tend to span up a sub-space of lower dimension than three (a plane or a line), there is evidence that the a,b,c-variables are dependent. In fact, it may then be possible to remove one of the variables within the triplets without much loss of information. On the other hand, if the N-samples show much scatter and form a set that extends over all three dimensions, the a,b,c-variables are likely to be independent. In that case none of the variables can be discarded without a significant loss of information.

To find out the degree of dependency within the variables a transformation of the data should be applied leading to a set of so called 'principal directions'. As a result, the projections of the data samples on these principal directions (i.e. the components within the new frame) will be mutually orthogonal. Then for each principal direction the variance must be determined. The variance can be seen as a measure for the 'energy' in the data. A larger 'energy' of a principal direction means that it contributes more to the total energy of the system (= total variance of the data set), and as a result this principal direction is more

'important' for the description of the data set. On the basis of their energy the principal directions can thus be ranked. In this ranking the first principal directions give directions in the data-space with the largest variability of the data, whereas the last principal directions are in directions with the smallest variations.

The principal directions and variance are obtained from the eigenvectors and eigenvalues of the cross-correlation matrix.

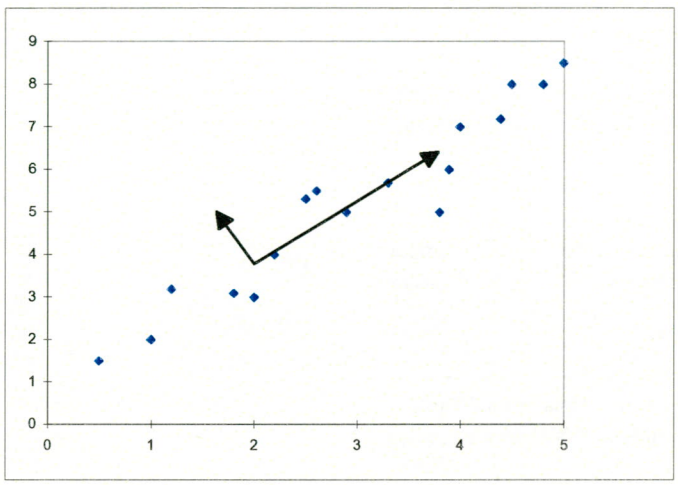


Figure 4.6 Two dimensional scatter plot in x,y-space

Figure 4.6 shows an example of a 2D scatter plot in the x,y-space. The PCA will generate two principal directions (two orthogonal vectors) in the two-dimensional space. The first principal direction will be along the direction of the largest variance. The second principal direction will have a direction where variance is smallest and orthogonal to the first direction.

4.2.4 PCA on abiotic data

A Principal Component Analysis is applied on N=1400 data samples of P=4 abiotic variables that do not contain any missing values. Table 4.2 presents some elementary statistics of the data set.

Table 4.2. Elementary statistics of data set.

Variable	Mean	Stand_dev	Variance	Min.	Max.
Salinity (‰)	16.26	6.771	45.85	1.290	31.61
Depth (m)	3.527	6.168	38.04	-2.200	37.70
Ebb-velocity (m/s)	0.6088	0.2973	0.08841	0.04	1.64
Median grainsize (μm)	167.6	84.38	7119	15.67	664.4

The results of the PCA denote the directions of variance in the cloud of plots in four dimensions r_1 salinity, r_2 depth, r_3 ebb-velocity and r_4 median grainsize. Every principal direction is normalised at length 1 (the sum of the squares of each component equals one). The first component explains the largest variance in the data set, the corresponding four components denote the ratio of importance between the variables. When for example the

component r_1 is large compared to the other components, this means that this variable explains the largest variance in the data set and that the other variables are independent of the first. On the other hand when all four components have a similar value this means that there is a mutual dependency between the parameters.

Table 4.3. Results of PCA.

```
Eigen Value 01    is  2.1342   , explaining  53.3556% of total variance
Cumulative variance is  2.1342   , explaining  53.3556% of total variance
Eigen Vector 01    is:
 0.30  0.58  0.58  0.48
 ------------------------------------------------------------------------
 Eigen Value 02    is 0.94832   , explaining  23.7080% of total variance
Cumulative variance is  3.0825   , explaining  77.0635% of total variance
Eigen Vector 02    is:
 0.87 -0.26 -0.36  0.21
 ------------------------------------------------------------------------
 Eigen Value 03    is 0.64302   , explaining  16.0754% of total variance
Cumulative variance is  3.7256   , explaining  93.1390% of total variance
Eigen Vector 03    is:
-0.38 -0.36 -0.14  0.84
 ------------------------------------------------------------------------
 Eigen Value 04    is 0.27444   , explaining   6.8610% of total variance
Cumulative variance is  4.0000   , explaining 100.0000% of total variance
Eigen Vector 04    is:
 0.12 -0.68  0.72 -0.12
```

The first principal direction explains 53% of total variance. The most important components in this vector are for depth (0.58) and ebb-velocity (0.58), which are partly dependent on median grainsize (0.48) and to a lesser extent on salinity (0.30). Overall, values are relatively close together, indicating a mutual dependency between these parameters. There is a positive correlation between the parameters, indicating that when depth increases, ebb-velocity increases, median grainsize increases and salinity increases. The fact that salinity increases with depth can be explained by the observation that average depth at the saline mouth of the Westerschelde is higher than in the brackish part of the estuary.

After the more or less equal contributions of the abiotic variables to the first principal direction, salinity is dominant in the second principal direction. The second principal direction explains 24% of total variance.

The third principal direction explains 16% of total variance and is roughly directed towards the median grainsize variable (0.84) with small values for the other variables. The fourth principal direction explains the remaining 7% of total variance.

It can be concluded that the majority of total variance is explained by the first three principal directions. The first principal direction explains about half of the variance and has depth and ebb-velocity as most contributing components, median grainsize to a somehow lower extend, followed by salinity. This indicates that depth, ebb-velocity, median grainsize and salinity are dependent parameters, but because together they only explain 53% of total variance, their correlation is not very large.

4.3 Description of the Ecomorphological module

The EcoMorphological Module is a software module developed in the LWI-program that computes the morphological development of the Westerschelde on a long time scale and the suitability of the habitats of benthic species under scenarios of dredging and dumping (Wang et al., 1997).

For the computation of long-term morphological changes in the Westerschelde estuary the one-dimensional morphological model Estmorf is used. Estmorf is a model for estuarine morphology. The semi-empirical algorithms in Estmorf are based on the assumption that after a disturbance in the morphological system the system tends to go to an equilibrium state by sedimentation and erosion. The hydrodynamic data required for defining the morphological equilibrium and for calculating sediment transport and the corresponding morphological changes are calculated with the one-dimensional hydrodynamic model Implic. The morphological changes in the estuary are used in a 1D to 2D post-processing to determine the changes in bathymetry and hydrodynamic condition on a 2D grid of 60 by 60 metres. Results of these hydro- and morphodynamic computations are bathymetry, minimum and maximum depth, current velocities and bed shear stresses. The abiotic environmental conditions determine the suitability of a location as a habitat for macrozoobenthos species. Based on suitability curves for species each grid cell is given a habitat suitability index.

Evaluation and analysis of morphological, hydrodynamic and ecological results for different dredging scenario's, may support decision-makers to optimise the dredging operation with respect to ecological damage.

4.4 Habitat suitability analysis

4.4.1 Univariate response curves

For four macrobenthos species, univariate response curves were applied on environmental conditions for salinity, depth, current velocity and median grainsize. Each variable is available on a spatial grid of 60x60 metres covering the Dutch part of the Westerschelde estuary:

- Depth: bathymetry of 1992 derived from WAQUA-model, Figure 4.7.
- Median grainsize: Interpolated coverage of median grainsize samples (McLaren data) in classes: 50μm, 100μm, 125μm, 150μm, 175μm, 250μm, >250μm, Figure 4.8.
- Salinity: interpolated yearly averaged, salinity range from 33.0 at Vlissingen to around 10.0 at Prosperpolder, Figure 4.9.
- Current velocity: maximum current velocity over an average tidal period derived from WAQUA-model, applied for 1992 bathymetry, Figure 4.10.

The variables Depth and Current velocity are computed within the EcoMorphological Module. Therefore it is possible to compute changes in these parameter values as a result of the dredging and dumping operation. The salinity gradient over the estuary is very simplified. Salinity is only relevant on a large spatial scale. This variable is not dependent on dredging-scenarios in the EcoMorphological Module. The median grainsize distribution is

derived from the McLaren samples taken in 1992-1993. This parameters also is not dependent on dredging scenarios in the EcoMorphological Module. In reality the salinity and median grainsize distribution may change due to dredging operations, but these changes are not implemented in the EcoMorphological Module.

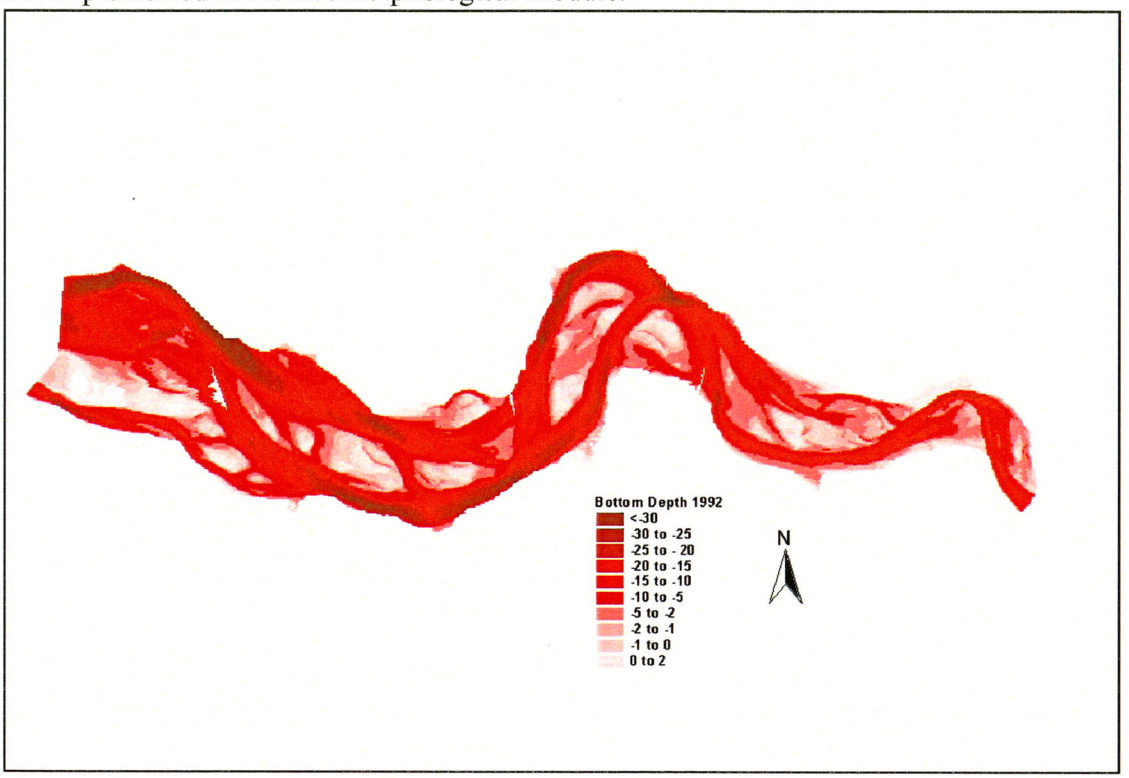


Figure 4.7. Bottom depth for 1992.

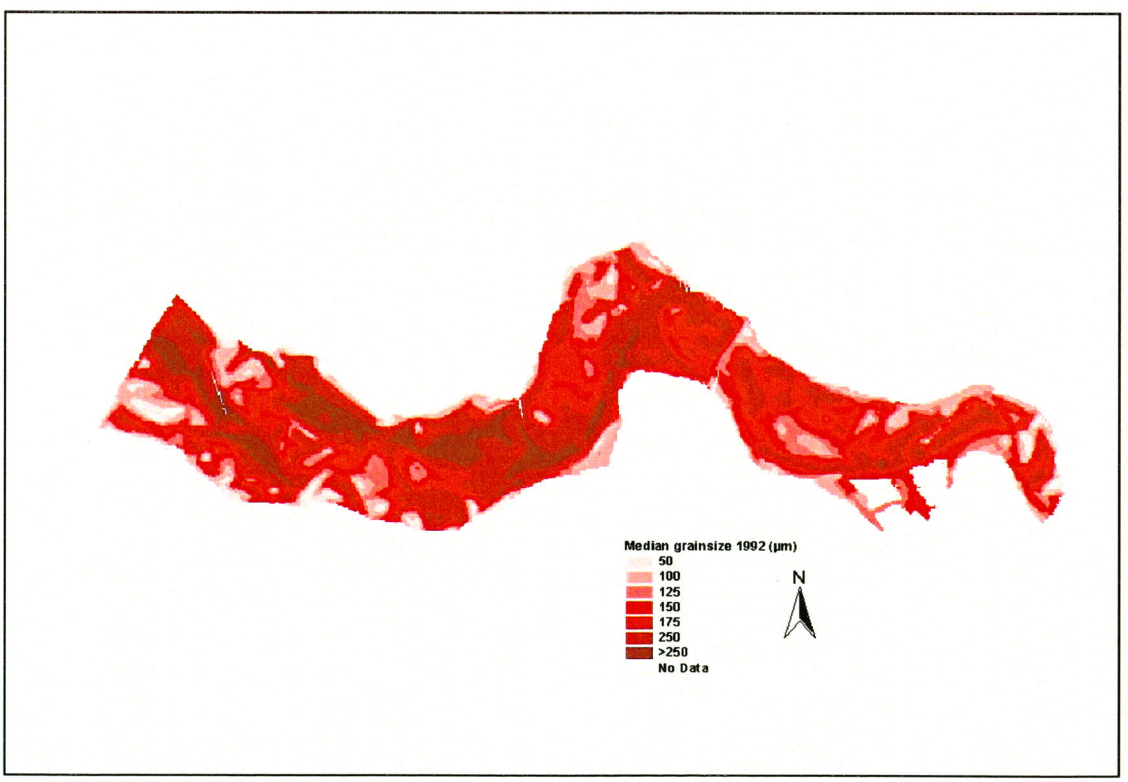


Figure 4.8. Median grainsize classes for 1992-1993 (McLaren data).

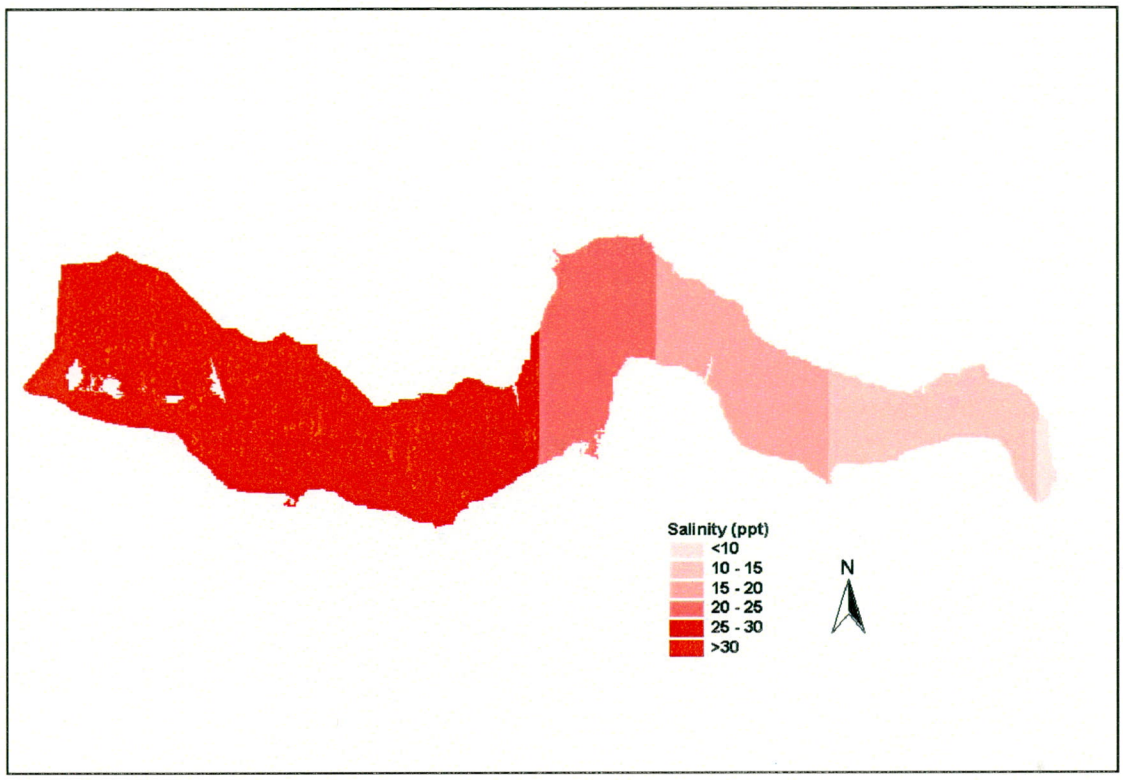


Figure 4.9. Salinity gradient.

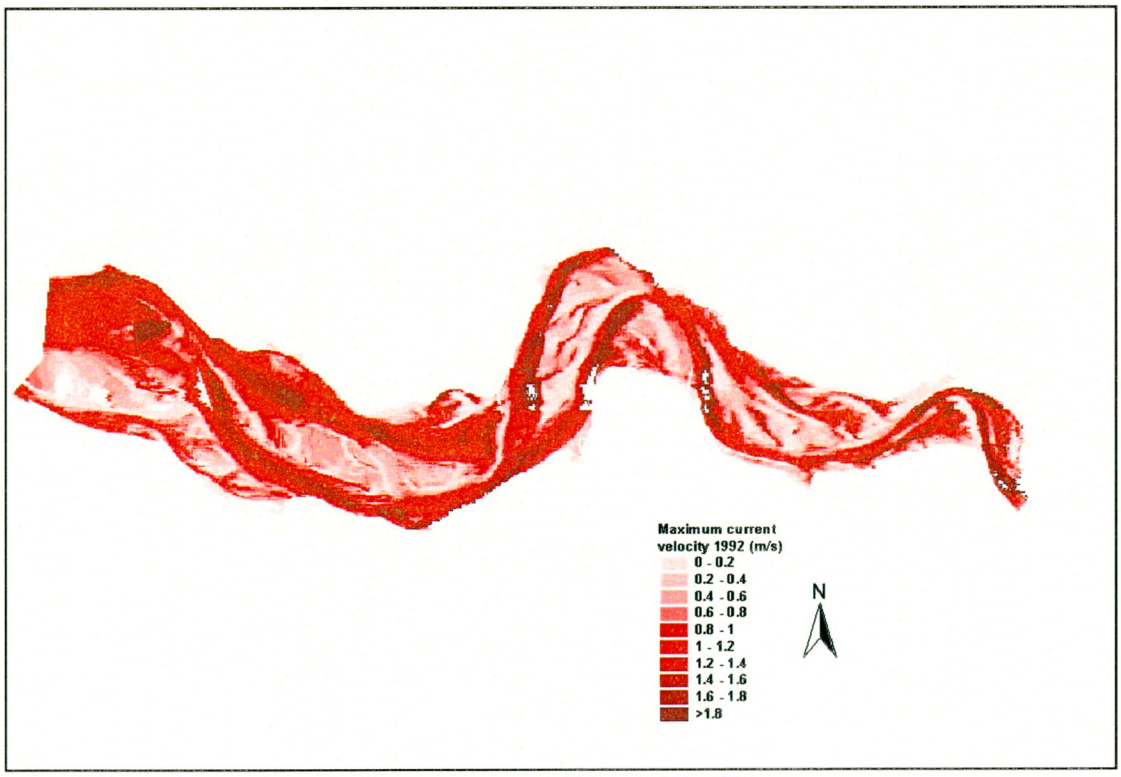


Figure 4.10. Maximum current velocity over an average tidal period for 1992.

4.4.2 Response Curves for Macrobenthos

The individualistic responses of some macrobenthic indicator species to environmental variables were analysed. For this aim the database of available macrobenthic samples of the Westerschelde estuary was analysed by means of stepwise logistic regression for presence or absence. This yields a probability score for presence at a certain environmental condition. The response curves were defined for univariate responses (one variable) as well as for multivariate responses (multiple variables together). This work was carried out by Tom Ysebaert and Patrick Meire of the Institute for Nature Conservation in Brussels, Belgium and is documented in Ysebaert & Meire (1999).

Formulation

The logistic regression equation that is used has the following form for a univariate regression:

$$P = \frac{e^{b0+b1x+b2x^2}}{1+e^{b0+b1x+b2x^2}}$$

In which:

P	= probability
x	= environmental variable
$b0$, $b1$, $b2$	= regression coefficients

A second order response curve will result in a symmetric bell-shaped (Gaussian) curve, while a first order response curve (b2=0) will result in a sigmoid shape. The equation can be expanded with other variables for multivariate regression.

Environmental variables that were used in the regression contained yearly averaged salinity (‰), depth (m to NAP), maximum current velocity (m/s) and median grainsize of the sediment (µm).

The univariate response curves are defined for four benthic species:

1. Edible Cockle *Cerastoderma edule*
2. Baltic Tellin *Macoma balthica*
3. *Nephtys cirrosa*
4. *Corophium volutator*

The response curves are presented in Figures 4.11 to 4.15

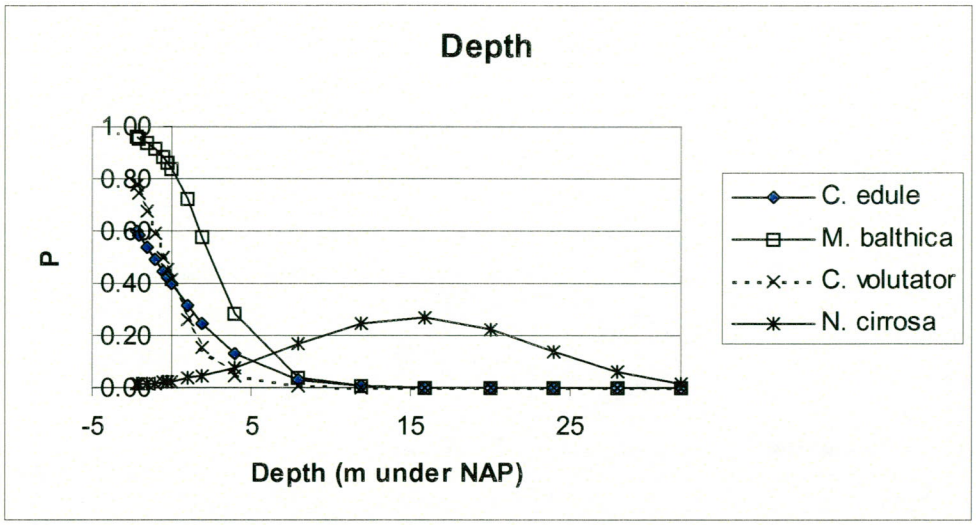


Figure 4.11. Response curves for depth.

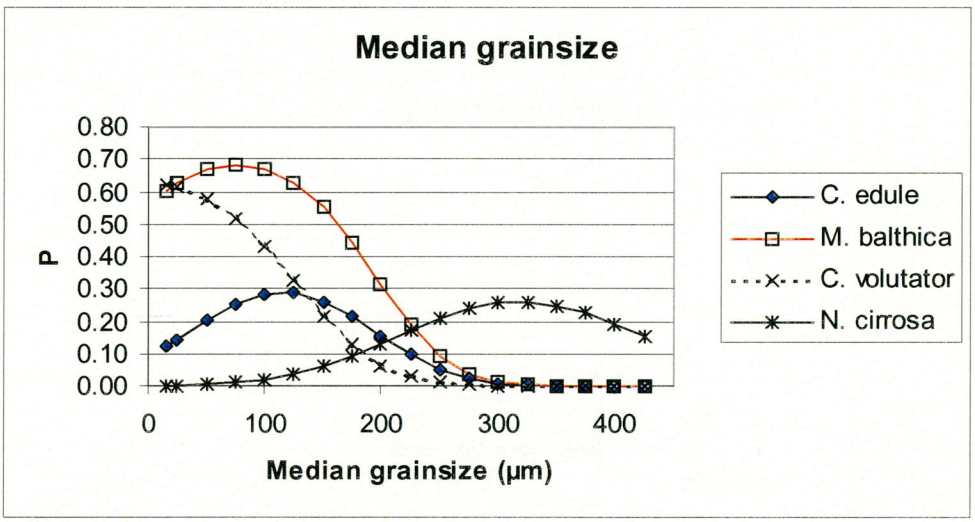


Figure 4.12 Response curves for median grainsize.

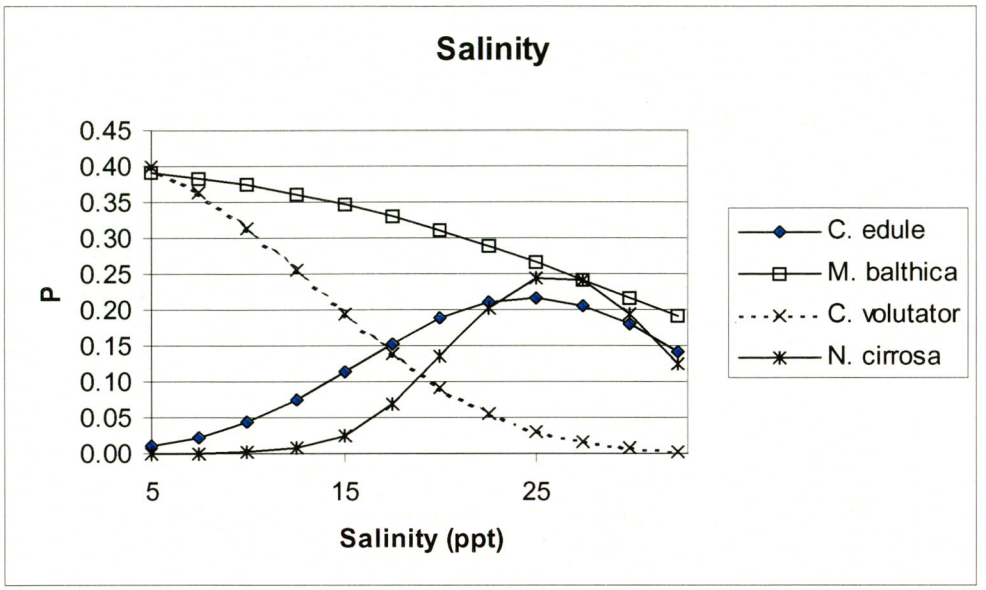


Figure 4.13 Response curves for salinity.

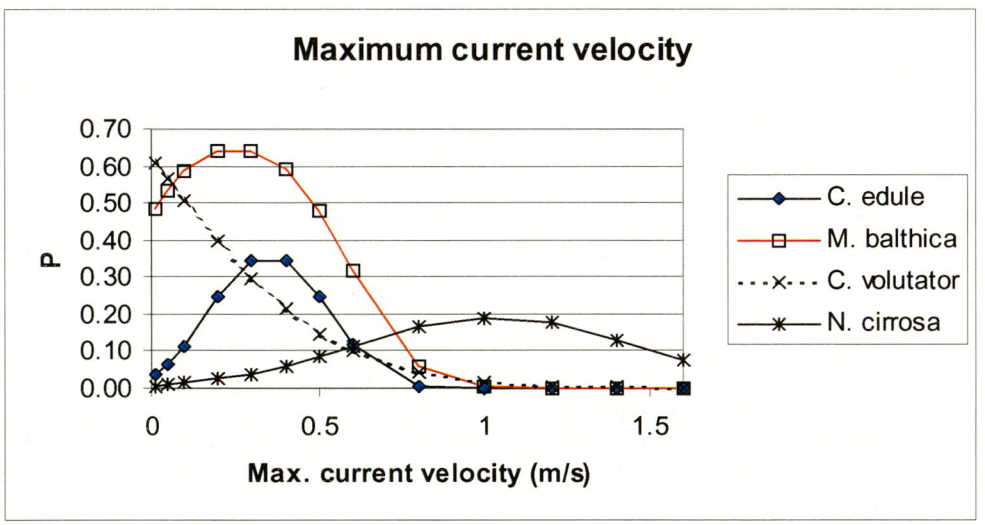


Figure 4.14. Response curves for maximum current velocity.

For all four species the univariate response curves were applied on the abiotic data. The results denote the probability P that the species is present as a function of a specific environmental parameter. As an example, the probability for the Cockle (*Cerastoderma edule*) as a function of depth is given in Figure 4.15 (for other species and parameters reference is made to a separate report, Baptist, 1999a).

The abiotic data for the Westerschelde that are used in the EcoMorphological Module are different than the data that are used to define the probability curves. The median grain size data for example stem from a different measurement campaign and the salinity is represented in a very simplified form. These differences may cause deviations in the habitat maps.

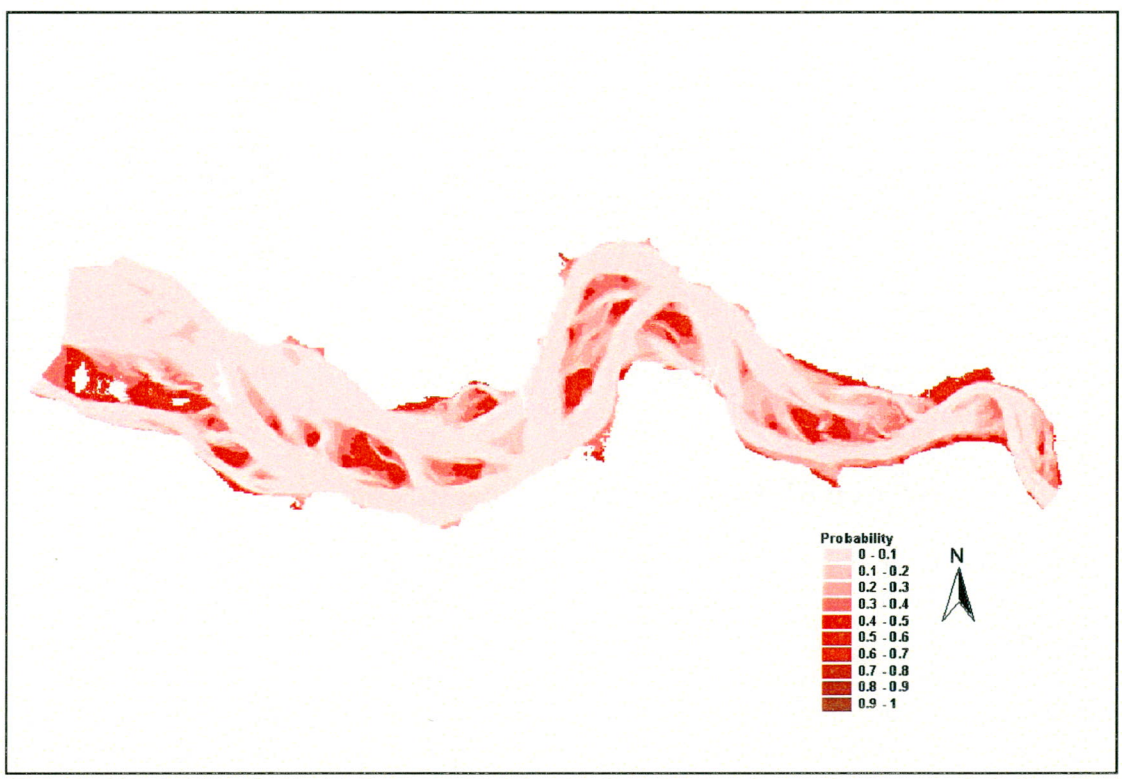


Figure 4.15. Probability that Cockle is present related to depth

4.4.3 Habitat Suitability Indices

The univariate response curves were used to define Habitat Suitability Indices (HSI) with an interval between 0 and 1. The probability that a species will be present is something different than the suitability of a location as a habitat. However, in this case it is assumed that at places where the probability is higher, the habitat conditions are better. The maximum probability P that result from the response curves was transformed to the value 1, when the maximum probability did not reach 1. These habitat suitability indices were applied on the abiotic data according to the 'minimum rule', i.e., the most limiting parameter (the lowest habitat suitability score) denotes the local habitat suitability. The resulting habitat map for one species as an example, the *Corophium volutator,* is presented in Figure 4.16.

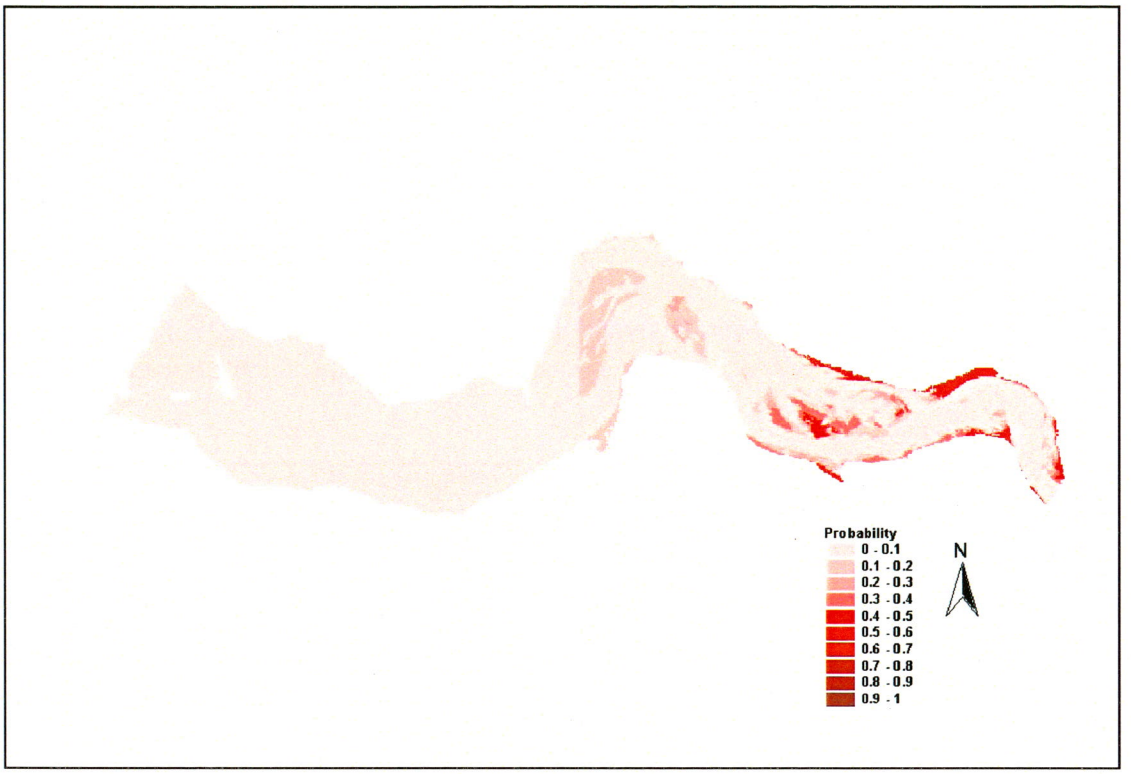


Figure 4.16. Habitat Suitability Index for *Corophium volutator*, for 1992.

Correlation of HSI with biomass measurements

The habitat suitability maps predict the HSI for each grid cell in the computational grid of the Westerschelde. In total there are 73,732 cells with a size of 60x60 metres. In this paragraph the predictions are compared to the measurements of biomass for each benthic species. The biomass measurements were first imported in ArcView to establish in what computational cells the samples are located. Subsequently the predicted HSI in a specific cell is compared to the measurement(s) in that cell. This way, 2469 biomass measurements could be compared to the predicted HSI. It is assumed that in locations with a high habitat suitability index the measured biomass is higher as well. For the Cockle the results of the comparison is presented in Figures 4.17.

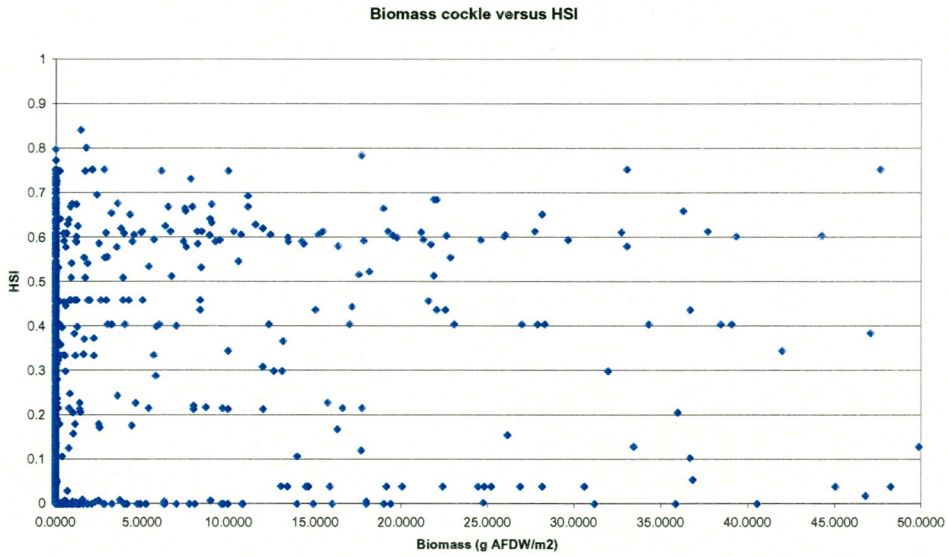


Figure 4.17. Measured cockle biomass versus predicted habitat suitability.

The correlation between the predicted HSI for a computational cell and the measurement is that cell is not very good. The correlation coefficient for *Cerastoderma* is 0.13. For the other species the correlation is in the same order, with *Macoma* having the highest (i.e. 0.30). Calculating the logarithm of the biomass values (slightly) improves the correlation. Particularly the correlation for the cockle improves (0.30). This may be due to the large variance in biomass values for cockles, which is decreased by calculating the logarithm. This implies that there is a logarithmic relationship between HSI and biomass rather than a linear relationship.

A multivariate response curve for the Cockle, based on all four abiotic variables together has also been applied. The results were compared to the results of the habitat suitability maps and it seems likely to conclude that the multivariate curve is more (too) sensitive to maximum current velocity values than to other abiotic variables. The correlation between the predicted multivariate probability and the (logarithm of the) biomass of Cockles was also computed and was not very good. The resulting habitat map for the Cockle is presented in Figure 4.18.

Figure 4.18. Habitat suitability index, using multivariate response, for the Cockle in 1992.

Historical analysis of the Cockle suitability of the Westerschelde

In the previous analysis, the response curves were applied on abiotic data from 1992. Abiotic data is also available for a period around 1970. The Ecomorphological Module contains the 1968 bathymetry and corresponding water movement. The median grainsize for this historic situation is derived from a map of median grainsize samples taken in the period 1973-1979. The data for salinity remain unchanged. These historic data were applied on the multivariate response curves for Cockle and the result is presented in Figure 4.19.

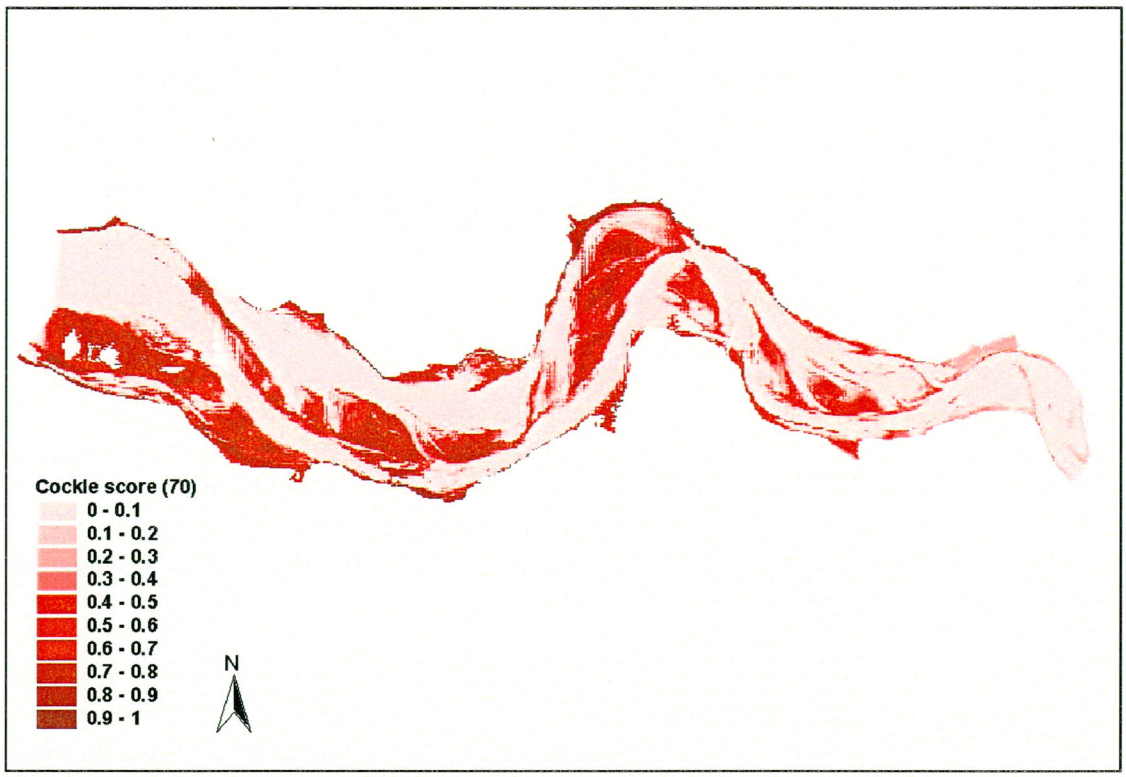


Figure 4.19. Habitat suitability index, using multivariate response, for the Cockle in the seventies.

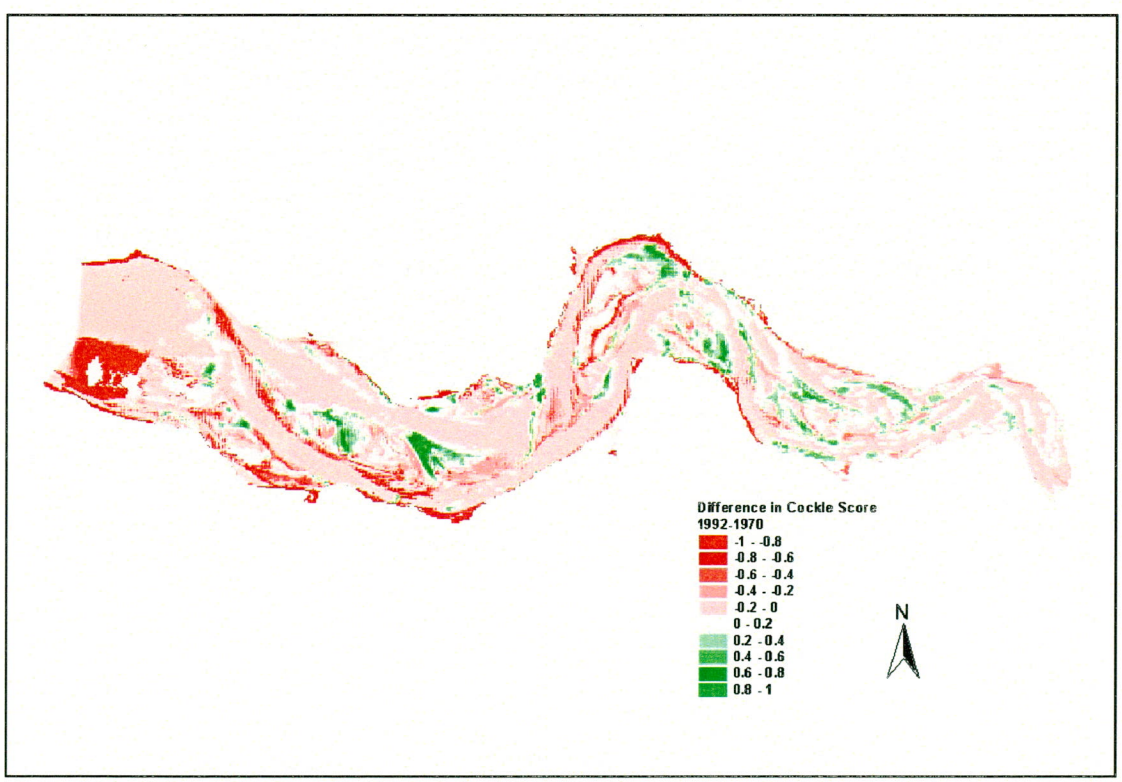


Figure 4.20. Difference between Cockle score for 1992 and 1970.

For each grid cell the difference was computed between the score in 1992 (Figure 4.18) and the score around 1970 (Figure 4.19), the result is presented in Figure 4.20. Note that the

westernmost part of the Westerschelde has got a score for 1970, but not for 1992, so the presented scores for the difference in this part are not correct.

Based on these results the net probability score is negative for the western half of the estuary (corrected for the westernmost part) and nil to slightly positive for the eastern half. Overall, the habitat conditions for the Cockle have worsened since 1970.

4.4.4 Conclusion and discussion

A number of reasons can cause the mismatch between the predicted values and the measured biomass. First, the biomass of macrozoobenthos is seasonally dependent. In summer a higher biomass is found than in winter. The *density* of a species is in general even more dependent of the season. The measurements of biomass are taken throughout the year and also in different years. The data set even contains measurements for one location that are collected in four seasons (Molenplaat). These measurements show four different values for biomass for one location.

Extreme events do also play an important role in the biomass of benthic species. It is known that Cockles are extremely sensitive to severe winters. After a severe winter the total biomass of Cockles can decrease dramatically. This may be followed by a successful spatfall in spring. Two years after, the Cockles are grown-up and a very high biomass may be the result.

In general Cockles occur in (sub)littoral cockle beds. This means that locally a very high density and biomass can occur. The measurements do contain biomass values over 300 g AFDW/m^2. The presence of these cockle beds makes the prediction of habitat suitabilities more difficult because there is a kind of binary "all or nothing" situation. Two locations with similar abiotic conditions might show no biomass or a high biomass. A species such as the *Macoma* has a more uniform and therefore less patchy distribution. The spatial variability of the *Macoma* can be predicted more accurately with habitat models.

Another concern is that of spatial scales. The measurements give a very local biomass. It can be observed in the field that the heterogeneity of biomass is very high. This means that a biomass measurement 10 metres further away can have a completely different result. The model results however provide the average abiotic condition in a computational grid of in this case 3600 m^2. Within this cell, homogeneous model conditions rule, but the heterogeneity of the biomass is large. This may lead to large differences between the predicted and the measured biomass.

Other reasons that may explain the small correlation may be found in biotic interactions (predation, competition) or in human impacts (fishery).

The univariate response curves that denote the probability for presence of a species under certain environmental conditions were used in this study to define Habitat Suitability Indices (HSI). The probability that a species will be present is something different than the suitability of a location as a habitat. However, in this study it is assumed that at places where the probability for presence is higher, the habitat conditions are better. The probability scores

were transformed to values between zero and one. When the probability for a certain environmental parameter reached a maximum of 0.70, it was assumed that the habitat suitability reached 1.0. The habitat suitability indices were subsequently correlated to the observed biomass. Unfortunately the correlation was not very high. Therefore, a different technique is tried: an Artificial Neural Network (ANN). An ANN is a generic regression method. One of the advantages is that it is not necessary to decide beforehand what kind of non-linear function is probably most suitable. An ANN is capable of finding any best multivariate non-linear fit between input and output parameters.

4.5 Neural networks method

For a more elaborate description on the application of artificial neural networks on Westerschelde data reference is made to a separate report, Baptist (1999b).

4.5.1 Introduction

Neural networks are based on research which views the brain as a parallel computer. The usual strategy has been to develop simplified mathematical models of brain-like systems and to study these to understand how various computational problems can be solved by this models. The human brain is said to have roughly 10^{10} neurons connected through about 10^{14} synapses. Each neuron is itself a complex device which compares and integrates incoming electrical signals and relays a non-linear response to other neurons. The brain certainly exceeds in complexity any system which physicists have studied. Nevertheless, there do exist many analogies of the brain to simpler physical systems (Domany et al., 1995).

Neural Network Architecture

A neural network is made of *neurons* and *synapses*. Other terms that are often used are *units* for neurons and *connections* for synapses. The most common models take the neuron as the basic processing unit. A neuron is characterised by an activation value, an output value, and a bias value. Each such processing unit is connected to other neurons by a set of input and/or output connections. Each connection has an associated weight which determines the effect of the incoming input on the activity level of the output. These weights can be positive or negative. The neuron compiles the weighted sum of the inputs and adds a bias to calculate the activation value. Then it determines the output value by a linear or non-linear transfer function of its activation value, usually a sigmoid, exponential or logistic function (Rumelhart et al., 1994).

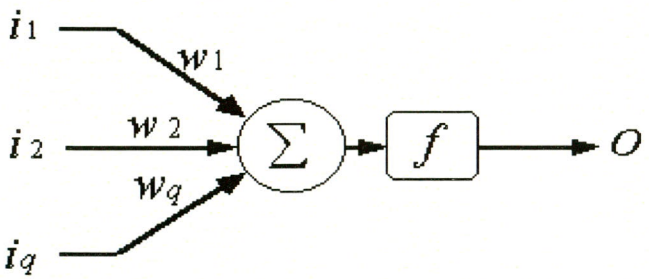


Figure 4.21. The processes in a neuron, i = input, w = weight, Σ = summation of weights, f = transfer function, o = output

The most common architecture of a neural network is the feedforward network, in which the direction of all synapses goes one way from input to output. The activity of a given unit thus cannot influence its own inputs. The set of input neurons are connected with the set of output neurons through a set of hidden neurons. Most often a single hidden layer is used.

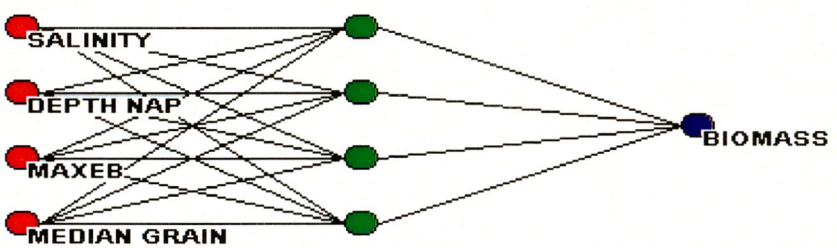


Figure 4.22. A typical feedforward neural network

The input neurons distribute the inputs over the hidden neurons. Each hidden neuron then evaluates the weights and inputs in the above mentioned non-linear way and sends an output value to the output neuron(s). Each output neuron compiles the weighted sum, which gives the output value.

Training Neural Networks

The major advantage of using a neural network is that it is able to learn, that is, it can arrive at a set of connections that estimate the desired set of output values, given the input values. This is done by *training* the neural network. Training of neural networks is the procedure to estimate the weights and biases in the network based on the minimisation of an error criterion. Most often the generalised delta rule by Rumelhart, Hinton and Williams is used, where the square of the error between output and desired output is minimised (Pao, 1989). The most common way to train a network is called the *back propagation method*. The network is provided with a set of example input- and corresponding output values, called a *training set*. The input values flow through the network generating a set of output values. The actual output is compared with the desired output and the goodness of the match is computed, using the least squares error. If the target output differs significantly from the

generated output a change is made to some of the connections. This process is iterated until the least squares error is minimised and the generated output and the desired output match (Rumelhart et al., 1994; Pao, 1989).

After this the network is fed another known set of input/output pairs, the *test set*. This is done to test the performance of the network. The configuration of connections, including their weights, represents the function it computes (Rumelhart et al., 1994).

4.5.2 Pre-processing of data

Biomass and density

The data on densities and biomass show a very large variance and contain a lot of zero-values. Zero-values are different from missing values in the data set. A zero-value may not be discarded because it means that the measured biomass or density was actually zero g/m^2 or n/m^2 respectively.

Prior to the neural network analysis the biomass- and density-values are scaled and transformed to logarithmic values by using:

$$^{10}\log\left(1+\frac{10\cdot x}{x_{\max}}\right)$$

where x_{max} is the maximum biomass or density in the data set and x is the biomass or density of a sample. When x=0, this yields log 1 = 0, and when x=x_{max}, this yield log 11 = 1.0414. Subsequently all biomass and density values are logarithmically scaled between 0 and 1.0414.

Sample date

The sample date gives the day, month and year of sampling. The relevant information for seasonal influences is the month of sampling. Therefore the sample month is used for further analysis. The sample month is a figure ranging from 1 to 12. This representation is not suitable for regression analysis. Suppose that there is a seasonal effect in which biomass in summer are high and biomass in winter and spring are low, according to Figure 4.23. Now intermediate values (6 to 9) for month correlate with a high biomass, whereas low values (1 to 5) as well as high values (10 to 12) correlate with low biomass. Linear regression is useless in this situation.

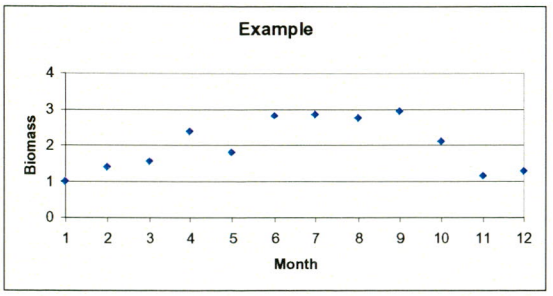


Figure 4.23. Example of biomass-seasonality

The sample month is decomposed into two functions:

$$SIN\left(\frac{2\pi \cdot (month - \frac{1}{2})}{12}\right) \text{and } COS\left(\frac{2\pi \cdot (month - \frac{1}{2})}{12}\right)$$

Each month is now represented by a sine or cosine-function scaled between 0 and 1.

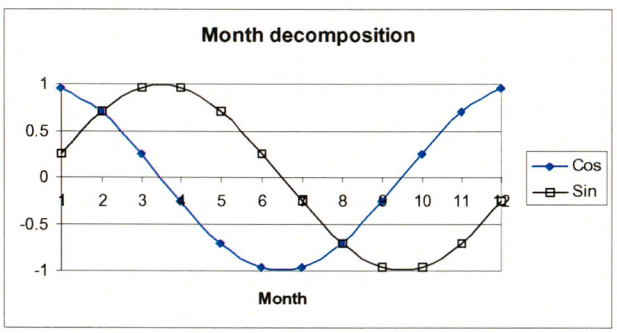


Figure 4.24. Decomposition of *month* into a sine- and cosine-function

When the value for month is decomposed into two components, seasonal variations show a better fit to one of the functions. In this case there is a linear fit to the COSINE value of month.

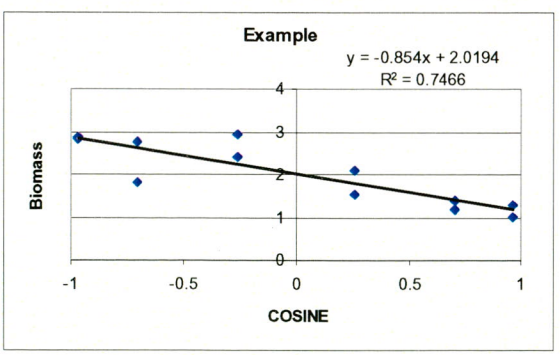


Figure 4.25. Example of linear regression through COSINE component

4.5.3 ANN Architecture

Number of neurons and layers

The database consists of four abiotic parameters (salinity, depth, ebb-velocity and median grainsize), a sample date (dd-mm-yy) and the density and biomass of five benthic species. The sample date is decomposed in two components. Each input neuron is fed with data from one abiotic parameter. The total number of input neurons therefore is six. For each benthic species, or for each functional group of species or for the parameter "total biomass", a separate network is configured with one output neuron. Each network is limited to one hidden layer, which has a variable number of hidden neurons.

Transfer functions

The neural network software that is used for this study is developed at WL | delft hydraulics. This software gives a choice between two transfer functions within the output neurons, a linear transfer function or an exponential transfer function. The hidden neurons always use a sigmoid transfer function. Usually, neural networks are applied on non-linear problems, so that non-linear transfer functions work best. In this study both transfer functions will be tried.

Bias weight

The sigmoid hidden neurons use a "bias" or "threshold" term in computing the net input to the unit. This bias term can be compared to an intercept for a linear regression model. A bias term can be treated as an extra connection from the input layer with a constant value of one. Hence the bias can be learned just like any other weight.

Standardising

All of the input data are standardised before they are fed into the network. The input values are standardised to "standard normal" variables with mean 0 and standard deviation 1. Standardising input variables tends to make the training process better behaved by improving the numerical condition of the optimisation problem.

4.5.4 NN runs on total biomass

In these runs neural networks were trained with the logarithmically scaled total biomass as output. This is the sum of the biomass of all five species *Cerastoderma edule*, *Macoma balthica*, *Heteromastus filiformis*, *Corophium volutator* and *Nephtys cirrosa*. Before scaling the total biomass ranged from 0.0 to 221.6 g AFDW/m^2. The completely filled data set (no missing values) contains 1349 samples.

Six input neurons are used:
1. Cosine function of sample month
2. Sine function of sample month
3. Salinity (ppt)
4. Depth (m to NAP)
5. Maximum ebb-velocity (m/s)
6. Median grainsize (μm)

One output neuron is used for the logarithmically scaled total biomass.

To test neural network behaviour and performance, various neural network configurations and both transfer functions are applied on the complete data set of 1349 samples. No use is made of a separate training set and testing set. Each training session consists of 300 iterations. Two training sessions are run each time, the best results of both are chosen.

4.5.5 NN runs on all data

First a variety of network configurations is tried on the complete range of output values between 0 and 1.0414. A linear transfer function as well as an exponential transfer function is used and the number of hidden neurons is gradually increased. For each configuration a synopsis of the results is described together with the correlation coefficient and the Root Mean Square (RMS) error between the target values and the NN output.

Linear network without hidden neurons

This configuration is similar to a linear, multivariate regression. Results show that the network is highly underestimating larger values; the maximum of NN output is 0.27. This network also predicts negative values, which is absolutely wrong. The zero-biomass values are predicted within a range of -0.1 and 1.6. Correlation between target and NN output is 0.50, RMS error is 0.1171.

Linear network with one hidden neuron

This network is underestimating higher values; the maximum of NN output is 0.54. Negative values are not predicted, but there is a lower boundary of 0.03; i.e. the network is not predicting zero values. Correlation between target and NN output is 0.63, RMS error is 0.1057.

Linear network with two hidden neurons

This network is underestimating higher values; the maximum of NN output is 0.50. The output shows many values of 0.10 and has 0.075 as lower boundary. Correlation between target and NN output is 0.69, RMS error is 0.0978.

Linear network with three hidden neurons

This network is underestimating higher values; the maximum of NN output is 0.51. Negative values are predicted. Correlation between target and NN output is 0.74, RMS error is 0.0914.

Linear network with four hidden neurons

This network is underestimating higher values; the maximum of NN output is 0.55. Negative values are predicted. Correlation between target and NN output is 0.76, RMS error is 0.0888.

Linear network with eight hidden neurons

This network is underestimating higher values; the maximum of NN output is 0.62. Negative values are predicted. Correlation between target and NN output is 0.76, RMS error is 0.0872. Results are shown in Figure 4.26.

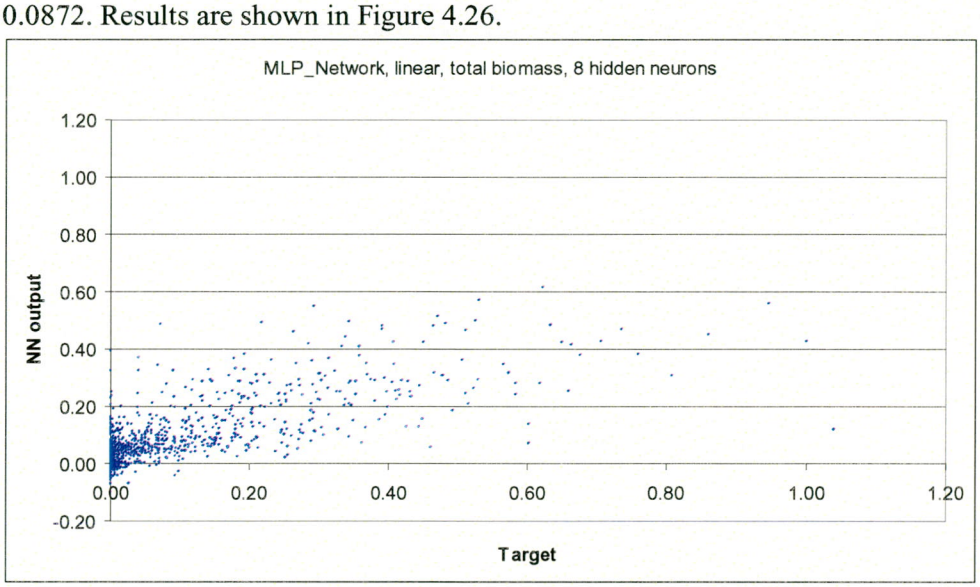


Figure 4.26. NN results for total biomass, exponential, 8 hidden neurons.

Exponential network without hidden neurons

In this network, the output is generated by an exponential function which has the weighted mean of the connection-weights as input. An exponential function is more capable to compute extreme values.

This network is underestimating higher values; the maximum of NN output is 0.62. There are no negative values predicted. Correlation between target and NN output is 0.62, RMS error is 0.1064.

Exponential network with one hidden neuron

This network is underestimating higher values; the maximum of NN output is 0.55. There are no negative values predicted, but there is a lower boundary of 0.03. Correlation between target and NN output is 0.63, RMS error is 0.1056.

Exponential network with two hidden neurons

This network is highly underestimating higher values; the maximum of NN output is 0.36. Correlation between target and NN output is 0.71, RMS error is 0.0950.

Exponential network with three hidden neurons

This network is underestimating higher values; the maximum of NN output is 0.45. Correlation between target and NN output is 0.76, RMS error is 0.0886.

Exponential network with four hidden neurons

This network is underestimating higher values; the maximum of NN output is 0.50. Correlation between target and NN output is 0.78, RMS error is 0.0850.

Exponential network with eight hidden neurons

In general this network is underestimating output values. The linear regression through the scatter plot of targets versus NN output is described by y=0.74x, with an R^2 correlation of 0.64. Ideally this should be y = x, with an R^2 of 1.

Correlation between target and NN output is 0.81, RMS error is 0.0795. Results are shown in Figure 4.27.

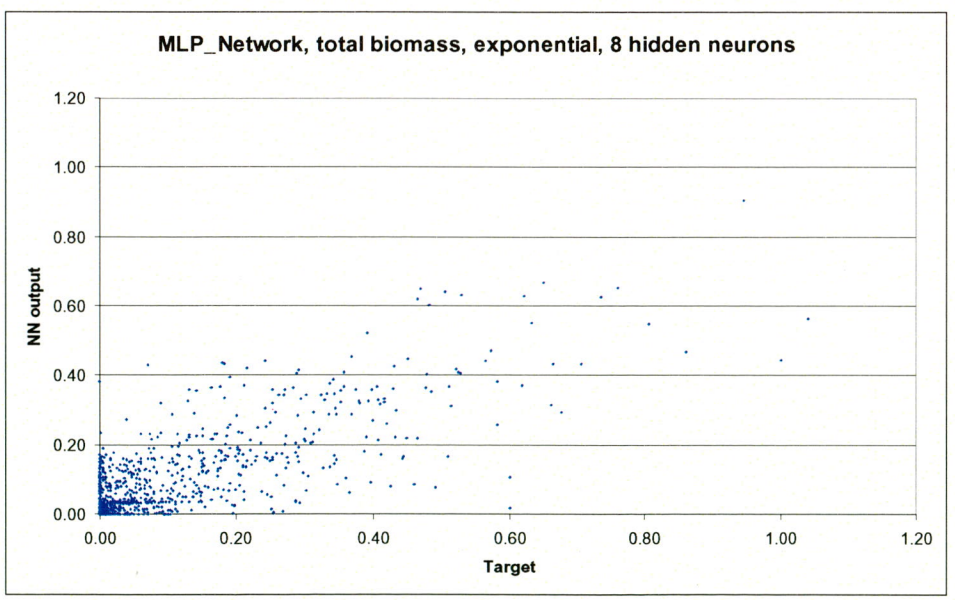


Figure 4.27. NN results for total biomass, exponential function, 8 hidden neurons.

Results and discussion

All network configurations suffer from underestimation of the output. The target values contain a lot of relatively small values and only few higher values, so the network weights are adjusted to the lower rather than the higher values. Increasing the number of hidden neurons helps, because the degrees of freedom in the network (the number of connections) increase fast and a more tailor-made output can be generated. However, this has a major disadvantage. When a neural network is trained on a data set using many hidden neurons, the network is more or less 'specialised' on the data set, and is no longer able to give good results on a new and unknown set of data, in other words, it cannot generalise anymore.

The networks with linear transfer functions were generating negative values for output. Furthermore they were highly underestimating the output. Even after logarithmic scaling of the data set a multivariate linear regression cannot be applied.

4.5.6 NN run on a limited range of output values

Although all samples contain important information on the distribution of macrobenthos, a neural network run is applied on a limited range of data. Table 4.3 presents the histogram data on the scaled total biomass. Of all samples, almost 23% have a value of zero, i.e. not one benthic species was present in the sample. Over half of the samples (54%) have a scaled value between 0 and 0.1. These are non-scaled total biomass between 0 and 5.74 g AFDW/m^2.

Table 4.3. Histogram data on total biomass

Bin	*Frequency*
0	309
0.1	735
0.2	132
0.3	74
0.4	44
0.5	26
0.6	12
0.7	9
0.8	3
0.9	2
1	1
More	2

A neural network run is applied on all total biomass larger than 1 g/m^2. This equals scaled values larger than 0.019. Of a total of 1349 samples, 569 are left. An exponential transfer function is used in a network with 8 hidden neurons

This neural network performs reasonably well. The linear regression through the scatter plot of targets versus NN outputs is described by y=0.81x, with an R^2 correlation of 0.50. Correlation between target and NN output is 0.78, RMS error is 0.0727. Results are shown in Figure 4.28.

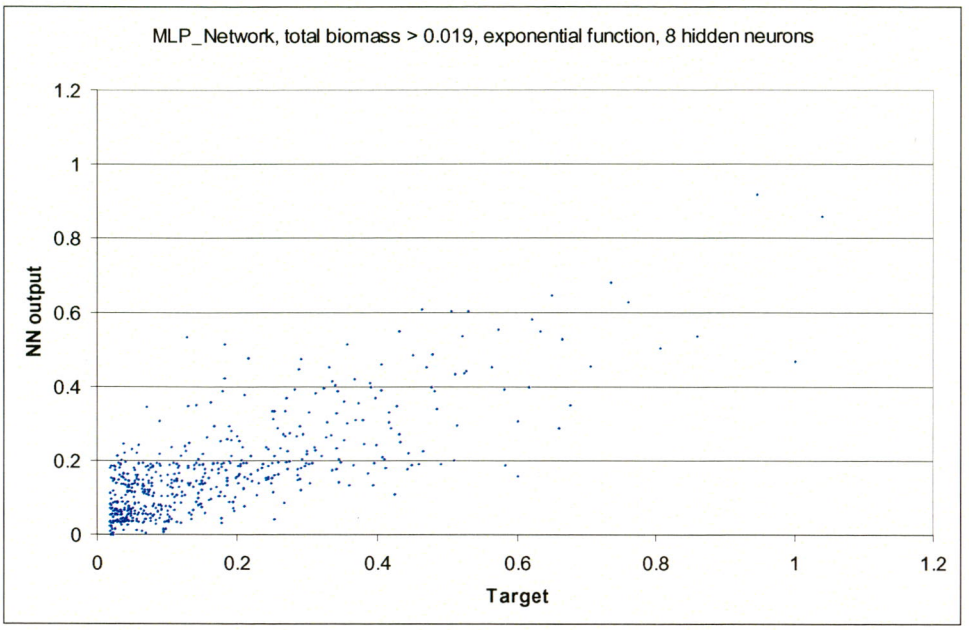


Figure 4.28. NN results for total biomass > 0.019 (>1 g/m2), exp. function, 8 hidden neurons.

4.5.7 Training and testing on all data

So far, all available data have been used for training the neural network. To test the general applicability of the neural network, part of the set can be used for training and the remaining part for testing. First, the data set is randomly mixed to exclude any trends in sample sequence. Then, two thirds of the samples are used for training (900 samples) and the remaining samples (449) are used for testing.

A neural network with an exponential transfer function and eight hidden neurons is used. Results are shown in Figure 4.29 and 4.30.

	Training	testing
RMS error	0.0755	0.1085
corr. coeff.	0.83	0.65

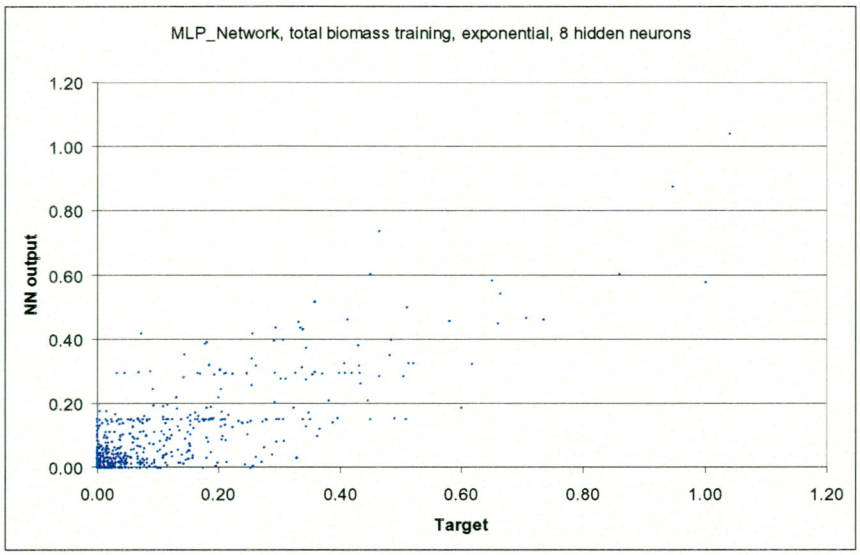


Figure 4.29. NN results for total biomass, 900 training samples.

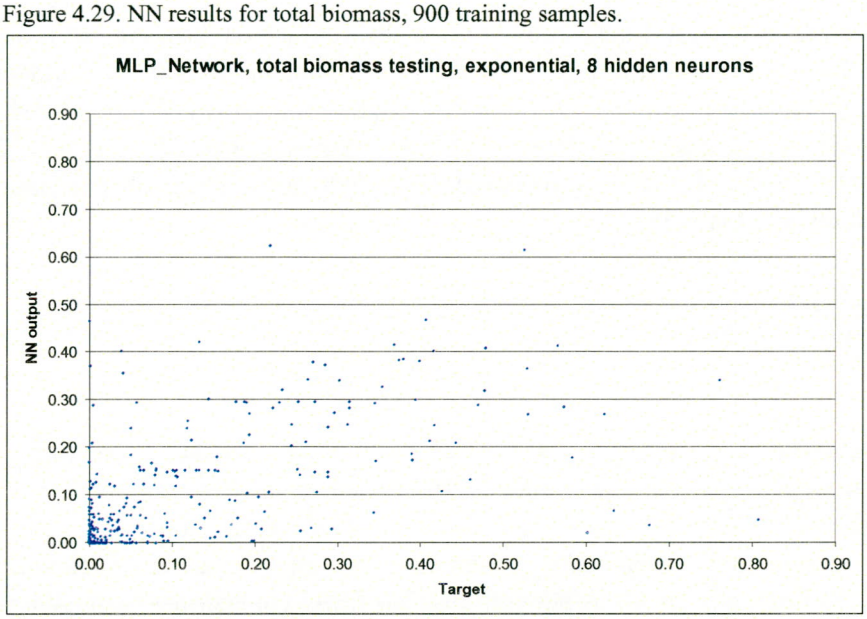


Figure 4.30. NN results for total biomass, 449 testing samples.

4.5.8 Limiting the number of input neurons

In the previous paragraphs, neural networks were applied on all available input data, i.e. four input neurons were used for abiotic local parameters and two input neurons were used for a representation of sample month. When this neural network is applied in a model that predicts benthic biomass distributed over the Westerschelde, one needs input data for all six input neurons. It could be worthwhile to limit the number of input neurons, in order to find a good prediction with less input variables.

One option is to omit the sample month as input variable. The neural network is than trained with the four local abiotic parameters as input. Seasonal effects however can disrupt the outcome of the network. In case the results are *not* a lot worse compared to the network which has the sample month included, this will indicate that the sample month is not a very important parameter.

Another option is to omit one of more of the local abiotic parameters. This can be considered when one or more of the parameters are dependent of each other. Based on the results of the PCA, depth, ebb-velocity and median grainsize are dependent parameters, but their correlation is not very large. Leaving one of these parameters out of the data set, will probably significantly worsen the outcome of the neural network.

NN runs on local abiotic parameters only

Sample month excluded

A neural network run is carried out on the local abiotic parameters only, so the two input neurons for sample month are disregarded. The neural network architecture has four input neurons, eight hidden neurons and one output neuron for all data values on total biomass. An exponential transfer function is used, and a distinction is made between 900 samples for training and 449 for testing, after randomising the data set.

Results for this run are not a lot worse compared to the run which has sample month included. The RMS error for the training set is 0.0849 (compared to 0.0755), and the correlation coefficient between target and NN output for the training set is 0.77 (compared to 0.83). The results have the tendency to underestimate higher target values. The test set even has an improved RMS error of 0.0995 (compared to 0.1085) at a improved correlation coefficient of 0.71 (compared to 0.65).

	Training	testing
RMS error	0.0849	0.0995
corr. coeff.	0.83	0.71

The worsened fit between target values and NN output may not be explained by omitting of the sample month alone, but may also be explained by the neural network architecture. The new neural network architecture has four input neurons instead of six, combined with an equal number of hidden neurons (eight). This means that the number of connections within the neural network has decreased by 16. A network that has fewer connections is, in general, less able to find a good fit between input and output.

Selection of sample month September

A second neural network run is carried out with a data set that is limited to those samples taken in September of any year. This data set contains 703 samples. All samples were used for training in a network with an exponential transfer function, four input and eight hidden neurons. Selecting September as a specific month for training does not improve the network results.

	Training
RMS error	0.0741
corr. coeff.	0.82

NN run on limited number of abiotic parameters

Salinity excluded

Based on the results of the PCA, salinity is the parameter that is most independent of other parameters. It may be expected that leaving salinity out of the data set will lead to worse results.

A neural network is trained on total biomass values, with an architecture of five input neurons (salinity is excluded) and eight hidden neurons. The training set contains 900 samples, 449 are used for testing.

	Training	testing
RMS error	0.0815	0.122
corr. coeff.	0.79	0.57

The results for the training set are slightly worse, but reasonably comparable to the results of the training set which has salinity included. The real test for a general applicable neural network is in the test set and those results are a lot worse, see Figure 4.30. It can be concluded that omitting the salinity as abiotic parameter leads to a network that is not generally applicable in the Westerschelde.

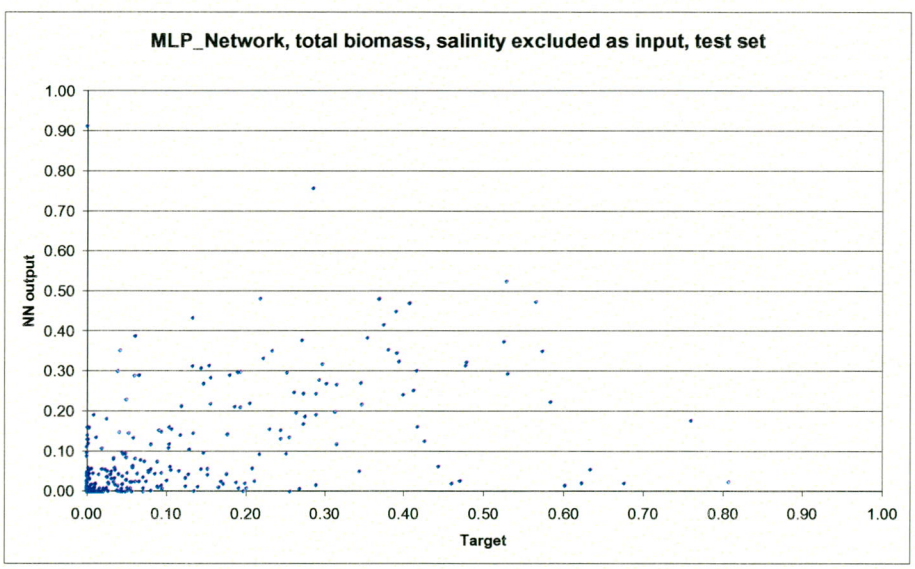


Figure 4.30. NN results for test set, total biomass, salinity excluded as input parameter

4.5.9 NN runs on ranges of input data

In the previous paragraphs, neural networks were trained on the complete range of input data that are available. It is possible that the selection of a range of input values for a certain parameter may improve network results. One of the parameters that can be considered is depth, because high biomass values only occur in shallow areas. A depth range is selected between -1.70 m and 1.00 m NAP (a negative value is above NAP). A neural network with six input neurons and eight hidden neurons is trained on 400 randomly selected training samples and 268 testing samples.

	Training	testing
RMS error	0.0800	0.146
corr. coeff.	0.86	0.58

The results for the test set are not very good. The neural network is not able to predict correct biomass values in this depth range.

4.5.10 Parameter ranges

Introduction

The neural networks applied so far were not able to find a very good fit between target values and NN output. A reason for this may be that at any combination of abiotic parameters, very high biomass as well as very low biomass can occur, and anything in between. An investigating data analysis is performed on the total biomass values and the corresponding ranges of abiotic parameter values.

Method

The logarithmically scaled values for total biomass were divided into seven biomass classes, see Table 4.4.

Table 4.4. Biomass classes for total biomass

Class number	Class boundaries, Logarithmically scaled	Class boundaries, not scaled (g/m^2)
0	0.000 - 0.001	0 - 0.05
1	0.001 - 0.1	0.05 - 5.7
2	0.1 - 0.2	5.7 - 13.0
3	0.2 - 0.3	13.0 - 22.1
4	0.3 - 0.4	22.1 - 33.5
5	0.4 - 0.5	33.5 - 47.9
6	>= 0.5	>= 47.9

For each biomass class the corresponding minimum and maximum value for every abiotic parameter was determined. Results for this analysis are shown in Figures 4.31 to 4.34.

Results

Results show that low biomass (including absence) occur at a very wide range of values for each abiotic parameter. Going from low to higher biomass, the wideness of the ranges of parameter values decreases. This analysis can conclude for example that total biomass over 22 g/m^2 (classes 4, 5 and 6) can only occur in certain limited ranges for parameters: depth between -1.7 and 1.0 m, salinity between 6.3 and 30.2‰, ebb-flow between 0.18 and 0.58 m/s and median grainsize between 28.9 and 193 µm. However, the relationship between abiotic parameter values and total biomass does not hold the other way around. At any value in between the above mentioned ranges one can find biomass between 0.0 and 200 g/m^2.

Although the higher biomass are restricted to these ranges, the majority of the measurements show lower biomass. A neural network that is trained on these data is more likely to predict low than high biomass values.

A conclusion that can be made is that biomass over 5.7 g/m^2 cannot occur at depths greater than 5.4 m.

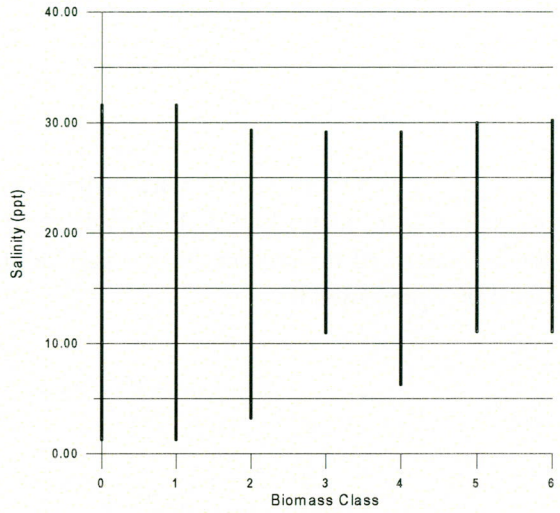


Figure 4.31. Min-Max range for salinity for seven biomass classes

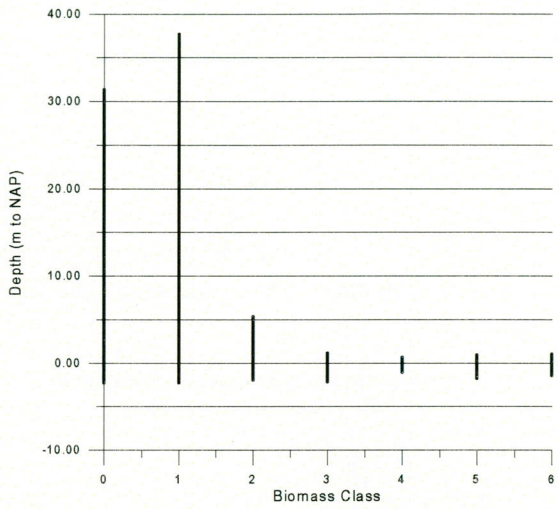


Figure 4.32. Min-Max range for depth for seven biomass classes

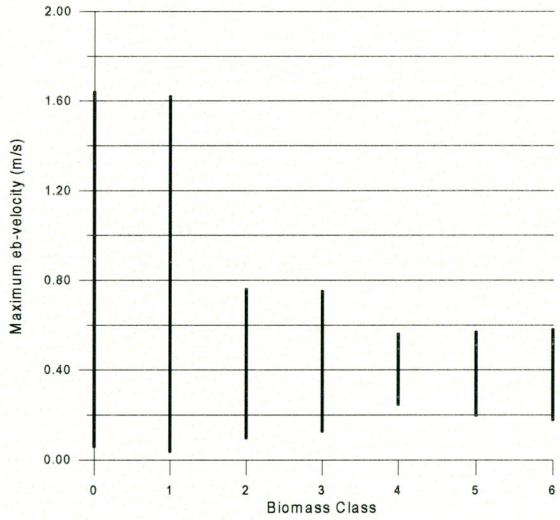


Figure 4.33. Min-Max range for maximum ebb-velocity for seven biomass classes

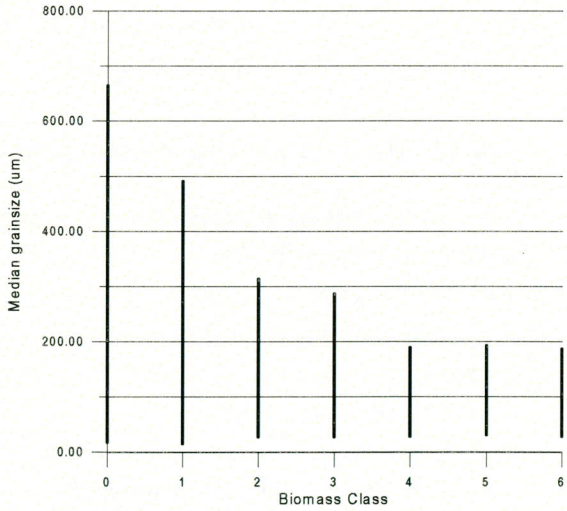


Figure 4.34. Min-Max range for median grainsize for seven biomass classes

Multivariate parameter ranges

A second analysis is carried out to establish parameter ranges for multiple parameters together. For each parameter the data range is selected that corresponds to the logarithmically scaled total biomass values greater than 0.3 (or 22 g/m^2):

- depth between -1.7 and 1.0 m
- salinity between 6.3 and 30.2‰
- ebb-flow between 0.18 and 0.58 m/s
- median grainsize between 28.9 and 193 µm.

Of a total of 1350 samples, 393 samples are selected. High biomass values are restricted to these ranges, but within these ranges low biomass can also occur. Figure 4.35 presents the histogram for all seven biomass classes and the frequency of occurrence of biomass values within each class. The upper boundary for each class is presented here.

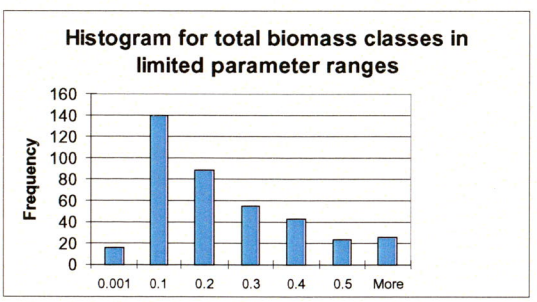


Figure 4.35. Histogram for total biomass classes corresponding to selected ranges for abiotic parameters.

It can be concluded that within the multiple parameter ranges that are suitable for high biomass, there are more occurrences of low biomass than of high biomass.

4.5.11 Results for the Westerschelde

The neural network for total biomass was applied on Westerschelde data. The input data that were used were available on a spatial grid that covered the complete Westerschelde in cells with a surface area of 60m^2. Input data were:

- Depth: bathymetry of 1992 derived from WAQUA-model.
- Salinity: linearly interpolated, yearly averaged, salinity range from 33.0 at Vlissingen to 10.0 at Prosperpolder.
- Current velocity: maximum current velocity over a tidal period derived from WAQUA-model, applied for 1992 bathymetry.
- Median grainsize: Interpolated coverage of median grainsize samples (McLaren data).

For each grid cell (in total 58.611 cells) the neural network was applied on the input values and the logarithmically scaled total biomass value was computed. The results are presented in Figure 4.36.

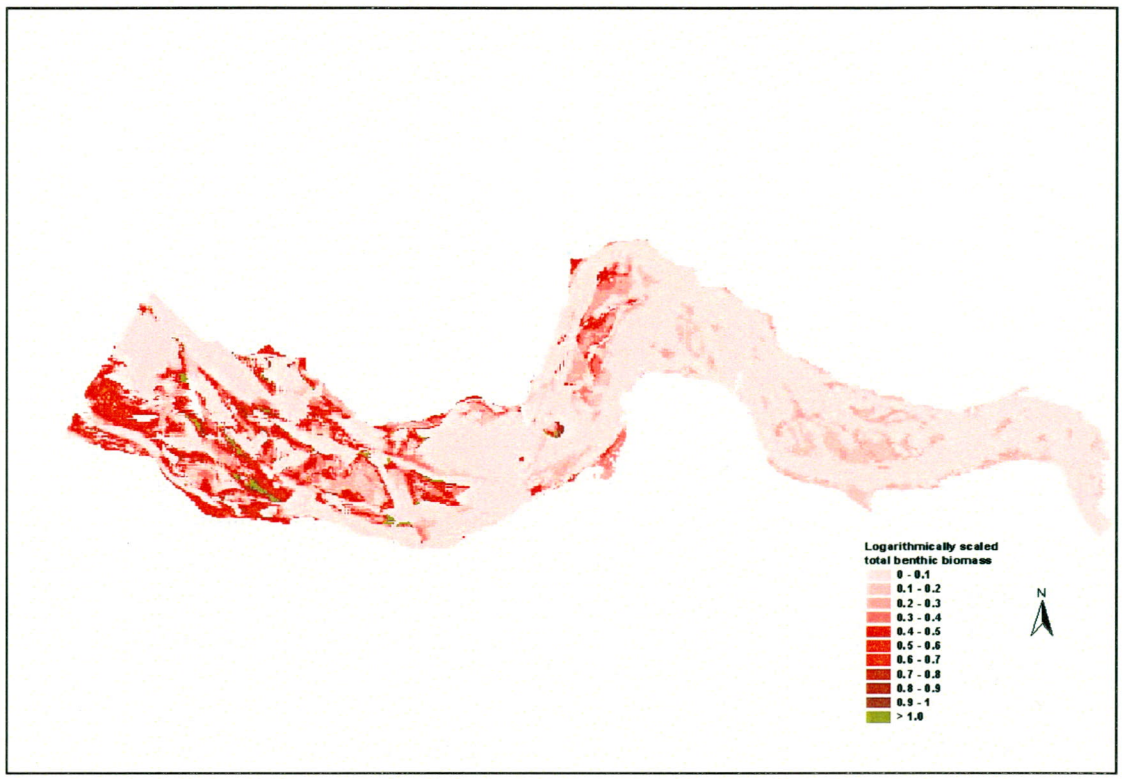


Figure 4.36 Predicted total biomass (logarithmically scaled) for the Westerschelde.

These results are compared to the measurements for total biomass as presented in Figure 4.37.

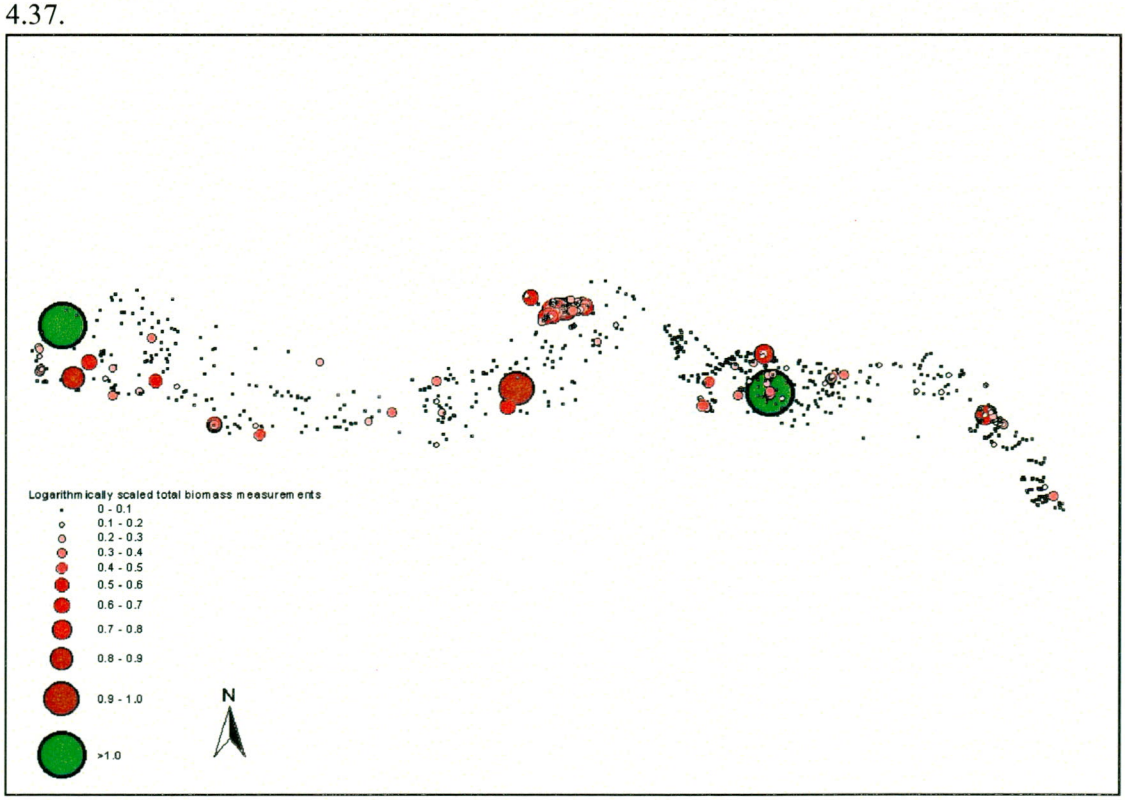


Figure 4.37. Measured total biomass (logarithmically scaled) for the Westerschelde.

The neural network results very clearly show the gradient in total biomass values as a result of the gradient in salinity over the Westerschelde, but this is not in full compliance with the