Identification of seafloor provinces - specific applications at the deep-sea Håkon Mosby Mud Volcano and the North Sea

Dissertation
zur Erlangung des
Doktorgrades der Naturwissenschaften
im Fachbereich Geowissenschaften
der Universität Bremen

vorgelegt von
Kerstin Jerosch

Bremen
2006

Content

List of figures

List of tables

Abbreviations

A	Area
AOM	Anaerobic Oxidation of Methane
AUV	Autonomous Underwater Vehicle
BGH	Biogeochemical Habitat
BMBF	Bundesministerium für Bildung und Forschung
CART	Classification and Refression Trees
DFG	Deutsche Forschungsgesellschaft
EEZ	Exclusive Economic Zone
eHMMV	entire area of the Håkon Mosby Mud Volcano
EUNIS	European Nature Information System
FFH	Floral and Faunal Habitat
GIS	Geographical Information System
GVM	Georeferenced Video Mosaic
HCR	High CH_4 bearing Region
HI	Habitat Indicator
HMMV	Håkon Mosby Mud Volcano
IBU	Inhaltsbasierte Bildanalyse von Unterwasseraufnahmen
IDW	Inversed Distance Weight
IK	Indicator Kriging
IMS	Internet Map Server
MA	Mean Area
MAE	Median Absolute Error
MarGIS	Marine Geo-Information System for Visualisation and Typology of Marine Geodata
MATISSE	Mosaicking Advanced Technologies Included in a Single Software Environment
MDBS	Marine DataBase Systems
MDM	Marine Data Model
MePE	Mean Percentage Error
MPE	Median Percentage Error
MPM	Mono-Parametric Map
MSDR	Mean Squared Deviation Ratio
MV	Mud Volcano
rDBMS	relational DataBase Management System
RMSE	Root Mean Square Error
RMSSE	Root Mean Square Standardised Error
ROV	Remotely Operated Vehicle
SDE	Spatial Data Engine
SSL	Suggested Source Location
TA	Total Area
USBL	Ultra Short Base Line

Summary

Recently, geological, geochemical, and biological data collection increased considerably in the marine environment together with ecological, economical, and scientific interests in marine coastal environments (e.g. North Sea) and ocean margins (e.g. Håkon Mosby Mud Volcano). The increasing amount of geodata results from new sampling devices, as *in situ* sensors, and mobile underwater platforms as ROVs (Remotely Operated Vehicles) and AUVs (Autonomous Underwater Vehicles), or from satellite-supported data transfer from moorings. Compared to the multitude of measured parameters and the quantity of information compiled during multidisciplinary research cruises, only few concepts and methods were developed for visualisation, distribution of data and thematic maps, efficient integration of the inhomogeneous data into existing database structures, management and spatial analysis of geodata.

The identification of distinct provinces is currently an emphasis of marine research geosciences. A typological approach combining geological, biological and chemical properties is accomplished by geostatistical, multivariate statistical, and GIS techniques (Geographical Information System). Besides scientific needs as surface-related balances of geological and geochemical cycles, seafloor provinces support management decisions related to upcoming economic use of the seafloor (e.g. such as installation of off-shore wind parks or the declaration of protection zones) and bear up to model spatio-temporal connections and changes of coastal regions.

Submarine mud volcanoes are considered as significant source locations for methane indicated by unique chemoautotrophic communities as *Beggiatoa* mats and pogonophoran tube worms. The Håkon Mosby Mud Volcano (HMMV) is located at the continental slope of the Barents Sea in a water depth of 1260 m. A large amount of georeferenced video mosaics and microbathymetric data, derived from a camera system and a multibeam echo sounding system mounted onto the ROV *Victor6000* (Ifremer), are basis for a morphological as well as biogeochemical regionalisation of the HMMV. This regionalisation is accomplished due to visual inspection of video mosaics, a defined classification scheme, concerning the distribution pattern of the two occurring chemoautotrophic communities and uncovered mud areas, and due to a GIS-supported overlay of 13 resulting surface maps which were calculated geostatistically using indicator kriging. Furthermore, microbathymetry and slope inclination were included, defining and calculating areas of five biogeochemical habitats. These habitats indicate graduated methane consumption of microbial consortia, consisting of sulphate-reducing bacteria and anaerobic methane-oxidising archaea. These consortia represent an efficient methane biofilter, e.g. at gas hydrate bearing continental margins, and are thus an important sink for the global methane cycle.

Approximately 16% of the flat centre of the HMMV is nearly void of any benthic communities. Therefore, this area is considered as a region of high methane discharge into bottom water.

Source location and drainage direction for current mud flows were identified by trend surface computation of the comparably flat crater area (1.58° average slope angle), and consideration of temperature data, as well as the distribution pattern of the chemoautotrophic communities. This suggests that a present mud flow ascends close to the northern edge of the flat unit of HMMV, and that the drainage pattern of mud flows shifted from a westward to a south-south-eastern direction.

The quality assessment of the surface maps is conducted by cross-validation evaluating the fit of the indicator kriging variograms by using statistical mean values of the deviations between estimated and measured values. Furthermore, the estimate was evaluated by a validation dataset of visual inspected analysis of video mosaics not included in the interpolation process, proving the interpolated surfaces independently.

The large amount of video mosaics requires the development of an image analysis technique for the automated detection and quantification of the spatial distribution of *Beggiatoa* mats. In a first step it is differentiated between *data, non-data* and *redundant* (overlapping) areas on the mosaics. In the second step the data areas are pre-segmented into disjunctive homogeneous regions by a watershed transformation. A probabilistic approach, relaxation labelling, then realises the assignment of these regions as *bacterial* or *non-bacterial* on the basis of spatial correlation and defined contrast thresholds (comparing the grey values of neighbour regions). Comparison of the data derived by visual inspection with the automated image analysis revealed similarities better than 90%.

Kriging methods were also applied and evaluated for selected parameters for the North Sea (bottom water measurements on salinity, temperature, silicate, ammonium, nitrate nitrite, phosphate as well as from punctual data on grain size ranges (0-20 μ, 20-63 μ, 63-2000 μ) creating surface maps from measured data as an assumption for multivariate statistics like Classification and Regression Trees (CART). The evidence of spatial autocorrelation by variogram analysis allowed calculation of raster maps by applying ordinary kriging.

After intersecting, these raster maps with punctual data on benthic epifaunal communities a classification system is derived to predict the occurrence of these communities within the study area. The classification system is calculated from the intersected data by producing decision trees using CART. Since these decision trees correspond to hierarchically ordered sets of decision rules, they are applied on the geostatistically estimated raster data to predict benthic habitats.

Zusammenfassung

Gegenwärtig ist eine drastische Zunahme der Informationsvielfalt an meeresgeologischen, geochemischen und biologischen Daten zu verzeichnen. Dies gilt auch für die Zahl der Projekte, die eine wirtschaftliche Nutzung oder eine wissenschaftliche bzw. ökologische Fragegestellung von Küstengewässern (z.B. Nordsee) und Bereichen des Kontinentalrands (z.B. Håkon Mosby Schlamm Vulkans) zum Ziel haben. Die Zunahme an Geodaten resultiert u.a. aus dem vermehrten Einsatz neuer Technologien, die eine kontinuierliche Datenerfassung ermöglichen. Die rasche Entwicklung auf diesem Sektor, wie z.B. *in situ* Sensoren, AUVs (Autonomous Underwater Vehicles), ROVs (Remotely Operating Vehicles) oder verankerten Messgeräte mit satelliten-gestützter Datenübermittlung werden die Verfügbarkeit meereswissenschaftlicher Daten weiter beträchtlich erhöhen. Der wachsenden Informationsvielfalt und Verfügbarkeit an Geodaten, die auf jeder meereswissenschaftlichen Expedition erhoben werden, stehen bislang vergleichsweise wenige Konzepte und Verfahren zur Visualisierung, Verbreitung der Daten und thematischen Karten, Integration der heterogenen Datensätze in bestehende Datenbankstrukturen, und damit der optimalen Weiternutzung erhobener sowie neu zu generierender meereswissenschaftlicher Datensätze gegenüber.

Die Identifikation von Provinzen ist seit einiger Zeit ein aktueller Forschungsschwerpunkt mariner Geowissenschaften. Der Anspruch der Typlogisierung des Meeresbodens, geologische, biologische und chemische Eigenschaften miteinander zu kombinieren, wird mit Hilfe von geostatistischer und multivariat statistischer Verfahren sowie GIS-Techniken (Geographisches Informationssystem) umgesetzt.

Die daraus resultierenden Raumgliederungen unterstützen neben wissenschaftlichen Aspekten, wie flächenbezogene Bilanzierungen geologischer und geochemischer Stoffkreisläufe, die durch intensive Nutzung erforderliche marine Raumplanung (z.B. bei Planung von Offshore Windparks und Ausgleichsflächen) und helfen, raum-zeitliche Zusammenhänge und Veränderungen, denen Küstenregionen unterliegen, zu modellieren.

Marine Schlammvulkane sind bekannt als bedeutende Methan-Quellen, dessen Vorkommen dort von einzigartigen chemoautotrophen Gemeinschaften (Bakterien, Genus *Beggiatoa,* und einer Gruppe von endosymbiontischen Röhrenwürmern, Pogonophora) angezeigt wird. Der Håkon Mosby Schlamm Vulkan (HMMV) befindet sich am Kontinentalhang der Barentssee in einer Wassertiefe von 1260 m. Eine große Anzahl von georeferenzierte Videomosaiken und hoch auflösende bathymetrische Daten, aufgenommem mit einem am ROV *Victor6000* (Ifremer) angebrachten Kamerasystem und einem Flachwasser-Fächerecholot, bilden die Datengrundlage für eine sowohl morphologische als auch biogeochemische Gliederung des HMMV.

Die Regionalisierung beruht auf einer visuellen Videomosaikauswertung, der Entwicklung eines Klassifikationsschemas, bezüglich der vorkommenden chemoautotrophen Organismen

und der unbesiedelten freien Schlickflächen, sowie einer GIS-basierten Verschneidung von insgesamt 13 Flächenkarten, die mit Hilfe von Indikator Kriging generiert wurden. Weiterhin wurden Mikrobathymetrie und Hangneigung bei der Definition und Berechnung von fünf biogeochemischen Habitate berücksichtigt. Diese Habitate indizieren graduell das Vorkommen mikrobieller Konsortien, die aus sulfat-reduzierenden Bakterien und anaerob methan-oxidierenden Archaea bestehen. Diese Konsortien bilden einen effektiven Methan-Filtermechanismus, z.B. an gashydratreichen Kontinentalhängen, und tragen so wesentlich zum globalen Methankreislauf bei.

Etwa 16% der von Methan beeinflussten Fläche des HMMV ist nicht von chemoautotrophen Organismen besiedelt und wird aufgrund dessen als Fläche mit dem höchsten zu erwartenden Eintrag von Methan in die Wassersäule angesehen. Basierend auf der Berechnung einer Trendfläche des vergleichsweise ebenen zentralen Kraters (1.58° mittlere Hangneigung), unter Berücksichtigung von hohen Sedimenttemperaturen und der Besiedlungsstruktur der chemoautotrophen Vergesellschaftungen wurden eine potentielle Austrittsfläche von frischen Schlick und dessen Abflussrichtung definiert. Demnach steigt der Schlamm am Nordrandes des flachen Kraters auf und fließt, anders als ein alter nach Westen gerichteter Ausstrom, jetzt in eine süd-süd-östliche Richtung.

Die qualitative Bewertung der Flächenkarten wurde mittels Kreuz-Validierungen durchgeführt, die mit Hilfe von statistischen Mittelwerten der Abweichungen zwischen geschätzten und gemessenen Werten die Anpassung des Semi-Variograms an die Verteilung der Daten untersucht. Unabhängig vom Interpolationsmodell wurden weiterhin die berechneten Flächen anhand eines Validierungsdatensatzes überprüft, bestehend aus weiteren visuell analysierten Videomosaiken, die nicht im Interpolationsverfahren integriert waren.

Die große Anzahl der Videomosaike erforderte besonders für weitere Untersuchungen die Entwicklung eines automatischen Auswerteverfahrens für die Identifikation und Quantifizierung von Bakterienmatten. Hierfür wird in einem ersten Schritt zwischen *Daten*-, *Nichtdaten*- und *redundanten* (Überlappungs-) *Bereichen* differenziert. Im zweiten Schritt werden die Datenbereiche mit Hilfe einer Wasserscheiden-Transformation in disjunkte, homogene Regionen vorsegmentiert. Ein Relaxationsverfahren attribuiert diese Regionen als *Bakterien* oder *Keine Bakterien* aufgrund von räumlicher Korrelation und definierten Kontrast-Schwellwerte (Vergleich der Grauwerte von Nachbarregionen). Die erzielten Ergebnisse hielten einem Vergleich mit visuell analysierten Ergebnissen mit einer Übereinstimmung von 90% stand.

Kriging Methoden wurden auch im Bereich der Nordsee auf verschiedene Bodenwasser-parameter angewendet und evaluiert (Messdaten zu Salinität, Temperatur, Silikat, Ammonium, Nitrat, Nitrit, Phosphat, sowie von Korngrößenverteilung des Sediments in 0-20 μ, 20-63 μ, 63-2000 μ). Die hieraus resultierenden Flächenkarten sind Voraussetzung für die Durchführung von multivariat statistische Methoden wie CART (Classification and Regression Trees).

Nach der Untersuchung der räumlichen Autokorrelation der Daten (Variogramm-Analyse), wurden mit Hilfe von Ordinary Kriging Flächenkarten berechnet. Nach der Verschneidung dieser Flächenkarten mit punktuell verteilten Epibenthos-Daten wurde ein Klassifizierungssystem abgeleitet, um das Auftreten dieser Gemeinschaften innerhalb des gesamten Untersuchungsgebietes vorauszusagen. Das Klassifizierungssystem wurde von den kombinierten Daten errechnet, indem Entscheidungsbäume (Dendrogramme) mit CART produziert werden, die sich anhand von hierarchischen Richtlinien aufgliedern.

Section A:
Concepts, regional settings, and methods
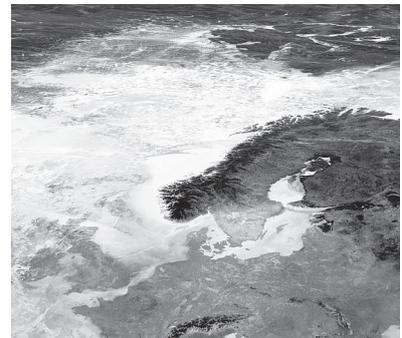
1 Introduction

The fast development of innovative devices changes the data acquisition, the data types, and the results in marine sciences considerably. High technology instruments produce data of different types, which requires either innovative evaluation procedures - like an automatic image analysis for georeferenced video mosaics, or the application of well-known terrestrial procedures, now transferable to the marine environment due to a significant increase of the data availability - applying GIS (Geographical Information Systems), geostatistical or multivariate statistical methods such as kriging and CART.

In marine research, sampling and data acquisition is conducted during cruises by research vessels and, to a lesser extend, by satellite or airborne remote sensing. During cruises, a variety of devices are applied for sediment cores obtained concerning geochemical analysis, *in situ* measurements and water samples collected for chemical analysis of nutrients or pollutants, plankton samples, or water currents. More technical equipment is used to catch fish, plankton, or benthic biota. Acoustic techniques as multi-beam, side scan sonar, or shallow seismic are used for bathymetric surveys, and investigation of sub-seafloor geology (Wright and Bartlett, 2000). High resolution still photographs and videos are recorded by towed devices and during dives by ROV (e.g. *Victor6000*, Ifremer), submersibles, or Crawlers like the MOVE-System (MARUM, Univ. Bremen) for visual identification of geological or biological features as well as for inspection of infrastructure as pipelines. From a geoinformatical perspective these data are points, multipoints, lines, areas, and time series reflecting e.g. sediment cores, water sampling at a site in different water depths, fishery tracks, seafloor images, and employment of moorings (Fig. 1-1).

An integrated approach, supporting spatial analysis of marine geodata, is required which allows compilation of all different datasets and data types within one database management system (DBMS). Basically, this encompass that - in best case - one query is able to retrieve a complex georeferenced data set which can mapped and analysed within a GIS.

This thesis *Identification of seafloor provinces - specific applications at the deep-sea Håkon Mosby Mud Volcano and the North Sea* deals with two marine regions; they are both particularly well-investigated; for both large amounts of data and a comprehensive data variety are available; and both are the focus of GIS-based identification of seafloor provinces, however, with respect to different scientific questions. In this study, regionalisation is useful to detect both provinces and predictive provinces of biogeochemical aspects linked to the methane cycle at the Håkon Mosby Mud Volcano and of benthic macrofaunal species, probably endangered due to anthropogenic use of the North Sea. The identification of predictive provinces is defined here as the development of a numerical or statistical model about the relationship among environmental variables: bottom water measurements on salinity, temperature, dissolved

Time Series Data	Point / Multipoint	Line, Polygon, Grid	All data types
$(x_{1..n}, y_{1..n}, z_{1..n}, t_{1..n})$	(x_1, y_1, z_1, t_1) / $(x_1, y_1, z_{1..n}, t_1)$	$(x_{1..n}, y_{1..n}, z_{1..n})$	(X_n, Y_n, Z_n) …
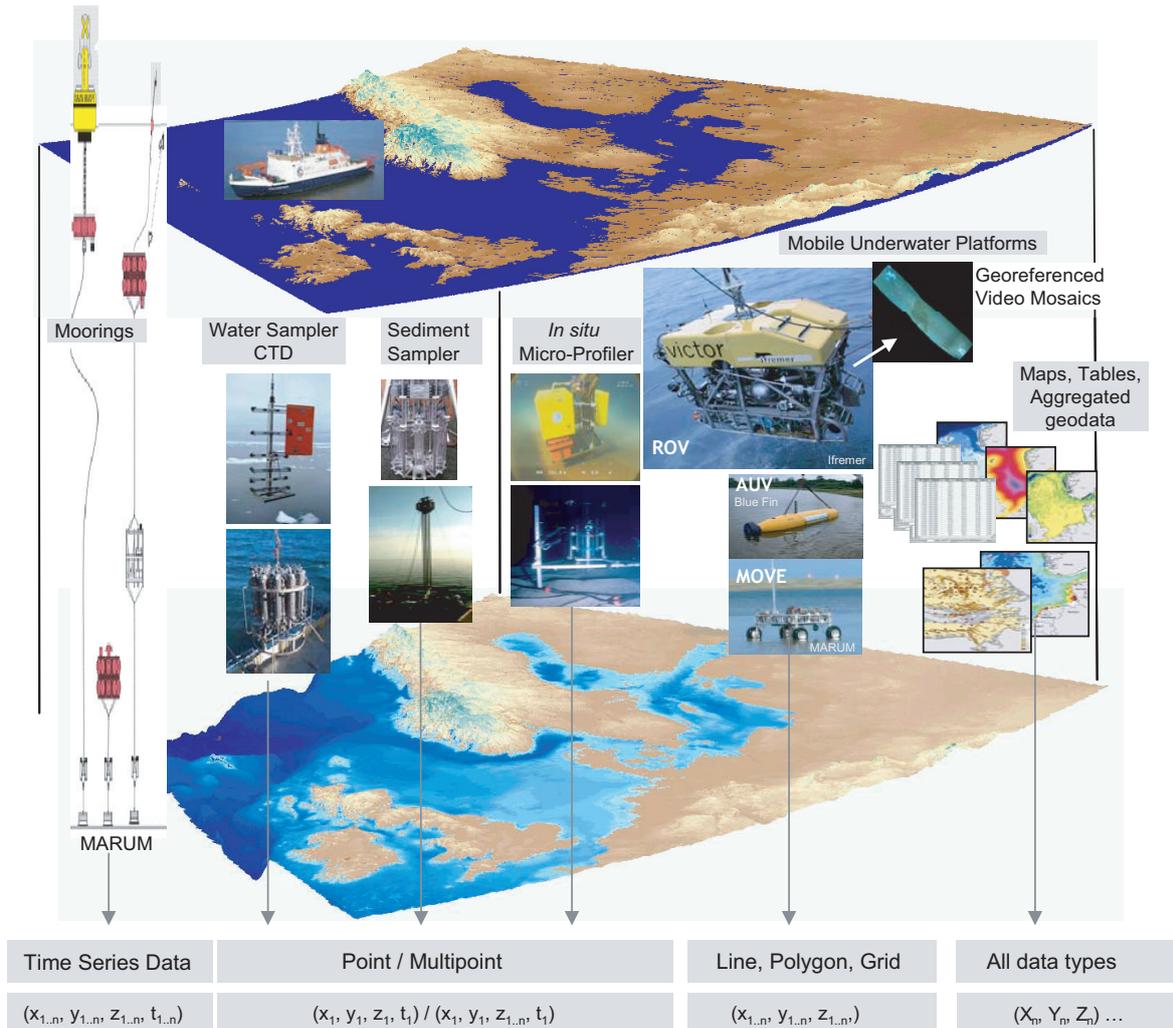
Figure 1-1: Data types of marine geodata. New underwater techniques as ROV, AUV, Crawlers as the MOVE system, and moorings (the two latter from MARUM, University of Bremen) build platforms for in situ analysers and acoustic sensor mapping. Video mosaics and microbathymetry data analysed in this thesis (chapter 4-9) are recorded by means of the ROV *Victor6000* (Ifremer).

components as silicic acid, oxygen, phosphate and nitrate; sedimentological data, and benthic biological and microbial communities. The availability of several information layers (parameters) is thereby assumption for a successful execution of such an interdisciplinary approach.

1.1 Identification of seafloor provinces

Today high activity in the identification of seafloor provinces can be observed. Several scientific investigations are conducted in this regard with different focus, for instance, the detection of clay mineral provinces of the South Atlantic using the kaolinite/chlorite ratio distribution (reflecting oceanographic current systems, transport and sedimentation processes) to mirror the distribution of southern and northern component water masses (Petschik et al., 1996); hydrographical provinces choosing the density as identification parameter, as a function of

temperature and salinity (Hooker et al., 2000); benthic deep-sea provinces classified by Greene et al. (1999) which are hierarchically organised on salinity, bathymetry, seafloor morphology, and substratum (derived from remote sensing geophysical and geological techniques); or GIS-based approaches to select priority areas for conservation (Woodhouse et al., 2000), organic carbon provinces of seafloor sediments (Seiter et al., 2004), or geochemical provinces with respect to methane consumption of microbial communities at the Håkon Mosby Mud Volcano (chapter 4).

An accomplished regionalisation is further applied to economy questions concerning, e.g. in commercial fisheries, of offshore wind park operators or oil industry. In response, national and international nature protection agreements (e.g. Flora Fauna Habitats, Natura 2000) and definition of international zones of responsibilities (e.g. Bonn Agreement for Co-operation in Dealing with Pollution of the North Sea by Oil and Other Harmful Substances, 1983) are always based on any kind of provinces.

After Longhurst (1998) provinces are defined as the "geography of the sea", suitable for standardised describing the distribution of all marine organisms. A consistent regionalisation did not exist despite a history of oceanographic research starting with the Challenger Expedition (1872-1876) caused may be by the excessive preoccupation of various specialists with their favourite taxonomic groups. However, the real reason is probably that developing a truly synoptic vision of the ocean was impossible before the advent of satellite-based oceanography. Longhurst is among the first to have realised the fundamental differences between ocean provinces using remote sensing. The inclusion of the subsurface temperatures and pigments, such as chlorophyll, and their saisonal fluctuations (Longhurst, 1995) led to an estimate of global marine primary production, based on a stratification of the world´s oceans into 57 biogeochemical provinces (Longhurst et al., 1995).

Until recently, assessment of marine seafloor provinces defined by geology (along with depth, chemistry, sedimentology, and associated benthic communities has been mostly limited to intertidal and subtidal (i.e. 0-30 m water depth) regions of the continental shelf (Greene, et al., 1999). Extensive characterisation, mapping and classification schemes have been developed for European shallow coastal biotopes, primarily using Scuba, video surveys, acoustic imaging and geological sampling (Connor et al., 2004a-f).

New technology development and geophysical methodologies, such as side-scan-sonar, multi or single beam bathymetry and seismic reflection profiling are now being used to investigate benthic sea floor provinces in deep water. Especially within the last two decades developments of new sampling devices, *in situ* sensors, and mobile underwater platforms as ROVs, (Remotely Operated Vehicles), AUVs (Autonomous Underwater Vehicles) and Crawlers (mobile, wheel driven underwater vehicles) provide new capabilities for marine research (Fig. 1-1). Underwater video camera systems, remote sensing, and in situ analysers on or by means of ROVs, AUVs,

occupied submersibles, and benthic sleds have made fine-scale surveys in the deep-sea more commonplace (Allais et al., 2004; Greene, et al., 1993; Yoklavich, 1997).

1.2 Mud volcanoes - submarine methane sources

Mud volcanoes - which come in different sizes, from knee high to as big as a mountain - are often shaped like normal volcanoes, but instead of lava they expel a mixture of mud, rocks and gasses. They are built by a mixture of hot water and fine sediment (mud and clay) that pours gently from a vent in the ground like a fluid lava flow. They can be seen as open pressure valves in the earths crust and as a result they produce large amount of debris in the form of sediment (J. Woodside, pers. comm.). Especially marine mud volcanoes are therefore different to the general association of a "volcano".

Whereas terrestrial sources of climate change factors, including methane sources and sinks, are well characterised, most submarine sources remain poorly quantified. Deep-sea gas seeps such as submarine mud volcanoes and gas hydrate deposits are the potential sources to affect the atmospheric methane budget. Submarine mud volcanoes are distributed on earth more extensively than their sub-aerial analogues. This is to be expected considering that seas and oceans cover about two thirds of the earth's surface (Milkov, 2000). At the present time,
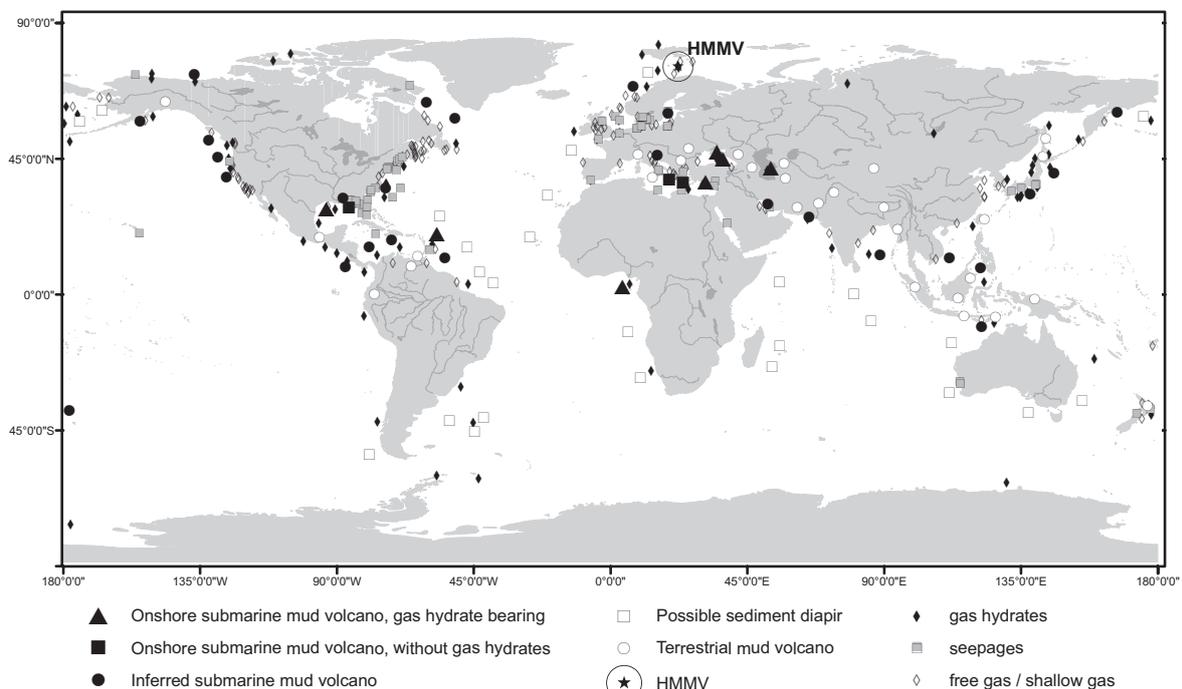
Figure 1-2: Worldwide distribution of onshore, offshore, and inferred submarine mud volcanoes including the distribution of terrestrial mud volcanoes (Milkov, 2000), gas hydrates (Kvenvolden and Lorenson, 2001), free gas occurrence (Fleischer et al., 2001), and seepages as submarine groundwater discharge (Taniguchi, et al., 2002) and offshore pockmarks (Hovland, 2002; Jørgensen, 1992).
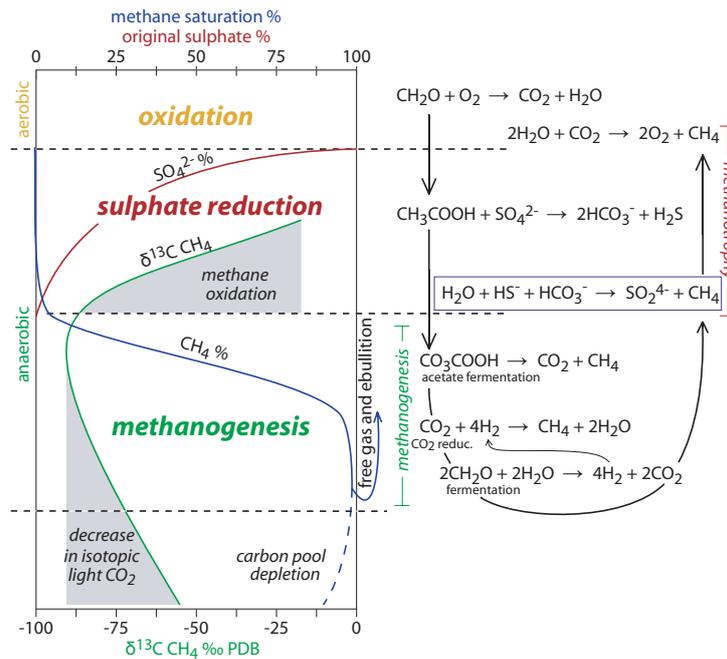
Figure 1-3: Scheme of methane forming and decomposing processes in the marine environment. Very important at cold seeps is the anaerobic oxidation of methane by microorganisms via sulphate reduction (modified after Greinert 1999, and Whiticar 1996).

evidence for submarine mud volcanoes has been found in all oceans (Fig. 1-2).

Marine mud volcanoes were observed worldwide (Milkov (2000); Kopf (2002)). They are often located at tectonically active areas of continental margins (Fig. 1-2). Upward flow of mud from deeper geological strata, enhanced temperatures, high inventories and fluxes of methane and other hydrocarbons, and occurrence of specific chemoautotrophic communities are characteristics for marine mud volcanoes (Hjelstuen et al., 1999; Kopf, 2002; Milkov, 2000). The methane may accumulate in the sediments and forms gas reservoirs such as gas hydrates (solid methane) (Bohrmann and Torres, in press; Egorov et al., 1999; Ginsburg et al., 1999). At mud volcanoes, pore water, gas, and mud is expelled from deep beneath the seafloor forming mounds and craters at the seafloor. Active mud volcanoes are therefore a seep for natural gas (methane) and are often densely populated by bacteria, tube worms, bivalves and other symbiotic organisms (Gebruk et al., 2003; Milkov et al., 2004; Pimenov et al., 2000; Sahling et al., 2005; Smirnov, 2000; Soltwedel et al., 2005).

The methane emanating from the seafloor is often very efficiently used by a symbiosis of archaeal and bacterial microorganisms, forming chemoautotrophic communities that are able to oxidise methane with sulphate which is abundant in seawater. The main microbial processes are anaerobic methane oxidation (AOM) (Boetius et al., 2000) coupled to sulphate reduction (Fig. 3-1) and aerobic or anaerobic oxidation of sulphide. Colonists are chemoautotrophic bacteria (dominant genus *Beggiatoa*) using the enzymatic oxidation of sulphide as a basic anabolic energy source or metazoan organisms, such as pogonophorans and bivalves, that

live in symbiosis with these bacteria (Sahling et al., 2003).

The Håkon Mosby Mud Volcano (HMMV) is an active mud, fluid, and methane-venting seep, located at 1250 - 1266 m water depth in the centre of the most recent Bear Island fan slide valley at 72°00.25'N 14°43.50'E. HMMV is considered as the only active mud volcano in an arctic area and since its discovery in the 1989-1990 SeaMARC II imagery (Crane et al., 1995) was studied by multidisciplinary research (Egorov et al., 1999; Ginsburg et al., 1999; Gebruk et al., 2003, Lein et al., 1999; Milkov et al., 1999, 2004; Smirnov, 2000; Soltwedel et al., 2005; Pimenov et al., 2000).

HMMV displays a caldera with a circular shape (~1.2 km in diameter) which can be divided into three concentrically arranged morphological units, recognisable on the microbathymetry map in figure 1-4: 1. The flat central zone (400-500 m diameter) consists of highly reduced, methane containing, non-stratified sediments (Lein et al., 1999). In the *thermal eye* of the
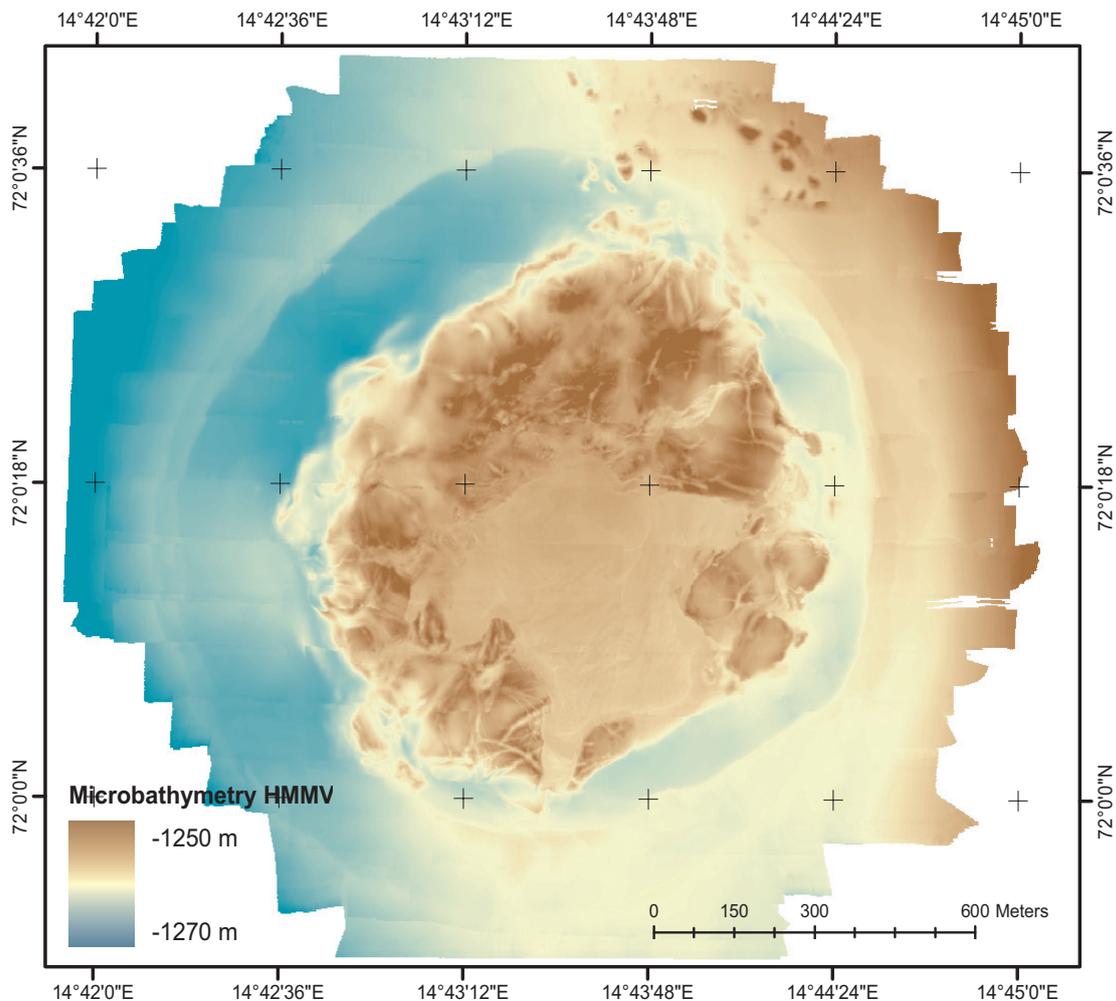
Figure 1-4: High resolution microbathymetry of HMMV recorded during two *Victor6000* (Ifremer) dives of 44 and 15 hours bottom time. Mainly by the Reson SeaBat 8125 high resolution data with a depth resolution of better than 0.1 m and a foot-print of 0.25 m² at the seafloor were derived.

volcano temperatures as high as 22 °C at 2 m depth and very steep gradients up to 3 °C/m were observed by Kaul and Heesemann (2004). This zone contains areas mostly non-colonised by epi/megafauna but in the southern and south western part also the regions with the highest density of bacteria coverage. 2. The hummocky periphery with slope angles greater than 20° and a relief of 8-10 m in height enclosing the centre. This second unit is minimum 10 to maximum 440 m wide and is mostly colonised densely by pogonophorans and sparsely by *Beggiatoa*. 3. A shallow depression surrounding the mud volcano, the moat area, builds the third unit, which is characterised by very small slope angles (mean value 1.58° and apparently no colonisation - neither by pogonophorans nor *Beggiatoa* (see chapter 4).

Major motile epi/megafauna components at HMMV are brittle stars, pantopods (*Pycnogonida*) and benthic fish (mainly *Zoarcidae*, *Liparidae*, and *Rajidae*; Milkov et al., 1999). The highest densities of fish were observed in the central part of the crater (Gebruk et al., 2003). Large molluscs, common representatives of many mud volcanoes and gas seeps, are completely missing at HMMV (Soltwedel et al., 2005).

At the HMMV, three main habitats have been described in previous studies (Milkov et al., 1999; Gebruk et al., 2003: 1. a central, barren area of sediment not colonised by sulphur-oxidising communities; 2. the zone of highest methane turnover which is indicated by the presence of white mats of giant sulphur-oxidising bacteria (*Beggiatoa*) on the seafloor (Fig. 1-2). These bacterial mats cover large areas around the centre of the HMMV; 3. pogonophoran tubeworm fields populating the hummocky part of the HMMV outside of the centre.

Microbially mediated anaerobic oxidation of methane is the major biological sink of methane in marine sediments. Hence, this process is crucial in maintaining a sensitive balance of our atmosphere's greenhouse gas content (Hinrichs and Boetius, 2002). Chemoautotrophic communities, i.e. organisms which are fuelled by the chemical energy of dissolved minerals, can indicate the presence of active gas seeps. Hence, at HMMV *Beggiatoa* and pogonophorans can be used as indicators for enhanced methane consumption. The investigation of their distribution and their density distinguishes between areas with high versus low methane discharge. This is important because there are still gaps in our understanding about the degree of methane oxidation in mud volcano sediments.

1.3 North Sea – an overview

The North Sea is a margin sea of the Atlantic Ocean, located between Norway and Denmark in the east, the United Kingdom in the west, and Germany, the Netherlands, Belgium and France in the south and is economically sub-divided into national zones e.g. the German Exclusive Economical Zone (EEZ). The North Sea is situated on the continental shelf of north-west Europe, opens into the Atlantic Ocean to the north and, via the Channel to the south-west, and into the Baltic Sea to the east. It has a surface area of about 750,000 km^2 and a volume of

Figure 1-5: Sediment map of the entire North Sea. For the aggregation of six analogue sediment maps different sediment classifications (e.g. Wentworth, 1922; Folk, 1954) had to be fitted within the MarGIS project.

German Bight (Figge, 1981)
Dogger Bank (BGS, 1977-1993)
Oyster Ground (BGS, 1977-1993)

Dutch Continental Shelf (RGD, 1986)
Denmark (GEUS, 1992)
International Quarternary Map of Europe (BA für Bodenforschung/UNESCO, 1970)
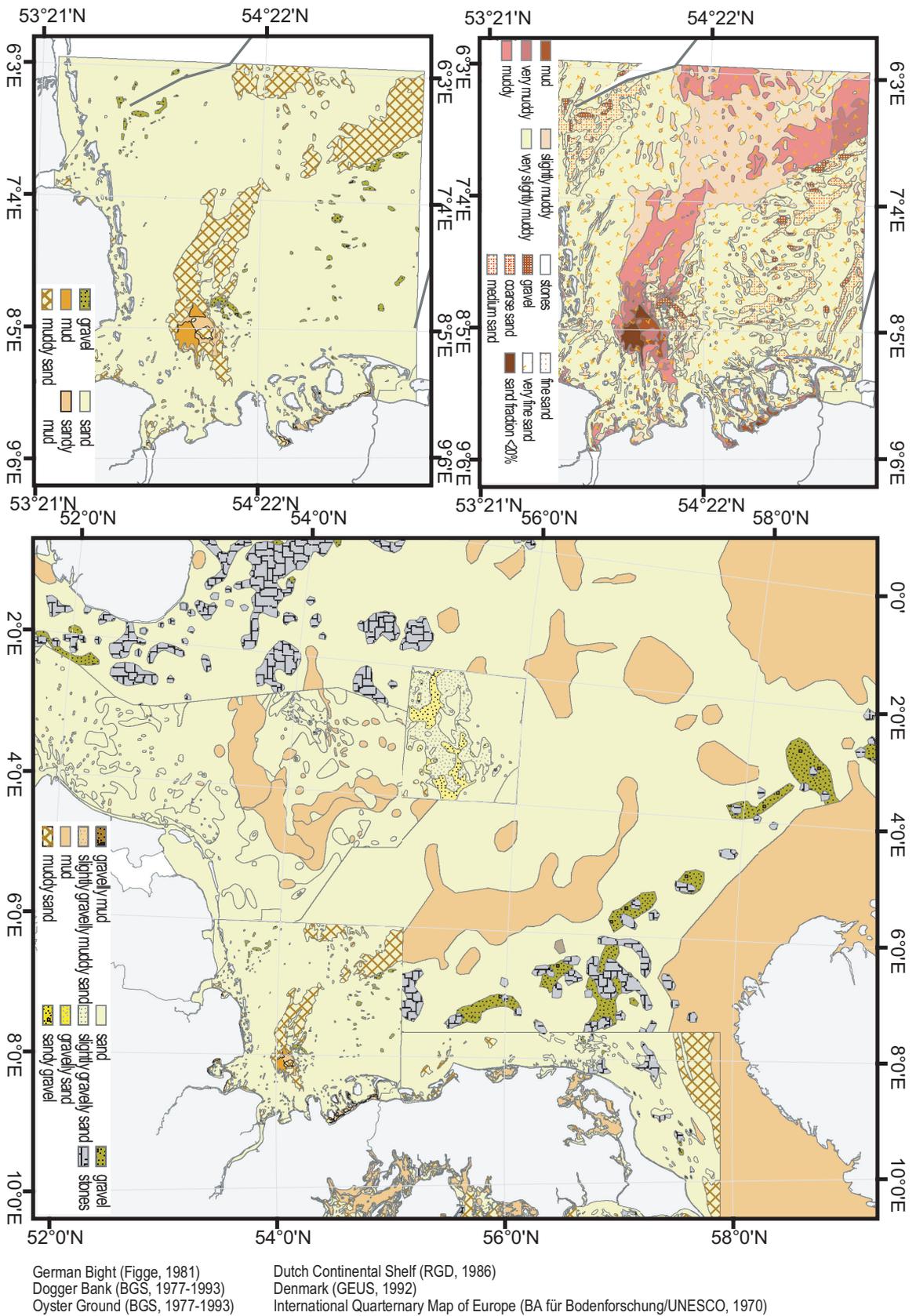
about 94,000 km^3 (OSPAR Commission, 2000).

The depth of the North Sea increases towards the Atlantic Ocean to about 200 m at the edge of the continental shelf. The Norwegian Trench, which has a sill depth of 270 m off the west coast of Norway and a maximum depth of 700 m in the Skagerrak, plays a major role in steering large inflows of Atlantic water into the North Sea (Saetre and Becker, 1990). The Channel is relatively shallow, and from a depth of about 30 m in the Strait of Dover deepens gradually to about 100 m in the west. Seabed topography shows evidence of river valley systems that were carved into the seabed during glacial periods when the sea level was lower (Streif, 2003). In the area between The Netherlands and Great Britain, extending northwards from the Channel to the Frisian Front (45 m), average depths are between 20 and 30 m. On the north-west side of the Dutch part of the continental shelf lies the shallow area of the Dogger Bank where depths can be less than 20 m.

The North Sea shelf area is an ancient continental drift depression with a general north-south axis. This depression is overlain by sedimentary deposits several kilometres thick originating from the surrounding land masses, and some of their strata contain large amounts of liquid and gaseous hydrocarbons, which are intensively exploited. During the glacial era, multiple invasions of Scandinavian and Scottish mountain glaciers spread over the North Sea causing large sea level changes and supplies of additional sediment into the North Sea basin. It also shaped the general style of the present underwater topography, for instance, elevations such as the Dogger-Bank and depressions like the Oyster Ground, the submerged part of the Elbe valley, and the Norwegian Trench (Pratje, 1951). Sea levels rose from the end of the last glaciations until about 6000 years ago, but since that time there have been only minor variations. The resulting hydrographical circulation, as well as the wave and tidal regime, created the sediment dynamics and the sediment distribution pattern seen today (Fig. 1-5). Mainly sand and gravel deposits occur in the shallower areas and fine-grained muddy sediments accumulate in many of the depressions (Streif, 2003). Tidal flats like the Wadden Sea receive their sediments directly or indirectly from rivers and from adjacent North Sea areas. The suspended particulate matter settles to form either sandy or muddy sediments according to its composition and the predominant local hydrodynamic conditions (Flemming and Bartholomä, 2003).

The water of the shallow North Sea consists of a varying mixture of North Atlantic water and freshwater run-off. The salinity and temperature characteristics of different areas are strongly influenced by heat exchange with the atmosphere and local freshwater supply. The deeper waters of the North Sea consist of relatively pure water of Atlantic origin, but they too are partly influenced by surface heat exchange (especially winter cooling) and, in certain areas, slightly modified through mixing with less saline surface water (Saetre and Becker, 1990). Several water mass classifications exist for the North Sea, based on temperature and salinity distributions or on residual current patterns or stratification. The circulation and distribution

Figure 1-6: Extensive anthropogenic use of the German Exclusive Economic Zone (modified after BSH, German Federal Maritime and Hydrographic Office, 2005).
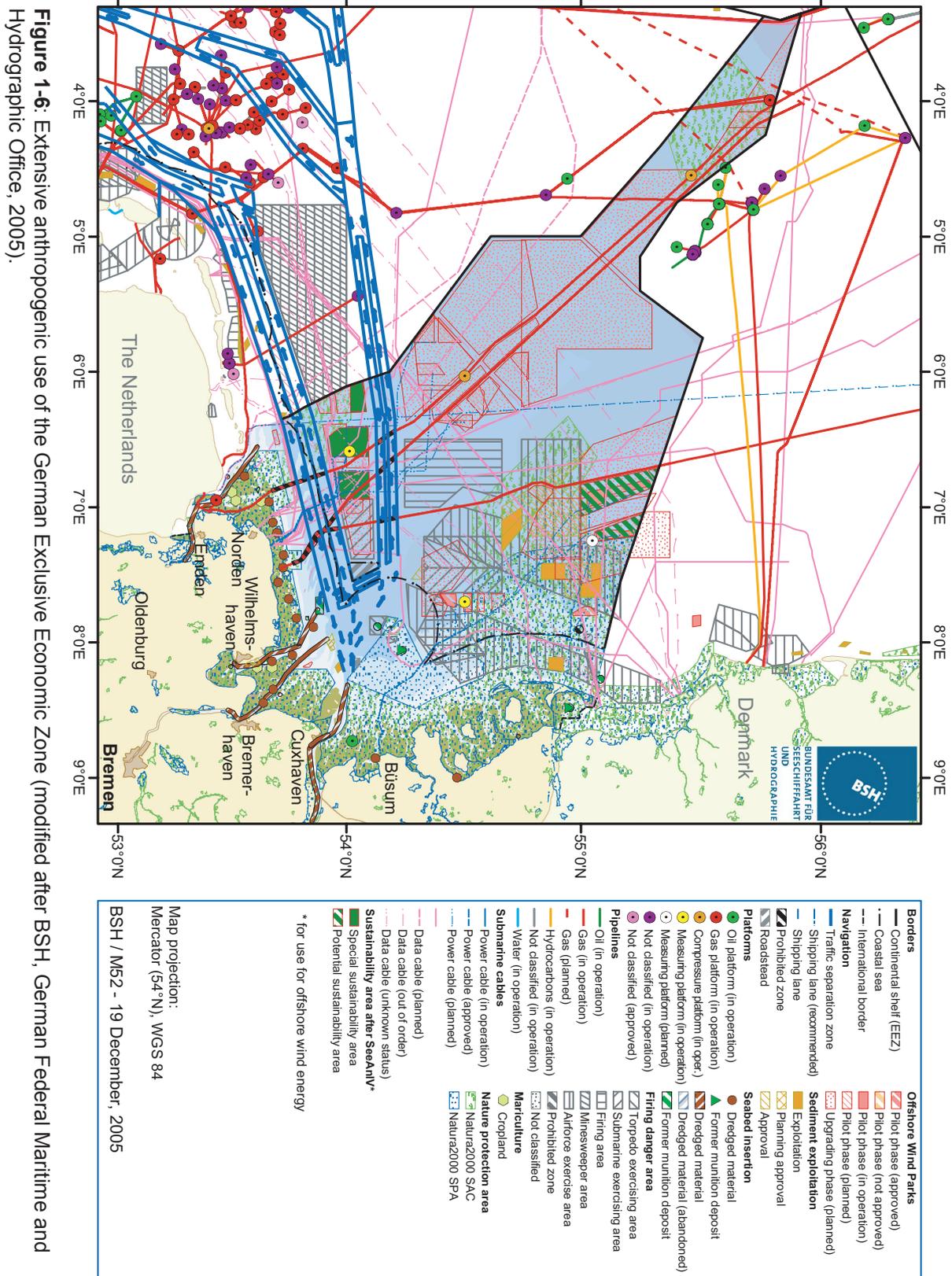
of these water masses are of the utmost importance in supporting biological productivity, transport and concentration of living (e.g. larvae) and non-living matter in the region (Lang, et al., 2003).

In coastal waters beyond estuaries and fjords, typical salinity ranges are 32 to 34.5 ‰, except in the Kattegat and parts of Skagerrak where the influence from the Baltic results in salinities in the ranges 10 – 25 ‰ and 25 – 34 ‰, respectively. In the open waters and especially in western parts of the North Sea, seasonal changes in sea surface salinity (around 35) are comparatively small (Becker, 1990). Large annual changes can be seen in the regional distribution of sea surface salinity, and long term salinity records of the North Sea also show significant variability. The high salinities are primarily caused by a combination of reduced freshwater input and vertical mixing, as well as increased influx of Atlantic water. The sea surface temperatures show a strong yearly cycle, with amplitudes ranging from 8 °C in the Wadden Sea to less than 2 °C at the northern entrances (Becker, 1990). The increasing amplitude towards the south-east is related to the greater proportion of low salinity coastal water and the reduced depth. In general, the sea surface temperatures seem to be rather stable. The long-term variability of the sea surface temperatures is closely correlated with the strength of the atmospheric circulation of the North Atlantic, the North Atlantic Oscillation. Light transmission through the water column is mainly limited by the presence of suspended matter and plankton. The spatial and temporal variability in concentrations of plankton and suspended matter results in high variability in light transmission. High turbidity associated with river outflow, high plankton concentrations and/or resuspension of bottom sediments. Such features are frequently observed in satellite images (Becker, 1990).

For the North Sea certain activities cause widespread impact or increasing trend, such as the impact of fisheries, inputs of nutrients from agriculture and rivers (Brockmann et al., 2003), and inputs of oil and chemicals associated with increased quantities of produced water from offshore oil and gas production. Concentrations of the antifouling agent tributyltin (TBT) still exceed safe levels in marine areas and an increasing number of synthetic compounds are being detected for which the ecological significance is not known. Though dredging impacts have diminished because of reduced contaminant loads, the quantities of dredged material are expected to increase in future, due to anticipated increases in the size of cargo vessels (OSPAR Commission, 2000).

Human impacts are greatest in the coastal zones. Many sensitive habitats with large ecological significance are disturbed or vanish due to a range of activities. It is foreseen that there may be significant impacts due to demographic developments and climate change. In conclusion, the intensive, and sometimes conflicting, use of the North Sea causes a number of problems in relation to a healthy ecosystem and sustainable use. There is as yet no real method for predicting the effects that climate change and extensive economical use (e.g. wind parks),

which is represented in figure 1-6, might have on the North Sea ecosystem. In order to better understand the mechanisms, it is very important, therefore, that long-term monitoring of key physical (e.g. temperature and salinity), chemical (e.g. dissolved CO_2 and oxygen) and biological (e.g. benthic species) variables are continued under the control of scientists and the relevant intergovernmental organisations. The availability of such data will make it possible to detect trends above the impacts due to the natural and anthropogenic (short-term) variability of the North Sea ecosystem.

1.4 Objectives and structure of this thesis

Nowadays, large interest exists to classify the marine environment by determination of abiotic factors and benthic organisms living at the seafloor (chapter 1.1). Time- and cost-intensive surface data are usually not available area-wide. Objectives as identification of seafloor provinces requires extrapolation of measurements obtained at distinct sites or along tracks. For these purposes geostatistical methods were applied (see chapter 2.3). Furthermore, important open questions can be answered by detailed characterisation of the seafloor combining sedimentological, geochemical, biological, and hydrological investigations by multivariate statistics as CART (see chapter 2.4). These include the following aspects:

• the definition of categorical biogeochemical habitats and a suggested source location of ascending mud at the Håkon Mosby Mud Volcano (HMMV) applying indicator kriging of marine data derived from georeferenced video mosaics combined with microbathymetric and temperature data,

• the estimate quality of the interpolated surfaces at the HMMV by means of statistical mean values derived from cross-validation and a validation dataset not included within the geostatistical model,

• the need to develop a system of fully automatical recognition and quantification of bacteria mats on georeferenced video mosaics because of the huge amount of recorded data during each expedition equipped with Remotely Operated Vehicles or Autonomous Underwater Vehicles,

• the requirement for a GIS-supported data management performing the integration of any kind of data type of marine geodata,

• the GIS-based generation of different thematic information layers (e.g. seafloor temperature) from measured data using ordinary kriging for the German Exclusive Economical Zone (EEZ) of the North Sea as basis for

• the calculation of seafloor provinces for the EEZ adapted from the geostatically estimated surface maps by means of multivariate statistical methods.

Chapter 2 generally represents the applied methodical procedures: geostatistics, multivariate statistics, and data management of marine geodata. Identification of seafloor provinces at the Håkon Mosby Mud Volcano (HMMV) is the issue of chapters 3 to 5 (section B). Section C regionally deals mainly with the German EEZ of the North Sea (chapters 6 to 8). Chapter 9 gives a general summary of chapters 3 to 8 with special emphasis on discussing the aspects given in this chapter.

2 Methodological overview

In this chapter, the applied methods at the HMMV and the North Sea are represented. Figure 2-1 illustrates the connection between the individual work steps and/or chapters of this thesis: data acquisition and management, georeferenced video mosaicing used for visual inspection and automatic image analysis, geostatistical and multivariate statistical methods as tools for the identification of sea floor provinces.

2.1 Georeferenced video mosaicing – an approach for advanced seafloor mapping

Quantitative analysis of seabed features in video images is of a great importance for numerous underwater applications. It is particularly interesting to produce mosaics from the seabed to create a video map whose extent is far larger than the size of a still camera image.

The MATISSE® system (Mosaicing Advanced Technologies included in a Single Software Environment developed by Ifremer) is based upon image and signal processing components and produces on-line and off-line georeferenced mosaics of the seabed by given video input and navigation data (Allais et al., 2004; Vincent et al., 2003). The resulting mosaics can be represented in a dedicated environment as a GIS enabling a further image analyses and linkages to other georeferenced data.

The whole algorithm to produce georeferenced mosaics is performed by different steps as lighting correction, camera self-calibration, and mosaic post-processing. Lighting correction is necessary because in case of deep underwater vehicle applications, videos are acquired
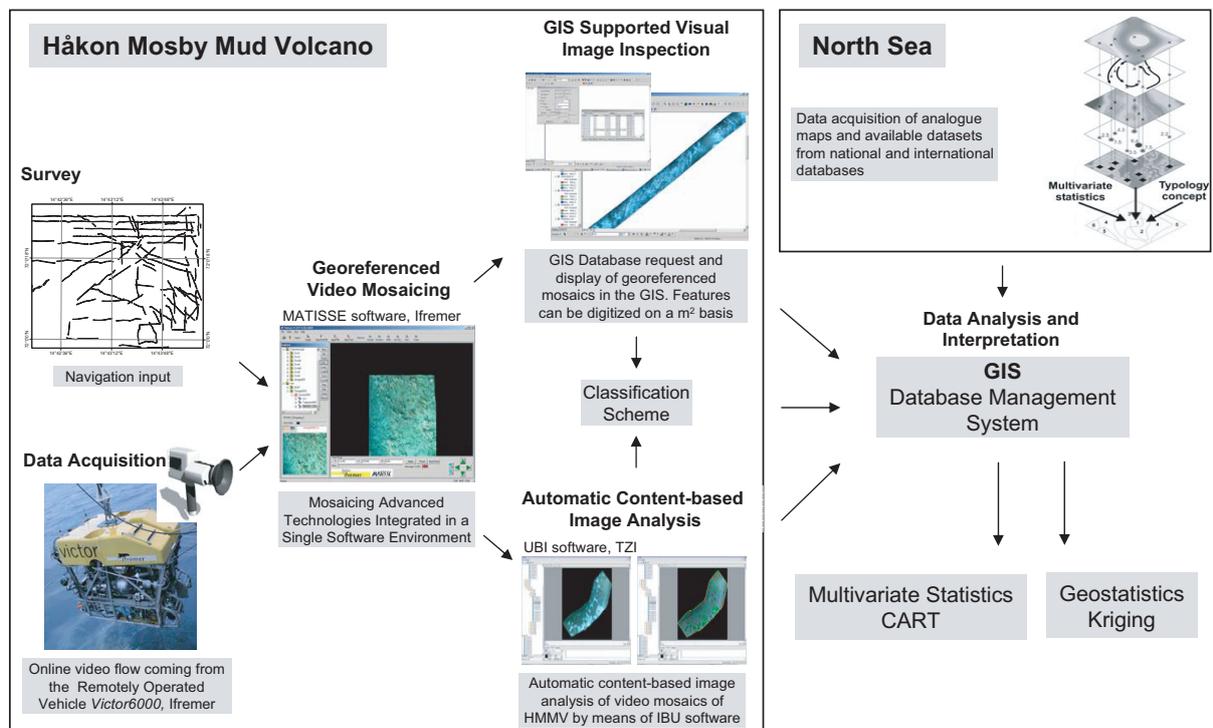
Figure 2-1: Methodological procedures: data acquisition, processing, and analysis.

using an artificial lighting. In general, a non-uniform intensity distribution is produced within the image and this can have negative effects on the mosaic processing and rendering. The intensity correction in an underwater environment was investigated in Borgetto et al. (2003). Since the marine environment is subject to varying optical conditions and since existing uncalibrated video data can be considered, tools for automated camera calibration without external means are required. A specific method has been developed by Pessel et al., (2003). This method needs only a sequence of few images of the seabed and is based upon the determination of the epipolar geometry between two successive images of the sequence. The epipolar geometry is represented by the fundamental matrix which links the image coordinates of the same point in two images. The method is then based upon the estimation of the fundamental matrix which is a function of the intrinsic and extrinsic parameters (Pessel et al., 2003).

At this stage, the mosaics can be built and enhanced by lighting correction. The next step introduces trajectory data into the processing of mosaic correction. Trajectory data of the camera are directly linked to vehicle navigation data provided by data from sensors such as Doppler velocity log, gyrocompass, inertial systems, which imply a long-term drift, or noisy acoustic positioning. A second method is based upon the trajectory of the image centre on the seafloor. A preliminary step allows correction of the image trajectory by blending vehicle navigation and image motion. Once the image trajectory on the seafloor is corrected, the mosaics are processed by warping algorithms (Jouffroy and Opderbecke, 2004).

2.2 GIS - Geographical Information Systems

Although many other approaches are important, conservation of biodiversity (e.g. Woodhouse et al., 2000), investigation of biogeochemical budgets (e.g. Schlüter et al., 2000; Seiter et al., 2004), or prediction of hazard zonation (e.g. Van Westen et al., 1997) will primarily be achieved by selection of key areas. For the analysis of the causative factors the application of geographical information systems (GIS) is an essential tool.

In the conception of GIS all available parameters represent information levels. GIS programs are capable to link these levels to each other, i.e. the put them in a relationship of mutual dependence (Bonham-Carter, 1996). GIS programs consist usually of a database system and a cartographic module and permits the processing of geographic information. The geoprocessing is one of the basic functions of a GIS. It provides a way to create new information by applying an operation to existing data, such as clip, select, or intersect datasets (McCoy, 2004). The surface overlay technique is the most important and most well-known GIS geoprocessing function for the production of new information (Fig. 2-2). Two or more spatially overlapping input data layers (e.g. x_1 and x_2) of geometrical type surface, line or point will be overlaid geometrically one above the other to produce a new original data layer y. Each knot and each

edge in x_1 and x_2 remain by overlay and further knots are produced at the cuts by edges from x_1 with x_2. At these existing edges at special sites are divided in order to be able to build them up to new geometry later. After the execution of the topology computations (cuts and new object class formation) the data content operations (appending the objects of x_1 and x_2) follow according to mathematical functions (e.g. Boolean algebra) (Bartelme, 2005). These logical operators based on Boolean algebra are due to a bivalent logic. They contain expressions as AND, OR, NOT and XOR (excluding OR). The functions can be conducted both with vector and with raster data.

The principle to show geographic data as levels or layers was applied for the first time in 1962 in the USA by the landscape architect McHarg. For the construction of a highway he included natural circumstances and produced a transparent map for every factor (topography, geology, vegetation, land use etc.). For evaluation purposes these maps were put on top of each other and thus it was relatively simply possible for McHarg to localise the best place for the highway (Conan, 2000). At that time merely analogous data were available. However, with the quick development in computing it became possible to collect many data, essential to manage and to represent them digitally as well as to exchange them. Thus, different software manufacturers started to discover the market in the beginning of the 80s and developed the first software for Geographic Information Systems (GIS). For the first time it became possible to handle a huge amount of data to work on and to represent them visually (Pfeil, 2005). In the end of the 80s these applications were accepted more and more by a growing group of users.

From a project of the CERN -European Organization for Nuclear Research-European Laboratory for Particle Physics in Berne the World Wide Web was developed in 1989. In the middle of the 90s the Internet was used and promoted by universities and public institutions with rising trend so the possibility to exchange data via Internet was strongly simplified and accelerated (Pfeil, 2005). Nowadays it is state-of-the-art in the GIS community to work in or to provide a Map Server environment. Up to now, several web-based Internet Map Server (IMS) were implemented to provide information about terrestrial environments (e.g. http://vitalgraphics. grida.no/arcticmap/). In contrast, in the marine context these developments seem to be just at the beginning (e.g. http://gisweb2.awi-bremerhaven.de/Website/margis/viewer.htm or http:// gisweb2.awi-bremerhaven.de/Website/metrol/viewer.htm).

2.3 Geostatistics – from point data to surface information layer

The representation and area calculation of irregularly distributed data from separated research fields is in the focus of all territorial geochemical balancing methods or definition of protection zones. The surface exactness of specific oceanic regions, as the estimate of global marine primary production (Longhurst, et al. 1995), benthic material flux (Zabel et al., 2000), organic carbon and sediments (Seiter et al., 2004) or the distribution of chemoautotrophic organism
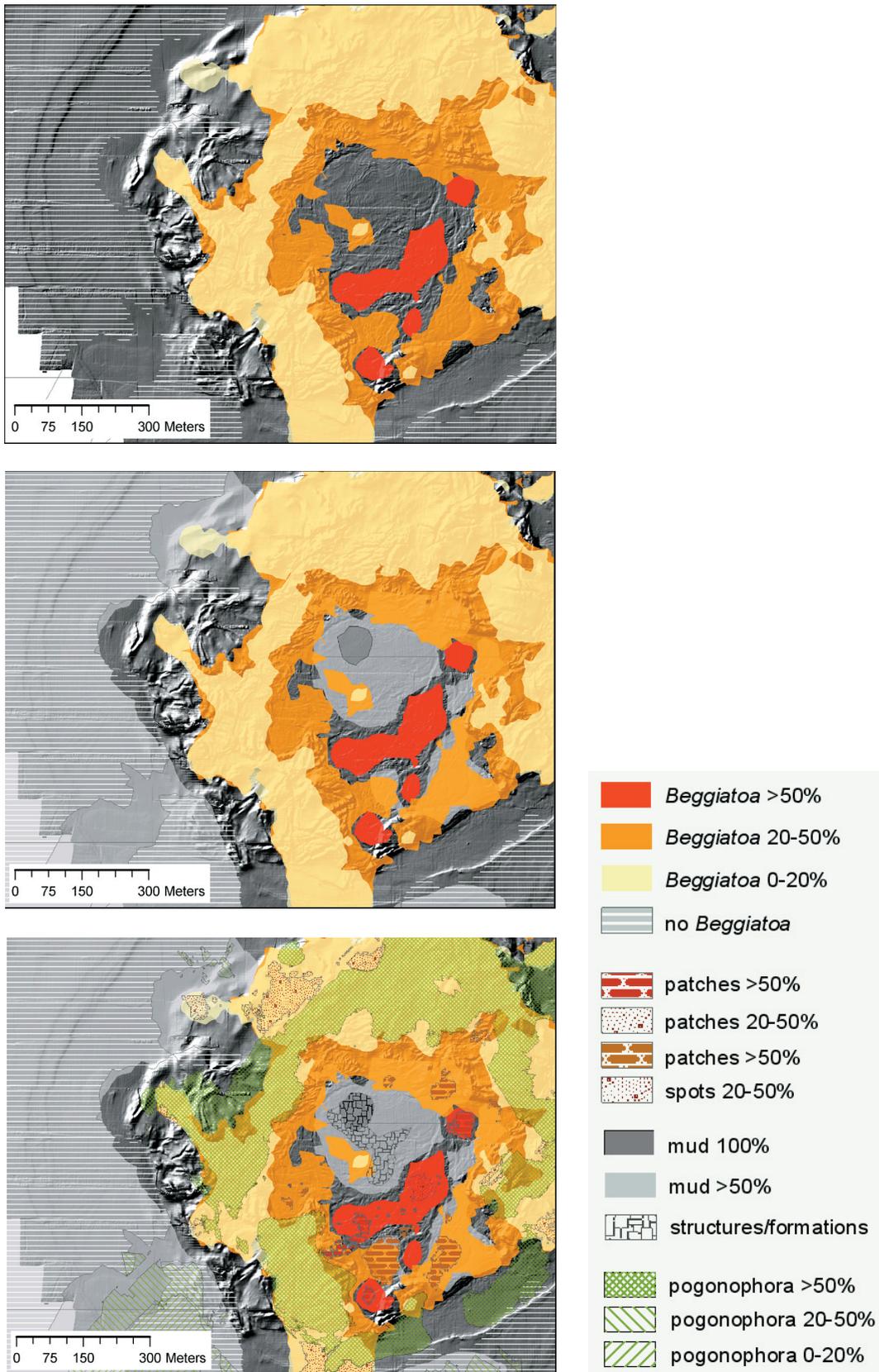
Figure 2-2: The principal of the overlay technique exemplified by Håkon Mosby Mud Volcano information layers.

(chapter 4) is always related to surface areas. Therefore, the interpolation of points into surfaces represents an important gain of information. Meanwhile large interest exists in multidiscipline research, since close relationships exist between the individual disciplines, for instance, global balancing of geochemical material cycles requires frequently the linkage of different information levels (layers). Multivariate statistical methods as CART (Classification and Regression Trees) in combination with GIS build the platform for the interdisciplinary view connecting the various information levels to each other (see chapter 2.4).

Formerly, before deploying autonomous underwater vehicles, the number of reliable and comparable measurements was often very low because most field studies were limited to selected ocean areas (coastal zones or surface oceanography). The low number of analysis and the extremely heterogeneity of the geographical distribution exclude en application of kriging procedures for regionalisation of and the generation of contour plots by means of GIS methods. Thus, frequently regression analysis was applied for interpolation of small datasets (e.g. Jahnke, 1996; Schlüter et al., 2000). Originally coming from geological research and applied to estimate mineral resources and reserves (Krige, 1951; Matheron, 1963), geostatistics are nowadays being used in various terrestrial and marine fields of research. As far as marine research is concerned geostatistical instruments were applied by various scientific disciplines, e.g. pollution research (Poon et al., 2000), geology (Chihi et al., 2000; Pehlke, 2005), biology (Harbitz and Lindstrøm, 2001; Jelinski et al., 2002), or marine geochemistry (Zabel et al., 2000; Schlüter et al., 1998). Compared to deterministic procedures like IDW (Inverse Distance Weighted method), geostatistical methods take into account the degree of spatial autocorrelation when predicting measurements. Geostatistics can be subdivided into two working steps: variogram analysis and kriging methods. With variogram analysis the autocorrelation structure of the underlying spatial process is examined and modelled. Variogram maps can be used to detect directional dependencies or so called anisotropies in the data field. The variogram models are used to predict measurement values by chosen kriging procedures. For more advanced insight see specialised literature (e.g. Akin and Siemens, 1988; Davis, 2002; Journel and Huijbregts, 1978; Schlüter, 1996; Wackernagel, 1998).

Kriging is an interpolation method for the computation of surfaces from regional distributed point data which is based on the concept of regionalised variables. A variable is called regionalised if its values are dependent on a location. Different possibilities exist describing these relations (autocorrelations) of spatial data. They all are based on the assumption that the autocorrelation of the data is not dependent on the absolute (geometrical) location of the sites, but on the relative spatial situation of the sites to each other in distance and direction (Isaaks and Srivastava, 1992). Thus, a regionalised variable is called autocorrelated if the characteristics of close points are more similar as these from distant points (Riek and Wolff, 1997). Kriging considers the spatial distribution of natural characteristics as combination of

a calculable deterministic (e.g. with a mathematical function) and a coincidental component (recordable with statistic methods). The spatial limit of autocorrelation can be quantified based on the computation of a variogram, identifying correlations of independence between measured points statistically (Schlüter, 1996, p. 239 therein). The variogram is a crucial basis for the execution of the kriging interpolation.

Variogram analysis can be performed following two principals: the first is based on determining semivariances for defined distance intervals and radial angle sectors (Fig. 2-3a), and the second on computation and visualisation of the semivariogram as so-called variogram maps. The new latter approach is presented here. Variogram maps give hints about anisotropies in the dataset assigning where spatial dependence not only changes with distance but also with direction. Such processes can be found, for example, in marine systems where ecological patterns strongly depend upon topography or water currents. The four work steps following Pesch et al. (in prep., chapter 9) are distinguished in modelling the spatial autocorrelation structure with the aid of variogram maps:

1. performing direction vectors also referred to as lag vectors from all point pairs of the considered point distribution (Fig. 2-3b);

2. connecting the origins of all lag vectors (Fig. 2-3c);

3. overlaying the result of 2. with a grid of a defined mesh size, assigning the semivariogram values for each cell, respectively (Fig. 2-3d) which results in a variogram map (Fig. 2-3e). The semivariogram values are weighted according to a Kernel-function depending on the distance of the ending of each lag vector to the centre of each cell (Johnston et al. 2001, pp. 254 therein). Variogram maps are point-symmetrical relative to their origin, because all lag vectors have a counterpart in the opposite direction;

4. assigning the calculated semivariances to a coordinate system defined by separation distance (x-axis) and semivariance (y-axis) resulting in the experimental variogram (Fig. 2-3e).

It is necessary to fit a defined variogram model to the experimental data. Three key parameters can be distinguished that allow describing the variogram model: range, sill and nugget effect in order to perform kriging (Fig. 2-3a). The range equals the maximum separation distance within which a distinct increase of semivariogram values can be observed. This is indicative of spatial autocorrelation. The sill corresponds to the semivariance assigned to the range. If anisotropies can be detected, both sill and range will vary with respect to direction. Small-scale variability or measurement errors may lead to high semivariances at nearby locations. The variogram model refers to this in terms of the so-called nugget effect, where the variogram model cuts the ordinate above the origin. A pure nugget effect indicates a complete lack of spatial autocorrelation.

The adaption of the variogram to the experimental data can be achieved by means of mathematical models fitted to the experimental variogram in terms of a least-squares regression
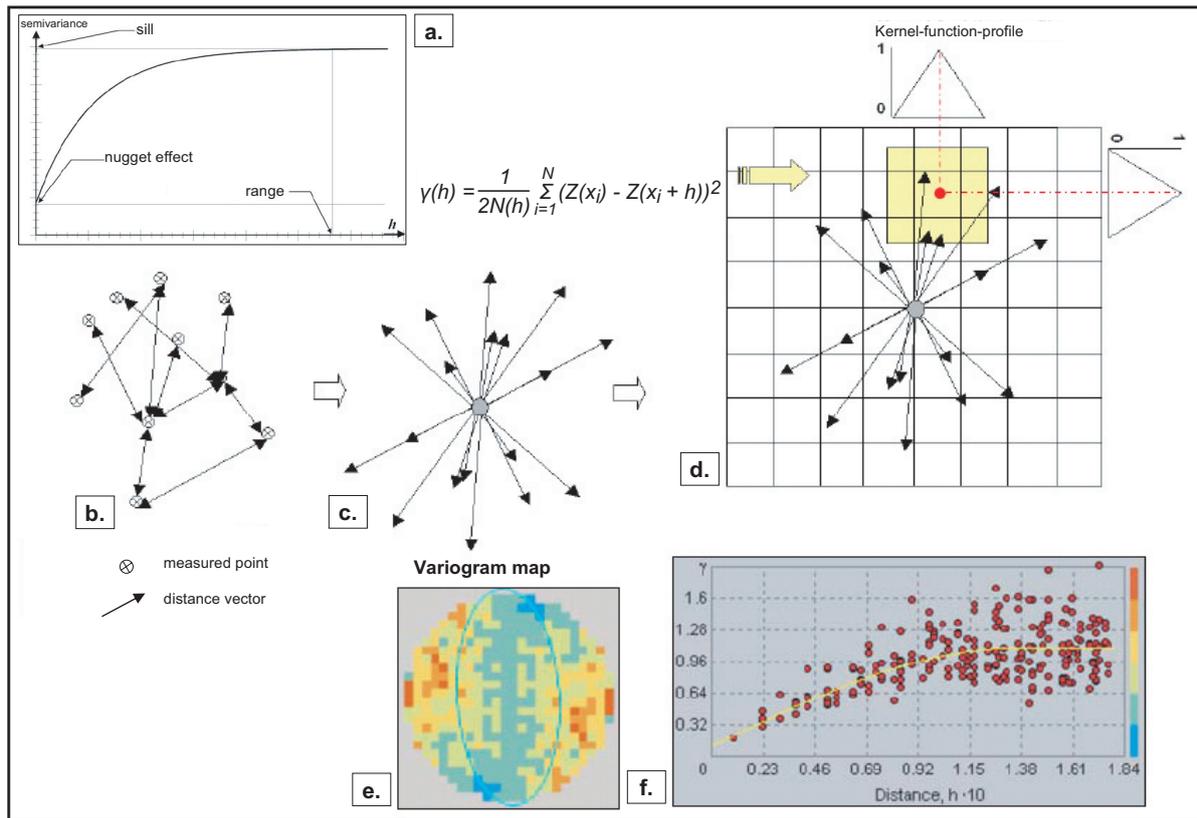
Figure 2-3: Operating sequence of variogram analyses by means of variogram maps (modified after Pesch et al., in prep., chapters 8 and 9). a. semivariogram, b. lag vectors from point pairs, c. lag vectors connected by their origins, d. overlay with a grid assigning the semivariogram values, e. variogram map, f. experimental variogram.

line. The ArcGIS software *Geostatistical Analyst* distinguishes eleven such models (Johnston et al., 2001), from which above all the spherical and the exponential models are used most frequently. For detailed mathematical descriptions, please refer to Webster and Oliver (2000) or Johnston et al. (2001).

Cross-validation can be used to choose the optimal variogram model. For this purpose, each measurement value is extracted from the dataset and estimated by means of the selected variogram model and the kriging method to be applied. By subtracting the measured value from the estimated value an estimation or kriging error can be calculated resulting in an error distribution for the whole dataset. Various key parameters can be calculated from the distribution to characterise the global quality of the chosen variogram model and the resulting surface estimation.

2.4 Combing information layers by multivariate statistics (CART analysis)

The methods used for predictive mapping vary widely from statistical approaches (including geostatistics) to more complex methods, such as expert systems and decision tree analysis (Kelly et al., 2005). CART (Classification and Regression Trees) is applied in various scientific disciplines to reveal hidden structures in complex data matrices, but in marine research CART

represents an innovative analysis method. CART is a so called tree-growing algorithm that produces decision trees to predict or classify the outcome of a certain feature (target variable) from a given set of meaningful predictor variables (Breiman et al., 1984). A major advantage of this technique is its ability to model non-additive and non-linear relationships among input variables. In contrast to most of the classification techniques as e.g. cluster analysis or classical regression analysis, CART handles very large sets of mixed, i.e. both categorical and parametric data without prior transformation of scale dignity. The central goal of the CART-algorithm is to produce homogenised classes with respect to the features of the target variable. Whether the target variable is of metric, ordinal or nominal scale dignity, different impurity measures exist. The Gini-index is commonly used when the target variable is categorical, although other options exist (Steinberg and Colla, 1995). CART does not make any assumptions on the distribution of the data and is extremely robust with respect to special cases as outliers or rare biotopes.

The CART analysis results in a decision tree grown in binary splits leading to a determined number of end nodes or classes, respectively. Each decision tree starts with one root node containing all observations of the sample. Following the dendrogram from up to down, stepwise increasing end nodes are observed in which one of the observations will be dominant. Since each of these end nodes is defined by a set of decision rules, the tree can be applied to predict the possible occurrence in percent of a certain observation at sites where no such information is available (chapter 9).

3 Outline of the manuscrips

Six manuscripts developed in the context of this thesis are described in this chapter. The manuscripts written by me as first author are based on own investigations and computations and were written independently. All manuscripts are linked to each other either methodically or content-based. The first three manuscripts are concerned to the evaluation or evaluation mechanisms of georeferenced video mosaics at the HMMV: first content-based interpretation; second estimation of the quality of the applied geostatistical method; third development of an alternative to visual inspection of marine image data, a content-based automatic image analysis. The next three manuscripts deal with already existing data for the area of the North Sea with focus on the German EEZ. The first of them presents the data management, the second the geostatistic method as one of the bases for the third manuscript, a multivariate statistically method for the identification of sea provinces.

Chapter 4

Spatial distribution of mud flows, chemoautotrophic communities, and biogeochemical habitats at Håkon Mosby Mud Volcano

K. Jerosch, M. Schlüter, J.P. Foucher, A.G. Allais, M. Klages, and C. Edy

A high resolution microbathymetry map and georeferenced video mosaics generated by means of the MATISSE software (Allais et al., 2004) and the Remotely Operated Vehicle *Victor6000* (Ifremer) of the HMMV, located in the deep sea at the slope of the Barents Sea (Vogt et al., 1999) are used to understand the spatial distribution of chemoautotrophic communities linked to the methane cycle benefiting from one of the products of anaerobe oxidation of methane (Boetius et al., 2000) and their relation to temporal to mud flow sequences. After application of indicator kriging interpolation the predicted surfaces of the individual biogeochemical habitats (BGH) were computed and assigned to a certain habitat indication class (BGH 0-4). Combing biological, morphological, heat flow (e.g. Crane et al., 1997; Kaul and Heesemann, 2004), and polynomial trend surface computation (Davis, 2000, p. 397-416 therein) of the microbathymetric data we suggest a source location of freshly ascending mud and assume a temporal dependence of the colonising structure by bacteria and pogonophorans.

Chapter 5

Spatial analysis of marine categorical information using indicator kriging applied to georeferenced video mosaics of the deep-sea Håkon Mosby Mud Volcano

K. Jerosch, M. Schlüter, and R. Pesch

The geostatistical methodology and GIS overlay technique applied previously is regarded more exactly in this chapter. The proven provinces of the HMMV are assessed by their estimate quality. For this purpose in a first step the indicator kriging interpolation method (Clark, 2000;

Davis, 2000; Krige, 1951; Olea, 1999) is examined for a working dataset (2310 analysed video mosaics) analysing variogram and cross-validation for each category. The estimate quality for the individual parameter category is then expressed on the basis of statistic mean values. Apart from the evaluation of the indicator kriging method itself, the results, mono-parametric habitat maps, are proven in a further step using a validation dataset consisting of the visual analysis of further 530 video mosaics.

Chapter 6

Automatic content-based analysis of georeferenced image data: detection of *Beggiatoa* mats in seafloor video mosaics from the Håkon Mosby Mud Volcano

K. Jerosch, A. Lüdtke, M. Schlüter, and G. Ioannidis

Due to the increasing amount of submarine image data and the deployment of new underwater technology, e.g. ROVs, new analysis methods for an automatic image analysis for the acquired image data are required. During six ROV dives more than 4200 video mosaics were compiled for the HMMV. A fully automatic detection and quantification of bacteria mats on the georeferenced video mosaics is developed. It applies watershed transformation for a pre-segmentation of the data areas on a video mosaic into disjunctive homogeneous regions (Beucher, 1991; Roerdink and Meijster, 2000) and relaxation labelling (Kittler and Illingworth, 1986) assigning these regions as *bacterial* or *non-bacterial* on the basis of spatial correlation and defined contrast thresholds. Difficulties arise particularly due to different illuminating of the seafloor and the heterogeneity of the bacteria appearance. The algorithm was compared to 2840 visually analysed mosaics and performed similar (precision better than 90%).

The programming work was accomplished by A. Lüdtke who was also a vital assistance in understanding automatic image analysis.

Chapter 7

MarGIS Marine Geo-Information-System for visualisation and typology of marine geodata

K. Jerosch, M. Schlüter, R. Pesch, W. Schröder, A. Köberle, and L. Vetter

This chapter focuses on the North Sea and gives an overview of the MarGIS project contents funded by the DFG and BMBF (Sonderprogramm Geotechnologien). A combination of Geo-Information-Systems (GIS) and research data build the framework for geostatistical (Clark, 2000; Davis, 2000; Olea, 1999) and multivariate geostatistical (Schröder and Schmidt, 2003; Thuiller, 2003; Wackernagel, 1998) techniques. Aim is the characterisation and identification of distinct provinces at the seafloor of the North Sea. MarGIS allows enhanced visualisation of multiple information layers by a multitude of biological, hydrological, sedimentological and bathymetric data, and improves the application of large environmental datasets (Perencsika

et al., 1999). MarGIS intends to support management issues as offshore wind power plants or the declaration of protection zones and is represented via the web by a GIS web-based viewer which permits a dynamic actualisation and access to metadata and additional information.

Chapter 8

Combining geostatistical methods and GIS to estimate temperature maps of the seafloor in the German Exclusive Economic Zone (EEZ) of the North Sea

R. Pesch and K. Jerosch

The geostatistical calculation of temperature maps for the sea floor of the EEZ for the summer months of the years between 1991 and 2000. Geostatistical methods (Clark, 1987; Davis, 2000; Krige, 1951; Olea, 1999) applying ordinary and indicator kriging are subject of this chapter. It is shown how variogram analysis and two different kriging techniques can be applied to spatially extrapolate measurement data to surface maps. Special emphasis was put on the optimization of the variogram model with respect to directional dependencies within the data field. Distinct anisotropies were detected in all ten years, resulting in search ellipsoids with the semi-major axis pointing in north-easterly direction (20° in 1992 to 84.4° in 1998) and anisotropy ratios ranging from 54% (in 1995) to 78% (in 1996). These directional dependencies allow for a realistic display of the temperature conditions near the sea floor since these tend to depend upon the distance of the location at the seafloor to the coast, or alternatively, the bathymetric conditions.

My contribution is due to the acquisition and selection of the data, as well as in the participation in the evaluation and discussion of the results.

Chapter 9

Using decision trees to predict benthic communities within and near the German Exclusive Economic Zone (EEZ) of the North Sea

R. Pesch, H. Pehlke, K. Jerosch, W. Schröder, and M. Schlüter

A methodological concept is introduced to predict the occurrence of benthic communities within the EEZ. Raster maps of biotic and abiotic measured data calculated applying ordinary point kriging (Krige, 1951). After intersecting the raster maps with punctual data on benthic communities (Rachor and Nehmer, 2003) a classification system is derived to predict the occurrence of these communities within the study area. This classification system is calculated from the intersected data by producing decision trees by the use of Classification and Regression Trees (CART) (Breiman et al., 1984; Huettmann and Diamond, 2001; Norcross et al., 1999). The results of the CART analysis resulted in a decision tree grown in nine binary splits leading to 10 endnodes or classes, respectively. As can be seen each decision tree starts with one root node containing all observations of the sample. By following the dendrogram from up to down

it can be observed that the portion of each benthic communities increases stepwise.

My contribution is due to the acquisition and selection of the data. Furthermore, I was considerably involved in the processing and production of the individual surface maps (e.g. sediments or bathymetry) and in the discussion of the CART results.

This thesis contains the original texts, figures, and tables of the six manuscripts. Three of them were submitted to international, peer-reviewed, journals in January 2006. The first manuscript (chapter 4) was submitted to *Marine Geology* (Elsevier Amsterdam), the second (chapter 5) to *Ecological Informatics* (Elsevier Amsterdam), the third (chapter 6) to *Computers and Geosciences* (Elsevier Amsterdam). The fourth manuscript (chapter 7) was accepted in international conference proceedings of the EnviroInfo2005 Brno, Czechia - Informatics for Environmental Protection. The fifth and sixth manuscripts are ready for submission to *International Journal of Geographical Information Science* (Taylor and Francis Group) (chapter 8) and to *Environmental Monitoring and Assessment* (Springer Heidelberg) (chapter 9). Due to its cumulative form, repetitions cannot be avoided in this thesis.

Section B:
Submitted manuscripts - Håkon Mosby Mud Volcano

4 Spatial distribution of mud flows, chemoautotrophic communities, and biogeochemical habitats at Håkon Mosby Mud Volcano

Kerstin Jerosch[1], Michael Schlüter[1], Jean-Paul Foucher[2], Anne-Gaelle Allais[2], Michael Klages[1], Christian Edy[2]

[1]Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
[2]French Research Institute for Exploitation of the Sea (Ifremer), France

4.1 Abstract

Marine mud volcanos are significant source locations contributing to the marine methane cycle. Enhanced heat flow, unique chemoautotrophic communities, occurrence of massive gas hydrates and large gas plumes are direct evidences of elevated methane concentrations and the dynamic environment of mud volcanoes. Related to the high concentrations and large inventories of CH_4 in surface sediments only a fraction of the methane is exported to the bottom water. This is mainly due to chemoautotrophic communities oxidizing methane and proving a "microbial filter" reducing CH_4 fluxes. Although these processes were studied for several mud volcanoes still little is known about the spatial pattern and the areas covered by chemoautotrophic communities or by present mud flows.

For this purposes the Håkon Mosby Mud Volcano (HMMV), which is located at the continental slope of the Barents Sea, was studied by several dives with the Remotely Operated Vehicle *Victor6000*. During these dives a high resolution microbathymetric map, with a footprint of 25 x 25 cm and a vertical resolution of better than 10 cm was derived. Furthermore, video streams of the bottom camera was converted into georeferenced mosaics, providing a detailed image about the spatial distribution of seafloor features as bacterial mats, pogonophorans, both indicating methane oxidation, or mud flows. Based on visual inspection of 2310 georeferenced mosaics covering an area of 46,160 m², different biogeochemical habitats were identified and quantified on a m²-basis. By application of geostatistic techniques as indicator kriging the distribution of different biogeochemical habitats were quantified and mapped for the entire HMMV.

Considering the flat and hummocky area of HMMV, approximately 16% (115,165 m²) of the flat centre is nearly void of any benthic communities. This area is considered as a region of high methane discharge into bottom water. An area of 5% (38,244 m²), located in the south-eastern

part, is densely inhabited by *Beggiatoa*. The hummocky outer part is colonised dominantly by pogonophoran tube worms (37.3%; 276,121 m²) and only occasional by *Beggiatoa.* Source locations and drainage directions for current mud flows were identified by computation of trend surfaces and consideration of temperature data. This suggests that present mud flow ascend close to the northern edge of the flat unit of HMMV, and that the drainage pattern of mud flows shifted from a westward to a south-south-eastern direction.

4.2 Introduction

Mud volcanoes (MV), whether they occur onshore or offshore, are major locations of mud and fluid transfer from deep geological layers to the earth's surface. To date, more than 1,700 MV have been recorded, of which about 800 are located offshore. It is estimated that more than 10,000 exist in deep marine waters (Dimitrov, 2002; Fleischer et al., 2001; Ivanov et al., 1996; Kopf, 2002; Milkov, 2000; Milkov et al., 2003; Pimenov et al., 2000). The formation of MVs is often linked to a tectonic compression, dehydration of clay minerals at depth, or rapid deposition of mass flows as slumps or turbidites (e.g. Kopf, 2002; Huguen et al., 2004; Kohl and Roberts, 1994; Vogt et al., 1999). They frequently form mud domes with diameters of up to a few kilometres and heights of several tens of meters above adjacent seafloor.

Observations such as recent mud flows, active seepage of fluids, enhanced heat flow, unique chemoautotrophic communities, massive gas hydrates, and large gas plumes are strong evidence for elevated methane concentrations in the sediments and water column (Damm and Budéus, 2003; Sauter et al., in press). Therefore, MVs are considered as significant source locations contributing to the marine CH_4 budget.

Up to now, the Håkon Mosby Mud Volcano (HMMV), located in the Norwegian Sea, is the only active MV reported for high northern latitudes (Fig. 4-1). This structure was found on the continental slope of the western Barents Sea fan (72°00.3'N and 14°44.0'E), within the submarine valley of Bear Island in a water depth of 1265 m (Hjelstuen et al., 1999; Vogt et al., 1997, 1999). Research cruises, including a survey by *NIS Akademik Mstislav Keldysh* equipped with the manned submersibles *MIR-1* and *MIR-2* studied the geological and geochemical settings (e.g. Milkov et al., 2004; Vogt et al., 1999). Multibeam and side-scan sonar surveys by Vogt et al. (1997, 1999) revealed the morphology and sediment textures. Biogeochemical, geophysical, and biological studies were conducted by Eldholm et al. (1999), Lein et al. (1999), Mienert and Posewang (1999), Milkov et al. (1999, 2004), Pimenov et al. (1999), and Vogt et al. (1999), among others.

At the HMMV, as on other offshore mud volcanoes, high concentrations of hydrogen sulphide and methane in the sediment offer a specific habitat for microbiological and macrofaunal associations (e.g. Boetius and Hinrichs, 2002; De Beer et al., in press; Levin et al., 2003; Pimenov et al., 1999; Sibuet and Olu, 1998; Sibuet and Olu-Le Roy, 2002; Soltwedel et al., 2005).
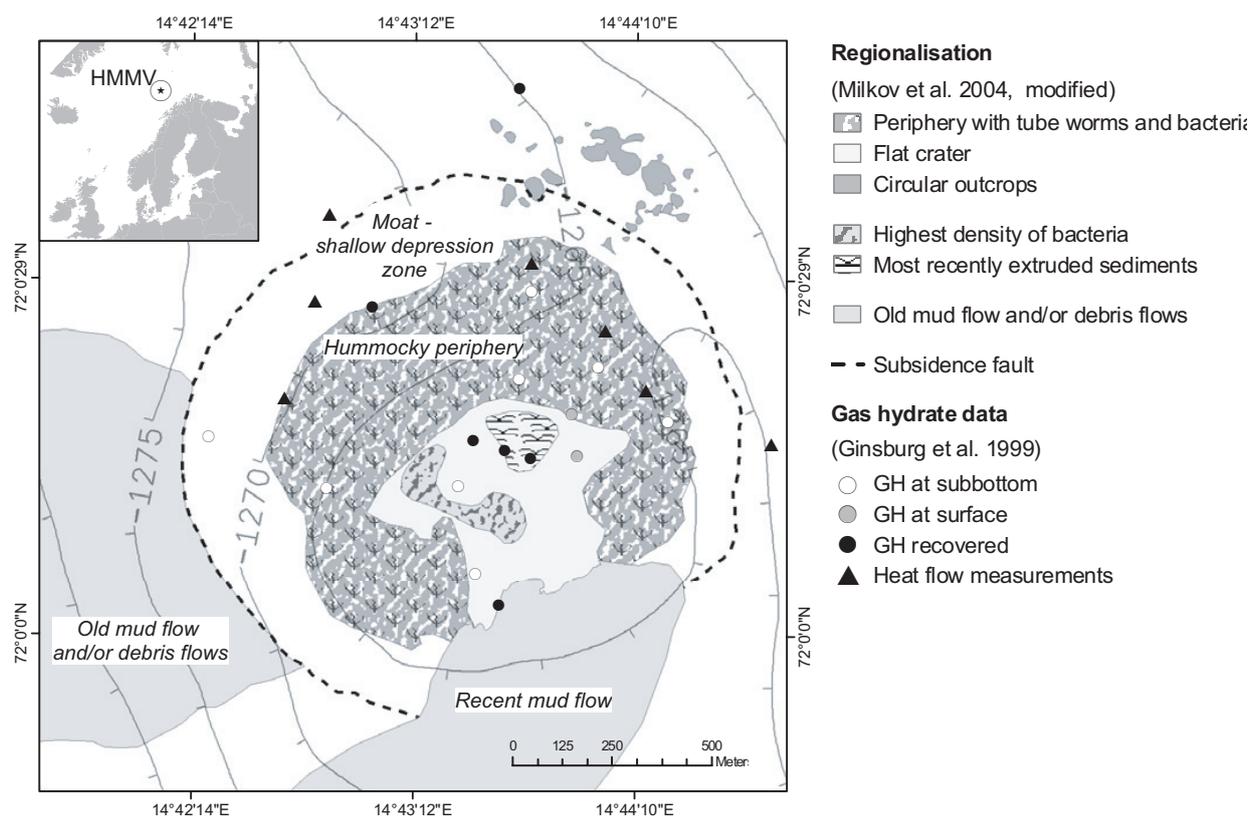
Figure 4-1: Map of the Håkon Mosby Mud Volcano (HMMV) located at the continental slope of the western Barents Sea. Mud flows, observations of gas hydrates, and distribution of bacterial mats and tube worms are indicated according to studies by, Eldholm et al. (1999), Ginsburg et al. (1999), Milkov et al. (2004), Smirnov (2000), and Vogt et al. (1999).

Sulphide-oxidising bacteria (dominant genus *Beggiatoa*, e.g. Boetius et al., 2004; De Beer et al., in press) and tube worms that rely on endosymbiotic bacteria (especially the pogonophoran species *Sclerolinum contortum* and *Oligobrachia haakonmosbiensis*) were found for example by Pimenov et al. (1999) and Gebruk et al. (2003). Sulphide mainly is produced by a microbial consortium of sulphate reducing bacteria and methane oxidising archaea (Boetius et al., 2000). This process is known as the anaerobic oxidation of methane (AOM) and provides an effective biological filter regulating the release of methane into bottom water. Subsequently, the terms *Beggiatoa* and pogonophorans are used as synonyms for chemoautothrophic communities, microbial consortia and associated symbionts dwelling form methane and reduced sulphur species dissolved in the pore water of sediments.

Due to anaerobic oxidation of methane, at most cold seeps and MVs methane fluxes into the water column are lower at sites covered by communities as *Beggiatoa* mats and pogonophoran tube worms if compared to the uncovered muds (Boetius et al., 2002; De Beer et al., in press; Hinrichs and Boetius, 2002; Sauter et al., in press). These chemoautotrophic communities are visual biogeochemical habitat indicators, detectable on videos and still photographs; quantification of their spatial distribution belongs to are the objectives of this study.

Although for the Håkon Mosby Mud Volcano the presence of these chemoautotrophic benthic communities was documented by towed video systems and observations by submersibles (Gebruk et al., 2003; Milkov et al., 1999), little is still known about the area covered by *Beggiatoa* mats and pogonophoran colonies, the mode of coverage (e.g. small spots versus patches or large mats), or their spatial relationship to mud flows and other morphological features.

For this purpose high resolution multibeam bathymetric data, still photographs, and video mosaics of the seafloor were recorded. The data were derived during dives with the remotely operated vehicle (ROV) *Victor6000* (Klages et al., 2004) and integrated into a Geographic Information System (GIS). This allows for the first time a detailed analysis of spatial patterns, the quantification of biogeochemical habitats (BGH) by geostatistical methods, and a investigation of mud sequences for a mud volcano like the HMMV.

1 km
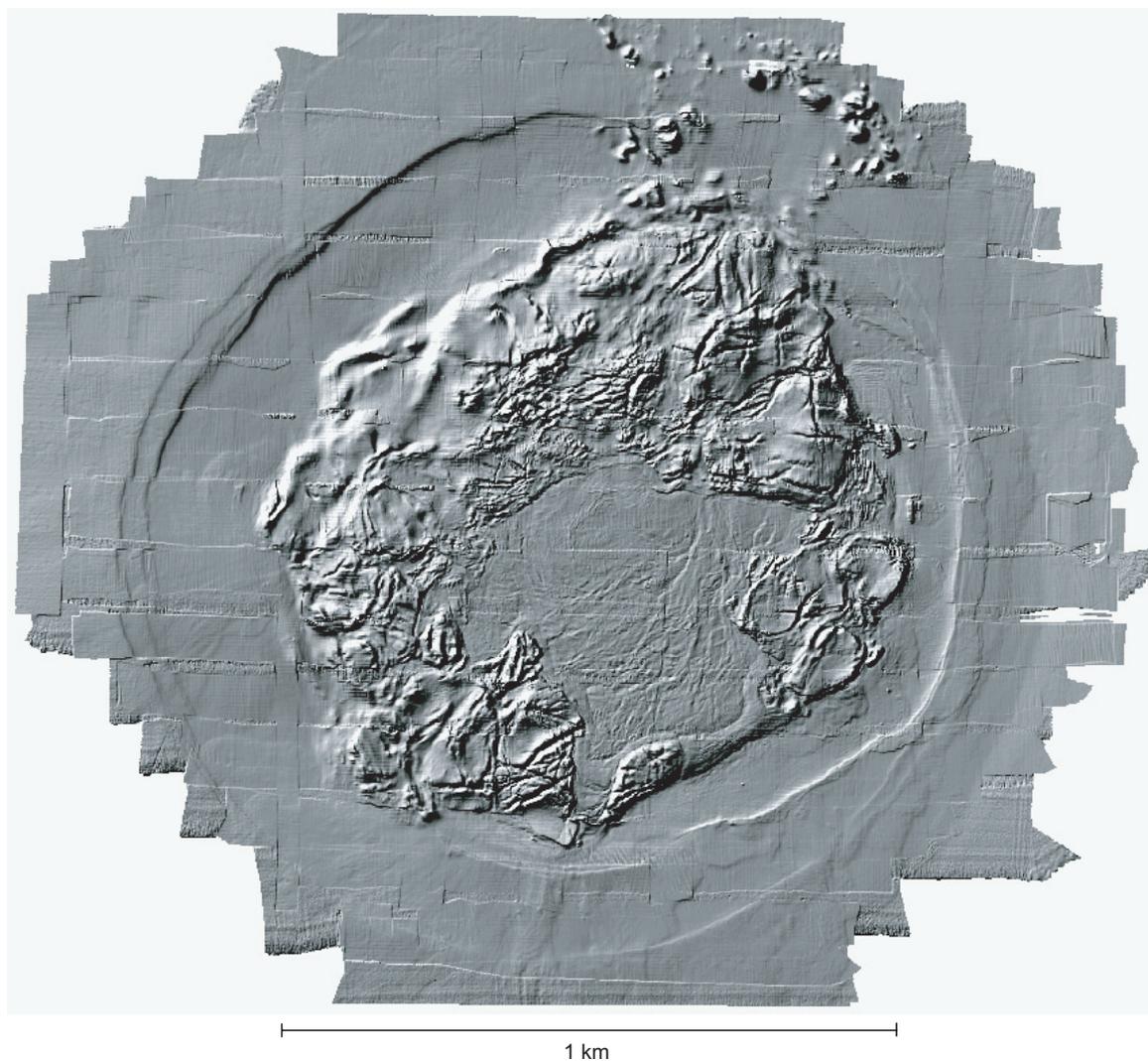
Figure 4-2: Morphology of the HMMV. Microbathymetry data derived by high resolution multibeam systems operated during dives of the ROV *Victor6000* are visualised as illuminated hill shading map. In total 250 million soundings were recorded. The vertical resolution of the map is better than 0.1 m and the foot-print at the seafloor is 0.25 m^2.

4.3 Material and methods

During the cruise ArcXIX3b (2003) of *RV Polarstern*, equipped with the remotely operated vehicle (ROV) *Victor6000*, the topography of HMMV was mapped, video observations of the seafloor were recorded, and sediment and water column samples were collected for geological and geochemical investigations. High precision underwater navigation using USBL (Ultra Short Base Line) were applied during the ROV dives and sampling. Further details on the ROV specification concerning water column and sediment sampling are reported in Klages et al. (2004).

4.3.1 Microbathymetric mapping and video mosaicing

During two ROV dives of 44 and 15 hours bottom time the high resolution multibeam systems Reson SeaBat 8125 (455 kHz) and Simrad EM2000 (200 kHz), respectively, were applied for microbathymetry mapping of HMMV. More than 250 million soundings were recorded, providing a full spatial coverage of the HMMV. Data were processed with the CARAIBES® software (Ifremer). Technical aspects of the multibeam survey and data processing are given in Edy et al. (2004) and Klages et al. (2004).

Mainly by the Reson SeaBat 8125 data a high resolution microbathymetry map with a depth resolution of better than 0.1 m and a footprint of 0.25 m^2 at the seafloor was derived. Based on the computed digital elevation model a hill shading map was generated for improved visualisation of seafloor features as mud ridges and flow patterns (Fig. 4-2).

Six dives of *Victor6000* were dedicated to video surveys. The ROV was navigated at about 0.3 m·s^{-1} and at a maximum altitude of 3 m above seafloor. This ensured a high image quality for the vertical camera installed at the bottom of the ROV tool sled. Since the aperture of the camera is 60°, the width of the mosaics is about the same as the altitude of the ROV during the survey.

The video stream generated by the vertical camera along the ROV transect was processed online by the MATISSE Software®. This mosaicing software, based upon image and signal processing components (Allais et al., 2004; Vincent et al., 2003), produces georeferenced mosaics (geotiffs) of the sea floor by combining the video input and Ultra Short Base Line navigation data. Each mosaic has a width of about 3 m. The entire length of the video profiles is about 17 km. The video mosaics were recorded and integrated into the Geo-Information-System ArcGIS (ESRI). Among the geotiffs, data published in journals and reports as well as unpublished data derived during previous cruises to HMMV were integrated into the GIS.

A total of 2310 georeferenced mosaics covering an area of 45,790 m^2 were analysed by visual inspection and GIS supported image analysis. This allows a detailed mapping of the spatial distribution of the different biogeochemical habitats which are related to the occurrence of *Beggiatoa*, pogonophorans, and uncovered mud. Examples of different biogeochemical
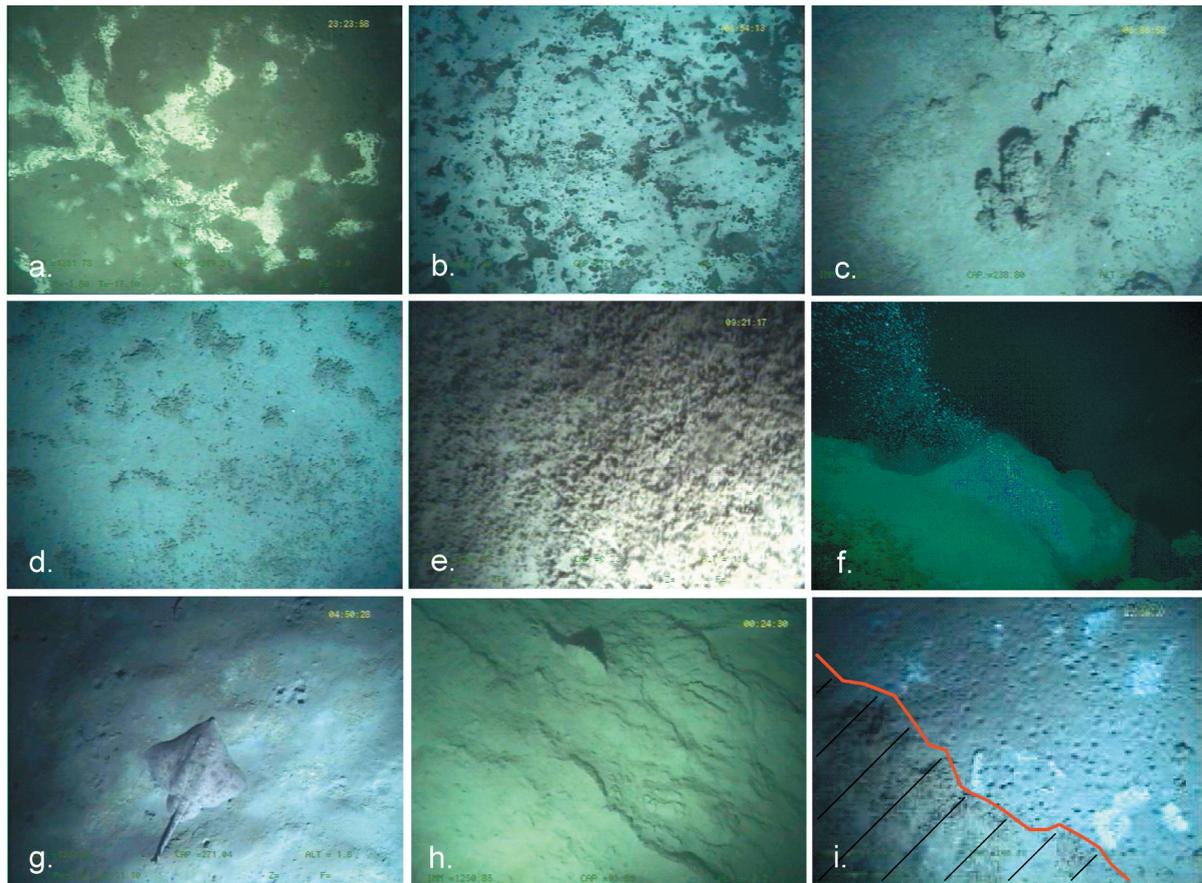
Figure 4-3: Still photographs of specific features at the seafloor of HMMV: a. *Beggiatoa* mats (light grey) distributed as patches with a spatial coverage of 20-50% of the depicted area, b. *Beggiatoa* patches (dense coverage of >50% of the area), c. authigenic carbonate crusts, d. pogonophorans (coverage <20%), e. pogonophorans (coverage >50%), f. active venting and release of free methane gas from the seafloor, g. smooth mud, with small holes (ray resting on seafloor), h. structured mud, and i. pogonophorans (lower left) and small holes (upper right).

habitats, characterised by the nature and density of benthic communities found on the seafloor as well as the sediment type are given in figures 4-3 and 4-4.

4.3.2 Classification scheme for biogeochemical habitats

The microbathymetric map derived by the multibeam data, the still photographs, and the video mosaics allow a detailed localisation and quantification of areas covered by the main spatial entities observed at HMMV. Special emphasis was placed on the determination of 1. white bacterial mats of sulphide-oxidising bacteria (dominant genus: *Beggiatoa*), 2. areas covered by pogonophoran tube worms, and 3. regions of apparently uncovered mud. As a result of this visual analysis, a classification scheme was developed (Tab. 4-1). This scheme considers the main habitat indicators (HI) as occurrence of *Beggiatoa*, pogonophorans, or uncovered mud, the degree of coverage by benthic organisms, as well as additional observations such as seafloor texture and specific features as small holes probably shaped by episodic gas release (Figs. 4-3f, i).

> 50% *Beggiatoa* patches

20-50% *Beggiatoa* patches
& > 50% smooth mud

< 20% *Beggiatoa* patches
& > 50% smooth mud

30 m

a.

<20% *Beggiatoa* patches
& 20-50% pogonophoran
& >50% smooth mud

<20% *Beggiatoa* spots &
>80% pogonophoran

<20% *Beggiatoa* patches &
20-50% pogonophoran

<20% *Beggiatoa* spots &
50-80% pogonophoran

>80% Pogon.
& Ray

<20% *Beg*. patches
& 50% smooth mud
& small holes & Fish

15 m

b.
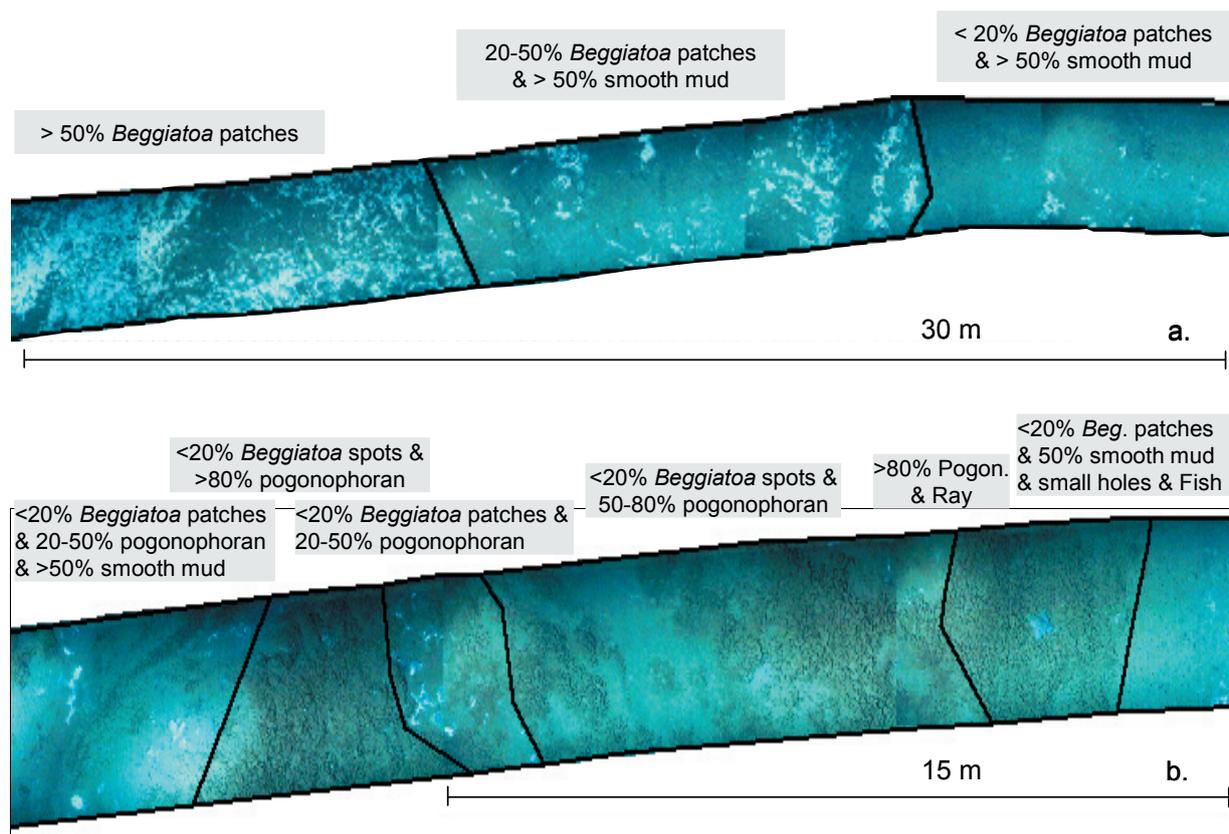
Figure 4-4: Video mosaics derived during ROV dives. a. mosaics presenting a sequence (left to right) from a dense coverage by *Beggiatoa* to essentially uncovered mud. b. mosaics showing a transition zone from *Beggiatoa* to pogonophorans and uncovered mud. By visual inspection of the video mosaics different sub-sections were identified following the classification scheme and indicated by polygons bordered by black lines within GIS.

GIS supported inspection of the video mosaics revealed that in some areas the chemoautotrophic organisms are arranged as small spots whereas in other regions large patches occur. Therefore, the category "*Beggiatoa* mats" was subdivided in narrow spots (Ø <30 cm) and patches (Ø >30 cm). The degrees of spatial coverage by pogonophorans or *Beggiatoa* mats on video mosaics were grouped into the classes <20%, 20-50%, and >50% related to sub-sections of the video mosaics (Fig. 4-4).

Mud regions essentially uncovered by benthic biota were observed for large areas of HMMV. The texture of the sediment surface is either nearly smooth or structured by ridges and holes (Fig. 4-3h). According to the texture, the mud areas were classified as smooth or structured. Accompanying observations such as the occurrence of carbonate precipitates or small holes were recorded and considered within the classification scheme (Tab. 4-1). The number of small holes (Ø 1-2 cm) varies considerably and in flat parts of the hilly periphery e.g. in the southeast of the HMMV reaches up to 137 holes per square meter.

The derived classification scheme relies on a small set of accurately identifiable features without loss of essential information. Based on the video mosaics, these classes were identified

Table 4-1: Classification scheme of habitat indicatores (HI), derived by visual inspection of video mosaics and still photographs (Figs. 3-3 and 3-4) for the HMMV.

Habitat indicator (HI)	Coverage degree [%]			
Pogonophorans	<20	20-50	>50	
Beggiatoa patches (Ø>30 cm)	<20	20-50	>50	
Beggiatioa spots (Ø<30 cm)	<20	20-50	>50	
Uncovered mud (smooth)			>50	100
Uncovered mud (structured)			>50	100
Accompanying observations (AO)	relief, large holes (Ø<15-20 cm), small holes (Ø<1-2 cm), carbonate crusts, demersal fish as eelpouts (Zoarcidae) and rays			

and digitized as polygons within the GIS. For each polygon, the geometry (e.g. diameter and perimeter), the area, and the habitat category is computed and stored. In total, more than 1722 polygons were identified and determined coming from 2310 analysed geotiffs.

4.3.3 Geostatistics: indicator kriging

Geostatistical techniques such as variogram analysis and kriging are commonly applied in oil, gas and mineral exploration, in groundwater research, and for marine or terrestrial habitat mapping (e.g. Bonham-Carter, 1996; Isaaks and Srivastava, 1998; Rivoirard et al., 2001; Schlüter, 1996). These techniques allow the computation of regional patterns and contour plots under consideration of the spatial distribution of the input data. Based on the visual interpretation of the entire set of video mosaics different, habitats were identified and digitised as polygons (Figs. 4-4 and 4-5). The dense coverage of video mosaics allows the application of indicator kriging (IK), variogram analysis and, therefore, the computation of surface maps covering almost the entire mud volcano.

The results of the geostatistical analysis were validated by calculation of standard errors of prediction and by cross-validation. In this statistical technique, raw data are removed from the entire data set. This artificially reduced data set is the input for a new Variogram analysis and IK, which now predicts values for coordinates where data (which were not part of the analysis) are available. Comparison between available raw data and predicted values express the accuracy of computed maps (Jerosch et al., subm-b.).

The result of the variogram analysis and IK are visualised as single maps for each of the three major habitat types: 1. areas of uncovered, structured and smooth mud sequences, 2. regions covered mainly by pogonophorans, and 3. regions mainly covered by *Beggiatoa* (Figs. 4-6 a-c). These three maps were aggregated to one by overlay technique within the GIS. This supports visualisation of the complex distribution of the distinct biogeochemical habitats (Fig. 4-6d). Combining spatial information as microbathymetry (Fig. 4-2), visual interpretation of video mosaics (Fig. 4-4), maps on slope direction (Fig. 4-5), and geostatistical analysis allows to estimate the coverage of HMMV by different biogeochemical habitats.

4.4 Results

4.4.1 Quantification of morphological units

The high resolution microbathymetric data (Fig. 4-2) allows a detailed identification and quantification of three major morphological features of HMMV: I. the flat area in the centre and the south east, II. the hummocky periphery, and III. the moat area including circular outcrops (Figs. 4-2 and 4-5). The latter is a shallow depression surrounding the hummocky periphery. The entire area of the HMMV (eHMMV), units I, II and III, is almost circularly limited by subsidence faults, has a diameter of 1365 m in N/S and 1380 m W/E direction, and covers a total area of 1,395,532 m^2.

The flat area (unit I) in the centre of HMMV is located in a water depth of 1255 m to 1259 m, has a diameter of 400 to 500 m, and is formed by recent mud flows. It is enclosed by a hummocky peripheral rim (unit II) with a maximum width of 440 m in the North, 350 m in the West, and 150 m in the East. In the SE the hummocky peripheral rim is rather narrow and has a minimum width of only 10 m.

The hummocky region shows a maximum height of only 9 m in relation to the centre and at
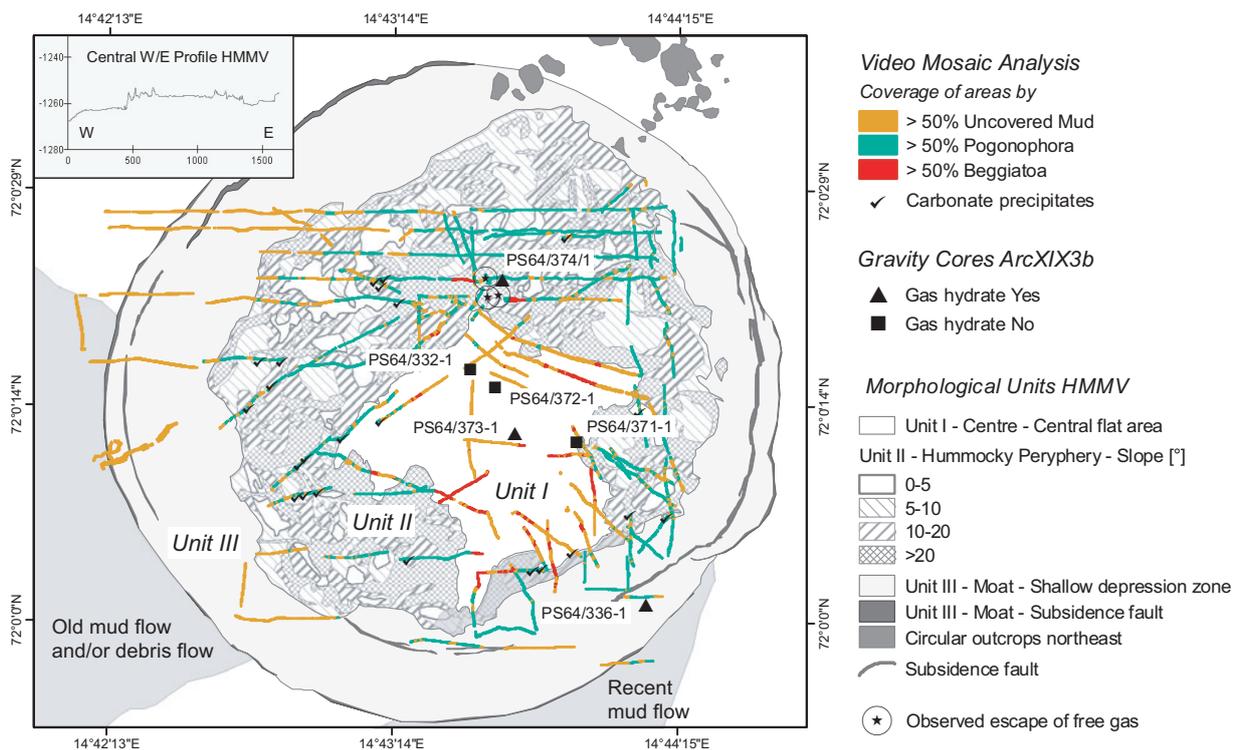
Figure 4-5: Major morphological units of the HMMV derived by microbathymetric data and computed hill shading map (Fig. 4-2): centre (unit I), hummocky peripheral rim (unit II) and moat are (unit III). The area of these three units is defined as the entire area of the HMMV (eHMMV). Subsidence faults (black, nearly circular lines) limit the eHMMV from the surrounding seafloor. Tracks of video mosaicing survey are indicated by coloured lines. The red, green, and brown lines indicate areas covered by more than 50% with *Beggiatoa* mats, pogonophorans or areas of unsettled mud, respectively. Unit I and II are reflected as the high CH$_4$ region (HCR).

most 16 m relative to the surrounding moat area of the mud volcano. In the southeast the fresh mud flow replaces the hummocky structure partially, indicating the direction of recent mud flows (Fig. 4-2). The third unit, a circular 2 m deep moat of 100 to 270 m width surrounds the hummocky unit. In the NE this moat is partly interrupted by small circular outcrops (10 to 30 m in diameter) exceeding the mud volcano structure.

Unit I and II are the regions where chemoautotrophic communities, mud flows, enhanced geothermal gradients, high CH_4 concentration in sediments, seepage of CH_4 bubbles, and gas hydrates were observed. This very active area, related to the CH_4 cycle, of unit I and II is referred as the high CH_4 region (HCR) of HMMV, subsequently. The HCR covers an area of 739,729 m² (53% of the eHMMV) and forms the reference area for the surface calculation in the following. In N/S and E/W direction the diameter of the HCR is 1000 m and 1050 m, respectively. The maximum diameter of the HCR, 1160 m, is observed in SW/NE direction.

Besides these characteristic units, the hill shading map visualises the general lineation pattern as rims and ridges on the sediment surface (Fig. 4-2). The dominant orientation of these lineations is in NE to SW direction. Several E to W oriented lineations are obvious in the northern part. The mud sequence in the central part of HMMV seems to be structured by distinct flow lines.

The accuracy of the high resolution bathymetry allows the computation of elevation profiles (Fig. 4-5) and calculation of slope angles for the different units (Tabs. 4-2 and 4-3). This spatial analysis reveals that the central flat region has an area of 192,527 m² (26% of the HCR or 13.8% of the eHMMV) and is characterised by low inclination values of generally less than 3° (mean value: 2.0°). Only in the southwest of unit I a few areas show values up to 5° (mean value: 3.5°) due to local undulations. Nevertheless, figure 4-2 reveals that these areas stand out clearly from the slope inclinations of the hummocky unit II and are thus assigned to unit I. The mud layer in the centre seemed to be subdivided by rims or terraces (Fig. 4-3h) caused probably by individual sequences of mud flows. We assume that these regions are indicating the temporal sequence and pathways of mud flows and their interaction with the hummocky landscape, for example, mud flows terminated at hummocky structures in the south and the south-east (Fig. 4-2).

The total area of the hummocky peripheral rim is 547,202 m² (74% of the HCR or 39.2% of the eHMMV), thus more than 2.5 times larger than the flat mud region in the centre of the HMMV. Large slope inclinations of more than 20° are typical. A maximum slope angle of more than 50° was found at lineations distributed regularly in the hummocky area. For both the flat area in the centre and the hummocky peripheral rim the calculated slope angles coincide well with the visual observations obtained during dives by *Victor6000*; especially in the northern part steep slopes and escarpments were obvious.

The third unit, the shallow depression zone surrounding the HMMV, has a size of 612,182 m²

Table 4-2: Slope inclinations and areas of the main morphologic units at the HMMV. For the moat region slopes of >5° (marked by *) were observed in the direct vicinity of fault lineations (Fig. 4-5). The HCR region comprises of HMMV units I and II, the eHMMV of units I, II and III.

	Unit I	Unit II	High CH_4 region (HCR) Unit I+II	Unit III		Entire HMMV (eHMMV) Unit I+II+III
Slope [°]	Central flat area [m2]	Hummocky periphery [m2]	Sum [m^2] HCR	Moat area [m^2]	Circular outcrops in NE [m^2]	Sum [m^2] eHMMV
0-5	144284	-	144284	612182	-	756466
5-10	48243	139808	188051	17485 *	-	205536
10-20	-	205021	205021	7065 *	9552	221637
>20	-	202373	202373	2037 *	16,584	220993
Sum [m2]	192527	547202	739729	629667	26,136	1395532
HCR [%]	26.03	73.97	100			
eHMMV [%]	13.8	39.21		45.12	1.87	100

(43.9% of eHMMV) and rather low inclinations of less than 1.5° (mean value: 1.4°). The abrupt transition between the hummocky region and the moat area was also visually confirmed during the ROV surveys.

4.4.2 Spatial distribution of biogeochemical habitats

Image analysis showed that 21,567 m^2 (47.1%) of the observed area (45,790 m^2) is densely, by more than 50%, colonised by pogonophorans. In contrast, with 2290 m^2 (5%) a much smaller area is densely (>50%) colonised by bacterial mats. The mud area, essentially uncovered by bacterial mats or pogonophorans, covers 21,933 m^2 (47.9%) of the investigated region. 6460 m^2 of these uncovered area is located outside of the HCR (Fig. 4-5).

Since the video surveys cover a significant part of the MV, these calculated areas are considered as robust estimates for the spatial distribution of biogeochemical habitats of the HCR. A step

Table 4-3: Summary of the microbathymetry analysis: minimum, maximum, and mean bathymetric values concerning the morphologic units, together with variances and the average slope inclination.

Morphologic units		Bathymetry [m]				Slope [°]
		Min	Max	Diff	Mean	Mean
Central flat area (unit I)	0-5°	-1258.7	-1254.9	3.8	-1256.8	2.0
	5-10°	-1259.	-1254.7	4.5	-1257.0	3.5
	entire I	-1259.2	-1254.1	4.5	-1257.0	2.4
Hummocky area (unit II)	0-5°	-1263.2	-1254.1	9.1	-1258.7	2.7
	5-10°	-1266.0	-1257.6	8.4	-1261.1	4.7
	10-20°	-1265.3	-1250.0	15.3	-1257.7	6.8
	>20°	-1265.5	-1250.0	15.5	-1257.8	13.6
	entire II	-1266.0	-1250.0	16.0	-1258.0	8.2
Moat area (unit III)	0-5°	-1266.3	-1257.2	9.1	-1261.8	1.4
	5-10°	-1264.7	-1257.3	7.4	-1261.0	5.6
	entire III	-1266.3	-1257.2	9.1	-1261.8	1.6
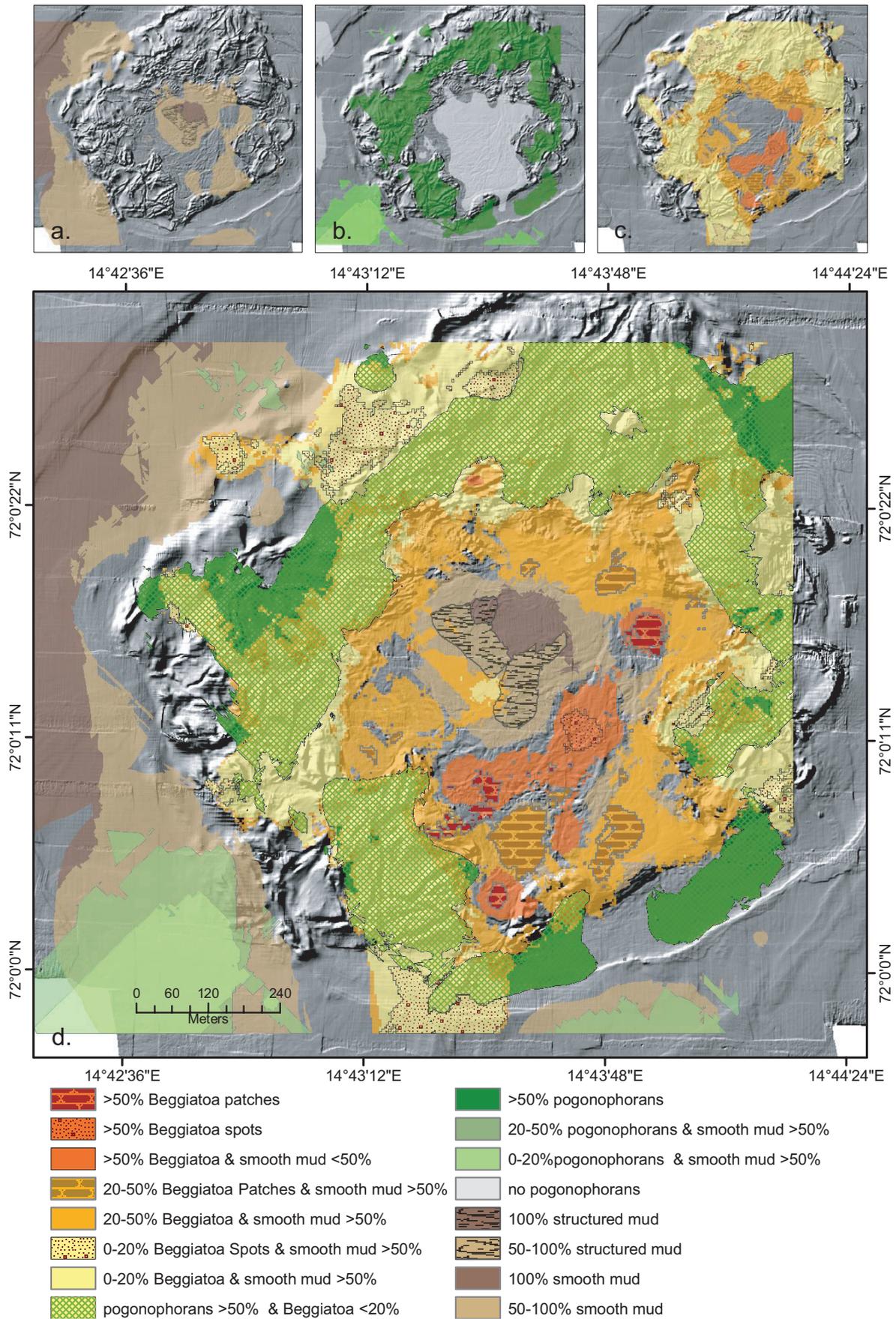
beyond interpretation of these observations is provided by geostatistical analysis. Variogram analysis and indicator kriging allow generating area-wide prediction especially for the HCR (Fig. 4-6, Tab. 4-5).

Figures 4-6 a-c show the spatial distribution of uncovered mud and of dense population by pogonophorans or bacterial mats. Visual inspection of the seafloor images reveals a high coverage by bacterial mats for the outer rim of the flat central area (Fig. 4-6c) whereas pogonophorans cover large regions of the hummocky periphery (Fig. 4-6b). The regions of uncovered mud can be subdivided into two settings: 1. the flat centre of the crater and 2. the moat region enclosing the hummocky periphery (Fig. 4- 6a). This general pattern, already reported by Milkov et al. (1999) and Gebruk et al. (2003), can be improved by the georeferenced video mosaics to a more detailed spatial analysis of the HMMV habitats. This spatial analysis allows to calculate areas as well as to decipher the degree of coverage by chemoautotrophic organisms dwelling on methane and sulphur fluxes from below.

Based on the three mono-parametric maps (Figs. 4-6 a-c) areas covered by chemoautotrophic communities and uncovered mud were calculated. These area calculations, referring to the three mono-parametric maps, are given in table 4-4. The spatial analysis revealed that 31,435 m^2 (4.2% of the HCR) are densely covered (>50%) by *Beggiatoa* mats and a considerable larger area of 275,958 m^2 (37.3% of the HCR) is densely covered by pogonophorans. In the centre of the mud volcano an area of 115,165 m^2 (15.6% of the HCR) is not or only very scarcely colonised by larger CH_4 indicating micro- and macro-fauna of *Beggiatoa* and pogonophorans. By combining the three information layers (Figs. 4-6 a-c) into a single map by GIS overlay technique the complex habitat distribution at the HMMV can be identified (Fig. 4-6d). Based on table 4-5 and figure 4-6d the combined populated areas (*Beggiatoa* and pogonophorans) can be quantified and assigned to sub-regions of the high CH_4 bearing region (HCR).

Mud areas essentially uncovered by any chemoautotrophic organism are observed in the centre and the SE of the HMMV (Fig. 4-6a). Whereas the major part of the mud surface is rather smooth and unstructured, an area surrounding the central part is structured by hollows and rims as in figure 4-3h. This structured mud is mainly stretching from the centre to the E and S of the centre (Fig. 4-6d).

Whereas a few regions in the central, rather flat part of the HMMV are covered densely by *Beggiatoa*, no pogonophorans were observed in this region (Fig. 4-6b). The pogonophorans are essentially restricted to the hummock periphery. Furthermore, this chemoautotrophic

Left page:
Figure 4-6: Spatial distribution of biogeochemical habitats at the HMMV. a. areas of uncovered, structured and smooth mud sequences. b. regions covered by more than 50% with pogonophorans and only sparsely (<20%) by *Beggiatoa*. c. regions covered by *Beggiatoa*. d. multi-parametric map representing the complex biogeochemical habitats at the HMMV. For example, dense coverage by bacterial mats (>50%) is observed for regions in the south east of the centre, whereas large areas in the hummocky periphery are only sparsely covered by the bacterial mats and dominated by pogonophorans.

Table 4-4: Areas covered by the different chemoautotrophic communities and uncovered mud regions according to a geostatistical calculation of three mono-parametric maps. The percentage of coverage is related to the HCR of the HMMV (units I and II: 739,729 m^2): a. *Beggiatoa,* b. pogonophorans, c. mud. The values in a-c of this table correspond with the areas in the mono-parametric maps in figures 4-6 a-c.

a.	Area [m^2]	% of Coverage	HI *Beggiatoa*
Beggiatoa	6809	0.9	>50% patches
	2357	0.3	>50% spots
	29078	3.9	>50%
	22699	3.1	20-50% patches
	43815	5.9	20-50% spots
	163077	22.1	20-50%
	250200	33.8	0-20%
	88099	11.9	no occurrence
∑ for >50%	38244	5.2	
∑	606134	81.9	

b.	Area [m^2]	% of Coverage	HI Pogonophorans
Pogono.	276121	37.3	>50%
	43295	5.9	20-50%
	36649	5.0	0-20%
	297746	40.3	no occurrence
∑ for >50%	276121	37.3	
∑	653811	88.5	

c.	Area [m^2]	% of Coverage	HI Mud
Mud	8286	1.1	100% smooth
	1182	0.2	100% structured
	93441	12.6	>50% smooth
	12256	1.7	>50% structured
∑ for >50%	*115165*	15.6	
∑	*115165*	15.6	

community is observed outside the moat subsidence faults in the S and SW of the HMMV. Except for the central part of the HMMV, where smooth or structured mud is observed, bacterial mats are spread almost over the entire HCR (Fig. 4-6c). Especially the region south of the structured mud is densely (by more than 50%) covered by *Beggiatoa* mats which occur on smooth mud areas.

The association "*Beggiatoa* mats with smooth mud" is also observed for the moat area in the S and NNW of the HMMV. Even in the hummocky periphery (morphological unit II) bacterial mats were observed. With 167,329 m^2 a large area (22.6%) of the hummocky area is settled by pogonophora (>50% coverage) and *Beggiatoa* (<20% coverage) (Fig. 4-6d, Tab. 4-5).

4.5 Discussion

The microbathymetry map derived by the high resolution multibeam systems (footprint of 0.25 m^2) and the video mosaics compiled during the ROV dives, provides a detailed insight into the morphology and distribution of mud flows and chemoautotrophic communities. This allows to identify seabed structures with only a few centimetres of elevation, the computation of a digital elevation model, or to derive a map on slope direction. Therefore, relations between geomorphologic structures and biogeochemical habitats, observed on georeferenced video mosaics, can be identified.

For spatial analysis of seafloor features such as the degree of coverage by benthic organisms or of sediment properties, video mosaicing provides several advantages compared to still photographs. Since they are georeferenced, their coordinates and extent - on a m^2 basis - is known, and the area of even small scale features can be measured accurately and related to bathymetry and other features. Whereas adjacent still photographs of the seafloor might overlap, bearing the risk of overestimating spatial extents, this overlap is eliminated by the

Table 4-5: Area and coverage degrees of the habitat indicators (HI) referred to the high CH$_4$ region (HCR) (739,729 m^2). The table gives a subset of the results after an application of GIS overlay technique using the mono-parametric maps (*Beggiatoa* and pogonophorans) and corresponds to figure 4-6d.

Habitat Indicator (HI)			Area [m^2]	% of Coverage
Beggiatoa		Pogonophorans		
>50% patches	and	no occurrence	6809	0.9
>50% spots	and	no occurrence	2357	0.3
>50%	and	no occurrence	22269	3
20-50% patches	and	no occurrence	21689	2.9
20-50%	and	no occurrence	35093	4.7
0-20% spots	and	no prediction	30185	4.1
0-20%	and	>50%	167329	22.7
0-20%	and	no prediction	82127	11.1
no occurrence	and	>50%	61013	8.3
no occurrence	and	0-20%	26436	3.6
no occurrence	and	20-50%	21333	2.9
no occurrence	and	no occurrence	*115165*	15.6
Σ			591803	80

video mosaicing algorithm. Since the mosaics provide a continuous cover of considerably larger areas than still photographs, regional transitions from scarce to dense population by *Beggiatoa* mats or from smooth to structured mud can be identified and quantified (Fig. 4-4). Compared to the size of the eHMMV (approx. 1.4 km x 1.4 km), which is comparable to other offshore mud volcanoes (Kopf, 2002; Milkov, 2000), the spatial coverage by video mosaics is rather high (Fig. 4-5). In addition to the interpolation of the mosaics, we applied variogram analysis and indicator kriging to derive maps about the distribution of different biogeochemical

habitats for the entire area of the HMMV. In general, spatial interpolation provides local estimates. Statistical techniques, maps on standard deviations, semi-variogram analysis, and cross-validation, allow to evaluate the accuracy of such linear estimates and are suitable to optimise the spatial interpolation process (Clark, 1982), one reason to apply kriging.

Furthermore, the applied IK deals with the data heterogeneity according to a scientific question to be studied (Goovaerts, 1997); for example, investigating and predicting the probability of the occurrence of an organism at a certain place at the seafloor. The determination of the areas shown in figure 4-6 is based on a defined threshold of occurrence probability of each HI, respectively (Jerosch et al., subm-a.). For our data set statistical mean values resulting from the cross-validation and the comparison between observed values and predicted values, revealed similarities of better than 83%. Particularly good results were obtained with the forecast of uncovered mud areas (>89%), regions with more than 50% bacteria coverage (>92%) and areas covered to 0-20% by pogonophorans (>90%).

The microbathymetry data and hill shading map (Fig. 4-2) allow interpretation of the sequence of mud flows. Mud flows identified by interpretation of side scan sonar images were heading in westerly (old mud flow) and southern (young mud flow) direction (Milkov et al., 2004; Vogt et al., 1997). This is confirmed by the multibeam microbathymetric data (Fig. 4-2), showing a lobe-shaped, 2 m high elevation within the western and the southern moat subsidence zone.

A more detailed consideration of flow direction and possible source location of current mud flows within the HCR can be derived by microbathymetric data and temperature measurements. For this purpose trend surface analysis was applied to bathymetric data obtained for the
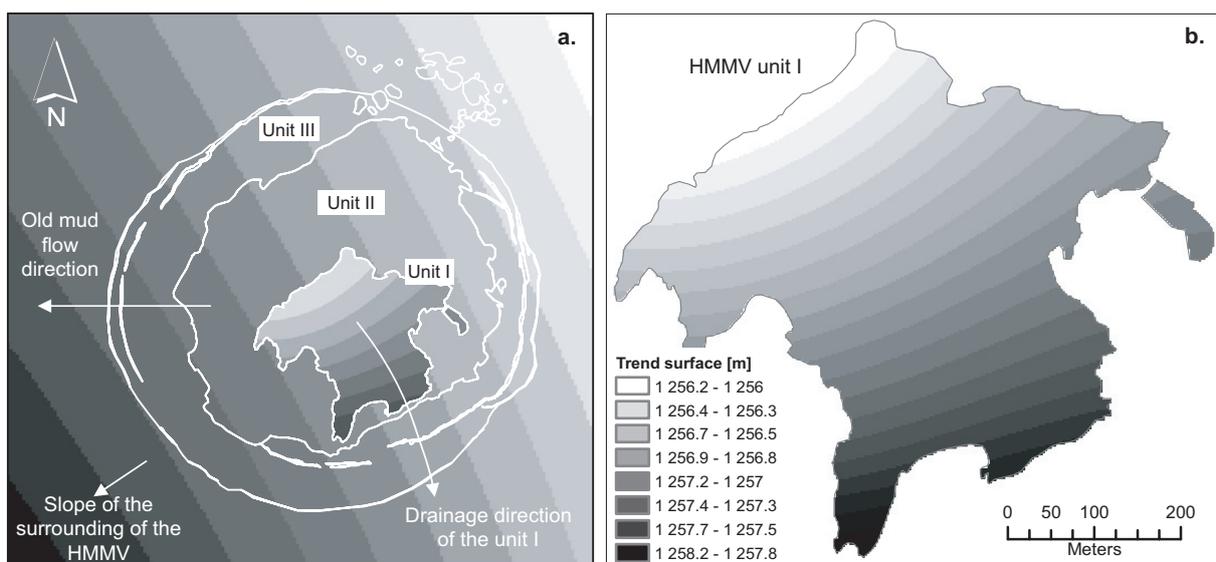
Figure 4-7: Polynomial trend surfaces a. using bathymetric data (*RV Polarstern*) from the HMMV and its surrounding and b. applying microbathymetry data (*Victor6000*) for the HMMV unit I. Composition of both analyses reveals that the slope direction of unit I is perpendicular to the slope of the surrounding of the HMMV.

surrounding, the western flank of the submarine valley, of HMMV and the microbathymetric data measured for unit I (Fig. 4-7). By trend surface analysis (Davis, 2002), small scale undulation can be leveled out to decipher the general trend of sediment layers. This technique, is used in oil and mineral exploration, hydrogeology, or structural geology for identification of strike or slip directions of subsurface strata or drainage patterns. The high resolution microbathymetry data (footprint of 0.5 x 0.5 m) allows calculation of trend surfaces without bias of undersampling and aliasing effects.

The surrounding of the HMMV, located in a submarine valley at the Barents Sea fan, has a general inclination in north-eastern to south-western direction (Fig. 4-7a). We assume that the old mud flow direction was influenced by this inclination.

In contrast, the slope tendency of unit I (Fig. 4-7b) clearly shows a trend from high topography in the north-west towards low topography in the south-eastern and southern part. Thus, at present the drainage direction of the flat centre (NW to SE) is perpendicular to the general inclination direction of the continental slope (Fig. 4-7a). Fresh mud flows, seeping from depths of several hundred metres to few kilometres (Eldholm et al., 1999) in the NW of unit I, would preferably follow the detected trend surface into south-eastern and southern direction (Fig. 4-7a). We suppose that the deposition of former mud flows shaped the morphology and caused the deviation in the direction of the current mud flows.

In the northern part of the central unit I we observed the highest trend surface elevation for unit I, the highest temperature and heat flows (Kaul and Heesemann, 2004), distinct lineations (Fig. 4-2), and the largest spatial distribution of uncovered mud areas. Therefore, we suggest this sub-region of the HMMV as the suggested source location (SSL) of fresh mud flows (Fig. 4-9). The recent flows are likely to have been stopped or re-directed by the interaction with the hummocky periphery (Fig. 4-6d).

Based on the spatial analysis (Figs. 4-6, 4-7 and 4-8b) and the hill shading image we assume a temporal relation existing between the colonisation structures of the chemoautotrophic organisms and the flow path of the fresh mud (Figs. 4-6, 4-7 and 4-8b). Figures 4-8a and 4-8b show a sequence of elevation profiles based on the microbathymetric data and oriented on the trend surface of unit I. Figure 4-8c presents the typical population sequence at the HMMV along the elevation profile 2 and includes the derived biogeochemical habitats and morphologic zones. The SSL is indicated by uncovered or sparcely covered mud, absence of gas hydrates due to high temperatures and highest elevation of unit I. From this site to the south we observe a zone of highest *Beggiatoa* density (>50%), changing into a zone with a bacteria density of 20-50% coverage, again changing into a zone with low bacteria density (0-20%) combined with a high occurrence of pogonophorans, and finally ending at a zone where only tube worms settle the area of the HCR (Fig. 4-8c).

Microbial consumption limits the flux and the release of methane through the sediment water
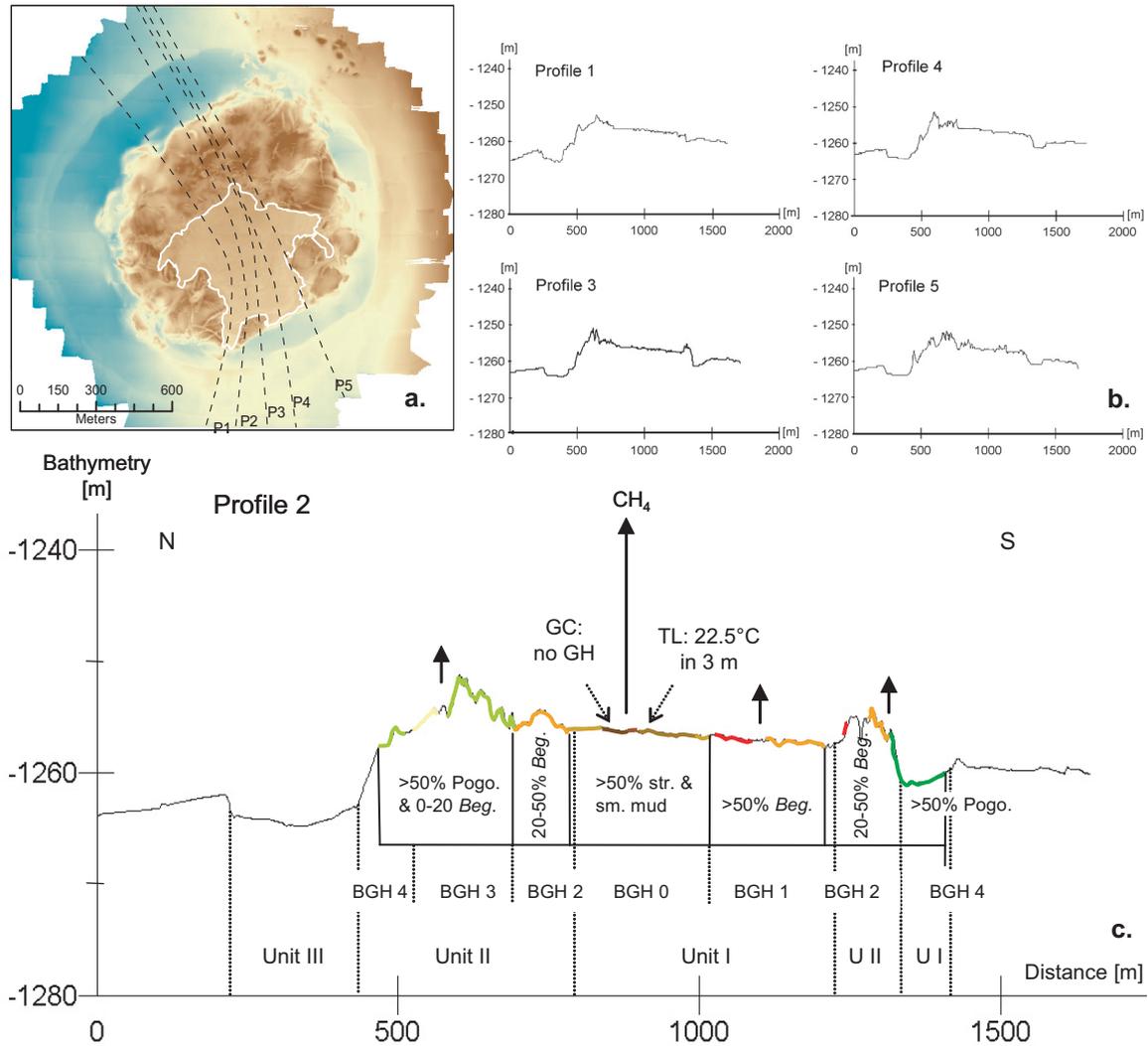
Figure 4-8: Elevation profiles. a. Overview of the orientation of five profiles. b. Elevation profiles P1, P3, P4, and P5 from east to west direction. c. Sequences of biogeochemical habitats (BGH) along the elevation profile 2.

interface. *Beggiatoa* and pogonophorans use primarily hydrogen sulphide resulted from the process of the AOM. If the zone of the AOM is near to the sediment surface, the rate of the sulphide synthesis and the sulphide gradients increases. The highly different vertical sulphide gradients below different communities strongly indicate the important role of sulphide availability in structuring community composition. Thus, high *Beggiatoa* density points on high AOM rate while pogonophorans indicate AOM in smaller extent; the tube worms are able to achieve to sulphide from deeper sediment layers than the *Beggiatoa* (Sahling et al., 2002).

Derived from the occurrence and spatial density of chemoautotrophic communities and uncovered mud areas, we define five biogeochemical habitats (BGH 0-4) represented in figure 4-9.

The central uncovered part, the BGH 0, is still not re-colonisation by *Beggiatoa* or pogonophorans, thus, one of the youngest mud flows (Fig. 4-6a). The zone around the SSL of current mud

flows, where probably no biofiltering AOM-active microbial consortia in the upper sediment layer occur, we expect as the region of the highest CH_4 flux (Fig. 4-9). This BGH 0 represents 15.6% of the HCR and 59.8% of the entire unit I (Fig. 4-6a). Different processes might be the reason why this central region is still uncovered: (i) not enough time has elapsed since the last mud flow event for re-colonisation by *Beggiatoa*; (ii) the chemical composition of pore water is still not suitable (iii) advective processes or sediment dynamics inhibit archaea development (compared to diffusive methane gradients). We exclude high temperatures as the limiting factor only for *Beggiatoa* (A. Boetius, pers. comm.); for pogonophorans a negative correlation
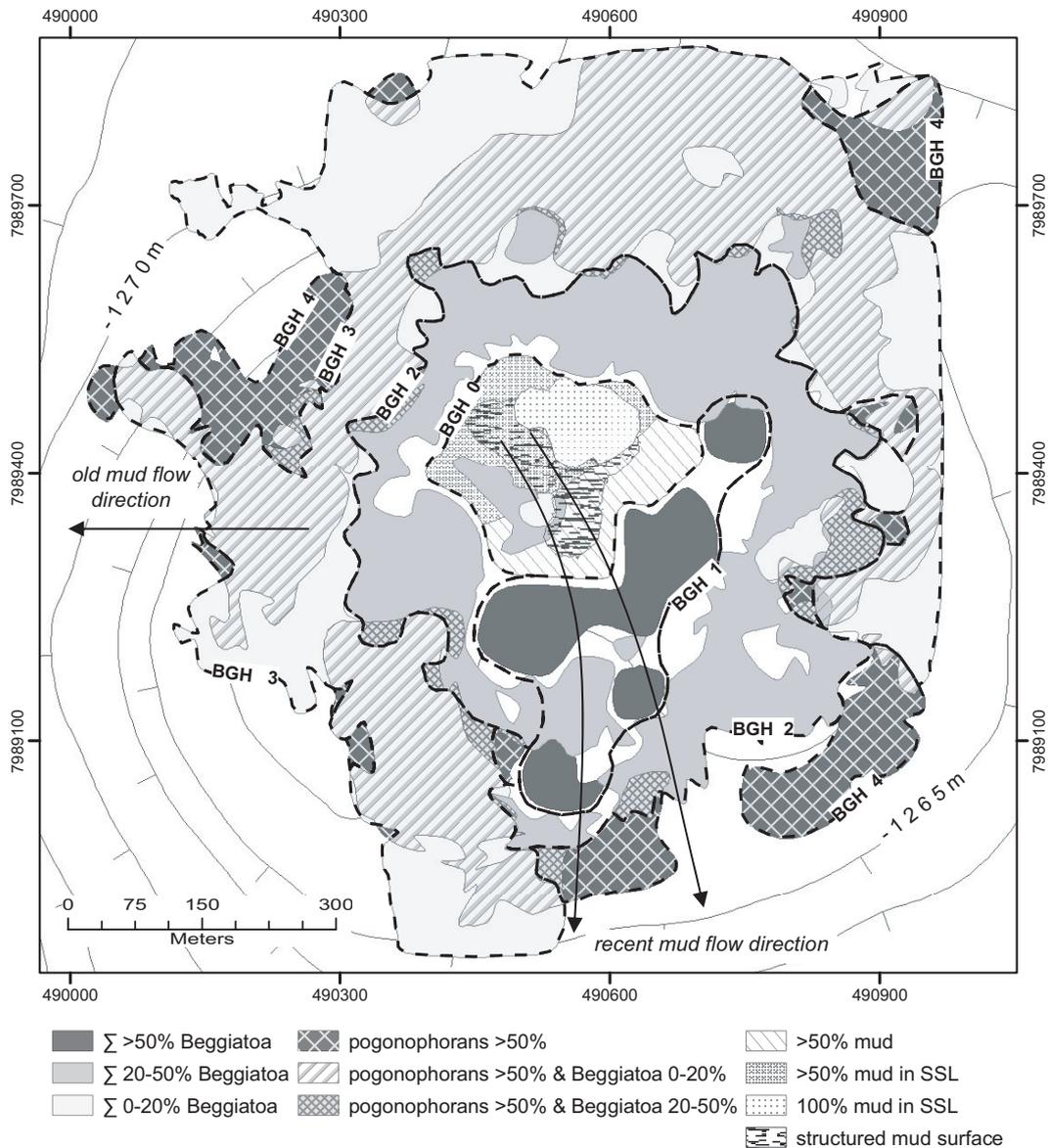
Figure 4-9: Schematic map of the major zones of release or anaerobic oxidation of methane. BGH 0 is the region of highest CH_4 release, BGH 1-4 are regions of bio-filtering communities from high to low CH_4 consumtion and indicated by occurrence of sulphide-oxidising *Beggiatoa* and pogonophorans. The suggested source location (SSL) of fresh mud flows is derived from the trend surface analysis, the locations of highest temperature (Kaul and Heesemann, 2004), and the regions of uncovered mud.

is more probable (Sahling et al., 2005).

BGH 1 is the region of highest population density by *Beggiatoa* on the sediment surface. *Beggiatoa* mats (>50%) cover 5.2% of the HCR or 19.9% of unit I (Fig. 4-6c). Where southern and eastern of the BGH 0 bacteria patches of more than 30 cm in diameter cover areas of up to several square meters, a total absence of pogonophorans is observed. This lobe-shaped area, stretching from the hummocky periphery in the west to the north-east, separates the uncovered mud sequence into a central, western, and southern sub-region (Fig. 4-6d). Since Beggiatoa and their symbionts are tolerant to high concentrations and steep gradients of reduced chemical components as hydrogen sulphide, this spatial distribution suggests that these organisms are pioneering the fresh mud flows.

The BGH 0 and 1, both located within the flat centre, are enclosed by the BGH 2. This BGH 2 is to be found in parts of the flat centre and in parts of the hummocky periphery and is characterised by a high probability of a bacteria density of 20-50%. Pogonophorans or other macrofauna, dominant in regions of lower rates of methane seepage and hydrogen sulphide formation (Levin et al., 2003; Sahling et al., 2002), were not observed for BGH 2.

BGH 3 is characterised by a rough topography (morphological unit II) and a surface elevation of up to 9 m with respect to the centre and 16 m with respect to the moat. Probably, it consists of material of old mud flows and could have been disrupted and pushed aside by the younger mud flows emitted at the centre. This area is mainly covered by pogonophorans (coverage >50%) while the bacteria diminish concentrically towards the periphery (coverage 0-20%) building a large transition zone of *Beggiatoa* and pogonophorans (BGH 3) until the *Beggiatoa* are completely replaced by the tube worms (BGH 4). In the SSW of the HMMV the BGH 3 the otherwise concentric zone is interrupted in the area of the recent mud flow pathway.

The pogonophoran BGH 4 which is excluding colonised by pogonophorans covers the area along the SSW outflow of the flat area into the moat area (Fig. 4-9). Apparently, as a consequence of the direction of the recent mud flow bacteria and tube worms occur with the largest distance to the suggest source location of fresh mud. Beside that, the moat area is apparently free of macro and mega fauna. From this perspective the distribution pattern of chemoautotrophic communities suggests that the last outflow of mud from the HMMV structure occurred through the SE outlet, where no pogonophorans are observed (Fig. 4-6b). In contrast to the SE outlet, today the outlet located in the south is already colonised by these communities.

The temporal pattern of mud re-colonisation by chemoautotrophic communities can be estimated, therefore, from their spatial distribution along the S and SSW direction pathway (Fig. 4-8c): observations of *Beggiatoa* distribution evolving from very localised narrow occurrences (spots) southeast of the uncovered mud area to densely distributed large mats (patches) south and east of the centre (Fig. 4-6d) might indicate the increasing age of the underlying mud flow.

4.6 Conclusions

Since chemoautotrophic communities as *Beggiatoa* and pogonophorans indicate efficient "biofiltering" microbial organisms, limiting the release of methane from sediments to the water column, their spatial distribution (Fig. 4-6) is a suitable detector to distinguish between areas of high versus low methane discharge at the HMMV and other offshore mud volcanoes. The spatial distribution and absolute values -on a m² basis- for the coverage by these chemoautotrophic communities and mud flows provide therefore a basis for budgets of methane oxidation and of methane release into the water column.

Deducing the position and the alignment of the old mud flow as well as the highest elevation of the unit I the SSL seems to remain locally constant. The recent mud flow, originating from further mud discharge, follows the arising new trend direction toward S and SSW. The combination of accurate navigation as USBL, high resolution bathymetry, and georeferenced video mosaics allows deciphering temporal changes as spreading of mud flows or of *Beggiatoa* mats and might be able to present a "baseline study" for further analysis of temporal-spatial changes at extreme environments like mud volcanoes.

Our results represent an extension of the distribution pattern described by Gebruk et al. (2003) and Milkov et al. (1999; 2004). Whereas the previous mapping approaches basically used the presence of either pogonophorans, or *Beggiatoa*, or mud sequences, the co-occurrence of these features (Fig. 4-6d) as well as the quantification of areas and high resolution bathymetry allow a more detailed consideration of biogeochemical habitats and their dynamics. Thus, the GIS techniques applied here did not only allow the identification of the active centre, but also support quantification of areas dominated by bacteria, and transition zones, where we expect subsequently decreasing diffusive methane gradients.

The applied video mosaicing analysis is a progress beyond the classical still photograph-approach due to (i) the larger areas covered in two-dimensional manner and (ii) the availability of geographic coordinates for each pixel and therefore each feature at the seafloor. This offers a number of GIS applications to describe the spatial distribution of any kind of object identified in the mosaics, as shown in this study (visual feature detection attached to a certain geographical location, geostatistical interpolation (revisable e.g. by cross-validation, overlay, trend surface analysis. Geostatistics also provide techniques to identify spatial correlation with other parameters, e.g. to identify the complex colonisation structures which are related to geochemical parameters such as methane and hydrogen sulphide concentration in sediments, occurrence of gas hydrates, seafloor morphology, temperature and the sequence of mud flows. Furthermore, in marine geosciences the image analysis attains more at relevance, due to progressive techniques and amount of data, so automatic feature detection would reduce the labour- and time-intensive visual image inspection.

4.7 Acknowledgements

5 Spatial analysis of marine categorical information using indicator kriging applied to georeferenced video mosaics of the deep-sea Håkon Mosby Mud Volcano

Kerstin Jerosch[1], Michael Schlüter[1], Roland Pesch[2]

[1]Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
[1]Institute for Environmental Science, University of Vechta, Oldenburger Str. 97, 49377 Vechta, Germany

5.1 Abstract

The exact area calculation of irregularly distributed data is in the focus of all territorial geochemical balancing methods or definition of protection zones. Especially in the deep sea environment the interpolation of measurements into surfaces represents an important gain of information, because of cost- and time-intensive data acquisition.

The geostatistical interpolation method indicator kriging therefore is applied for an accurate mapping of the spatial distribution of benthic communities following a categorical classification scheme at the deep-sea submarine Håkon Mosby Mud Volcano. Georeferenced video mosaics were obtained during several dives by the Remotely Operated Vehicle *Victor6000* in a water depth of 1260 m. Mud volcanoes are considered as significant source locations for methane indicated by unique chemoautotrophic communities as *Beggiatoa* mats and pogonophoran tube worms. For the detection and quantification of their spatial distribution 2840 georeferenced video mosaics were analysed by visual inspection. Polygons, digitised on the georeferenced images within a GIS, build the data basis for geostatistically interpolated mono-parametric surface maps. Indicator kriging is applied to the centroids of the polygons calculating surface maps.

The quality assessment of the surface maps is conducted by leave-one-out cross-validation evaluating the fit of the indicator kriging variograms by using statistical mean values. Furthermore, the estimate was evaluated by a validation dataset of the visual inspection of 530 video mosaics not included within the interpolation, thus, proving the interpolated surfaces independently. With regard to both validating mechanisms, we attained satisfying results and we provided each category applied for the identification of biogeochemical habitats with a percentage probability value of occurrence.

5.2 Introduction

In limnology and marine research, environmental and ecological studies are mainly based on datasets obtained at distinct sites (points) or along track lines gathered during cruises by research vessels. Examples for data collection at distinct sites are water samples acquired for chemical analysis of nutrients or pollutants, plankton samples, or geochemical analysis of sediment cores (Fig. 5-1). From a geoinformatical perspective these data are of the type point (x_1, y_1, z_1) or multipoint $(x_1, y_1, z_1...z_n)$. Sampling by bottom trawls or dredges for fishery or petrography are examples for line features $(x_1...x_n, y_1,...y_n, z_1...z_n)$. Only, investigations by multibeam systems (e.g., applied for bathymetric mapping), by side scan sonar, or video
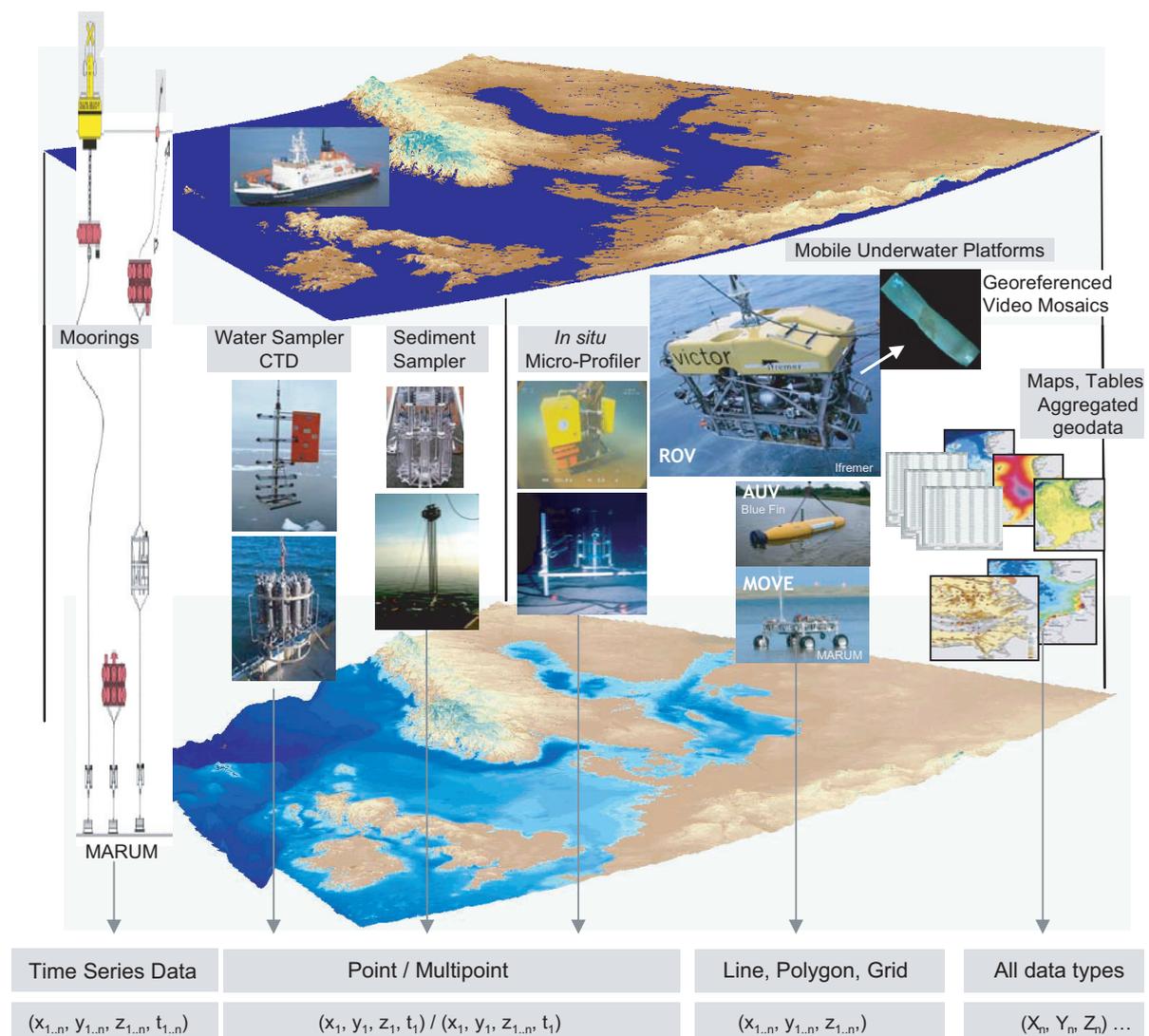
Figure 5-1: Examples for sampling methods and data types used in marine research. New techniques as AUVs, ROVs, Crawlers. or moorings (the two latter from MARUM, Univ. Bremen) provide underwater platforms for in situ analysers, acoustic sensors or video systems and for mapping of the seafloor. AUVs are unmanned, self-propelled vehicles designed to carry out measurements along pre-programmed courses and water depths, generally launched and recovered by a surface vessel. ROVs are connected by a cable to a surface vessel and are usually equipped with manipulators (robot arms) for sampling and experiments at the seafloor.

surveys are able to cover strips (polygons) of the seafloor with a width of a few meters to a few hundred meters. Even such surveys are unable to provide a dense coverage of larger areas of the coastal zone or the ocean due to time and financial restrictions.

New underwater technologies, as "Autonomous Underwater Vehicles" (AUVs), "Remotely Operated Vehicles" (ROVs) or Crawlers (Fig. 5-1), operated by offshore industry and a few research institutes, provides a step towards a refined spatial coverage of sampling sites and multi-parameter mapping of the seafloor. These underwater vehicles serve as platforms for in situ analysers (chemical and biological), for acoustic sensors, and underwater imagery by e.g., high resolution video cameras. A major advantage for multi-parameter mapping by ROVs and AUVs is the very accurate navigation by Ultra Short Base Line (USBL) or inertial navigation during the dives.

Due to the large amount of geodata compiled during cruises and dives by ROVs and AUVs marine environmental and ecological studies requires concepts for data management of large volumes of heterogeneous datasets. Spatial analysis has to consider and to combine different types of geodata (point, line, and polygon features or time series) and scales. This includes metric scales as measurements of temperature or chemical concentrations, relative scales derived by acoustic mapping techniques, or categorical data associated to the occurrence/absence of benthic organisms or geological features. Geostatistical techniques for neighbourhood analysis as well as for computation of spatial distributions and maps are applied.

Especially for mapping of geological or biological data, as occurrence of different rocks or sediment types or of benthic organisms and habitats, visual observations by underwater still photography or video imagery are very promising techniques. For example, video surveys by an ROV provides, in contrast to most towed systems, accurately georeferenced images and video sequences. The later can be converted to georeferenced video mosaics (GVM) by mosaicing algorithms (Allais et al., 2004). Visual inspection of GVMs allows to identify specific morphological, geological and biological features and to classify different sub-regions at the seafloor.

This study is focused to the application of geostatistical techniques and a geographical information system (GIS) to georeferenced video mosaics for computation of spatial distributions of ecological entities at the seafloor. For this purpose video mosaics derived during several dives by the ROV *Victor6000* at the Håkon Mosby Mud Volcano (HMMV) were considered. The HMMV is located in a water depth of 1265 m at the continental margin of the Barents Sea (Northern North Atlantic). The very dense data coverage, compared to the majority of related marine investigations, and the accuracy of navigation data supports the application of variogram analysis and Kriging. At marine mud volcanoes seafloor ecology is characterised by specific chemoautotrophic organisms associate to high methane concentrations derived by mud flows ascending from deeper subsurface strata.

5.3 Area of investigation and geochemical and ecological background

Worldwide more than 1700 mud volcanoes (MV) are reported for onshore and offshore environments (Milkov, 2000). It is estimated that more than 10,000 exist in deep marine waters (Dimitrov, 2002; Fleischer et al., 2001; Kopf, 2002). They are major locations of mud and fluid fow (water, brine, gas, or oil) or erupt from deeper strata to the earth's surface. Submarine MVs are generally confined to shelves, continental and insular slopes and to the abyssal part of inland seas. They are closely linked to high sedimentation rates and tectonic compression, dehydration of clay minerals at depth, or rapid deposition of mass flows. Frequently they form mud domes with diameters of up to a few kilometres and heights of several tens of meters above adjacent seafloor.

MVs are significant sources for the transfer of methane from the geosphere to the hydro- and biosphere. Thus, investigations of MV and cold seeps are in the focus of ongoing geological, biological, and geochemical studies due to their significance as considerable source locations for the transfer of methane and trace elements to the ocean, as indicators for active transport pathways linking deeper geological strata with the surface, and as environments settled by unique chemosynthetic benthic consortia (e.g. Sibuet and Olu, 1998; Pimenov et al. 1999; Sahling, 2002; Levin et al., 2003; De Beer et al., in press). Based on these studies the link between the occurrence of macrobenthos and bacterial mats and geochemical environments,
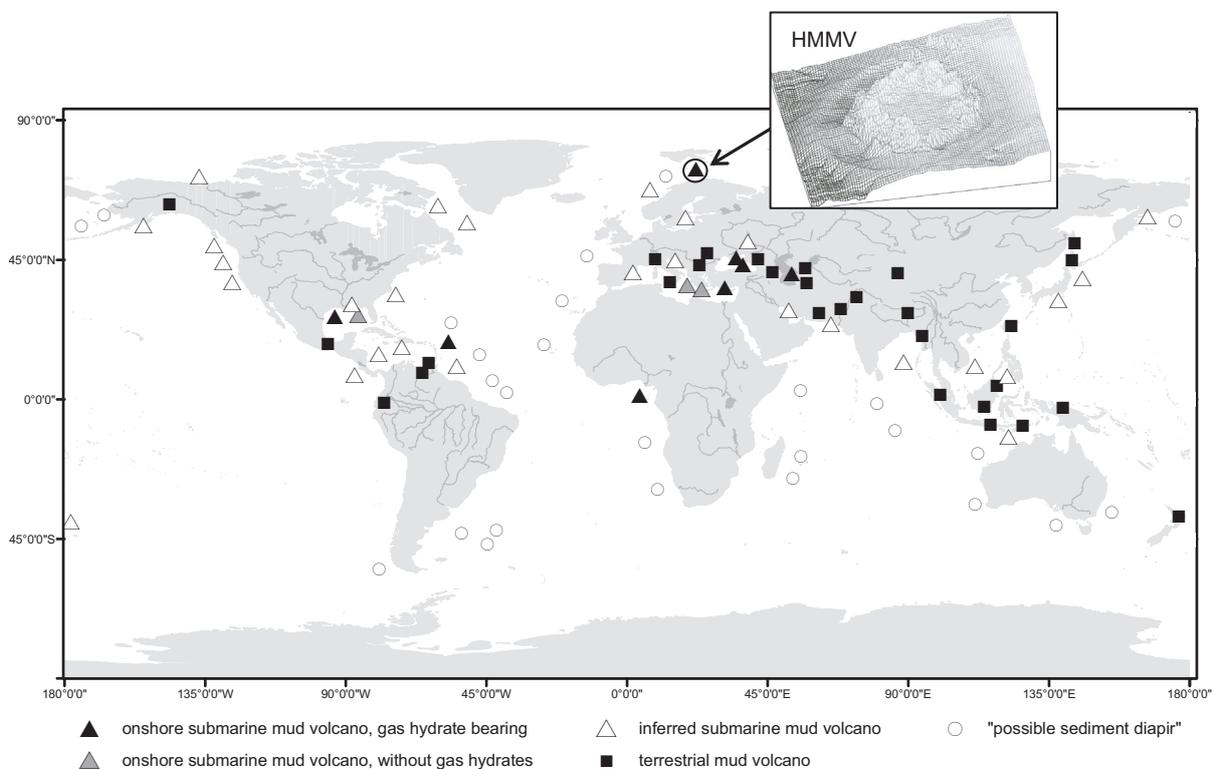
Figure 5-2: Worldwide distribution of onshore and offshore submarine mud volcanoes (modified after Milkov, 2000) and geographical location of the HMMV, the only well-know mud volcano in polar regions.

with respect to consumption and release of methane or the sulphur cycle, is well documented. A microbial symbiosis has been detected (Boetius et al., 2000), which is able to consume methane (anaerobic oxidation of methane, AOM). Sulphide, as one of the products of this reaction, is used as energy source by the chemosynthetic organisms as giant sulphide-oxidising bacteria like genus *Beggiatoa* (Boetius et al., 2004; De Beer et al., in press), and symbiotrophic pogonophoran tube worms *Sclerolinum* sp. and *Oligobrachia* sp. (e.g. Gebruk et al., 2003; Pimenov et al., 1999; Smirnov, 2000). Due to this "biofilter", related to the microbial consumption of methane, only a certain fraction of the methane is transferred through the sediment-water interface. In order to gain a better understanding of the global methane budget it is of relevance to quantify the amount of CH_4 and other greenhouse gas which enters the ocean. The sulphide-oxidising communities *Beggiatoa* and pogonophorans, in the following also referred as chemoautotrophic communities, indicating methane consumption are thus the objects of interest applying geostatistical techniques.

The HMMV is located at the Norwegian-Barents-Svalbard continental margin (~72°00.3'N and 14°44.0'E) (Fig. 5-2), which is characterised by major submarine slides, large-scale mass wasting, gas hydrates and smaller seafloor features (Vogt et al., 1999). It is situated in a submarine valley on the Bear Island submarine fan, a large complex composed of glacial sediments, covering the entire continental slope and reaches a thickness of more than 3 km beneath the mud volcano (Hjelstuen et al., 1999). The HMMV is about 1.4 km in diameter and rises up to 10 m above the seafloor, to water depths of 1250-1266 m (Jerosch et al., chapter 4). It is a concentric morphologic structure with highly gas-saturated sediments. A flat central zone of grey fluid sediments with a high geothermal gradient (Eldholm et al., 1999) is surrounded by an area with bacterial mats. This central region (crater) is surrounded by elevated sediment features (hummocky periphery) populated by pogonophorans (Pimenov et al., 1999).

5.4 Basic principals of interpolation using indicator kriging and estimate quality

Kriging is a geostatistical method that has been successfully applied to several field of research. This method, termed by Matheron (1963) and designated after the mining engineer D.G. Krige, was developed and is applied to compute maps from regularly and irregularly distributed data. Since then different types of kriging as ordinary, block or indicator kriging were developed, which allows considering trends or categorical data types (Isaaks and Srivastava, 1989; Davis, 2002; Olea, 1999).

All kriging types depend on mathematical and statistical models and predict values at non-sampled sites. The addition of a statistical model that includes probability separates kriging methods from deterministic interpolation methods. The Inversed Distance Weight (IDW) method for example uses a simple algorithm based only on the distances between sampling points. Kriging uses a semivariogram model developed by analysing the spatial structure of the data.

Kriging procedures transform data measured or observed at discrete sites to continuous distributions maps. Basically the procedure consists of four steps: 1. statistical data exploration, 2. variogram analyses, 3. kriging and 4. evaluation of results.

In a first step the data set is examined with regard to outliers, type of statistical distribution and trends. For identification of outliers, descriptive statistics and statistical tests as well as expert knowledge may be applied. By variogram analyses distance and direction-controlled spatial autocorrelation structures can be examined and modelled. Anisotropies or direction structures may be identified using variogram maps and can be integrated into the interpolation process. Different variogram models using different settings in terms of the range, the sill and the nugget effect have to be compared to each other by help of cross-validation (see below). The range equals the maximum separation distance within which a distinct increase of semivariogram values can be observed. This is indicative of spatial autocorrelation. The sill corresponds to the semivariance assigned to the range. If anisotropies can be detected, both sill and range can vary with respect to direction. Small-scale variability or measurement errors may lead to high semivariances at nearby locations. The variogram model refers to this in terms of the so-called nugget effect. A pure nugget effect indicates a complete lack of spatial autocorrelation. Further mathematical details are described e.g. by Atkin and Siemens, (1988), Isaaks and Srivastava, (1989), Olea (1999) or Webster and Oliver (2000).

In this study cross-validation (a procedure where measured values are removed from the entire dataset and to this reduced dataset variogram analysis and kriging is applied to compute estimates for sites were values were removed) was applied to improve and to control the quality of the applied geostatistical model and thus the results of the spatial analysis. The differences between measured and predicted values, summarised in terms of percentages and statistical mean values, provide a quality control for the model of computation for the habitat map covering the entire HMMV structure.

Characteristic key values were derived from the difference between measured and estimated values obtained by cross-validation: mean error, root mean square of standardised error, and the median of percentage error. The mean error (ME) is computed from the average deviations between measured and estimated values and refers to over- or underestimations of the variogram models. The ME ideally tends to 0. The root mean square of standardised error (RMSSE) corresponds to the average relation between the squared deviation of the measured and estimated values at the places x_i and the minimum estimation variances (kriging variances) computed there. This value thus reflects the relationship between experimental and theoretical variances. If the RMSSE value is higher than 1, the kriging variances are underestimated; if it is lower than 1, they are overestimated. In optimal case the RMSSE value is 1 (Johnston, 2001). The RMSSE signifies the level of the relative deviations between measuring and estimated values. If the respective measured value is defined as 100%, the

difference between measured and estimated value can be indicated in percent by the median of percentage error (MPE). A low MPE generally indicates a low deviation between measured and predicted values and therefore a reliable estimation model.

Based on the results of the data exploration and the variogram analyses, the measured data are converted to surface maps. The suitable kriging method depends on the data type, the data distribution, the data properties, and the scientific question. Universal kriging for example is to be applied if comprehensible deterministic trends can be detected.

Spatial generalisation of geological or biological information (e.g. Rivoirard et al., 2000), were categorical data as the occurrence/absence of sediment types, of benthic epifauna, or of chemoautotrophic communities is studied, is often performed by indicator kriging (IK). IK is derived from ordinary kriging. In case of the occurrence of the certain species or sediment type the variable 1, and in case of its absence the variable 0 is assigned to the measuring location. Categorical data are thus converted into binary data. With indicator kriging spatial probability maps are computed for the occurrence of the respective species or sediment type. The interpolation results can also be interpreted as the percentage probabilities; for example: "with a probability of 85% muddy sediment will occur".

5.5 Material and methical realisation

5.5.1 Data acquisition: Georeferenced video mosaicing

During the *RV Polarstern* expedition ArcXIX3b (Klages et al., 2004) a video system was attached at the bottom of the ROV *Victor6000* (Ifremer) (Fig. 5-1). During the dives the position of the ROV was determined by highly accurate Ultra Short Base Line navigation system. The ROV was operated at a maximum altitude of 3 m above seafloor (in a water depth of 1255 m to 1265 m) with a speed of 0.3 m·s^{-1}. The video system had an optical aperture of 60°. At an altitude of 3 m above seafloor the video observations by the ROV has a width of 3 m, therefore. This ensured a high image quality and allowed real-time video mosaicing of the seafloor by the software MATISSE (Mosaicing Advanced Technologies Integrated in a Single Software Environment; Ifremer). During the mosaicing process, one mosaic arises each half a minute merging about 500 images recorded by the video camera. Each mosaic covers an area of about 3 m of width and 6-7 m of length.

In total about 24 km of video mosaicing were recorded during six ROV dives (35 hours dive time). MATISSE (Allais et al., 2004; Vincent et al., 2003) produces online digital merging georeferenced mosaics (geotiffs) using the video input and the high precision underwater navigation. A small displacement between two consecutive images is required by the video mosaicing algorithms based either on feature tracking (Shi and Tomasi, 1994) or robust optic flow methods (Odobez and Bouthémy, 1994). The georeferenced mosaics are visualised and analysed within a Geographical Information System (GIS).

A total amount of 2840 mosaics was visually inspected; the different biogeochemical habitats were identified and mapped as polygons on the video mosaics with the GIS. Variogram analysis and kriging was applied to a so-called working dataset consisting of 2310 images. The remaining 530 mosaics were used for validation purpose which allows comparing the estimates derived by the geostatistical analysis with an additional independent dataset.

5.5.2 Data analysis

The work flow for the development of three mono-parametric maps (MPM) from image data which form the basis for a detailed habitat map of the HMMV is represented in figure 5-3. As a first step the raster data consisting of 2840 georeferenced video mosaics are imported into the GIS and areas of *Beggiatoa*, pogonophorans and uncovered mud are digitised as polygons (regionalisation). To each polygon the attribute (category) according to the classification scheme (see below) was assigned. For a transformation of polygons to point data the centroids coordinates of the polygons and the assigned attributes are used for the interpolation by means of indicator kriging. By these means one raster map was generated for each classification category. Finally, a mono-parametric map is created with help of GIS techniques, by a combination of all category maps of one parameter (e.g. *Beggiatoa*). A further overlay of the mono-parametric maps (*Beggiatoa*, pogonophorans and uncovered mud) can result in a detailed predictive multi-parametric habitat map for the HMMV. The MPMs are used for an overlay to a multi-parametric habitat map, and area computations are conducted with respect to the source location of mud flows and pattern of biogeochemical habitats allowing conclusions on the biogeochemistry of the HMMV (Jerosch et al., chapter 4).
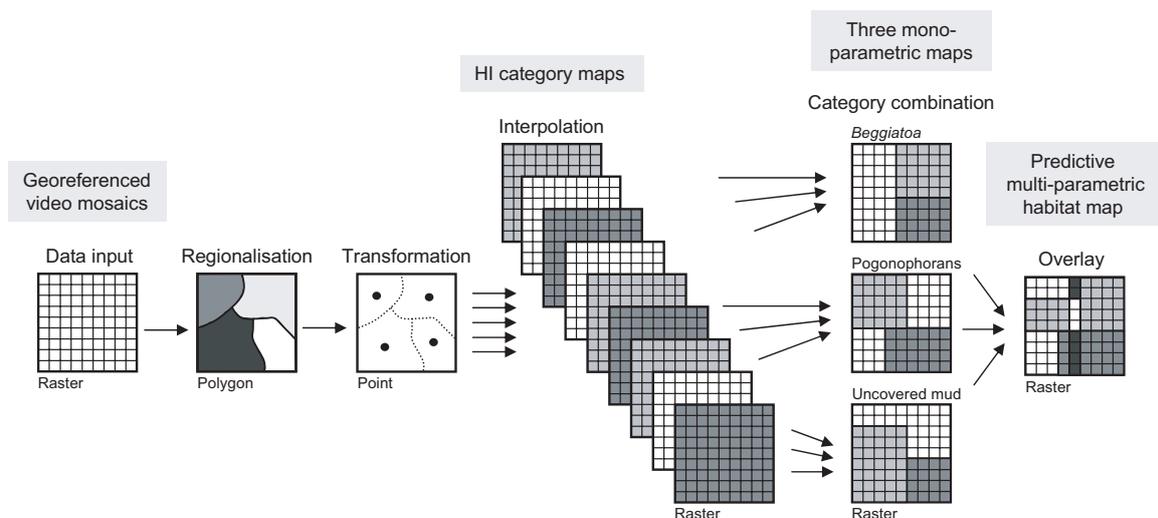
Figure 5-3: Schematic description of the GIS-based work flow.

5.5.3 Habitat indicator coding and GIS implementation

By the video surveys a significant part of HMMV was covered which allowed to identify and map sedimentological and biological features at the seafloor and the distribution of different biogeochemical habitats as *Beggiatoa*, pogonophorans or mud flows.

Basically, the HMMV can be grouped into three major spatial entities according to the habitat indicators: (i) uncovered mud of either a very young mud flow, not yet inhabited by *Beggiatoa*, or old mud flow; (ii) occurrences of *Beggiatoa* which indicates intense methane consumption (anaerobe oxidation of methane - AOM) taking place beneath the seafloor; (iii) pogonophorans indicating a lower CH_4 consumption and older mud flow. For the uncovered mud highest methane release into the bottom water is expected due to the missing "biofilter" by indicating *Beggiatoa* or pogonophoran population.
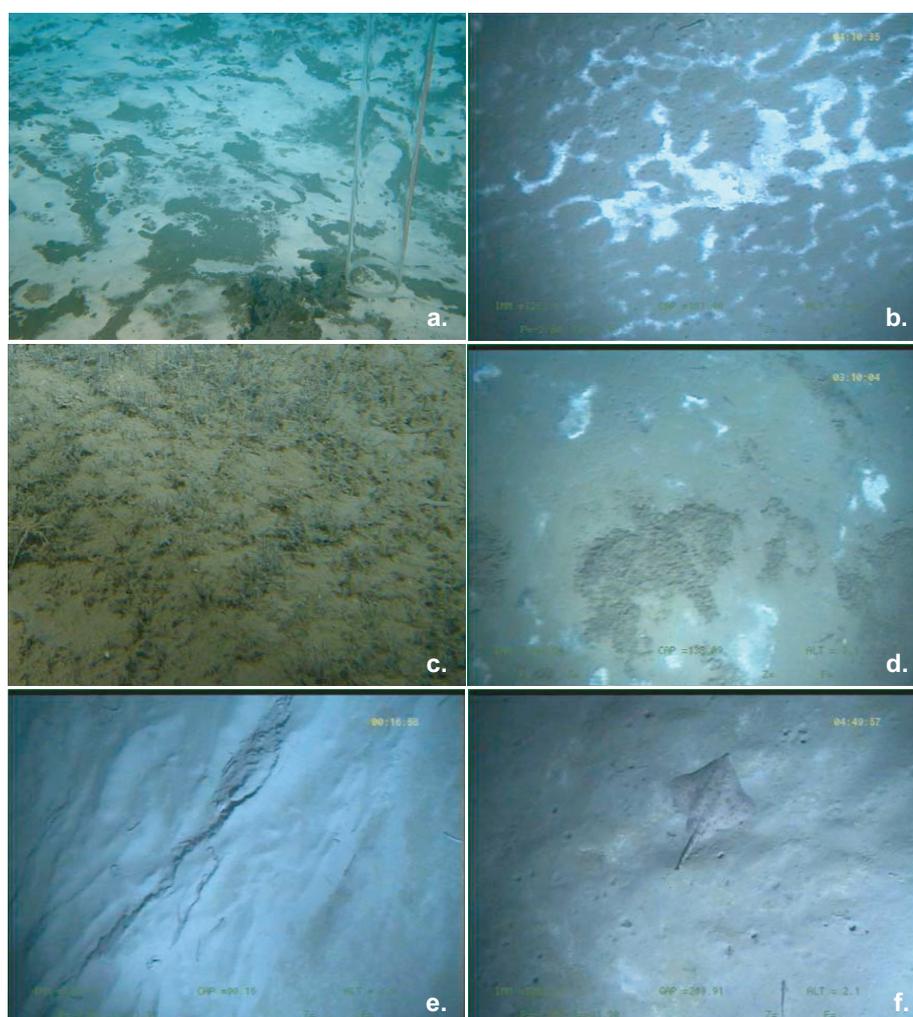
Figure 5-4: Exampled for different habitat categories at the HMMV: a. *Beggiatoa* patches >50%; b. *Beggiatoa* patches 20-50%; c. pogonophorans >50%; d. *Beggiatoa* spots 0-20% and pogonophorans 0-20%, e. 100% structured mud, and f. >50% smooth mud and 0-20% *Beggiatoa* spots (please notice: small holes in the mud, fish, and ray). Still photographs are recorded by *Victor6000*.

Table 5-1: Categories applied for mapping of biogeochemical habitats of the HMMV.

Beggiatoa 0-20%	Pogonophorans 0-20%	Mud smooth >50%
Beggiatoa 20-50%	Pogonophorans 20-50%	Mud structured >50%
Beggiatoa spots 20-50%	Pogonophorans >50%	Mud smooth 100%
Beggiatoa patches 20-50%		Mud structured 100%
Beggiatoa spots >50%		
Beggiatoa patches >50%		

These three parameters determine the biogeochemical habitats at the HMMV. In addition *Beggiatoa* and pogonophorans occur together and form transition zones covered by e.g. *Beggiatoa* (coverage of 0-20%) and pogonophorans (coverage of >50%). The attribution for example assigning only the category pogonophorans 20-50% to an area, signifies automatically a 50-80% uncovered mud fraction. Non-colonised mud, *Beggiatoa*, and pogonophorans (examples are shown in figure 5-4) were classified based on their degree of coverage (Tab. 5-1).

The categories given in table 5-1 are coded binary related to their occurrence in the dataset and surface maps are calculated by means of IK for all 13 categories. The combination of the category maps of one parameter (e.g. *Beggiatoa*) yields into a mono-parametric map. The MPMs in general are the basis for multi-parametric overlay maps. Two or more spatially overlapping input data layers of geometrical type polygon, line or point will be overlaid geometrically producing a new multi-parametric data layer (Fig. 5-5). A GIS-based overlay of all MPMs would reveal e.g. in a transition zone of *Beggiatoa* and pogonophorans and thus in a complex view of the habitat distribution of the HMMV.

About 24 km (17 km working dataset and 7 km validation dataset) on 2840 georeferenced video mosaics were analysed by visual inspection. The analysis of the images was done by measuring the extent of specific items as spots (<30 cm) and patches (>30 cm) of bacteria, pogonophorans or mud within the GIS. Areas are digitised as polygons (Fig. 5-6) with respect
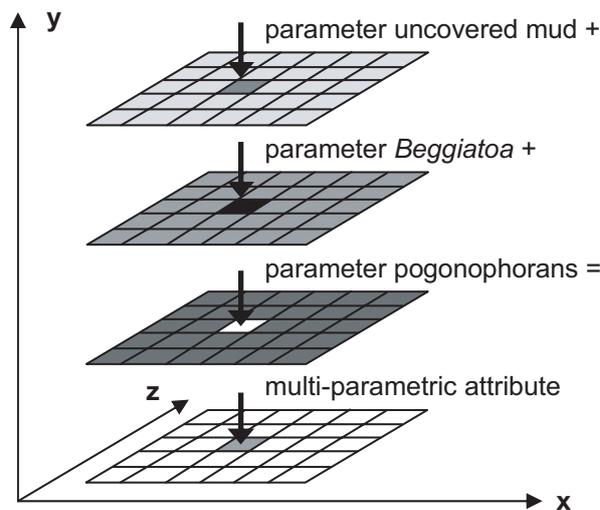
Figure 5-5: GIS-based overlay technique producing multi-parametric surfaces.

GIS data transformation
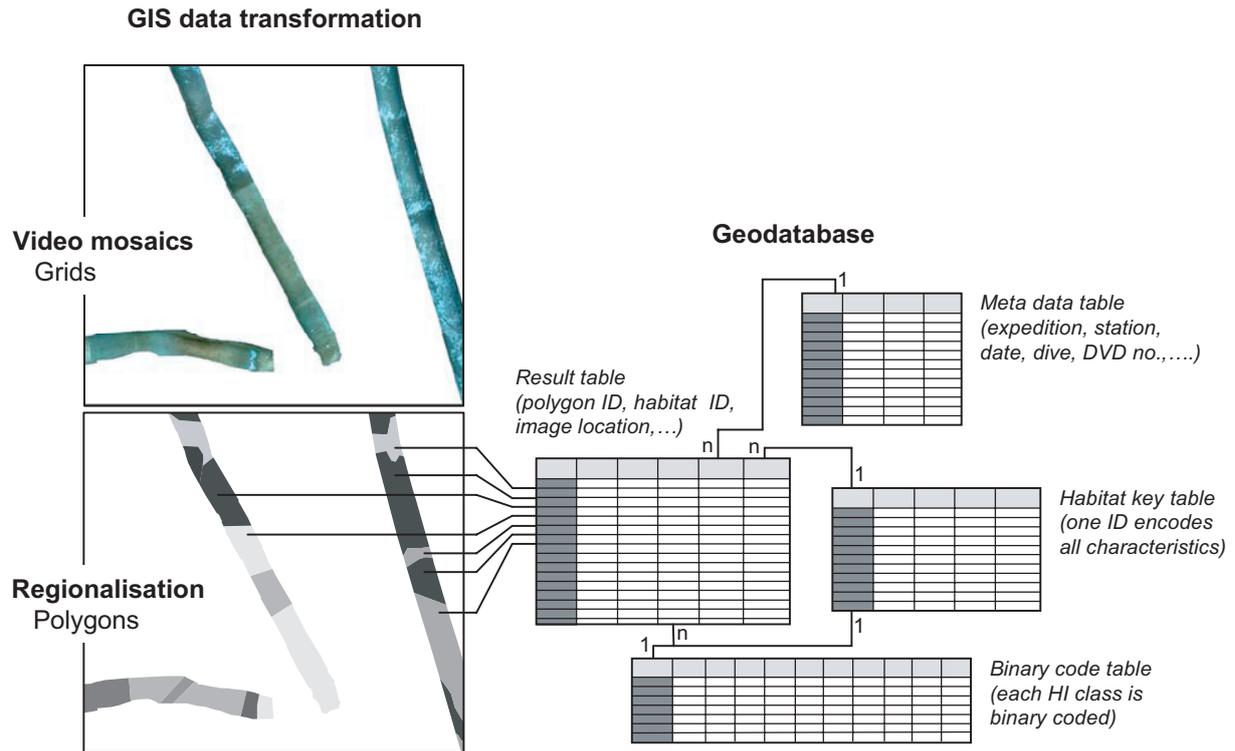
Figure 5-6: GIS based data simplification and data management.

to the defined categories, and are provided with keywords concerning their habitat indicating characteristics and metadata. For the indicator kriging procedure each category is to be coded binary regarding to its occurrence (yes = 1 or no = 0). All the information was stored within a relational geodatabase system (Fig. 5-6).

5.6 Results

The results of this study are presented in two steps: first the presentation of the resulting working dataset, the generated MPMs including variogram analysis and the visual control of the interpolated surfaces using the working datasets. The second step consists of a validation of the quality of the MPMs which is performed by cross-validation of the indicator kriging models and visual inspection using a validation dataset of further analysed video mosaics.

5.6.1 Predictive mono-parametric habitat maps at the HMMV

The working dataset is made up by 2310 of the 2840 visually inspected video mosaics. They are analysed with regard to the coverage degree of the parameters *Beggiatoa*, pogonophorans and uncovered mud. The resulting dataset consisting of 1578 polygons is used for the compilation of the three mono-parametric maps within GIS. The working dataset corresponds to a total area of investigation (TA) of 45,790 m^2 (15.3 km ROV transect x 3 m image width). The distribution of the data assigned to the respective parameter categories is given in figure 5-7.
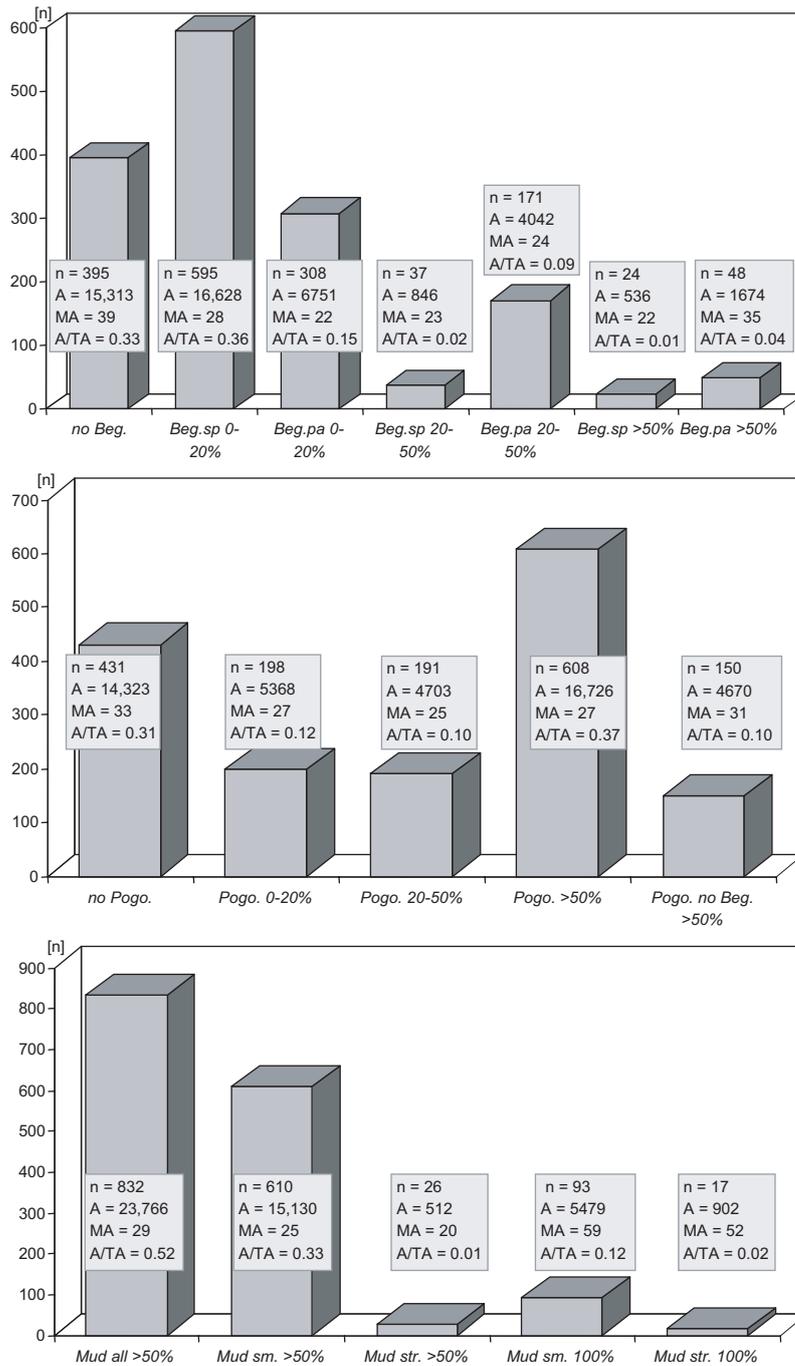
Figure 5-7: Data distribution of the parameters *Beggia*toa, pogonophorans and uncovered mud corresponding to polygons of the working datasets (Fig. 5-8d). The number of polygons (n), the area (A) which is covered by the parameter, the averaged area of one polygon (MA), and the ratio areas of coverage to the total area of investigation (A/TA) with regard to the parameter. Areas are given in m^2.

For example, the category *Beggiatoa spots 0-20%* is represented most frequently (number of polygons (n) is 595) covering an area of 6751 m^2 (A). The average area (MA) of one polygon of this category covers an average area of 28 m^2 (*Beggiatoa* spots, defined as <30 cm, are clustered into areas). Furthermore, this category is assigned to 36% (A/TA = 0.36) of the total

area of investigation. Including the categories "no *Beggiatoa*" (33%) and *Beggiatoa* patches 0-20% (15%) these three categories represent the majority (84%) of the analysed surface. The spatial distribution of the working dataset is shown in figure 5-8. The polygons, representing the video mosaicing tracks as shown in figure 5-6, there appear as lines with respect to the scale of the map.

According to Journel and Huijbregts (1978) the variogram calculations were performed exclusively for distances below half of the maximum horizontal extension of the area under investigation (here ~ 0.7 km). Cross-validation was used as the method to select the optimum variogram model: it was intended to minimise MPE and concomitantly approximate ME toward 0 and RMSSE toward 1. For the surface calculations the mean distance of each observation site to its nearest neighbour was used as the cell (lag) size. The number of lags allocated to each lag size was adjusted to ensure that significant autocorrelation (range) became clearly visible in the variogram window. If the semivariances displayed on the variogram map indicated anisotropies in the data field, different ranges for different directions (to account for anisotropies) were compared with each other.

For the kriging calculations the raster was set according to the average mean distance of each measurement site in relation to its nearest neighbour. The searching or variogram window within measured values were included to estimate a certain point was adjusted to the range of the variogram model. A four-sector neighbourhood was defined to avoid directional bias.

After the IK interpolation the combination of the category maps results in three MPMs (Fig. 5-3). The MPMs contain the predictive surfaces at the HMMV that are covered by the parameter, respectively (Fig. 5-8 a-c). Each MPM is colour coded distinctively according to the different categories (e.g. yellow: *Beggiatoa* 0-20% coverage, orange: *Beggiatoa* 20-50% coverage and red: *Beggiatoa* 50% coverage). The degrees of occurrence probability is then gradually arranged (light yellow corresponds to a smaller probability than dark yellow, etc.). Blank areas do not imply no occurrence of the parameter, but a probability occurrence less than e.g. 70%, respectively to the minimum thresholds given in table 5-2. No occurrence areas are indicated apart.

Visual control of the interpolated surfaces can be performed by overlaying the polygons of the working dataset with the raster maps. For this purpose, the working dataset is coloured equally to the parameter categories (Fig. 5-8 a-c). If no significant differences can be recognised between the appearing lines and the interpolated surfaces, the variogram model is well fitted to the data (encircled area in figure 5-8a). At other sites - apparently at all MPMs - a visual impression of underestimate arises, and it seems that blank regions should have been assigned to a parameter category. This impression results from the ambitious defined minimum threshold of occurrence probability (Tab. 5-2). Blank area are thus not assigned to no category but to a probability of occurrence less than a of 70%, for example, for the mud classes (note legends in figure 5-8).

The IK interpolation produces probability values for the occurrence of a feature between 0
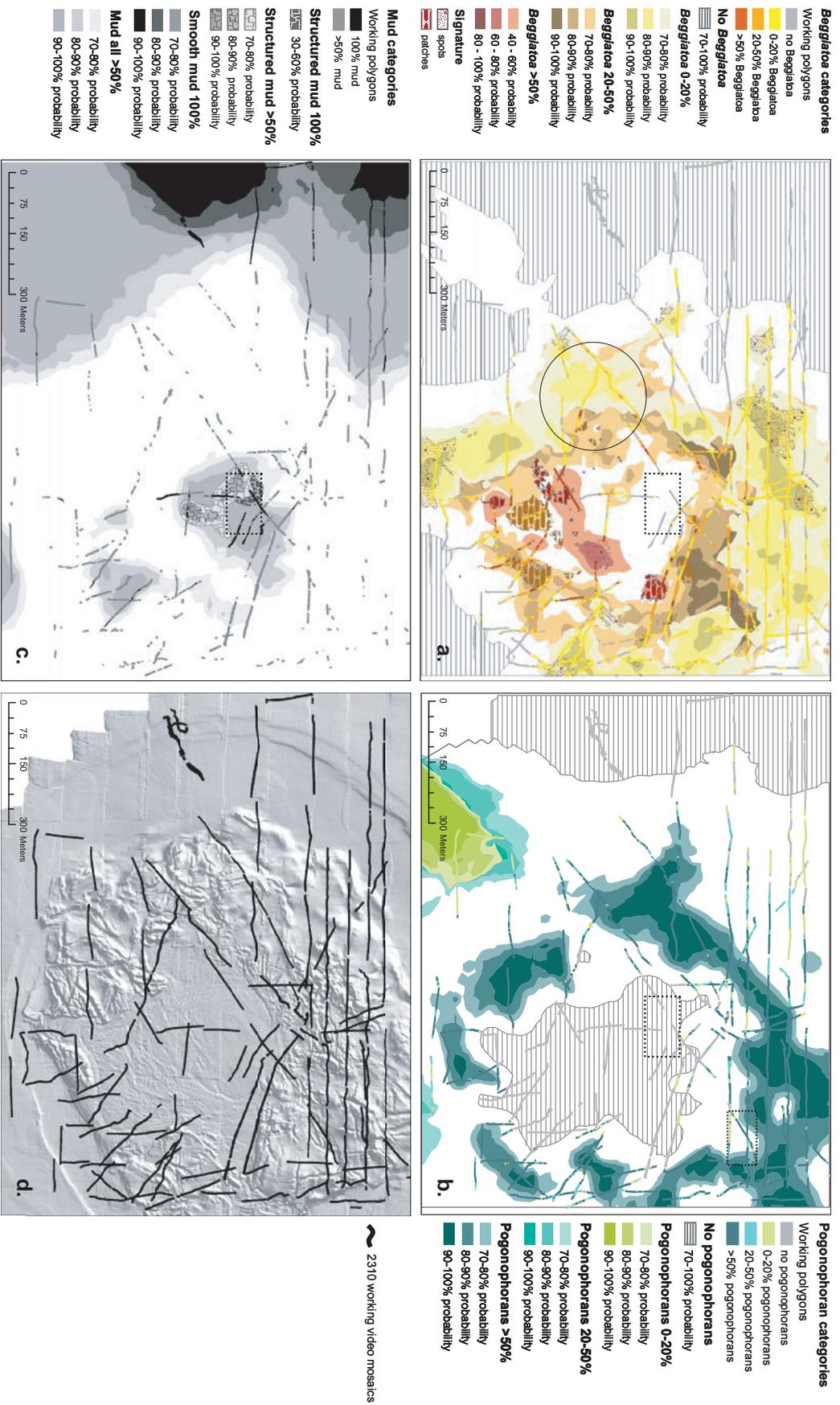
Figure 5-8 a-d: Distribution of analysed polygons and predictive surfaces after indicator kriging: a. *Beggiatoa*, b. pogonophorans, and c. uncovered mud. Colours are arranged gradually concerning their probability of occurrence (Tab. 2). Polygons appear as lines with respect to the scale of the map. If no significant differences can be recognised between the appearing lines and the interpolated surfaces, the variogram model fits well. Blank areas indicate a probability of occurrence less than the lowest probability value given in the legend. d. Data distribution of the entire working dataset. The encircled area provides an example for a reliable fit, whereas within in the rectangles the correspondence between observed and predicted categories is weaker (for details see text).

Table 5-2: Absolute and relative values of the interpolated surfaces at the HMMV. The mean probability of occurrence of each parameter category is calculated by the GIS. The minimum and maximum thresholds are user defined and correspond to the areas shown in the figures 5-8 a-c. The deviating total areas are caused by the individual category variograms used for the interpolation.

HI coverage degree	Interpolated areas [m²]	Percent of coverage	Probability of occurrence in %		
			User defined		*GIS calculated*
			min	*max*	*mean*
Pogonophorans no	297746	24.49	70	100	92
Pogonophorans 0-20%	36649	3.01	70	100	90
Pogonophorans 20-50%	43295	3.56	70	100	84
Pogonophorans >50%	276121	22.71	70	100	93
Beggiatoa not predicted	304424	26.68			
Beggiatoa no	318745	27.93	70	100	87
Beggiatoa 0-20%	250200	21.92	70	100	80
Beggiatoa 20-50%	163077	14.29	70	100	83
Beggiatoa spots 20-50%	43815	3.84	70	96	76
Beggiatoa patches 20-50%	22699	1.99	70	100	76
Beggiatoa >50%	29078	2.55	40	86	69
Beggiatoa spots >50%	2357	0.21	40	50	41
Beggiatoa patches >50%	6809	0.60	40	100	44
Mud all >50%	689563	58.39	70	100	96
Mud smooth >50%	276917	17.02	70	100	85
Mud structured >50%	12256	23.45	70	100	92
Mud smooth 100%	200974	0.10	40	45	41
Mud structured 100%	1182	1.04	40	99	76

(The left-most label column groups: Pogon. / *Beggiatoa* / Mud)

and 1. They were transformed into percentages (Tab. 5-2). Both surfaces with high probability of occurrence as well as surfaces with high probability of exclusion of the parameter category can be assigned. Minimum and maximum threshold values are defined manually oriented on the calculated mean probability values. The average value gives the overview of the general estimation quality; the minimum and maximum threshold values are important because they correspond directly with the areas and their graduated colours in the MPM (Fig. 5-8 a-c). Table 5-2 contains the three values associated to each parameter category and gives, therefore, information about the interpolated surfaces.

In general reasonable results in nearly all analysed parameter categories were achieved, except in the categories of „*Beggiatoa* >50%" and "smooth mud 100%". Here the results contain a certain degree of uncertainty.

5.6.2 Validation of the results applying cross-validation and a validation dataset

When performing validation, two datasets are used: a working and a validation dataset. The working dataset contains the measured locations on which the variogram analysis applied for prediction is based. Calculated statistics, resulting from leave-one-out cross-validation, serve as diagnostics that indicate whether the model and its associated parameter value are reasonable.

The validation dataset is used to prove the interpolated predictions independently from the variogram analysis by the results of the visual inspection of further 530 video mosaics.

5.6.2.1 Evaluation of the prediction models (variograms) derived from cross-validation

The so-called leave-one-out cross-validation involves subtracting a single observation from the working dataset and calculating this single observation from the remaining observations. This is repeated such that each observation in the working dataset is used once as the single observation. Such a cross-validation was applied for each of the 13 binary coded categories (Tab. 5-1) predicting each of the polygons of the working dataset by using the specific variograms. This allows comparison of results derived by the reduced dataset with the results derived by the complete working dataset. From the cross-validation errors the percentage conformance (and/or deviation) were expressed as single values for each parameter category; their median value is defined as the median percentage error (MPE). Furthermore, three averaged key parameters are calculated from the distribution to characterise the quality of the chosen variogram models (Tab. 5-3).

The graphs in figures 5-9 a-c show the percentage error of the single values for selected categories grouped by parameter. These single values result from a full cross-validation, predicting each of the 1578 polygons (x-axis) by the help of the reduced dataset and the percentage differences between measured and predicted values. The percentage of conformance is given then on the y-axis. For instance, 1361 of the 1578 polygons of the category *Beggiatoa spots >50%* (blue line) are forecast accurately (100%). The graph intersects the ordinate at an x-value of 1545 polygons, signifying a conformance of these polygons of 80% with the variogram model.

Table 5-3: Quality of estimation by means of statistical average values resulting from cross-validation.

Parameter	ME * 10^{-3}	RMSSE	MPE
Beggiatoa no	-1.588	1.009	16.12
Beggiatoa 0-20%	-0.354	1.004	36.52
Beggiatoa 20-50%	0.184	0.998	8.02
Beggiatoa spots >50%	0.080	1.090	2.11
Beggiatoa patches >50%	-0.014	1.047	0.00
Pogonophorans no	0.145	0.990	4.85
Pogonophorans 0-20%	-2.880	1.001	9.56
Pogonophorans 20-50%	-5.836	0.997	30.34
Pogonophorans >50%	0.618	1.008	30.17
Mud all	-1.681	1.038	31.16
Mud smooth >50%	-2.044	1.002	35.85
Mud structured >50%	-0.572	0.986	0.00
Mud smooth 100%	0.029	0.989	0.00
Mud structured 100%	0.074	0.973	0.00
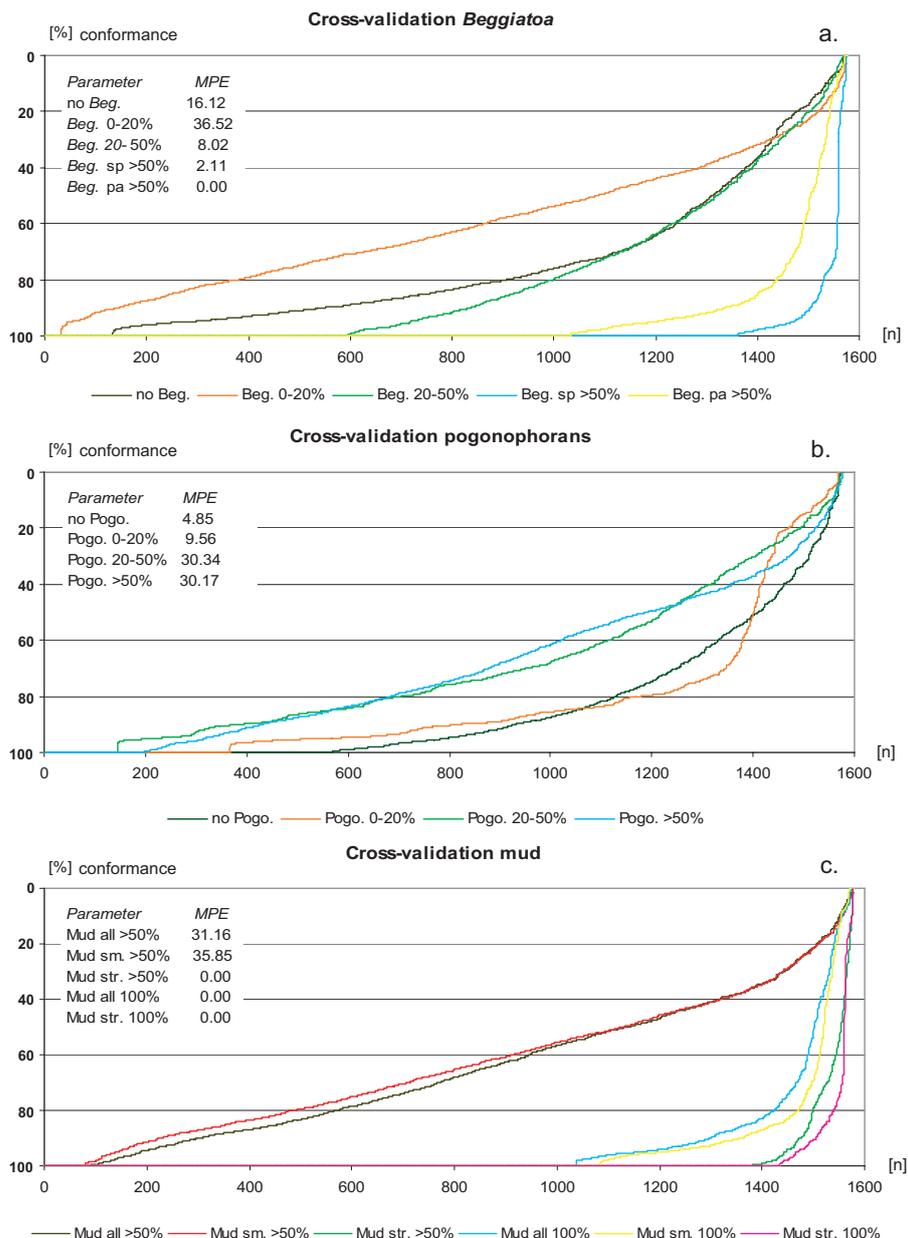
Figure 5-9 a-c: Results computed by cross-validation: visualisation of percentage errors and median percentage errors (MPE). The abscissa gives number of polygons and the ordinate the conformance between measured and predicted values after a full cross-validation of 1578 polygons (note the scale of the ordinate is inverted).

Considering graphs in figures 5-9 a-c the variograms fit particularly well with the categories *Beggiatoa >50%* and *mud 100%* coverage. In these categories the majority of the polygons (more than 1000 of 1578) coincide to 100% comparing examined and predicted values: 1361 polygons of the category *Beggiatoa spots >50%* are predicted precisely, as well as 1033 polygons of *Beggiatoa patches >50%*, 1030 polygons of *mud all 100%*, 1085 polygons of *smooth mud 100%*, and 1435 polygons of *structured mud 100%*. For the category *Beggiatoa spots >50%* further 71 polygons are predicted with up to 90% conformance, i.e.

Beggiatoa categories

Validation polygons

no Beggiatoa

0-20% Beggiatoa

20-50% Beggiatoa

>50% Beggiatoa

No Beggiatoa

70-100% probability

Beggiatoa 20-50%

70-100% probability

Beggiatoa 0-20%

70-100% probability

Beggiatoa >50%

40-90% probability

Pogonophoran categories

Validation polygons

no pogonophorans

0-20% pogonophorans

20-50% pogonophorans

>50% pogonophorans

No pogonophorans

70-100% probability

Pogonophorans 0-20%

70-100% probability

Pogonophorans 20-50%

70-100% probability

Pogonophorans >50%

70-100% probability

Mud categories

Validation polygons

100% mud

>50% mud

Structured mud 100%

30-60% probability

Structured mud >50%

70-80% probability

80-90% probability

90-100% probability

Smooth mud 100%

70-80% probability

80-90% probability

90-100% probability

Mud all >50%

70-80% probability

80-90% probability

90-100% probability

530 validation video mosaics

2310 working video mosaics

Figure 5-10 a-d: Validation dataset combined with simplified mono-parametric maps. a. *Beggiatoa*, b. pogonophorans and c. uncovered mud. d. distribution of the entire validation dataset. Significant deviations of the validation polygons to the MPMs are highlighted by rectangles.

1432 of 1578 polygons (90.8%) are well determined (Fig. 5-9a). Furthermore, the quality of the tube worm estimation model appears visibly in a reliable way for the categories "no occurrence of pogonophorans" and "pogonophoran coverage 0-20%", while the other two categories (20-50% and 50%) are represented rather moderately (Fig. 5-9b).

The performed averaged key values are summarised in table 5-3, which demonstrates among others the ME, RMSSE and the MPE values. Both the ME and the RMSSE indicate neither crucial under- nor overestimation and, therefore, no bias in the surface estimations: ME shows that the average cross-validation errors equal almost zero and RMSSE equals almost 1 in all cases. The MPE can be observed to be low for the categories: mud structured >50%; mud smooth 100%; mud structured 100%; *Beggiatoa* 20-50%; *Beggiatoa* spots >50%, and *Beggiatoa* patches >50%. Highest MPE are found for *Beggiatoa* 0-20% and mud smooth >50%.

5.6.2.2 Estimation of the predictive mono-parametric maps using a validation dataset

Subsequent to the statistic estimate of the quality of the respective IK model, the final control of the resulting IK surfaces follows with help of a validating dataset. This dataset was not used for the determination of the geostatistical data structure and thus gives pure information whether the interpolated surfaces approximate to further analysed data.

The validating dataset consists of an examined surface of 9358 m^2, which is transformed into 675 polygons, resulting from the visual analysis of 530 video mosaics. Therefore, a fifth of the 2840 examined mosaics were used as controlling dataset. In figures 5-10 a-c the results of this investigation are intersected with the kriged surfaces of the MPM enabling a direct comparison. Figure 5-10d shows the spatial distribution of the working dataset in black and the validation dataset in white colours.

Deviations become visible for all three parameters, but the agreements outweigh and confirm the general spatial structure of the data, and/or the spatial distribution of the probability of occurrence. In case of *Beggiatoa* it is noticeable that particularly in the north and north-east of the HMMV centre (see rectangles in figure 5-10a) the interpolated surfaces represent rather an overestimation of the bacteria occurrence. In these areas it becomes clear - on the basis of the validating data record - that pogonophorans are certainly outbalanced there (rectangles figure 5-10b). However, the exclusion surfaces of these two parameters are predicted particularly well. The validation polygons of the parameter mud (Fig. 5-10c) particularly underline occurrences of uncovered surfaces in the centre of the HMMV. Here, the polygons clarify that an underestimation of this parameter is probably the case. Note that at the blank areas the mud categories are not excluded, but attributed to a probability of occurrence of less than 70%.

5.7 Discussion

The representation and area calculation of irregularly distributed data is in the focus of all territorial geochemical balancing methods or definition of protection zones. The surface exactness of specific oceanic regions, as the estimate of global marine primary production (Longhurst, 1998), benthic material flux (Zabel et al., 2000), or the distribution of chemoautotrophic organism is always related to surface areas. Therefore, the interpolation of points into surfaces represents an important gain of information. Resulting mono-parametric maps are the basis for complex multi-parametric maps. For this purpose the quality assessment of the MPMs is an important factor. Within this study statistical (IK and cross-validation) and visual (working and validation polygons) methods were used to generate surface maps describing the spatial distribution of three parameters occurring at the HMMV. These methods enable to evaluate the quality of the mono-parametric maps with regard to the tables 5-2 and 5-3 as well as in the figures 5-8 to 5-10.

The average probability values in table 5-2, calculated within the GIS, describe very credible values for all *Beggiatoa* categories except for "spots and patches >50%". Figure 5-8 correlates directly with table 5-2, i.e. all represented *Beggiatoa* surfaces (Fig. 5-8a) except "spots and patches >50%" are predicted with high values and thus with high probability of occurrence. "Spots and patches 50%" surfaces represented in figure 8a are predicted with less probability than the remaining *Beggiatoa* categories. Even these surfaces are provided with reliable MPE values indicating high conformance between measured and predicted values (Tab. 5-3 and Fig. 5-9a) and a high quality of the applied IK models. This combination suggests that the prediction of these surfaces is limited regionally; assigning a colonisation density closely related to certain geochemical conditions. This assumption is confirmed by the regional distribution of the *Beggiatoa* categories >50% also within the validation dataset illustrated in figure 5-10a.

On the basis of the graphs in figure 9a the quality of the estimation model appears for each category. It can be determined which of the IK models worked in a confidential way by the percentage errors derived from cross-validation. Such a representation of the single percentage errors helps to make the interpolated maps more transparent. All *Beggiatoa* categories achieve satisfying MPE values, only the MPE for the category 20-50% coverage turned out comparatively high (Tab. 5-3). This class is the most frequently occurring class within the working dataset (Fig. 5-7), which is spatially distributed on a large area. This could explain the high probability average value of occurrence (Tab. 5-2) for the areas in figure 5-8a, but also the fuzziness of the IK of model for that category.

The distribution of the forecast surfaces which are not covered by the bacteria is limited to the region outside of the HMMV. This could be confirmed neither by visual video mosaic analysis (concerned surface are highlighted by a rectangle in figure 5-8a) nor during the ROV dive experiences of scientists. It is known that in the centre of the HMMV a surface exists completely

uncovered by *Beggiatoa* or pogonophorans. On the result map *Beggiatoa* are not excluded which would be then assigned by shaded surfaces (rectangle figure 5-8a). Nevertheless, areas are blank resulting from the adjusted minimum threshold in table 5-2. Regarding to this region concerning the two other parameters, pogonophorans are also excluded there (rectangle figure 5-8b), but for mud a probability of over 70% is calculated (rectangle figure 5-8c). Concluding reversely, the probability of occurrence for *Beggiatoa* is only 30% maximum.

The category *Beggiatoa* 20-50% attains both a good estimate for the interpolated probability of occurrence values and for the cross-validation average values. Furthermore, this category achieves applicable results also after the control using the validation polygons.

For all pogonophoran surfaces in figure 5-8b reliable probability of occurrence average values could be determined with help of cross-validation procedures (Tab. 5-2). However, the IK model is of a better quality in the categories no pogonophorans and 0-20% coverage than for the other two classes (20-50% and >50%). Thus, the surfaces of the two latter categories are predicted more uncertainly than those of the first mentioned. Regarding the pogonophoran distribution within the working dataset, a conspicuous region is observed in the northwest of the HMMV where a too small spreading of the interpolated surface is probable. This impression is affirmed by the validating dataset (rectangles figures 5-8b and 5-10b).

The validating dataset shows that the data distribution of the working dataset used for IK led to further insufficient estimates in the interpolated surfaces northeast of the mud volcano centre (rectangles figure 5-8b and 5-10b). There, it becomes apparent that the pogonophoran surfaces >50% should probably have been predicted more extensively to the centre of the mud volcano, while the *Beggiatoa* 20-50% surfaces were probably computed too expanded.

Evaluating all applied control mechanisms (visually predicted on the basis the figures 5-8b and 5-10b and statistically on the basis the tables 5-2 and 5-3), the kriged surfaces which are not colonised by pogonophorans are predicted reliably: both outside of the HMMV (reasonable by missing geochemical characteristics of the sediment pore water) and in the central mud volcano area (justified by high temperatures, fresh mud flows and a temporal hierarchy in the colonisation structure of the chemoautotrophic organisms *Beggiatoa* and pogonophorans) (Jerosch et al., chapter 4).

A special ecological meaning is attached to the uncovered and almost uncovered surfaces in the active centre of the mud volcano, because the highest methane release into the hydrosphere is expected due to the absence of biofilter indicating chemoautotrophic communities (Boetius et al., 2000; Damm and Budéus, 2003; De Beer et al., in press). The data distribution of the completely uncovered mud surfaces (mud 100%), in addition, the category mud all 50% is well converted into the IK models (Tab. 5-2), while the low probability of occurrence does not permit a supra-regional interpretation of surfaces in the MPM (similar to the *Beggiatoa* categories >50%). Also the graphs of the percentage errors

of these five classes are quite similar (Fig. 5-9a and 5-9c). The mud categories show that the MPE is not always meaningful enough, as to be seen by the example of the comparison of the categories „50% structured, 100% smooth, und 100% structured". All three MPE are 0 (Tab. 5-2); only the individual percentage error values classify a qualitative order, which is also reflected in the ME values. Hence, the IK model of the category smooth mud 100% is adapted best to the data distribution.

The data density (high and sporadically arising) has different weight of impact due to the IK function in the area (e.g. Richmond, 2002; Webster and Oliver, 2001). According to visual impressions of the validating evaluation, the mud surfaces seem to be slightly underestimated in the centre of the HMMV (see rectangle in figure 5-10c). Conducting an interpolation with both datasets would probably be more expanded and the individual central mud surfaces would be then connected to one large area as, for instance the polygon drawn manually in figure 5-10c.

5.8 Conclusions

Based on a rather dense dataset information about non-sampled seafloor areas is predicted with help of the geostatistical method indicator kriging. This could be conducted successfully due to a large amount of exclusive video mosaics taken up with the ROV *Victor6000*, which were provided furthermore with geographical coordinates. Data acquisition with a ROV is a time- and cost-intensive work. The approach was to extract categorical information from image data, to transform them into binary coded (0/1) discrete variables and to extrapolate them GIS-based into mono-parametric surface maps.

Interpolation procedure represents a gain in information over unsampled areas, however, with the restriction that the results always remain a degree of uncertainty; therefore interpolation cannot replace real measured values and the result of an interpolation becomes ever better, the more largely the used data density. However, this uncertainty can be made more transparent by the presented approach identifying the weak points but also the strengths of interpolated maps (Atkinson and Lloyd, 1998; Rufino et al., 2005). Recapitulatory, the quantity of the analysed data allows the partitioning into a working and a validating dataset and bears a cross-validation. We achieved reliable results and evaluated only small deviations in our results using sophisticated high minimum probability values.

Even the refined analysis procedure (transforming raster data into regionalisation polygons and points, use of geostatistics, and the GIS overlay technique) contains acceptable deviations the results supply the first surface maps of the HMMV. They are used in Jerosch et al. (chapter 4) in form of a predictive multi-parametric map after a GIS-based overlay of *Beggiatoa*, pogonophorans and uncovered mud. This map is used for area computations with respect to the source location of mud flows and pattern of biogeochemical habitats. Discussing

the quality of surface maps is thus also tool to evaluate following continuative studies.

The maps developed in this study are considered as a guideline for further expeditions to the HMMV. For earlier habitat studies at the HMMV please refer to e.g. Gebruk et al. (2003) and Milkov et al. (1999).

Submarine mud volcanoes occur worldwide. Together with submarine asphalt volcanism sites (MacDonald et al., 2004) they provide the geochemical conditions (e.g. gas hydrates, methane release) creating habitats for chemosynthetic life (bacteria and tube worms). For example the Milano Mud Volcano of the Central Mediterranean Ridge (Huguen et al., 2005) and the Chapopote Asphalt Volcano in the Campeche Knolls, Gulf of Mexico, (González et al., 2005; MacDonald et al., 2004) are potential study areas for similar evaluation procedures as conducted at the HMMV.

5.9 Acknowledgements

6 Automatic content-based analysis of georeferenced image data: detection of *Beggiatoa* mats in seafloor video mosaics from the Håkon Mosby Mud Volcano

K. Jerosch[1], A. Lüdtke[2], M. Schlüter[1], G. T. Ioannidis[2]

[1]Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
[2]Center for Computing Technologies - TZI, University of Bremen, PO Box 330 440, D-28334 Bremen, Germany

6.1 Abstract

The combination of new underwater technology as Remotely Operating Vehicles (ROVs), high resolution video imagery, and software to compute georeferenced mosaics of the seafloor provides new opportunities for marine geological or biological studies and applications in offshore industry. Even during single surveys by ROVs or towed systems large amounts of images are compiled. While these underwater techniques are now well-engineered, there is still a lack of methods for the automatic analysis of the acquired image data. During ROV dives more than 4200 georeferenced video mosaics were compiled for the Håkon Mosby Mud Volcano (HMMV). Mud volcanoes as HMMV are considered as significant source locations for methane characterised by unique chemoautotrophic communities as *Beggiatoa* mats. For the detection and quantification of the spatial distribution of *Beggiatoa* mats an automated image analysis technique was developed, which applies watershed transformation and relaxation-based labelling of pre-segemented regions. Comparison of the data derived by visual inspection of 2840 video images with the automated image analysis revealed similarities, a precision better than 90%. We consider this as a step towards a time efficient and accurate analysis of seafloor images for computation of geochemical budgets and identification of habitats at the seafloor.

6.2 Introduction

The dynamic development of high-technology underwater vehicles as Remotely Operated Vehicles (ROVs), Autonomous Underwater Vehicles (AUVs) or submersibles - designed for marine scientific applications and offshore industry - has led to a significant increase of geodata derived by chemical, optical, and acoustic sensors. These underwater vehicles can be navigated and positioned very accurately and are often equipped with high resolution digital

cameras and video systems. This allows generating georeferenced images and video mosaics by combination of the video stream and navigation data. Such image data is mostly analysed manually at present, which is a very labour-intensive and time-consuming task. The increasing application of ROVs and AUVs for scientific objectives or for maintenance of pipelines and platforms by offshore industry requires time and cost efficient techniques for analysis of georeferenced underwater images. As a step in this direction, a software platform and specified modules for the automatic content-based analysis and attribution of georeferenced image data from the underwater domain – the IBU software - was developed.

In this study we introduce the IBU software as a generic module-based system and its application for the automatic detection of bacterial mats living at the seafloor of organic-rich coastal sediments or at offshore mud volcanoes. For this purpose we used video mosaics from the Håkon Mosby Mud Volcano (HMMV) acquired during six dives of the ROV *Victor6000* (IFREMER) with the video mosaicing software MATISSE (Allais et al., 2004; Vincent et al., 2003) during a cruise by *RV Polarstern* (Klages et al., 2004).

6.2.1 Scientific background

Marine mud volcanoes have been observed worldwide (e.g. Kopf, 2002; Milkov, 2000). They are, like the Håkon Mosby Mud Volcano, often located at tectonically active areas of continental margins (Fig. 6-1). Upward flow of mud from deeper geological strata, enhanced temperatures, high inventories and fluxes of methane and other hydrocarbons, and occurrence of specific chemoautotrophic communities are characteristics for marine mud volcanoes (Hjelstuen et al., 1999; Kopf, 2002; Milkov, 2000). The methane may accumulate in the sediments and forms gas reservoirs such as gas hydrates (solid methane) (Bohrmann and Torres, in press; Egorov et al., 1999; Ginsburg et al., 1999). At mud volcanoes, pore water, gas, and mud is expelled from deep beneath the seafloor forming mounds and craters at the seafloor (Fig. 6-2d, f). Active mud volcanoes are therefore a seep for natural gas (methane) and are often densely populated by bacteria, tube worms, bivalves and other symbiotic organisms (Gebruk et al., 2003; Milkov et al., 2004; Pimenov et al., 2000; Sahling et al., 2005; Smirnov, 2000; Soltwedel et al., 2005).

The methane emanating from the seafloor is often very efficiently used by a symbiosis of archaeal and bacterial microorganisms, forming chemoautotrophic communities that are able to oxidise methane with sulphate which is abundant in seawater. The main microbial processes are anaerobic methane oxidation (AOM) (Boetius et al., 2000) coupled to sulphate reduction and aerobic or anaerobic oxidation of sulphide.

Colonists are chemoautotrophic bacteria using the enzymatic oxidation of reduced compounds as a basic anabolic energy source or metazoan organisms that live in symbiosis with these bacteria, such as pogonophorans (with bacterial symbionts) and bivalves (Sahling et al.,
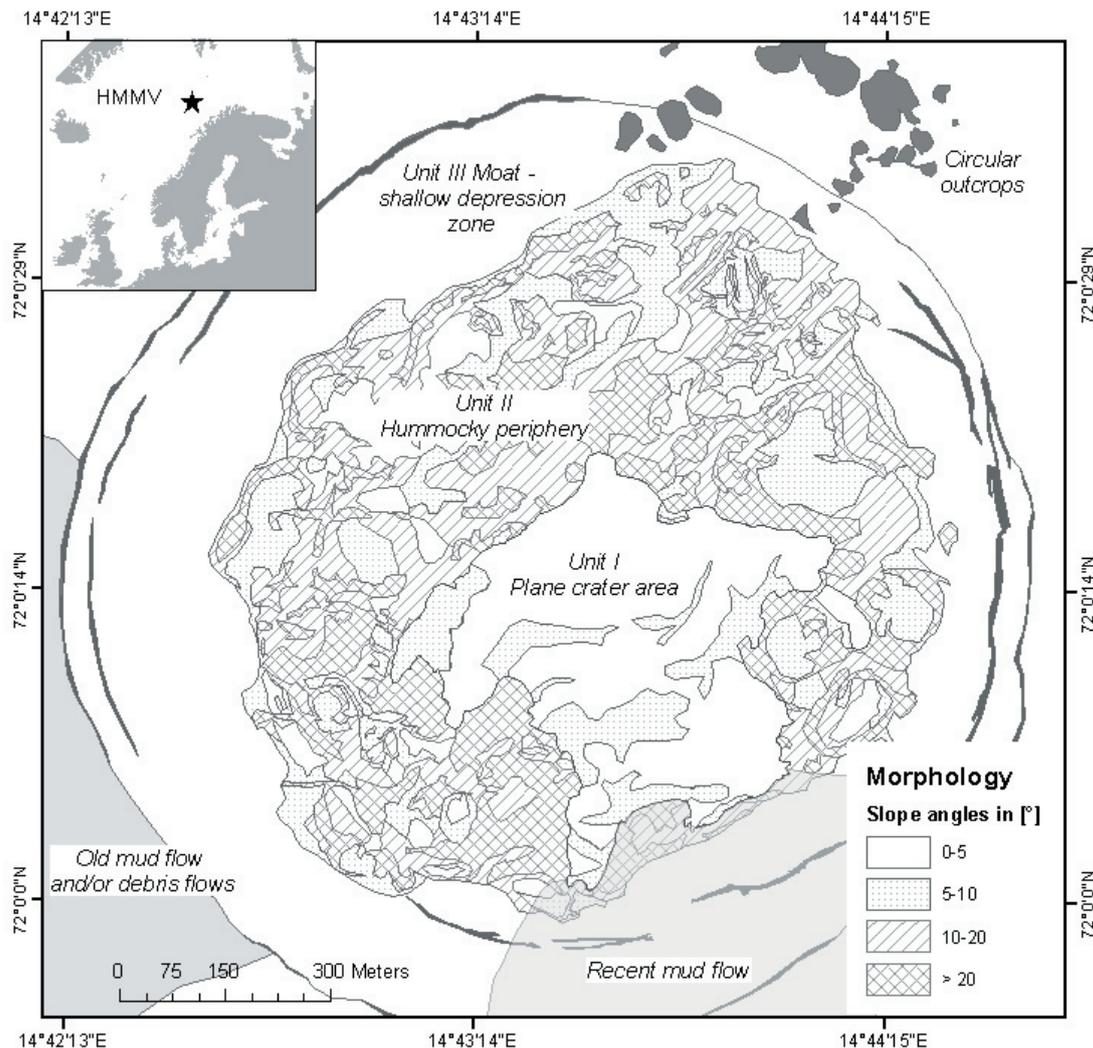
Figure 6-1: Regional setting of the HMMV and its morphological structure. The mud volcano is arranged into three morphologic units, which refer to differences in slope angles. These angles were derived from the microbathymetric investigation of the mud volcano (Edy et al., 2004; Jerosch et al., 2005b).

2003). Microbially mediated anaerobic oxidation of methane (AOM) is the major biological sink of methane in marine sediments. Hence, this process is crucial in maintaining a sensitive balance of our atmosphere's greenhouse gas content (Hinrichs and Boetius, 2002).

At the HMMV, three main habitats have been described in previous studies (Milkov et al., 1999; Gebruk et al., 2003; Jerosch et al., chapter 4): 1. a central, barren area of sediment not colonised by sulphur-oxidising communities; 2. the zone of highest methane turnover which is indicated by the presence of white mats of giant sulphur-oxidising bacteria (*Beggiatoa*) on the seafloor (Fig. 6-2). These bacterial mats cover large areas around the centre of the HMMV; 3. pogonophoran tubeworm fields populating the hummocky part of the HMMV outside of the centre.

Chemoautotrophic communities, i.e. organisms which are fuelled by the chemical energy of dissolved minerals, can indicate the presence of active gas seeps. Hence, at HMMV *Beggiatoa*

and pogonophorans can be used as indicators for enhanced methane consumption. The investigation of their distribution and their density distinguishes between areas with high versus low methane discharge. This is important because there are still gaps in our understanding about the degree of methane oxidation in mud volcano sediments.

6.2.2 Investigation area HMMV

HMMV is an active mud, fluid, and methane-venting seep, located at 1250-1266 m water depth in the centre of the most recent Bear Island fan slide valley at 72∘00.25'N 14∘43.50'E (Fig. 6-1). HMMV is considered as the only well-known active mud volcano in an arctic area and since its discovery in the 1989-1990 SeaMARC II imagery (Crane et al., 1995) it has been studied by multidisciplinary research (Egorov et al., 1999; Ginsburg et al., 1999; Gebruk et al., 2003; Lein et al., 1999; Milkov et al., 1999; 2004; Smirnov, 2000; Soltwedel et al., 2005; Pimenov et al., 2000).

HMMV displays a caldera with a circular shape (ø 1 km in diameter) which can be divided into three concentrically arranged morphological units (Fig. 6-1): 1. The flat central zone of 400-500 m diameter with highly reduced, methanecontaining non-stratified sediments (Lein et al., 1999). In the thermal "eye" of the volcano temperatures as high as 22 °C at 2 m depth and very steep gradients up to 3 °C/m were observed (Kaul and Heesemann, 2004). This zone contains areas mostly uncolonised by epi/megafauna but in the southern and south western part also the regions with the highest density of bacteria coverage. 2. The hummocky periphery with slope angles greater than 20° and a relief of 8-10 m in height enclosing the centre. This second unit is minimum 10 to maximum 440 m wide and is mostly colonised densely by pogonophorans and sparsely by *Beggiatoa*. 3. A shallow depression surrounding the mud volcano, the moat area, comprises the third unit, which is characterised by very small slope angles (mean value 1.58° and apparently no colonisation – neither by pogonophorans nor *Beggiatoa* (Jerosch et al., chapter 4).

Major motile epi/megafauna components at HMMV are brittle stars, pantopods (*Pycnogonida*) and benthic fish (mainly *Zoarcidae*, *Liparidae*, and *Rajidae*; Milkov et al., 1999). The highest densities of fish were observed in the central part of the crater (Gebruk et al., 2003). Large molluscs, common representatives of many mud volcanoes and gas seeps, are completely missing at HMMV (Soltwedel et al., 2005).

6.2.3 Data acquisition and processing

Multidisciplinary investigations of methane fluxes and related processes at the Håkon Mosby Mud Volcano were conducted during the cruise ARK XIX3b of *RV Polarstern* in summer 2003. By means of the Remotely Operated Vehicle (ROV) *Victor6000* (operated by IFREMER) the topography of HMMV was mapped and video observations of the seafloor were recorded.
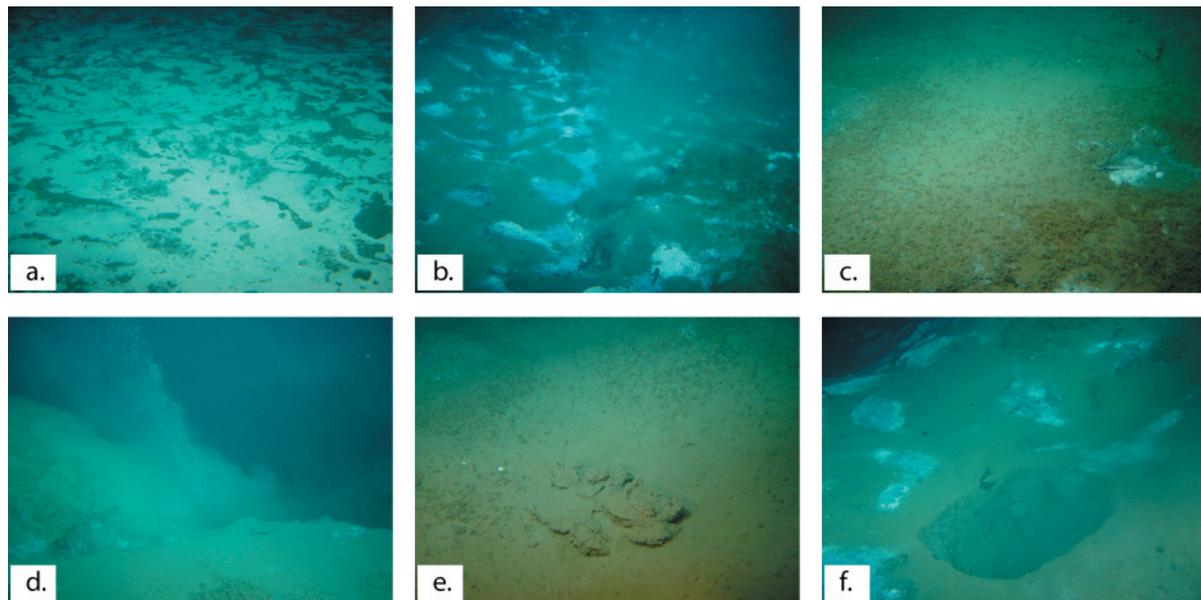
Figure 6-2: Observations of chemoautotrophic communities indicating AOM in the sediment layer, and other features revealing gas and fluid discharge at the HMMV recorded during dives of *Victor6000*. a. Greater than 50% coverage of *Beggiatoa* occurring in the southern and southeast part of the crater. b. 20-50% coverage of *Beggiatoa* in an area of fluid discharge. c. 0-20% *Beggiatoa* coverage and more than 50% pogonophorans, a combination covering a majority of the relief-rich periphery. d. Free gas escape. e. Authigenic carbonate precipitations as a result of AOM. f. Fluids coming through a 30 cm diameter hole in the mud, colonised by *Beggiatoa* and fish.

Further details of the ROV specification concerning water column and sediment sampling, not directly related to this study, are reported in Klages et al. (2004). On six dives of the ROV *Victor6000* video mosaicing was performed during more than 50 hours dive time. The ROV navigated at an altitude of 3 m above the seafloor and with a constant speed of about 0.3 ms^{-1}. This ensured a high image quality for the vertical camera installed at the bottom of the ROV's tool sled. Since the aperture of the camera is 60° the width of the generated video mosaics is about the same as the altitude of survey. The entire length of the video profiles is about 35 km (more than 4200 video mosaics).

The video stream generated by the vertical camera along the ROV transect was processed in real-time using the MATISSE software (Mosaicing Advanced Technologies Integrated in a Single Software Environment, developed by IFREMER). This mosaicing software, based upon image and signal processing components (Allais et al., 2004; Vincent et al., 2003), produces georeferenced mosaics by combining the video input and high precision underwater navigation, as Ultra Short Base Line (USBL). The video mosaics were recorded and integrated into the Geographical Information System ArcGIS 9.0 (ESRI). In total 2840 georeferenced mosaics have been analysed by visual inspection. For each mosaic spatial entities as occurrence of *Beggiatoa* mats, pogonophorans or uncovered mud were digitised as polygons within the GIS, described by specific attributes, and stored within a geodatabase. This allowed a detailed spatial analysis focusing on the distribution of different biogeochemical habitats at HMMV (Jerosch et al., chapter 4). The manually obtained geoinformation

served as validation data during development and for the application of the IBU Software.

6.3 The IBU software

The IBU software is a generic framework for the automatic analysis of georeferenced image data. It is generic in the sense that it provides an interface to link and import additional analysis modules. Therefore, it is adaptable to different scientific objectives concerning the analysis of georeferenced image data. It offers functionalities to access different image formats, to read metadata associate to images (e.g., world file used for georeferencing), for coordinate transformation or basic rendering of maps using the OpenMap™ toolkit (http://www.openmap. org), and provides specific image processing functionality. Besides automated image analysis a component for manual image annotation by keywords and a classification scheme defined and selected by the user is implemented. In the studies presented herein, the software is applied to and compared with the spatial analysis of the distribution of bacterial mats at HMMV derived by visual inspection of georeferenced video mosaics.

6.4 Image analysis

During dives by ROVs, submersibles and AUVs large data volumes of video mosaics and other geotiffs are recorded (just one campaign of six dives by *Victor6000* generated more than 4200 mosaics and geotiffs, Fig. 6-3). Due to this increasing deployment of image recording devices the multitude of image data are a demand for geodatabase systems and spatial analysis by GIS. To cope with this demand the concepts presented herein consist of a data reduction and data analysis component. Data reduction includes the automatic identification of data and non-data regions within the video mosaics. By this means the data region is enclosed by a polygon and separated from the background (Fig. 6-4 a). This polygon can be exported to the
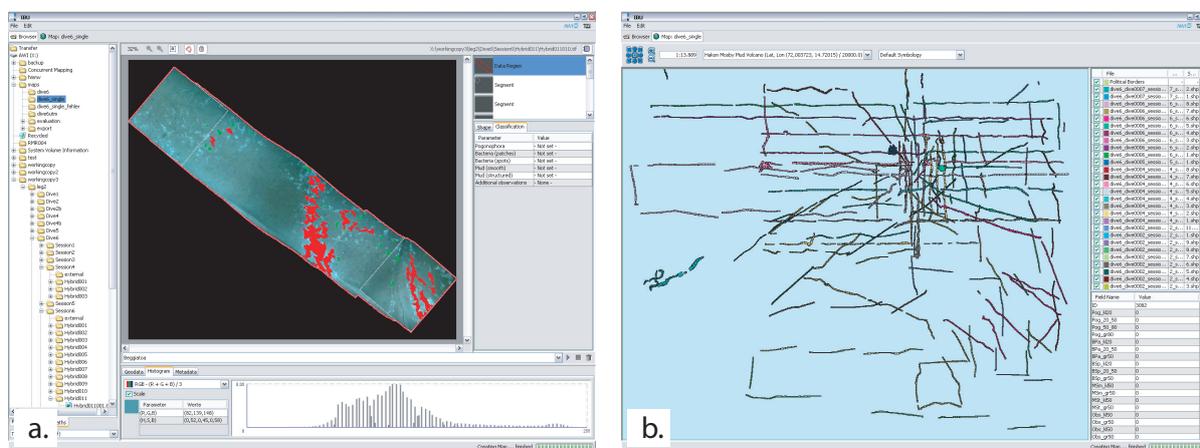
Figure 6-3: a. Analysis view and b. map view of the IBU application. The map view shows amount and distribution of the georeferenced video mosaics derived during the ROV dives.
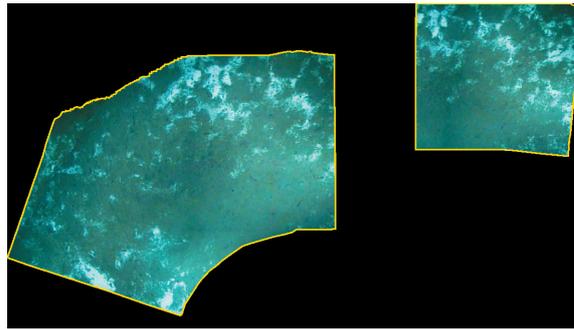
Figure 6-4: Identification of the data containing region within a video mosaic and enclosure of this area by a polygon (yellow line).

Geographical Information Systems (GIS) for further processing as the visual identification and quantification of bacterial mats. and other features.

6.4.1 Extraction of mosaic data regions

Basically georeferenced and other images are composed of a rectangular array of pixels. During video surveys a strip of this rectangular array is covered by data whereas a background color is assigned to the non-data region. Non-data regions are marked by black pixels (RGB={0,0,0}) by the MATISSE software (Fig. 6-4).

In the first step of our automated image analysis these non-data regions are extracted from the video mosaic. For this purpose, the input image is binarised by taking the maximum value of the three (R,G,B) channels and applying a threshold of 0. Then the regions in the binary image are extracted using a Grassfire-like transform (Ballard and Brown, 1982) and represented by their boundaries transformed to geographic coordinates (in term of geotiffs so called world coordinates). Figure 6-4 displays a mosaic produced by the MATISSE system and the extracted data regions (marked yellow in Fig. 6-4). The data containing polygon can be converted into a shape file and integrated into the GIS. By this means the area - on a meter-basis - covered by image information can be derived and attributes describing the content (e.g. occurrence of *Beggiatoa*) can be assigned.

6.4.2 Extraction and handling of data region overlaps

Associated to surveys by underwater vehicles and the mosaicing process, successive mosaics might partially overlap by a small extent (Fig. 6-5). Therefore, two neighbouring mosaics contain similar data for the small overlapping area. This might cause a slight overestimation of the entire area. This redundant information is eliminated by an automated analysis step. Within a single *treatment* (a row of consecutive images, Fig. 6-5), the positioning is accurate enough to identify such overlapping segments. We further divide the area covered by the mosaics of a
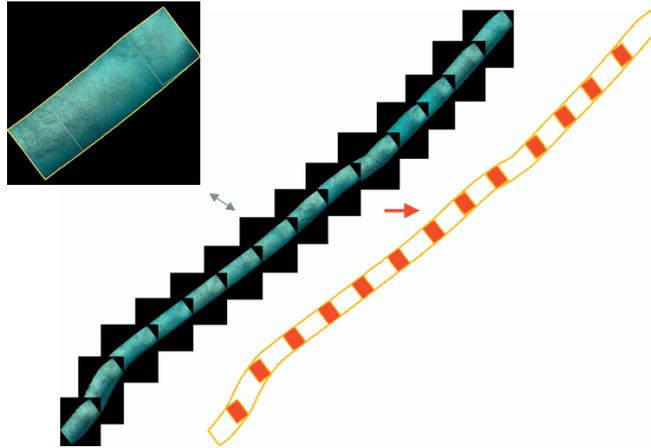
Figure 6-5: Overlapping data regions in consecutive video mosaics (overlaps marked red).

treatment into segments associated with a distinct set of images that we call *atomic segments*. These *atomic segments* are the basis for the quantification of bacterial coverage in the export of results described in chapter 6.4.4.

Let *W* be the set of the extracted data regions for a single treatment. Then *W* is divided into a set of *atomic segments* A using the following algorithm:

```
A := W while (segments s1,s2 in A intersect)
     remove s1 and s2 from A
     add intersecting segments to A
     add rest of s1 and s2 to A
end while
```

6.4.3 Segmentation of bacterial mats

The segmentation process for identification of bacterial mats (*Beggiatoa* mats) is mainly divided into two steps:

(1) an initial region-based oversegmentation gained by a watershed transform (Beucher, 1991; Roerdink and Meijster, 2000),

(2) a relaxation-based labelling (Kittler and Illingworth, 1986) of the resulting image regions into bacterial and non-bacterial regions.

6.4.3.1 Region-based pre-segmentation by watershed transformation

In the first analysis step the input image is converted to greyscale, smoothed by a 3×3 median operator and pre-segmented by the application of a watershed transformation (Roerdink and Meijster, 2000; Beucher, 1991). The main idea is to partition the image into a set of disjoint
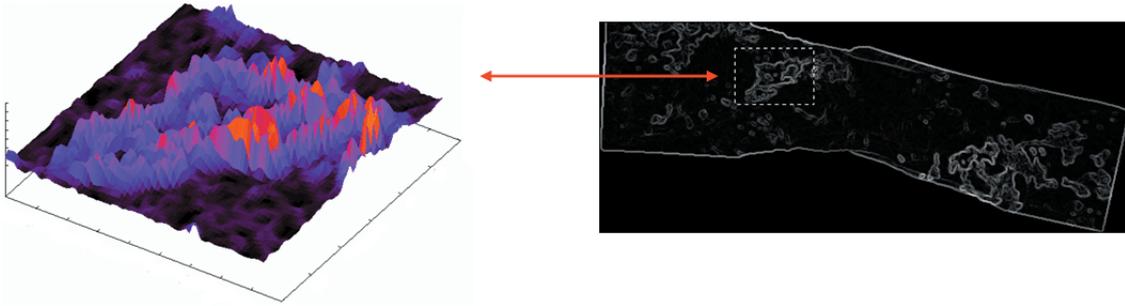
Figure 6-6: Gradient image gained by the application of a Sobel (Davies, 1990) edge detector and interpretation as a topographic surface.

homogenous regions $R = r_0, ..., r_n$ based on grey value discontinuities. Therefore the gradient image - e.g. the first derivative of the input image - is interpreted as a topographic surface (see figure 6-6) and iteratively flooded from local minima. The result is a region-based oversegmentation of the input image. Figure 6-7 shows an example image with *Beggiatoa* mats and the resulting regions of the initial segmentation step. The set of resulting image regions is the basis for the following classification into bacterial and non-bacterial regions.

6.4.3.2 Relaxation-based labelling of pre-segmented regions

In the labelling step we utilise the spatial correlation of the bacterial image regions. We apply a probabilistic approach - relaxation labelling (Kittler and Illingworth, 1986) - to assign labels in $L = \{l_0, l_1\}$ (where l_0 : bacterial region, l_1 : non-bacterial region) with probabilities $P_i(l_k)$ to regions r_i gained by the initial segmentation. We start with a rough initial classification into bacterial and non-bacterial regions and confine the classification in an iterative manner during the relaxation process.

Let $S^{(0)}$ be the set of regions r_i with a contrast to a neighbouring region r_j greater than a given contrast threshold c_0:

$$S^{(0)} = \{ r_i \in R : \exists r_j \in N(r_i) : m(r_j) \geq c_0 \} \qquad (1)$$

where $m(r_i)$ is the mean grey value of region r_i and $N(r_i)$ the set of neighboring regions for region r_i. Every region $r_i \in S^{(0)}$ is pre-classified as a bacterial region. We use the regions in $S^{(0)}$ as start regions for a "*region growing*"-like (Ballard and Brown, 1982) process. Iteratively, neighbouring regions of regions classified as bacteria with a greater mean grey value are also preclassified as bacterial regions. So the set of regions pre-classified as bacterial regions in the n-th iteration ($n \geq 1$) is given by:

$$S^{(n+1)} = S^{(n)} \cup \{ r_i : r_j \in S^{(n)} \wedge r_i \in N(r_j) \wedge m(r_i) \geq m(r_j) \} \qquad (2)$$
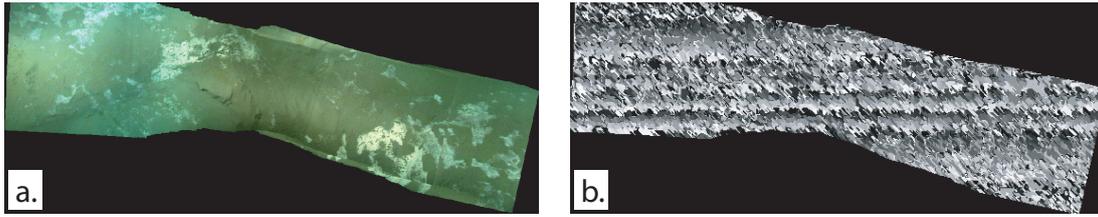
Figure 6-7: a. Sample image with *Beggiatoa* mats (light areas). b. Initial watershed-based segmentation (11,518 regions).

The termination condition is met if $S^{(n+1)} = S^{(n)}$.

Let now $S^{(n_0)}$ be the final state of the process described above. Then the initial label probabilities $P(r_i, l_k)$ for label l_k assigned to region r_i, the compatibility coefficient $C((r_i, l_i), (r_j, l_j))$ that measures the compatibility of label l_i assigned to region r_i given that label l_j is assigned to region r_j, and the impact parameter $W(r_i, r_j)$ measuring the impact of region r_j on region r_i for the relaxation labelling process (Kittler and Illingworth, 1986) are given as displayed in figure 6-8. Iteratively the update rule

$$P^{(n+1)}(r_i, l_k) = \frac{P^{(n)}(r_i, l_k)(1 + \Delta P^{(n)}(r_i, l_k))}{\sum_{l \in L} P^{(n)}(r_i, l_k)(1 + \Delta P^{(n)}(r_i, l_k))} \tag{3}$$

is applied, where

$$\Delta P^{(n+1)}(r_i, l_k) = \sum_{r_j \in R} W(r_i, r_j) \left[\sum_{l \in L} C((r_i, l_k), (r_j, l)) P^{(n)}(r_j, l) \right] \tag{4}$$

is the change in confidence for $P^{(n)}(r_i, l_k)$ in the $(n + 1)$-th iteration. The termination condition is met if

$$\forall r \in R, l \in L : | P^{(k)}(r, l) - P^{(k-1)} | < t_0 \tag{5}$$

for a given threshold t_0.

Finally, connected bacterial regions are extracted and further classified into patches and spots. Regions with a diameter greater than 40 cm are classified as patches, other regions are classified as spots according to the classification scheme used by Jerosch et al., chapter 4. The different categories indicate a spatial-tempolral developement of *Beggiatoa*. Figure 6-9 shows the binary result of the region labelling and result of the classification into patches and spots. Figure 6-10 additionally shows segmentation results for other images of the dataset.

Initial probabilities:

$$P^{(0)}(r, l) = \begin{cases} 0.75 & : (l = l_0 \wedge r \in S^{(n_0)}) \vee (l = l_1 \wedge r \in S^{(n_0)}) \\ 0.25 & : \text{otherwise} \end{cases}$$

Compatibility coefficient:

$$C((r_i, l_i), (r_j, l_j)) = \begin{cases} 1.0 & : l_i = l_0 \wedge l_j = l_0 \wedge m(r_i) \leq m(r_j) \\ 1 - \delta & : l_i = l_0 \wedge l_j = l_0 \wedge m(r_i) > m(r_j) \\ 1.0 & : l_i = l_0 \wedge l_j = l_1 \wedge m(r_i) + c_1 \leq m(r_j) \\ \delta & : l_i = l_0 \wedge l_j = l_1 \wedge m(r_i) + c_1 > m(r_j) \\ 1.0 & : l_i = l_1 \wedge l_j = l_0 \wedge m(r_i) \geq m(r_j) + c_1 \\ \delta & : l_i = l_1 \wedge l_j = l_0 \wedge m(r_i) < m(r_j) + c_1 \\ 1.0 & : l_i = l_1 \wedge l_j = l_1 \wedge m(r_i) \geq m(r_j) \\ 1 - \delta & : l_i = l_1 \wedge l_j = l_1 \wedge m(r_i) < m(r_j) \end{cases}$$

for $\delta = \dfrac{\min(|m(r_i) - m(r_j)|, c_1)}{c_1}$, $m(r_i)$: mean grey value of region r_i and a given contrast parameter c_1 .

Impact parameter:

$$W(r_i, r_j) = \frac{1}{|N(r_i)|}$$

where $N(r_i)$: set of neighbouring image regions for region r_i .

Figure 6-8: Parameters of the relaxation labelling process: initial probabilities $P^{(0)}$, compatibility coefficient C, and impact parameter W applied in the relaxation labelling step (Kittler and Illingworth, 1986).
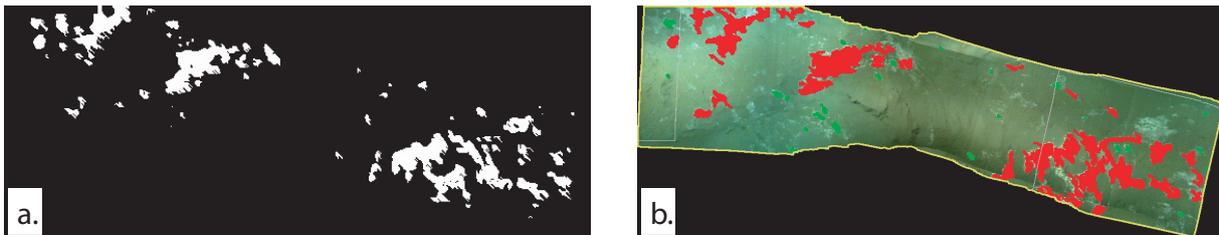
Figure 6-9: a. Binary result of region labelling b. Classification of bacterial regions into patches (connected segments with a diameter ≥ 40 cm, red) and spots (green).
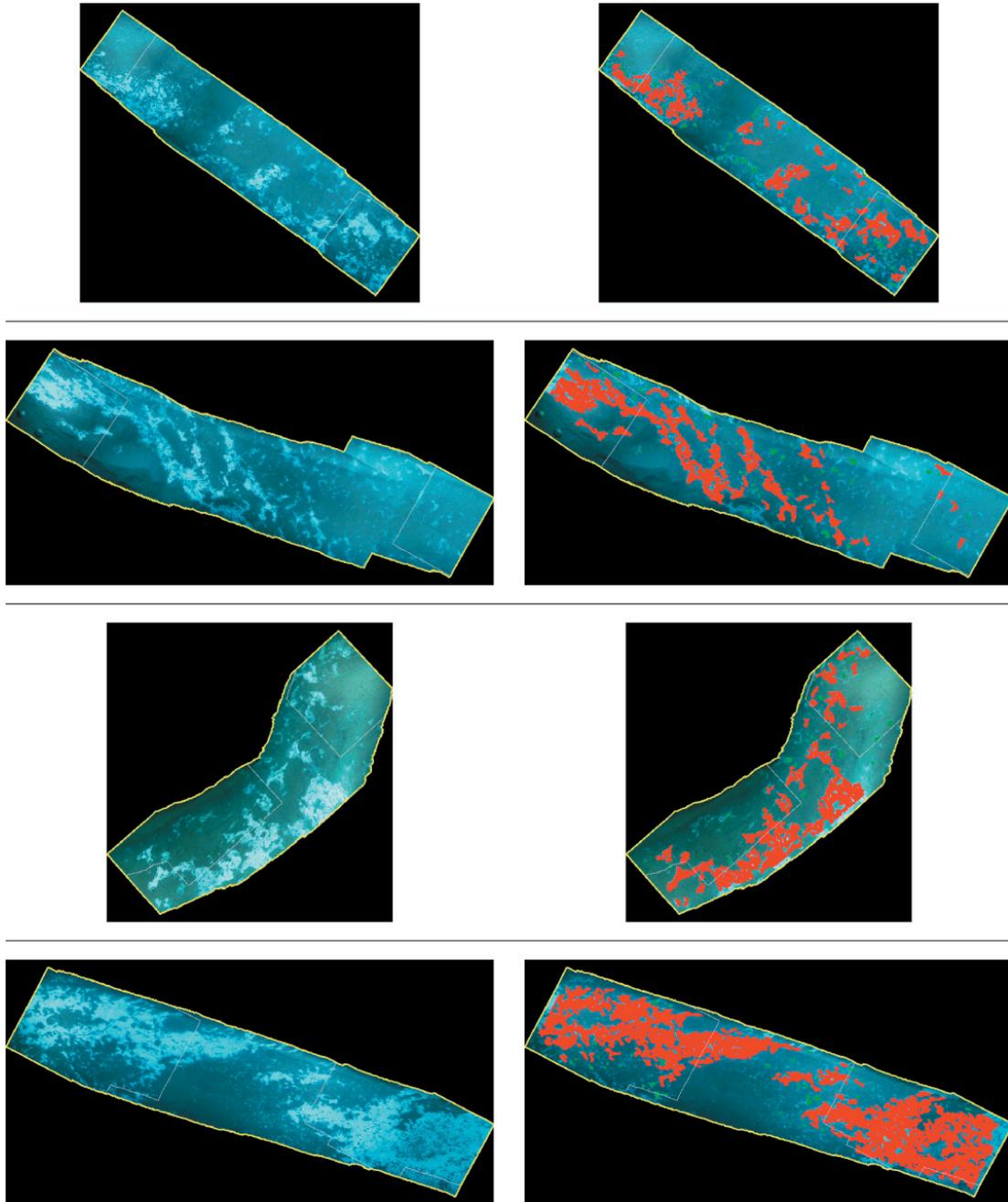
Figure 6-10: Example segmentation results for images with *Beggiatoa* coverage of the dataset.

6.4.4 Result export to Geographical Information Systems

The analysis results are exported to Geographical Information Systems (GIS) *treatment*-wise, where a *treatment* again is a row of consecutive mosaics, so the positioning is accurate enough to consider overlaps of the mosaic data regions. In the export the area covered by the data regions of the images of a treatment is further divided into small segments to gain precise local information about the spatial distribution of the detected bacterial coverage. Therefore the

union of the mosaic data regions is cut into segments of adjustable length (the default is 2 m) along its major axis. Figure 6-11 displays the resulting segments for the images of a single *treatment*.

The export of the detected bacterial coverage is then based on the *atomic segments* extracted as described in chapter 6.4.2. As the union of *atomic segments* covers the complete *treatment* data and vice versa, every export segment is completely covered by such segments. For every export segment *e* the set of intersecting *atomic segments A(e)* is determined. Let now inter*(e, a)* be the intersection of an export segment *e* and an *atomic segment a*, *I(a)* the set of images associated with an *atomic segment a*, and cov*(i, r)* the degree of bacterial coverage of a region *r* in image *i*. Then the bacterial coverage cov*(e)* of export segment *e*, taking the average bacterial coverage for overlapping image segments, is computed as follows:

$$\text{coverage}(e) = \sum \frac{\text{area}(\text{inter}(e,a))}{\text{area}(e)} \left[\frac{1}{\# I(a)} \sum \text{cov}(i, \text{inter}(e,a)) \right] \qquad (6)$$

Figure 6-12 shows the exported results of the automatic analysis for the complete dataset (all 4200 video mosaics). Currently analysis results are exported to GIS via ESRI shapefiles (ESRI, 1998) - the de facto standard for geospatial data exchange used by various GIS systems. Further integration into the ESRI GIS is planned for the future.

6.5 Quality assessment by comparison with visually inspected field data

This chapter contains an evaluation of the proposed algorithm including extensive tests accomplished in a gradual procedure. Tests are based on a set of 4200 georeferenced video mosaics obtained during expedition ARK XIX3b of *RV Polarstern* in 2003 providing information about the distribution and coverage degree of chemoautotrophic bacteria on the seafloor at the HMMV. Part of this dataset (2840 video mosaics) was analysed manually by visual inspection, assigning four classes of coverage degree to sub-regions of each video mosaic:

(1) no *Beggiatoa* coverage
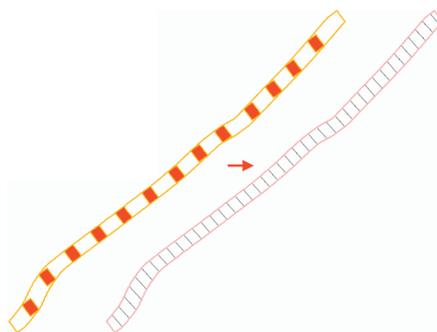
Figure 6-11: Image data regions for a single treatment (overlaps marked red) and resulting export segments.
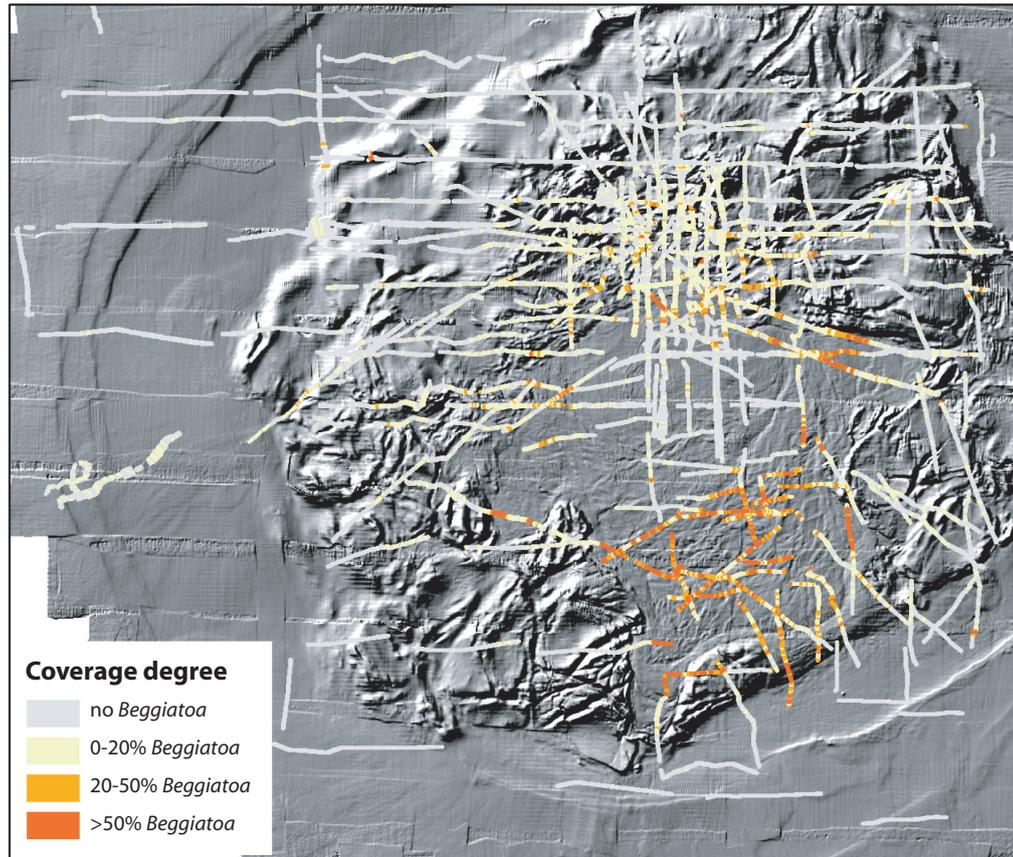
Figure 6-12: Spatial distribution of bacterial coverage derived by the automated image analysis of 4200 video mosaics. The results were exported into the GIS and overlaid on the topography (visualised as hill shading map) of the HMMV.

(2) > 20% *Beggiatoa* coverage

(3) 20 - 50% *Beggiatoa* coverage

(4) > 50% *Beggiatoa* coverage

A subset (related to the 2840 visually inspected mosaics) of 2310 video mosaics served as a working data set and was used during development and in experiments for adjustment of parameters like threshold values applied by the programmed algorithm. The other 530 visually examined mosaics served as a validation data set for tests of the algorithm applying the optimised parameterisation. The quality and performance of the image analysis was measured in terms of *precision* and *recall*, where *precision* measures the rate of correct detections of *Beggiatoa* according to the detections overall and *recall* measures the rate of correct detections according to the overall coverage (e.g., 20-50% coverage) estimated by visual inspection (Cleverdon et al. (1966)). Since for the visually examined mosaics coverage classes were applied (see above), whereas single values were computed for each subregion covered by *Beggiatoa*, we assume a correct result of the automatic analysis if the result lies in the accordant interval. Otherwise we assume the *measurable error*, e.g. an error of 5% if a region is manually classified with
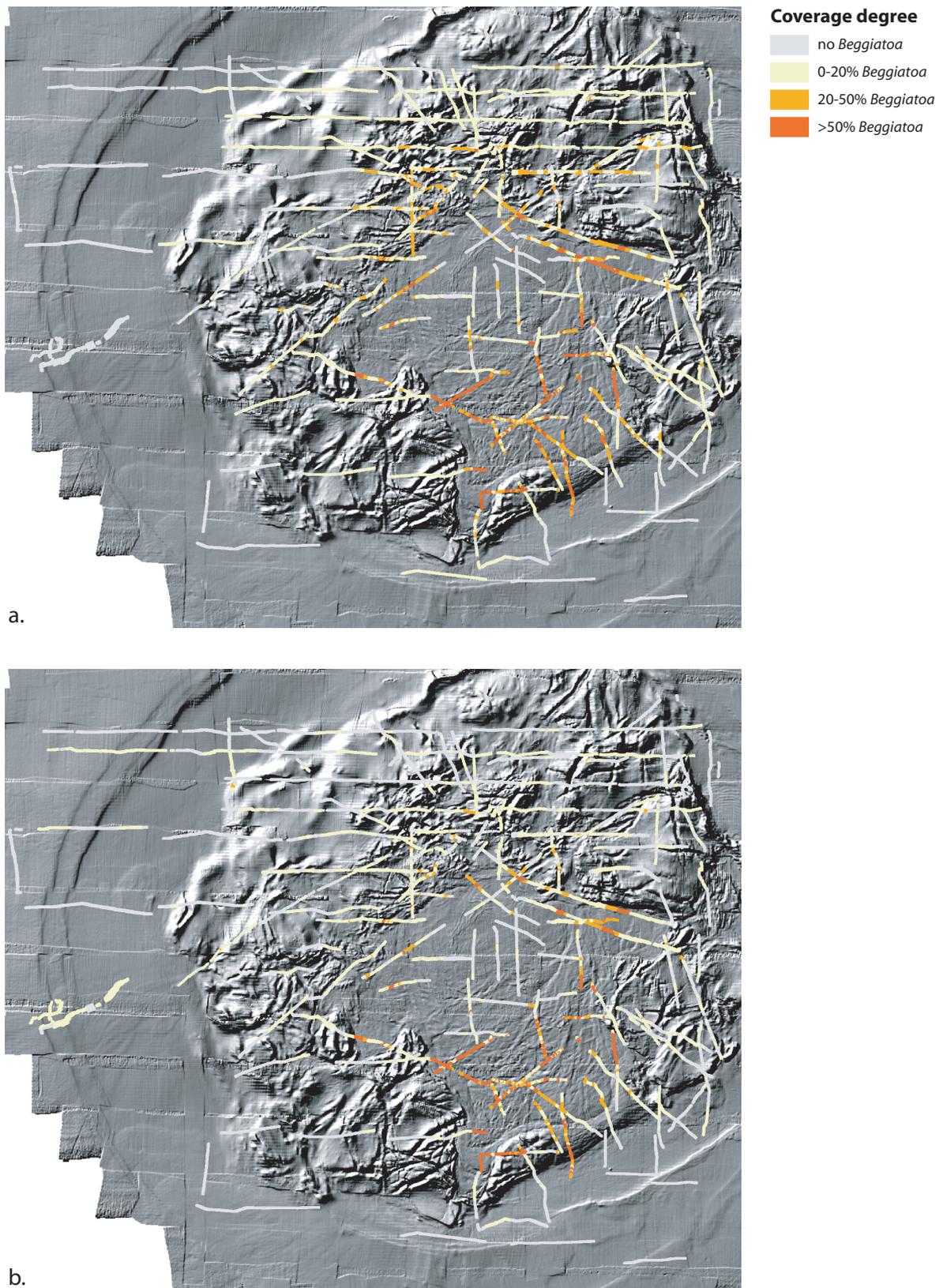
Figure 6-13: a. Coverage degree obtained by visual analysis of video mosaics. b. Results obtained by automatic image analysis. The spatial coverage of *Beggiatoa* is overlain onto the bathymetry of HMMV visualised as hill shading map.

20 - 50% and the result of the automatic analysis predicts 55% of coverage. In the evaluation against the visually examined dataset the algorithm performed as follows:

	working set	validation set	complete set
precision	92.001%	90.64%	91.38%
recall	76.07%	90.22%	81.84%

The average absolute *measurable error* (see above) per annotated region of the manually analysed dataset was 2.06%. Further divided into the coverage classes obtained by visual inspection the average absolute *measurable error* per class over the complete set was the following:

coverage class	#regions[1]	Ø abs. measurable error
no *Beggiatoa* coverage	599	1%
< 20% *Beggiatoa* coverage	1314	0.44%
20 - 50% *Beggiatoa* coverage	344	8.76%
> 50% *Beggiatoa* coverage	122	5.79%
all classes	2379	2.06%

[1]The regions annotated by the domain expert do not correspond to the mosaic data regions, so the numbers of regions and mosaics differ.

Figure 6-13 shows a comparison of the results obtained by visual inspection and by automatic image analysis. As can be seen in the figure and from the recall of ≈ 80% the automatic detection of bacterial coverage tends to produce a slight underestimation of the bacterial coverage. This is mainly due to very shallow bacterial coverage or insufficient illumination in certain image regions, where bacterial coverage has not been detected automatically but has been annotated by the domain expert during the visual analysis. Misdetections are mostly due to overexposure and reflections and light structures or objects similar to bacterial coverage on the seafloor. Examples for causes of undetected or misdetected bacterial coverage are displayed in figure 6-14.

The computation time for the analysis of the complete dataset (all 4200 georeferenced video mosaics) was ≈ 4 hours and 18 minutes on standard PC hardware (Athlon XP2600+ with 512MB memory installed) . The computation time for the analysis of a single image varies, mainly because the run-time of the relaxation step increases when more bacteria is detected. The computation time for the analysis of a single image did not exceed 12 seconds. The algorithm can therefore be regarded as real-time enabled for habitat mapping purposes as in our case with the MATISSE (Allais et al., 2004) system.
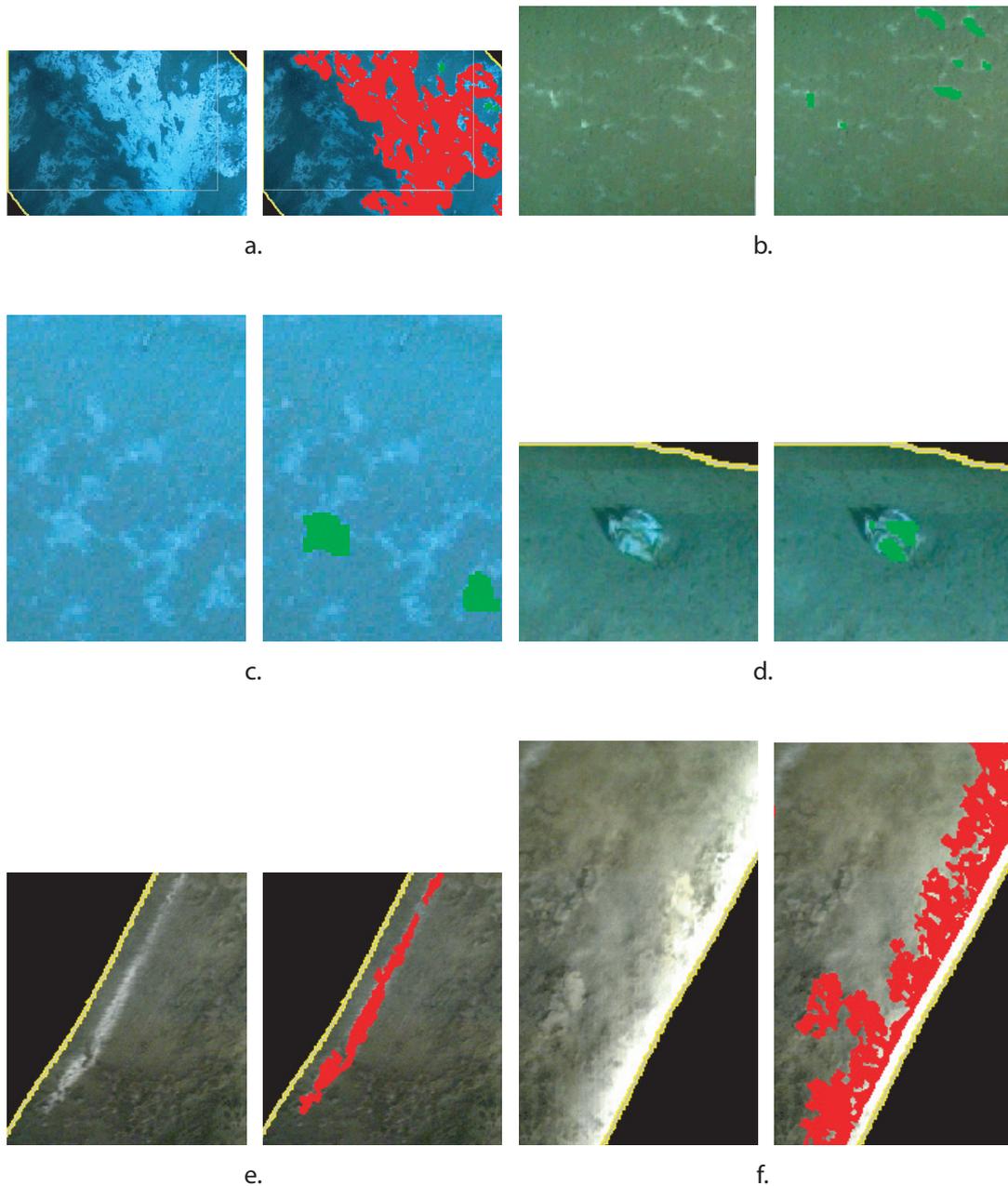
Figure 6-14: a. Undetected bacterial coverage due to insufficient illumination. b. and c. Undetected shallow bacterial coverage. d. Misdetection caused by object (presumably waste) on the seafloor. e. and f. Misdetections due to overexposure.

6.6 Conclusions

The developments in the field as mobile underwater vehicles (e.g. ROVs and AUVs), digital image and video systems, as well as software combining video streams and navigation data into georeferenced video mosaics, is coupled to a fast increase of the number of studies and applications investigating spatial patterns at the seafloor or the identification of technical allignements as pipelines. To cope with the large number of video mosaics recorded during single dives of these vehicles semi-automated or even automated image analysis and feature

detection is required. Related to investigations of the methane cycle at the Håkon Mosby Mud Volcano an approach for a fully automatic detection and quantification of bacterial coverage at in seafloor by analysis of video moasaics was developed. The proposed algorithm is based on techniques of digital image processing. A watershed transform (Roerdink and Meijster, 2000) is applied to partition video mosaics recorded during the ROV dives into sets of disjoint homogenous regions. These regions are then labelled as bacterial or non-bacterial using relaxation labelling (Kittler and Illingworth, 1986). The developed algorithm was tested against a visually examined dataset of 2840 seafloor video mosaics and performed with comparable results. The evaluation dataset was acquired using the ROV Victor6000 and the video mosaicing software MATISSE (Vincent et al., 2003). Automatic means of analysis like the proposed approach enable scientists rapidly to gain insight into newly acquired data without a time-consuming manual analysis by domain experts, however manual visual inspection can not be totally replaced. This is particularly useful for the purposes of operations planning during an ongoing expedition. Furthermore, the availability of an automatic on-line analysis of image sensor data could enable unmanned fully autonomous devices to make decisions based on analysis results in the future, e.g. to further explore a site where interesting features (like bacterial seafloor coverage) have been detected. The GIS export of the analysis results allows further data processing. As investigations of geochemical processes - in our case the CH_4 cycle - are often concerned with budgets, this is useful for subsequent data processing like interpolation and spatial area calculations.

6.7 Acknowledgements

Section C:
Manuscripts North Sea
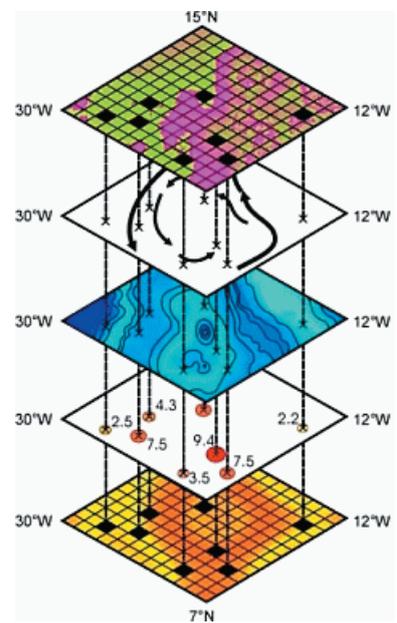
Section C: National and international habitat mapping activities

The implementation of geodata into habitat maps are the fundament for research objectives, requirement for management of offshore resources, and for execution of national and international regulations as the Water Framework Directive or natural conservations issues as the Flora-Fauna-Habitat guideline (FFH). Numerous marine habitat mapping activities are in operation at international and national level. Recent activities are summarised in Busch (2005) and spatially visualised in figure C-1.

Already in 1987 the extensive program MNCR (Marine Nature Conservation Review) was initiated in Great Britain by the JNCC (Joint Nature Conservation Committee), consisting of a committee of nature conservation representatives from England, Wales, Scotland and Ireland.

In 1992 the BioMar program was established by the JNCC in accordance to MNCR. The goal of BioMar was to create a habitat classification as an understandable base for other mapping tasks and to test remote research methods (Connor et al. 2004a-f). BioMar serves with his extensively described habitats as a basis for the habitat classification of EUNIS (European Nature Information System). The purpose of EUNIS is to determine European marine habitats and to make them comparable with each other (Davies et al., 2004).

In 2004, the most extensive international program up to now, the MESH project (Mapping
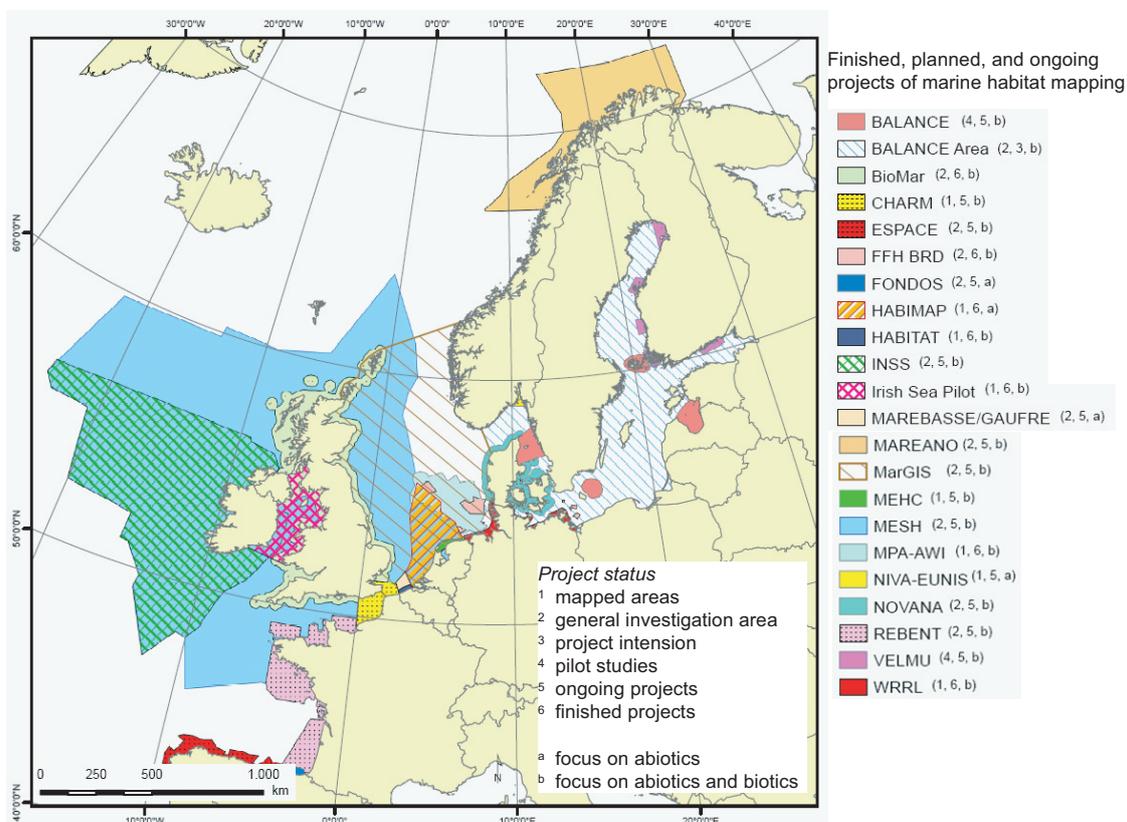
Figure C-1: National and international habitat mapping activities (modified after Busch, 2005).
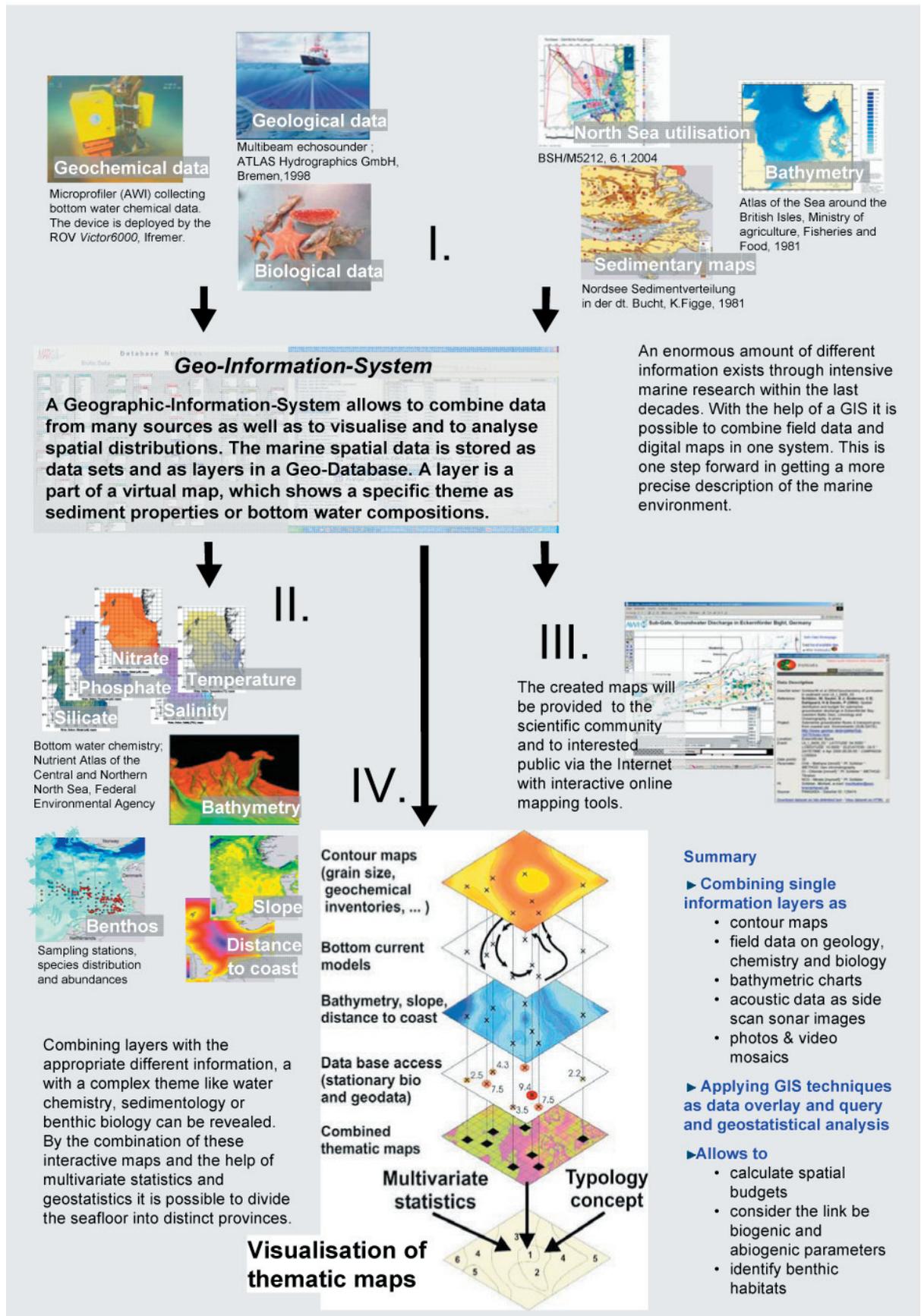
Figure C-2: Work flow of MarGIS - supporting the integration of any kind of data marine type geodata.

European Seabed Habitats), started in Great Britain managed by the JNCC. Within 3 years data from earlier researches shall be gathered and harmonised and an extensive database shall be implemented. As a result new habitat classes shall be developed and standards for the data elevation be delivered. Since 2005 the program BALANCE (Baltic SeA Management Nature Conservation and sustainable development of the Ecosystem through spatial planning) exists for the Baltic region involving 20 partners from 10 states. BALANCE should also collect and harmonise data of earlier studies as well as produce maps of marine habitats finally. At national level in particular Australia, Canada, the USA, Great Britain and Ireland are very active. In the meantime the European coastal states have become active with own projects of researching and mapping marine habitats (ICES, 2005).

However, even other states pursue with different interest and different intensity the mapping of marine habitats. In Germany activities started, boosted by the FFH guideline and the Federal Nature Conservation Act. Thus, a study in the AWI which has produced extensive new data about marine habitats in the North Sea was finished in 2003. The results of this research were needed to decide about protective areas in the North Sea. Areas have to be announced in the context of the German NATURA 2000 activities regulated by the European Union (Rachor and Nehmer, 2003). Since 2002 the MarGIS project, supporting the integration of any kind of data marine type geodata, is realised co-ordinated by the AWI and in co-operation with the University Vechta and the University Neubrandenburg (chapter 7).

7 MarGIS Marine Geographical Information System for visualisation and typology of marine geodata

Kerstin Jerosch[1], Michael Schlüter[1], Roland Pesch[2], Winfried Schröder[2], Andreas Köberle[3], Lutz Vetter[3]

[1]Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
[2]Institute for Environmental Science, University of Vechta, Oldenburger Str. 97, 49377 Vechta, Germany
[3]Department of Geomatics, Fachhochschule Neubrandenburg, 17041 Neubrandenburg, Germany

7.1 Abstract

MarGIS intends the combination of Geographical Information Systems (GIS), research data, and multivariate geostatistical techniques for characterisation and identification of distinct provinces at the seafloor of the North Sea. Such a typological approach, the analysis of classification based on types, is one prerequisite for management issues related to coastal seafloor. It provides a framework for improved application of large environmental data sets, allows enhanced visualisation of multiple information layers, and supports modeling of temporal and spatial interrelations of coastal and ocean regions. To tackle the problem of presentation of large data sets MarGIS uses a web-based viewer which allows a clear visualisation of the information for the general public. The viewer allows a dynamic actualisation and access to metadata and additional information.

7.2 Introduction

In recent years environmental, economic and scientific interests in marine coastal environments and ocean margins increased considerably. Key words are e.g. benthic habitats, fishery, wind energy or offshore oil and gas. Compared to the increasing amount of data, only few concepts were developed for efficient distribution of data and thematic maps to the research community and general public. MarGIS intends to fulfil these requirements by application of Geographical Information Systems and geostatistical techniques to target areas of the North Sea. This allows characterising distinct provinces of the seafloor by combination of geological, biological and chemical properties. Besides scientific needs, such a typological approach supports management decisions related to upcoming economic use of the seafloor.

Compared to the increasing amount of data and information about marine research, only a few concepts and techniques are applied for efficient visualisation and optimal utilisation of present and upcoming data sets. There is for example a considerable need for a generalised analysis and synthesis of seafloor data, clustering the multitude of detailed information. This includes spatial budgets of geological and biogeochemical cycles and characterisation of provinces at the seafloor based on the combination of several information layers. Furthermore, a capable distribution of data and thematic maps to the research community and general public should be supported via internet.

Especially the classification (typology) of the seafloor into provinces, an approach well-established in terrestrial geosciences and documented in form of geological maps, soil maps and other thematic maps, often is a prerequisite for management needs and is the main objective of this proposal. The typological approach, combining by multivariate statistics and geostatistical means several information layers for the assignment of areas of the seafloor to types, allows comparisons of geographical different regions. Therefore, this approach is relevant for assessment and modelling of temporal and spatial changes of the marine environment. This is one of the key issues of the BMBF research program "Information Systems in Earth Management" which funds MarGIS. Besides scientific needs, the typological approach supports management decisions related to upcoming economic use of the seafloor as offshore wind power plants, offshore platforms and pipelines, seafloor cable deployments, sand and gravel dredging, and declaration of protection zones.
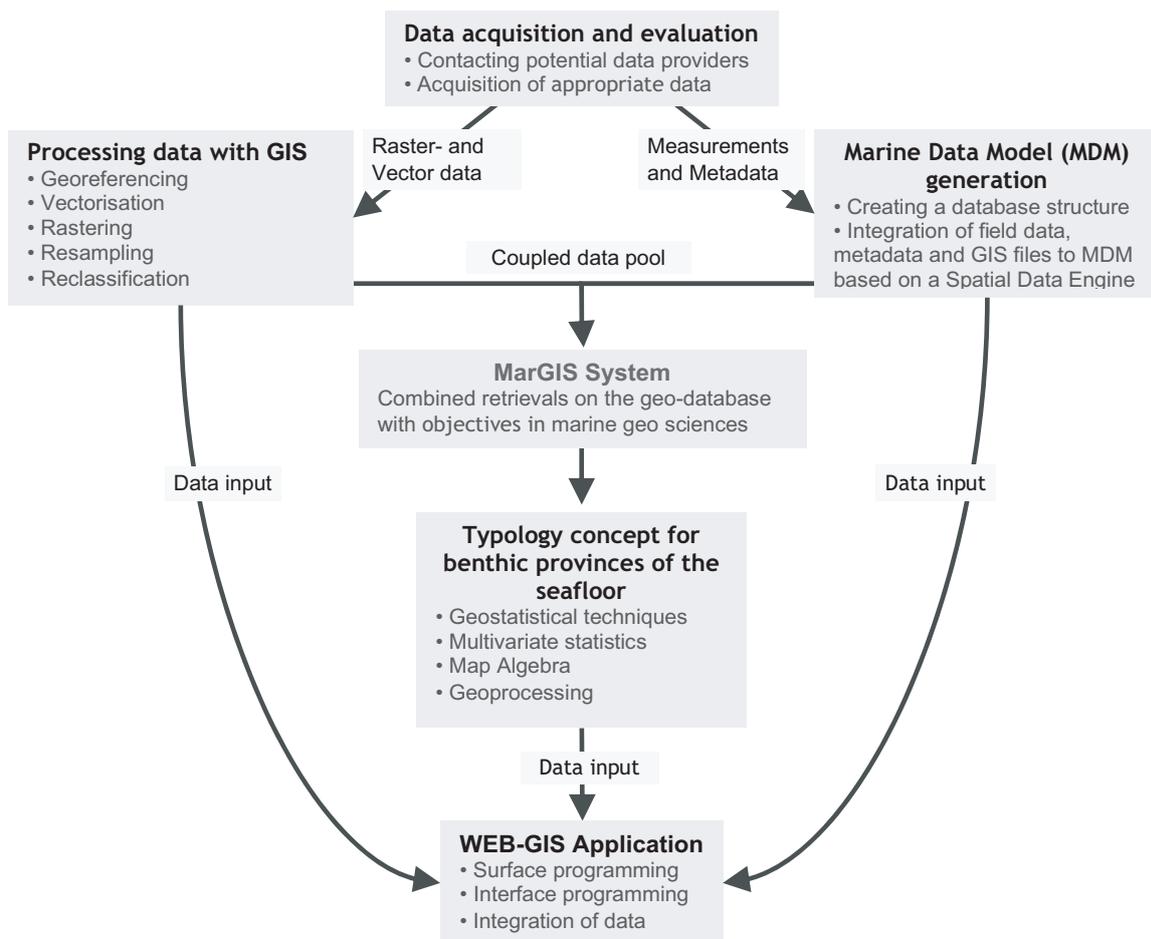
Figure 7-1: Work flow of MarGIS to derive sea floor provinces by means of GIS and statistical methods.

7.3 Methods and tools

The MarGIS project has four closely related tasks (Fig. 7-1): 1. field data acquisition and GIS data processing of analogue maps, 2. generation of a Marine Data Model (MDM) in shape of a relational database management system (rDBMS) located on a Microsoft SQL Server, coupling of the MDM to a geodatabase server based on ArcSDE (ESRI™) and integration of point and polygon information and metadata about the marine-geodata, marine-biological, and bathymetric data into the GIS-supported MDM, 3. spatial subdivision of the seafloor into distinct provinces based on measured data, GIS technology, multivariate statistics, and geostatistics and 4. development of a user-friendly Web-GIS application based on ArcIMS9.0 (ESRI™).

7.4 Results

7.4.1 Data acquisition: measurements and geodata

The data derive from an intensive recherche of published literature, reports and maps in close co-operation with scientists from various research disciplines and marine data base systems (MDBS) (Tab. 7-1). A considerable amount of published data was retrieved from marine data base systems established by international initiatives or by Federal Hydrographic and Oceanographic Agencies. Prominent examples as the data base systems operated by ICES

Table 7-1: Heterogeneous data sets (field data) provided by national and international databases and institutes compiled within the Marine Data Model.

Institutes	Period	Object of study	Description	No. of sites	No. of samples	No. of data
AWI	2000	benthos	Benthic Biol.	184	184	184
AWI	2000	benthos	Sedi., Chem., Biol., Bathy.	180	180	180
EC	2000	benthos	Benthic Biol., Sedi., Hydro., Bathy.	269	270	7694
CEFAS	2000	abiotics	Sedi., Chem., Bathy.	survey rectangle	survey rectangle	59
SBS/UWB	2000	benthos	Benthic Biol.	270	270	7653
SBS/UWB	1999	benthos	Benthic Biol.	241	241	7699
ICES	1976-2002	abiotics	Chem., Hydro., Bathy.	40764	40820	40820
ICES	1999-2002	fish	Fish, Hydro., Bathy.	survey rectangle	survey rectangle	57730
ICES	1985/1986	fishes	Fish, Hydro., Bathy.	1047	1047	109217
ICES	1985/1986	benthos	Benthic Biol.	306	506	21386
ICES	1999-2002	fishes	Fish, Hydro., Bathy.	3007	3007	388052
IFMHH	1984-2000	abiotics	Chem., Hydro., Bathy.	3810	3811	3811
BFA/IFOE	1981-1997	diseases	Fish, Hydro., Bathy.	2175	2175	158855
BFA/IFOE	1984-1999	species	Fish larvae	5952	1029	595
BFA/IFOE	1984-2002	deformations	Fish larvae	1141	1141	1357
BFA/ISH	2003	fish	Fish, Hydro., Bathy.	94	47	4854
BFA/ISH	1986	fish	Fish, Hydro., Bathy.	256	128	14097
BSH	1982-2000	abiotics	Chem., Hydro., Bathy.	27884	48477	48691
			∑	87580	103333	943452

(International Council for the Exploration of the Sea) or the Marine Environmental Database (MUDAB) initiated and operated by the germane Federal Maritime and Hydrographic Office (BSH) and the Federal Environmental Agency (UBA) hosting North Sea data should be mentioned. All the data sources and contributors are referenced in the section Partners and Contributors of our homepage and described on an institutional level in the metadata system of the GIS.

In total information, not only field data but also analogue maps about the following parameters are collected: bathymetry, salinity, temperature, concentrations of oxygen, ammonium, nitrate, nitrite, phosphate, silicate, and suspended matter, data on benthic biology as epibenthic and endobenthic organisms, fish populations, fish ages and lengths and on the geology and geochemistry of the sediments. The later includes sediment maps, distribution of gas rich deposits, fault zones on earth quakes and about distinct features at the seafloor as pockmarks, seeps and reefs. Furthermore, data about the use of the seafloor as pipelines, platforms, protected areas and sand and gravel mining are compiled.

Apart from the measurement data further geo-information was gathered if available. The mapping procedure of the sediment map should be mentioned here: the data density of this important parameter was not sufficient for an interpolation. Therefore, several existing maps had to be joined to cover the entire area of the North Sea. Although very detailed and high resolution sedimentological maps were generated and made available by the BSH (Federal Maritime and Hydrographical Office), GEUS (Geological Survey of Denmark and Greenland) as well as other institutes and authorities, overview maps about the sedimentology of the entire North Sea or beyond the economic zones are still scarce. To derive such an overview, data from maps and raw data on grain size distribution were compiled und converted to the same map projection within the GIS. Varied sediment classification systems and different qualities of raw data (e.g. spatial distribution or counts of sample sites) required a merging of different information levels concerning sediment attribute data. The maps are shown in the lowest common scale, the Folk sediment classification (Folk, 1954; 1974) aggregated into one sediment map for the North Sea. A nomenclature describing size distributions is important to geologists because grain size is the most basic attribute of sediments. Traditionally, geologists have divided sediments into four size fractions that include gravel, sand, silt and clay, and classified these sediments based on ratios of the various proportions of the fractions. Due to joint visualisation a reduction of information was necessary. More detailed information is maintained within the geodatabase according to different sediment classification systems.

The aggregation of heterogeneous geodata obtained from very various sources required a rather laborious harmonisation procedure and a refined data base model. This was one prerequisite for the integration of data and metadata into the geodatabase linked to the Geo-Information-System ArcGIS (ESRI™) (see below). Specific emphasis was given to the metadata stored conform to ISO 19115.

7.4.2 Data management

Marine data storage is organised historically file-based in expeditions. Thus, it is not possible to compare easily new raised data with old ones. In addition, the data are kept thematically, i.e. there is a lack of query possibilities of bio tables combined with abiotic parameters - spatially or temporally. The requirements of the MarGIS data model were thus the integration of bio tables and abiotic information, but also the requirement of marine data in general: point measurements (at the sea floor and, additionally, in the water column), area measurements (e.g. number of animals per surface), track line measurements (by e.g. dragged techniques or time series generated by anchored systems).

Designing such a rDBMS is a fundamental process that requires planning and revision (Böttcher and Teich, 2003). The database structure had to be specified to the way of marine data collection and all properties of the heterogeneity of data types. The compilation of marine data is yet made by *expeditions* of research vessels and often organised within international projects. In a research area *stations* are determined according to scientific aspects which are visited by the vessel and its crew. At the stations *samples* are taken and very often the samples, e.g. sediment cores, are subdivided into *series of measurements* analysing diverse *parameters*. For each parameter one *measurement* is required. Each sample is accurately defined by the location (position), the point of time (time and date) and the sampling depth. One of the most important tasks with regard to the use of the spatial database (visualisation of queries within the GIS) is the positioning. E.g. the database has to mirror the fact that fishery data often have two pairs of coordinates (heave down and up the gear) belonging to the same sample. In another case only one pair of coordinates is given. Beside this it is nearly impossible to sample the same sample site (position) twice at sea.

The following key words build the columns of the marine database: expedition, station, sample, series of measurements, parameter and measurements. Up to the sample level all data are contiguous, and then the data are split thematically. Arriving to the sample level it is possible to include biotic as well as abiotic data and bring them back together easily to a certain station (query function). The metadata (persons in charge, literature, ...) and the phylogeny establish two further blocks.

The integration of different data sets from several institutes implicates a lot of formatting work and questions of contents, which frequently is underestimated. Attributes had to be converted into the common language of our data management, the tables had to be created and added to the existing tables. Finally, the data must be evaluated and in case of redundancy taken out of the data base. Beside the organisation of the large measurement data sets a further goal of MarGIS was their integration into a geodatabase - the platform, which unites measuring data and geo data – the marine data model (MDM).

A geodatabase of ArcGIS (ESRI™) (Perencsika et al., 1999) is realised on a spatial data

engine (ArcSDE) and supports an object-oriented vector data model. In this model, entities are represented as objects with properties, behaviour, and relationships. Support for a variety of different geographic object types is built into the system. These object types include simple objects, geographic features (objects with location), network features (objects with geometric integration with other features), annotation features and other more specialised feature types. The model allows defining relationships between objects, together with rules for maintaining the referential integrity between objects comparable to the rDBMS in MS Access. After a good data model design and the database tuning within the ArcGIS the geodatabase is ready to start as a multiuser GIS system. ArcCatalog has various tools for creating and modifying the geodatabase schema, while ArcMap has tools for analysing and editing the contents of the geodatabase.

This system builds a stable platform also with large datasets and allows a comprehensive analysis altogether with digitised maps (sediments, bathymetry, currents, geological processes, ...) and field data in one query and to optimise the use of the available data. Furthermore the geodatabase establishes the basis for the implementation of geostatistical and multivariate statistical applications and finally the definition of distinct provinces at the seafloor. Also the internet map server ArcIMS accesses the geodatabase providing the compiled data, created maps and metadata via the Internet for the scientific community and for interested public with interactive online mapping tools.

7.4.3 Typology concept

The ecological regionalisation often is a prerequisite for marine planning and management needs, such as installation of off-shore wind power plants or the declaration of protection zones (Hughes 1997; Moog et al., 2004; Reiniger, 1997). The ecoregionalisation approach used
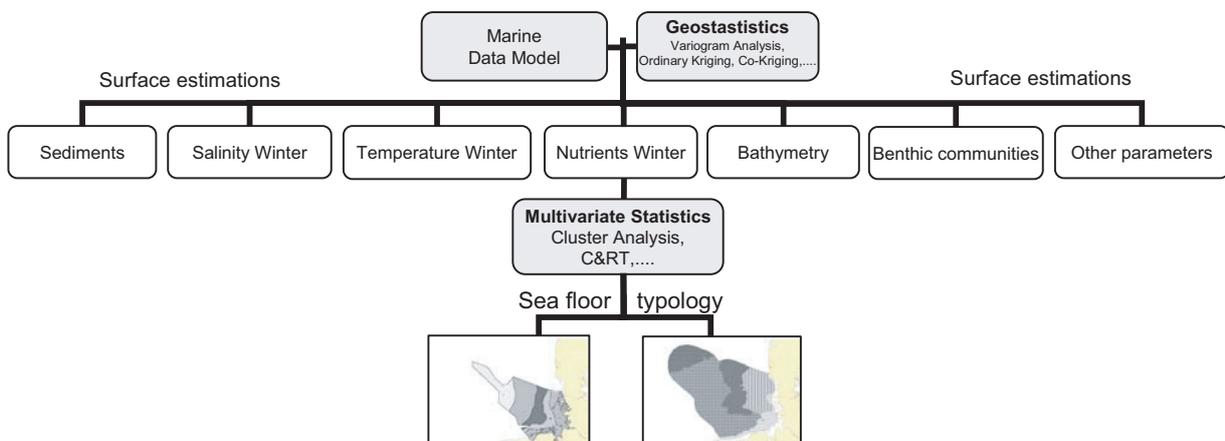
Figure 7-2: Schematic representation of the typology concept: first step is the production of mono-parametric surface maps based on the measurement data coming from the data model by geostatistical methods, in a second step sea provinces are defined by means of multivariate statistic methods as e.g. the CART analysis.

in MarGIS consists of two main work steps (Fig. 7-2): 1. by applying geostatistical methods such as variogram analysis and kriging, surface maps are calculated from measurement data (Goovaerts, 1997; Krige, 1951); 2. multivariate statistics like Classification and Regression Trees (CART) and GIS techniques are then used to calculate sea floor provinces from the kriging grid maps. Since the beginning of the project, kriging maps on temperature, salinity and nutrients (phosphate, nitrate and ammonium) were calculated and used to derive a habitat map for the German Exclusive Economic Zone (EEZ). A detailed execution to these geostatistic and multivariate statistic procedures is described by Pesch and Jerosch (in preparation; chapter 8 of this thesis) and Pesch et al. (in preparation; chapter 9 of this thesis).

7.4.4 Web-GIS application

In general the user of MarGIS is not a GIS expert but for instance planner, so he needs a user friendly environment which can easily manage these large amounts of marine geo-information.

To tackle the problem of presentation of large data sets MarGIS uses a web-based viewer which allows a clear presentation of information.This HTML-viewer was developed within MarGIS and is based on the mapserver ArcIMS. The MarGIS viewer enables the user to view the maps interactively in his browser and includes the following functions: pan, zoom, different types of queries, measurement, view layers in different categories, download of metadata (Fig. 7-4).

On the basis of cascading the information, the different layers are grouped in thematic blocks as for example geology, temperature, chemistry, fishery data and others. With the help of pull
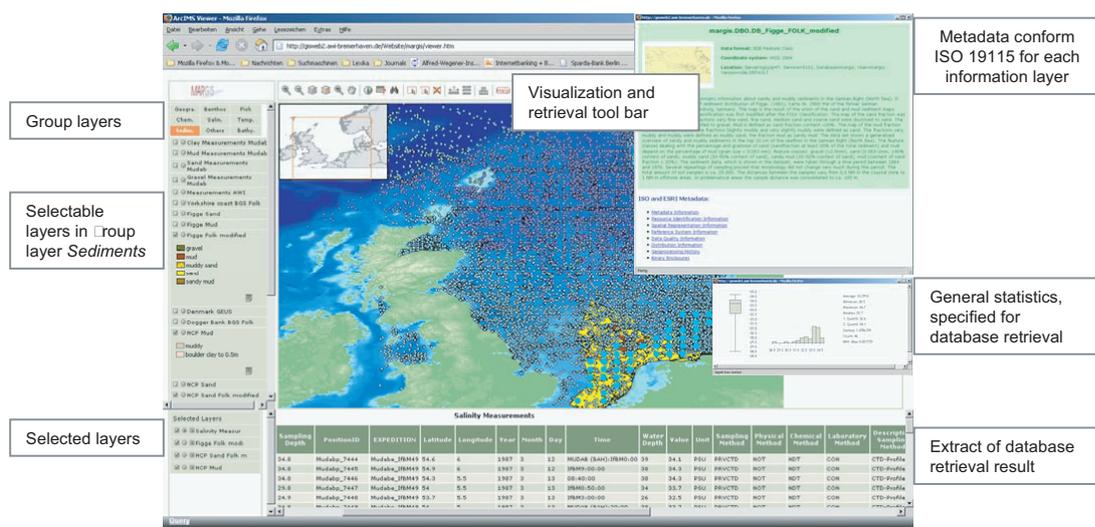
Figure 7-3: Functionality of the MarGIS web viewer based on ArcIMS.

downs the user can select the appropriate layer. To each layer metadata information (after ISO 19115) are available (Fig. 7-3). The document of the metadata information is also located on the database server which makes easier the updating process. There are different possibilities of querying the data: as data to a certain object (point, line, polygon, pixel, grid), to a certain region or with the help of SQL commands. A result of such a request (highlighted spots) is presented in figure 7-3 for the salinity layer. Information layers which include quantitative data e. g. the concentration of salinity, descriptive statistical parameters as box and whisker plots, average and others can be generated (Fig. 7-3).

The mapserver: a mapserver is a programme to disseminate maps in the Internet. Based on different geodata and inputs of the user, a mapserver creates pixel images or vector data and sends them to the client of the user. Furthermore, a mapserver can start different queries on the geodata. The client of the MarGIS project is a HTML-client, who runs in most modern browsers. The input of the user will be sent to the mapserver by HTML-forms and the response of the mapserver will be analysed and showed by the use of java script. For example, if the user pans the map in the viewer, the coordinates of the new view will be sent to the mapserver. The mapserver creates the image of the new coordinates and saves it on the server. Then a response with a link to this image will be created and sent to the client by the mapserver. The client analyses the response and replaces the old image in the view by the new one. Creating a query is working in the same way. The user creates a SQL-query in a special form or a spatial query on the map. The query will be sent to the mapserver which uses them on the geodata. The response is sent back to the client, where a java script analyses and displays the data. The client and server communicate via a special XML called AXL. The XML has the advantage that due to standardisation it is easy to process by the java script.

7.5 Summary

With the help of a Geo-Information-System (GIS) it is possible to combine field data and digital maps in one system. The marine spatial data are stored as data sets and as layers in a geodatabase. The marine data model was developed. The MDM was built at the same time together with the MDM of the ESRI group (D. Wright and OSU Webworks) (Wright and Bartlett, 2000), which was still in development. With MarGIS one step forward is done in getting a more precise description of the marine environment of the North Sea. Combining single information layers as contour maps, field data measured at separated sites or bathymetric charts with GIS techniques and geo-statistics allows deriving aggregated maps. By these means provinces at the seafloor of coastal areas will be derived. This provides a frame of reference for calculation of spatial budgets, consideration of benthic habitats and for upcoming use of the seafloor (Bundesamt für Bauwesen und Raumordnung, 2004; Jarass, 2002). The method was presented in a group of international experts (Working Group of Marine Habitat Mapping of the
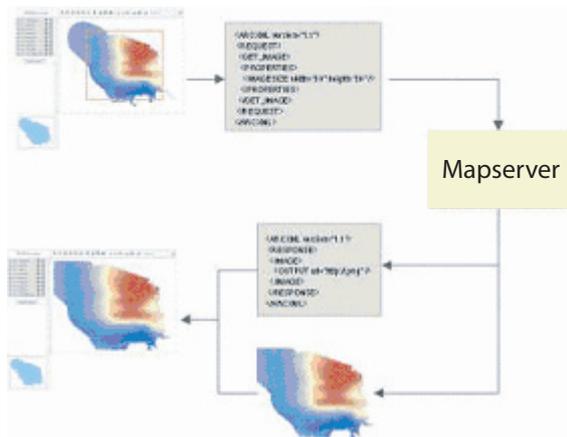
Figure 7-4: Implementation of a zoom operation.

International Council for the Exploration of the Sea).

MarGIS supplies the need for a generalised analysis and synthesis of seafloor data of marine geodata. Finally the multitude of detailed information was organised within a geographical information system which will be accessed easily by users. The amount of measurement data and meta information about marine research required the generation of a relational database management systems (rDBMS).

The geodatabase provides services for geographic data and supports a model of topologically integrated feature classes and other object-oriented features. Geographic data include all data with coordinates: raster data sets e.g. remote sensing data or georeferenced images as mosaics as well as field data and planar topologies e.g. digitized maps based on vectors. Within the MarGIS project the ESRI ArcGIS applications are used. ArcSDE (Spatial Data Engine) defines an open interface between the database system and the applications of ArcMap, ArcCatalog as well as ArcIMS. The complete system allows managing geographic information on a variety of different platforms, generating results for research and planning interests and presenting the data via internet in a user-friendly way.

7.6 Acknowledgement

8 Combining geostatistical methods and GIS to estimate temperature maps of the seafloor in the German Exclusive Economic Zone (EEZ) of the North Sea

R. Pesch[1], K. Jerosch[2]

[1]Institute for Environmental Science, University of Vechta, Oldenburger Str. 97, 49377 Vechta, Germany
[2]Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

8.1 Abstract

The subject of this investigation is the geostatistical calculation of temperature maps for the sea floor of the German Exclusive Economic Zone (EEZ) of the North Sea for the summer months of the years between 1991 and 2000. The data used to calculate these surface maps were taken from the database established in the project MarGIS. We applied geostatistical methods on an extensive dataset, which was acquired from different international authorities like the International Council for the Exploration of the Sea (ICES). It is shown how variogram analysis and two different kriging techniques can be applied to spatially extrapolate measurement data to surface maps. The calculated maps will then be described with an outlook on local and global quality criteria. Special emphasis was put on the optimisation of the variogram model with respect to directional dependencies within the data field. Distinct anisotropies were detected in all ten years, resulting in search ellipsoids with the semi-major axis pointing in north-easterly direction (20° in 1992 to 84.4° in 1998) and anisotropy ratios ranging from 54% (in 1995) to 78% (in 1996). All calculated variogram models furthermore produced very low nugget-sill ratios, ranging from 0 (in 1990) to 0.28 (in 1995). This is indicative of low small-scale variabilities as well as a strong autocorrelation of the measurement values in all ten years. Ordinary and indicator point kriging was then applied in order to estimate temperature maps for the EEZ. The results achieved represent a very good example of such an application, since spatial trends of the temperature conditions on the seafloor directly relate to the autocorrelation structure derived by means of variogram analysis. These directional dependencies allow for a realistic display of the temperature conditions near the sea floor since these tend to depend upon the distance of the location at the seafloor to the coast, or alternatively, the bathymetric conditions.

8.2 Introduction

While the amount of information in marine research steadily increases, only few concepts and techniques are applied to achieve an optimal utilisation of the available datasets. The development of a general concept for the analysis of spatial data, including a typological approach suitable to identify different provinces on the seafloor, is the central objective of the project MarGIS, which is financially supported by the German Federal Ministry of Education and Research. MarGIS pursues a combination of Geographical Information Systems (GIS), research data, and advanced statistical techniques to characterise and identify distinct seafloor provinces in marine and coastal zones. The latter often is a prerequisite for marine planning and management requirements like the installation of offshore wind power plants or the declaration of protection zones (Hughes, 1997; Moog et al., 2004; Reiniger, 1997). The typology approach used in MarGIS consists of two main working steps: (1) Application of geostatistical methods like variogram analysis and kriging to calculate surface maps from measured data. (2) Multivariate statistics including cluster analysis, Classification and Regression Trees (CART), and GIS techniques are then applied to aggregate these raster maps to seafloor provinces.

The objective of this study is to address the first working step exemplarily focussing on temperature conditions measured near the seafloor of the German Exclusive Economic Zone (EEZ) of the North Sea between 1991 and 2000. A calculation of temperature maps of the North Sea was already performed by Brockmann and Topcu (2002) who used deterministic procedures to chart surface and near-bottom temperature maps of the central and northern part of the North Sea including the winters and summers of 1984 - 2000. We applied geostatistical methods on a far more extensive dataset, which was acquired from different international authorities like the International Council for the Exploration of the Sea (ICES). It will be shown how variogram analysis (including variogram maps) and two different kriging techniques can be applied to spatially extrapolate measurement data to surface maps. The calculated maps will then be described with an outlook on local and global quality criteria.

8.3 Basic concepts of geostatistical analysis

Originally coming from geological research applied to estimate mineral resources and reserves (Krige, 1951; Matheron, 1965; 1971), geostatistics are nowadays being used in various terrestrial and marine fields of research. As far as marine research is concerned geostatistical instruments were applied by various scientific disciplines, e.g. pollution research (Poon et al., 2000), geology (Chihi et al., 2000) and biology (Harbitz and Lindstrøm, 2001; Jelinski et al., 2002). The latter exhibits a multitude of examples of geostatistical applications serving the interests of various fisheries: Lembo et al. (1999) applied geostatistical techniques to map the spatial distribution of the deep-water Rose Shrimp in the central-southern Thyrrhenian Sea. González-Gurriarán et al. (1992) and Maynou et al. (1998) proceeded similarly with regard to

three Brachyuran species populating the Galician continental shelf (NW Spain) and the Norway Lobster, respectively. Other examples relating to the abundances of certain fish species can be found in the publications of Petitgas (1996; 1997), Lembo et al. (2000) as well as Rivoirard et al. (2001).

In the following some of the fundamentals of geostatistics will be outlined with special emphasis on the functions introduced by the software extension 'Geostatistical Analyst' of the ESRI product ArcGIS 9 which was used to perform the analyses presented here.

8.3.1 Basic assumptions

Geostatistical procedures are based on the theory of regionalised variables, first introduced in the doctoral thesis of Matheron (1965): "Les variable régionalisées et leur estimation". A regionalised variable z(x) is understood as a variable whose respective values depend on the location x. The changes of the values of this variable are "... neither completely statistically detectable nor deterministically recordable by mathematics in terms of exact formulae" (Akin and Siemens, 1988, pp. 27). In geostatistics the realisations of $z(x_1, x_2, x_3, ... x_n)$ are therefore seen as purely random in a way that they result from a set of random variables $Z(x_1, x_2, x_3, ... x_n)$. Each of these random variables follows a probability distribution with a mean μ and a variance σ^2. By disregarding this random component the spatial variability of the occurring values may depend on an underlying spatial process referred to as spatial autocorrelation. Spatial autocorrelation expresses the notion that objects that are close to each other tend to be more alike than if they were farther apart. Such spatial relationships between two random variables $Z(x_1)$ and $Z(x_2)$ separated from one another by their distance h can be expressed in terms of the covariance $c(x_1, x_2)$ (Webster and Oliver, 2000, pp. 52 therein):

$$c(x_1, x_2) = E [(Z(x_1) - \mu_1)(Z(x_2) - \mu_2)] \hspace{2cm} [1]$$

Even if the realisations of $Z(x_1)$ and $Z(x_2)$ are available, equation [1] cannot be solved since their mean values μ_1 and μ_2 are unknown. Therefore certain assumptions concerning stationary conditions are made, referred to as second-order stationarity (= weak stationarity). It is assumed that:

(a) the mean values of all random variables are constant $\mu = E[Z(x)]$ at all locations, and

(b) the covariance $c(x_1, x_2) = E[(Z(x_1) - \mu)(Z(x_2) - \mu)]$ exists and only depends on distance h between x_1 and x_2.

According to Matheron's intrinsic hypotheses it is only assumed that the expected differences between $Z(x_1)$ and $Z(x_2)$ equal zero, and that the increment $Z(x_1) - Z(x_2)$ has a finite variance independent of location x:

$$Var [Z(x_1) - Z(x_2)] = E [(Z(x_1) - Z(x_2))^2] = 2\gamma h \hspace{2cm} [2]$$

Equation [2] represents the variogram, *γ(h)* the semivariance of distance *h*. The semivariogram therefore describes the variances of the increments $Z(x_1)$ and $Z(x_2)$ as a function of distance *h*. Following Matheron (1963) the semivariogram can be estimated from any empirical dataset by computing half of the mean squared differences of the measured values of pairs of points *(γ)* within given distances *(h)*:

$$\gamma(h) = \frac{1}{2\,N(h)} \sum_{i=1}^{N} (\,Z(x_i) - Z(x_i+h))^2 \qquad\qquad [3]$$

where *N(h)* = number of sample pairs separated by distance *h.*

8.3.2 Variogram analysis

There are two ways to perform a variogram analysis: (1) By determining semivariances for defined distance intervals and radial angle sectors, and (2) by using so-called variogram maps. The variogram analyses reported in this article were based on the second approach. Variogram maps automatically give hints about so-called anisotropies in the data field. Anisotropies are properties of the underlying spatial process where spatial dependence not only changes with distance but also with direction. Examples of such processes can be found, for example, in ecology where ecological patterns strongly depend upon topography. The following work steps are distinguished in modelling the spatial autocorrelation structure with the aid of variogram maps (Fig. 8-1):

(1) Forming direction vectors also referred to as *lag vectors* from all point pairs of the considered point distribution;

(2) Connection of the origins of all *lag vectors*;

(3) Overlaying the result of (2) with a grid of defined mesh size and assigning semivariogram values for each cell respectively *bin* according to equation [3] (The semivariogram values are weighted according to a Kernel function depending on the distance of the ending of each lag vector to the centre of each cell (Johnston et al., 2001, pp. 254)) → resulting in a *variogram map*;

(4) Assigning the calculated semivariances to a coordinate system defined by separation distance (abscissa) and semivariance (ordinate) → resulting in the *experimental variogram*.

Note that variogram maps are point-symmetrical relative to their origin because all lag vectors have a counterpart in the opposite direction (Fig. 8-1).

It is necessary to adapt a defined variogram model to the experimental variogram in order to perform kriging. This can be achieved by means of mathematical models fitted to the experimental variogram in terms of a least-squares regression line. The software Geostatistical Analyst distinguishes eleven such models (Johnston et al., 2001), from which above all the

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N} (Z(x_i) - Z(x_i + h))^2$$

(3)

(1)

(2)

(x)

⊗ measurement point
▼ distance vector

(x) For readability reasons only an extract
of all possible distance vectors is displayed
in this figure.

(4)
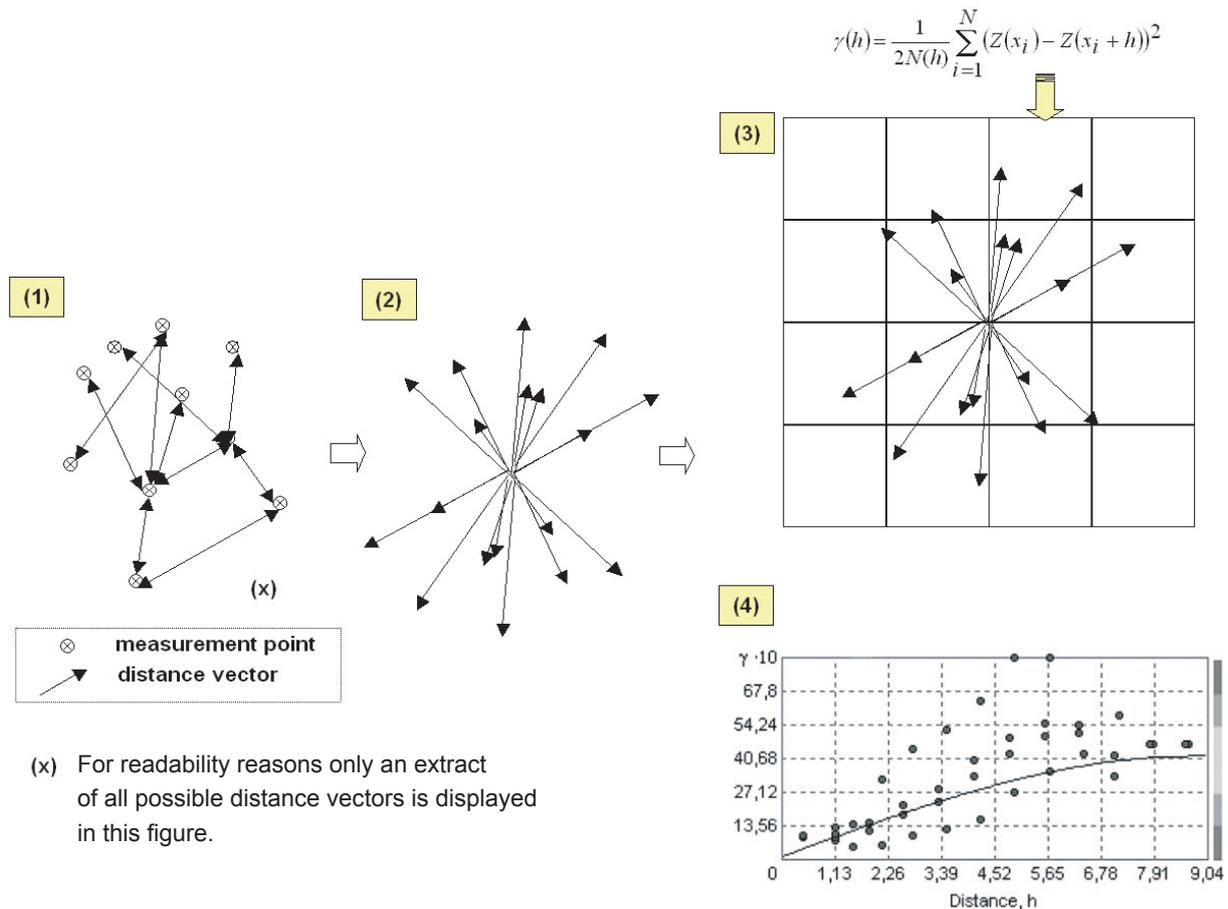
Figure 8-1: Variogram analysis based on variogram maps.

spherical and the exponential models are used most frequently. For detailed mathematical descriptions, please refer to Akin and Siemens (1988), Olea (1999), Webster and Oliver (2000) as well as Johnston et al. (2001). Three key parameters can be distinguished that allow to describe the variogram model: range, sill and nugget effect. The range equals the maximum separation distance within which a distinct increase of semivariogram values can be observed. This is indicative of spatial autocorrelation. The sill corresponds to the semivariance assigned to the range. If anisotropies can be detected, both sill and range will vary with respect to direction. Small-scale variabilities or measurement errors may lead to high semivariances at nearby locations. The variogram model refers to this in terms of the so-called nugget effect, where the variogram model cuts the ordinate above the origin. A pure nugget effect indicates a complete lack of spatial autocorrelation.

Cross-validation can be used to choose the optimal variogram model. For this purpose, each measurement value is extracted from the dataset and estimated by means of the selected variogram model and the kriging method to be applied. By subtracting the measured value from the estimated value an estimation or kriging error can be calculated resulting in an error distribution for the whole dataset. Various key parameters can be calculated from the distribution

113

to characterise the global quality of the chosen variogram model and the estimation as a whole. Examples of such parameters are the Mean Error (ME), the Mean Squared Deviation Ratio (MSDR), and the Mean Squared Error (MSE) (Webster and Oliver 2000):

o ME equals the average value of the kriging errors and describes the tendency of the variogram model to over- / underestimate the measurement values. At best, ME should equal 0.

o MSDR corresponds to the ratio of (experimentally derived) mean squared kriging errors and the (theoretically derived) kriging variances. Ideally, MSDR should equal 1.

o MSE equals the average of the mean squared deviation between measured and predicted values. MSE should be as small as possible.

8.3.3 Kriging

The term "kriging" was invented by Matheron (1963) in order to recognise the contribution of the mining engineer D.G. Krige to improve the estimation of metal concentrations in geological deposits (Webster and Oliver, 2000). By applying a variogram model to the available information of the spatial variation pertinent to a certain property kriging can be used to make predictions for blocks (block kriging) or points (point kriging). In point kriging predictions are made for a defined point raster that is then transformed into a grid raster. A large number of kriging options are available. Only two commonly used kriging algorithms, i.e. ordinary kriging and indicator kriging, will be briefly introduced in the following.

Upon applying ordinary kriging it is assumed that the mean values of all random variables μ are constant at all locations, although unknown. Ordinary kriging is a linear kriging method in which all measurement values within a defined search window are averaged with respect to the distance of the respective measurement point to the point to be predicted. The weighting factors are calculated so they add up to one, so that the estimation error by average equals zero and the estimation variance is minimised (Akin and Siemens, 1988).

Indicator kriging proceeds just as ordinary kriging but only with binary variables, taking on values of either 0 or 1 (Johnston et al., 2001). Such values may indicate the presence (1) or the absence (0) of events, for example, the presence of certain fish or crab species. Indicator kriging may also be applied to continuous data when a certain threshold can be defined as:

$$Z(x) = I(Z(x) > c) \qquad\qquad [4]$$

Here, I is an indicator function that equals 1 if the condition $Z(x) > c$ is true, or 0 if it is false (Johnston et al., 2001). The predicted map which results can be seen as a probability map, in which each predicted point indicates the probability of its value exceeding the defined threshold.

Assuming that the variogram model is true for the whole considered sample area, and the kriging errors follow a Gaussian distribution, the minimised estimation variance - also referred to as the kriging variance - may account for local prediction accuracy. The square root of the kriging variance, the kriging standard deviation, can be seen as an estimate-centred confidence interval containing the true value with a 68% probability (Olea, 1999). Taking the double standard deviation the resulting confidence interval contains the true value with a relative probability of 95.5%.

8.4 Methodological procedure

8.4.1 Data preparation

The data used to calculate bottom water temperature maps for the EEZ was taken from the database that has been established since the onset of the MarGIS project in October 2002. Along with thematic surface data (e.g. sediment data maps) most of this data consist of biotic (e.g. data on benthic organisms) and abiotic measurements of the lower water column. The datasets were provided by several national and international databases and projects and integrated in a relational Database Management System (rDBMS). Table 8-1 gives an overview of all abiotic measurement data which have been integrated into the rDBMS until now. Approximately 235,000 abiotic measurement data sets have been collected from the International Council for the Exploration of the Sea (ICES), the Marine Environmental Data Base (MUDAB) of the German Federal Maritime and Hydrographic Office (BSH) and the German Federal Environmental Agency (UBA) as well as the Institute of Marine Research (IfM), University of Hamburg.

Table 8-2 lists the annual number of temperature data measured within the EEZ in the time between 1991 and 2000. To ensure that the areas near the border of the EEZ were estimated properly, the sampling area was extended by a 40 km buffer zone. A four-month period between May and August was chosen for which to extract temperature data from the MarGIS database (Tab. 8-2). Limited to these four months a minimum of 30 measurement datasets per year evenly distributed within the EEZ were provided, some locations being sampled more than once. As geostatistical analysis only functions with one measurement value per sampling site, the mean value was considered in further analysis.

8.4.2 Variogram analysis

Variogram analysis was performed on the basis of variogram maps according to the procedure described in chapter 8.3.2. Cross-validation was applied to optimise the selection of a suitable variogram model. According to Journel and Huijbregts (1978) the variogram calculations were performed exclusively for distances below half of the maximum horizontal expansion of the area under investigation (here ~ 235 km). Cross-validation was used as the method to select

Table 8-1: Abiotic bottom water measurement data derived from the MarGIS database.

Parameter	MUDAB	ICES		IfM
	1982-2000	1976-1998	1999-2002	1984-2000
Alkalinity	29		65	
Amonium	293		1615	1678
Dissolved Oxygen	20051		2190	
Nitrate	8744		3110	2922
Nitrite	9951		2434	2901
Phosphate	11764		3109	3682
Salinity	34351	27744	12314	3718
Silicate	372		3117	3732
Temperature	31222	28114	12362	3738
Sum	*116767*	*55858*	*40316*	*22371*
Sum (all)	**235212**			

MUDAB: Marine Environmental Data Base
ICES: International Council for the Exploration of the Sea
IfM: Institute of Marine Research, University of Hamburg

the optimum variogram model: As explained in chapter 8.3.2 it was intended to minimise MSE and concomitantly approximate ME toward 0 and MSDR toward 1. In summary, three definite working steps can be distinguished:

(1) Initially, it was necessary to define cell (lag) size and the number of adjacent cells lying between the centre of the variogram map and its edges. For the calculations performed the mean distance of each measurement site to its nearest neighbour (Tab. 8-2) was used as the lag size. The number of lags allocated to each lag size was set so that the distance of significant autocorrelation (range) became clearly visible in the variogram window.

(2) In a next step we tested various variogram models, including different types of models, sills, ranges, and nugget effects.

(3) Finally, if the semivariances displayed on the variogram map indicated anisotropies in the data field, different ranges for different directions (to account for anisotropies) were compared with each other.

8.4.3 Kriging procedures

Ordinary and indicator point kriging was applied in order to estimate temperature maps for the EEZ. Whereas ordinary kriging was applied on all ten datasets, indicator kriging was only performed with the summer data of 2000 in order to demonstrate an alternative to ordinary kriging. Here, the median of all measured values was taken as the threshold value. The mean distance of each measurement site to its nearest neighbour was calculated for all ten years to

Table 8-2: Number of bottom water temperature measurements applying to the EEZ (annually and during the summer month from May to August) from 1991 to 2000.

year	year	summer	
	n	n	mnn [m]
1991	1961	841	2318
1992	579	171	8077
1993	654	93	10951
1994	1546	1214	1395
1995	2144	957	2381
1996	882	346	8569
1997	754	198	8589
1998	1260	314	2063
1999	1180	354	2895
2000	788	266	9248

n = number
mnn = mean nearest neighbour

find an adequate point raster shown in table 8-2. For comparability reasons the average of all ten such distances was set yielding in a 5649 * 5649 m².

The kriging window, which includes the measurement values to estimate a certain point, was adjusted to the range of the variogram model. Since anisotropies were detected in each year under study, the search windows had both semi-major and semi-minor axes (Johnston et al., 2001), hence corresponding ellipses (Figs. 8-3 and 8-4). A four-sector neighbourhood was defined to avoid directional bias.

Two other key parameters, primarily suited for empirical scientists to assess the quality of the estimations performed, were calculated from the cross-validation results: The correlation coefficient after Spearman (Ca) and the Mean Percentile Error (MPE):
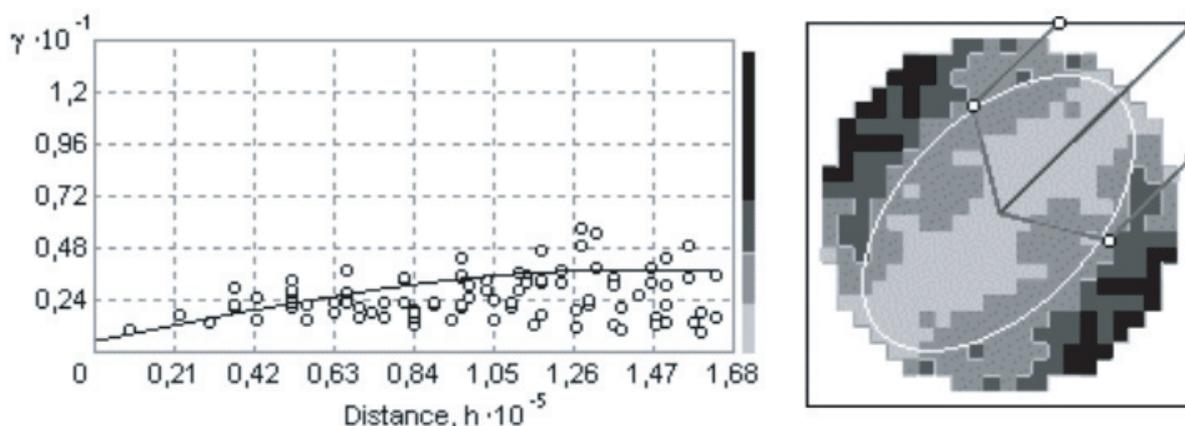
Figure 8-2: Semivariogram values (NE direction).
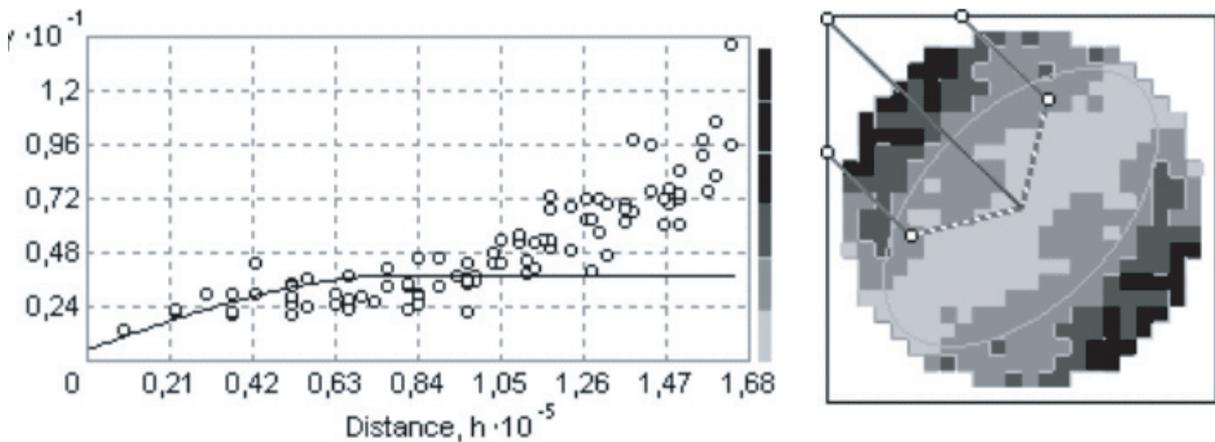
Figure 8-3: Semivariogram values (NW direction).

o   The correlation between measured and estimated values may be described by the coefficient of correlation after Spearman Ca. In case of an ideal correlation Ca should be equal to 1. For the cross-validation results of indicator kriging Cs was not calculated because the measured values only produces values of 0 and 1.

o   By setting the measured value to 100% the kriging error may be quantified relative to the measured value. MPE is the average value of all percentile kriging errors in the dataset.

Table 8-3: Results of variogram analysis of temperatures in the EEZ from 1991 to 2000.

Year	A [°]	A-ratio	ns-ratio
1991	25.0	0.72	0.14
1992	20.0	0.72	0.23
1993	50.0	0.72	0.06
1994	30.0	0.75	0.03
1995	47.9	0.54	0.28
1996	81.2	0.78	0.17
1997	81.5	0.67	0.03
1998	84.4	0.73	0.07
1999	41.9	0.67	0.00
2000	44.6	0.57	0.15
2000 (Ind)	44.2	0.56	0.24

A [°] = rotation angle of the search ellipse
A-ratio = ratio of semi-major and semi-minor axes of the search ellipse
ns-ratio = nugget-sill ratio
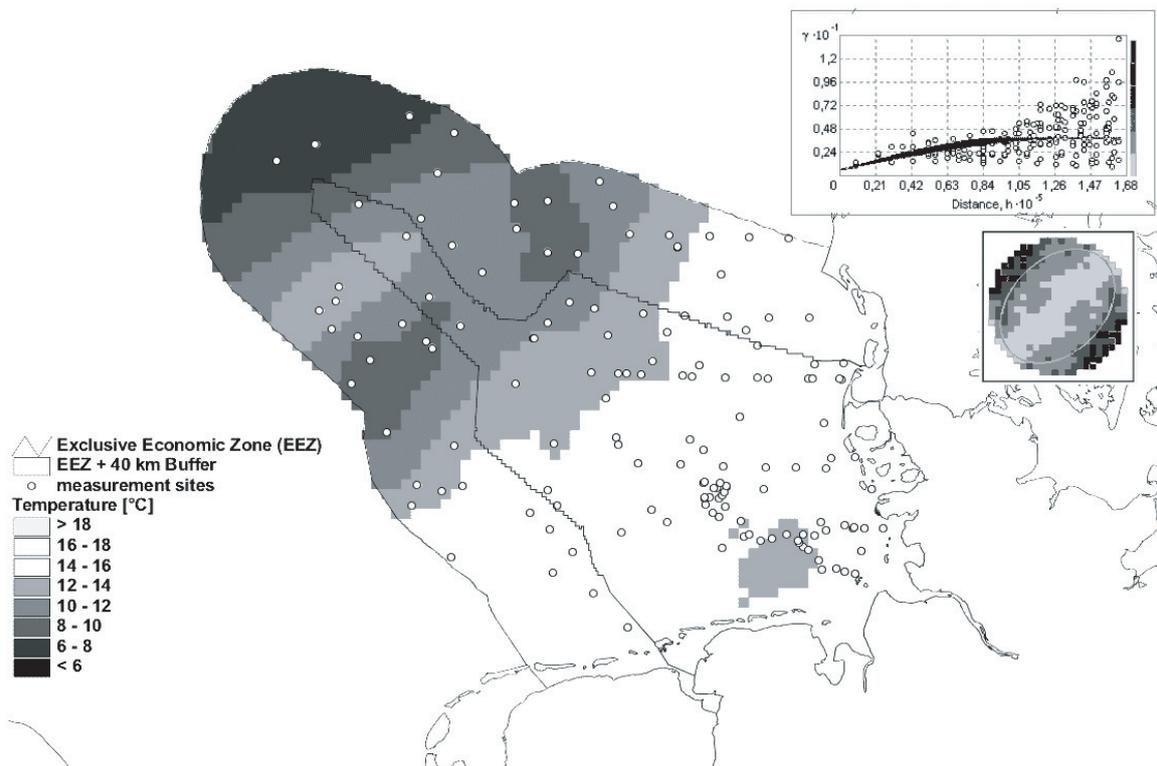
Figure 8-4: Surface estimation of temperatures at the sea floor in the EEZ in the summer of 2000 obtained by ordinary kriging.

8.5 Results

As the example of the variogram map calculated for the summer of 2000 shows the spatial autocorrelation pattern of the temperature distribution within the EEZ is characterised by strong anisotropies: The semivariogram values in NE direction (44.6°) reach their sill at a distance of 140 km (Fig. 8-2). In NW direction the maximum autocorrelation range amounts to 80 km (Fig. 8-3), resulting in a semi-minor and semi-major axis or anisotropy ratio of 57%. All these ratios including the direction of the semi-major axis for all ten years are displayed in table 8-3. It is obvious that similar observations can be made for all variogram calculations: Distinct anisotropies were detected in all ten years, resulting in search ellipsoids with the semi-major axis pointing in north-easterly direction (20° in 1992 to 84.4° in 1998) and anisotropy ratios ranging from 54% (in 1995) to 78% (in 1996). All calculated variogram models also produced very low nugget-sill ratios, ranging from 0 (in 1990) to 0.28 (in 1995). This is indicative of low small-scale variabilities as well as a strong autocorrelation of the measurement values in all ten years. In all ten experimental variograms the spherical variogram model showed the best cross-validation results, and was therefore used in kriging interpolation.

Figure 8-4 shows the result of the ordinary kriging interpolations for the bottom water

Table 8-4: Quality measures obtained by cross-validation.

Year	ME [°C]	MPE [%]	Cs
1991	0.00	5.6	0.78
1992	-0.03	7.4	0.86
1993	-0.10	4.5	0.89
1994	-0.02	2.6	0.94
1995	-0.04	11.0	0.75
1996	0.04	14.4	0.77
1997	0.02	1.9	0.91
1998	0.01	2.8	0.91
1999	-0.10	0.6	0.97
2000	0.03	4.2	0.83
2000 (Ind)	0.00	6.2	

ME [°C] = Mean Error
MPE = Mean Percentile Error
Cs = Coefficient of Correlation (Spearman)

temperature conditions in the EEZ in the summer of 2000. Except for a small area within the inner German Bight, where values below 13 °C were estimated, a continuous decrease of temperature values was observed to range from the coastal areas to the outer part of the EEZ. A similar spatial trend was observed as a result of the indicator kriging calculations (Fig. 8-5). Here, the probability of areas exceeding the median of all measured values (= 14.72 °C) is less than 10% at distances of more than 150 - 180 km from the coast.

To address the global quality of the surface estimations performed key parameters were extracted from the results of cross-validation. These are summarised in table 8-4, which also demonstrates the ME, MPE as well as the C_S values. ME shows that the average kriging errors equal almost zero, indicating neither underestimation or overestimation, nor any bias of the measurement values. MPE is seen as low in all three years, with the lowest values at 0.6% in 1999, the highest at 11% and 14.4% in 1995, and 1996, respectively. The C_S values indicate highly significant correlations ($\alpha = 0.01$) exceeding a value of 0.75 in each of the ten years. The best correlation between measurement and estimations was found in data from 1999 ($C_S = 0.97$).

The standard error maps were calculated for each of the ten years to assess the local quality of estimation. Standard error maps represent the confidence interval of the centred estimate value within which the true value is found with a probability of 68%. The standard errors of the example relating to the year 2000 are shown in figure 8-6 (above). Here, standard deviation values vary between 0.3 and 2.56 °C. The latter values were found farthest from the sampling sites.
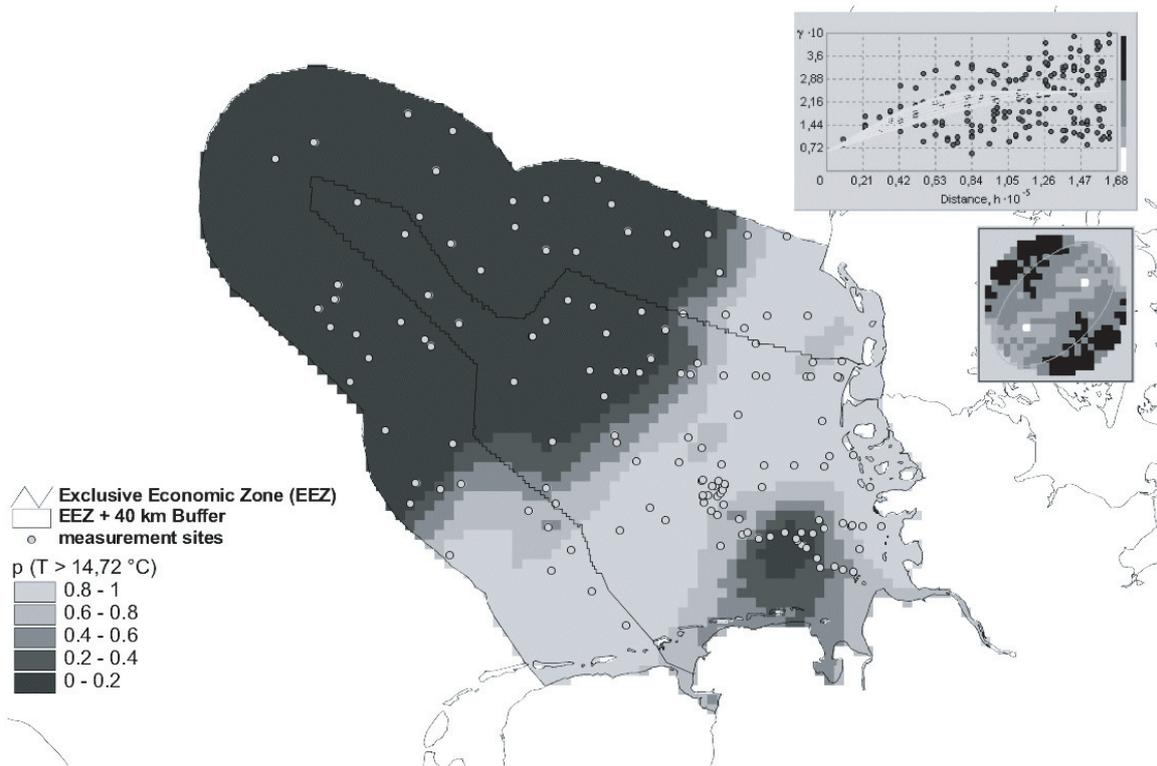
Figure 8-5: Surface estimation of bottom water summer temperatures in the EEZ in the year of 2000 obtained by indicator kriging (threshold value = median of value distribution - 14.96 °C).

8.6 Discussion

The study submitted exemplarily presents the use of geostatistical methods to produce surface estimations from marine measurement data. The results achieved represent a very good example of such an application, since spatial trends of the temperature conditions on the seafloor directly relate to the autocorrelation structure derived by means of variogram analysis. In all years under investigation, spherical variogram models were found to be best for modelling the autocorrelation structure. Strong autocorrelations could be found wherever nugget-sill ratios were very low. Furthermore, with the aid of variogram maps anisotropies could also be detected with a slower decrease of semivariances in north-easterly direction, depending upon the year of measurement. These directional dependencies allow for a realistic display of the temperature conditions near the sea floor since these conditions tend to depend upon the distance of the location at the seafloor to the coast, or alternatively, the bathymetric conditions.

In addition, each map can be described in terms of global or local quality criteria: As can be seen from the results of cross-validation these estimations revealed rather good quality measures derived from the distinct autocorrelation of variogram analysis. When comparing the MPE values in table 8-4 with the nugget-sill ratios listed in table 8-3, a highly significant
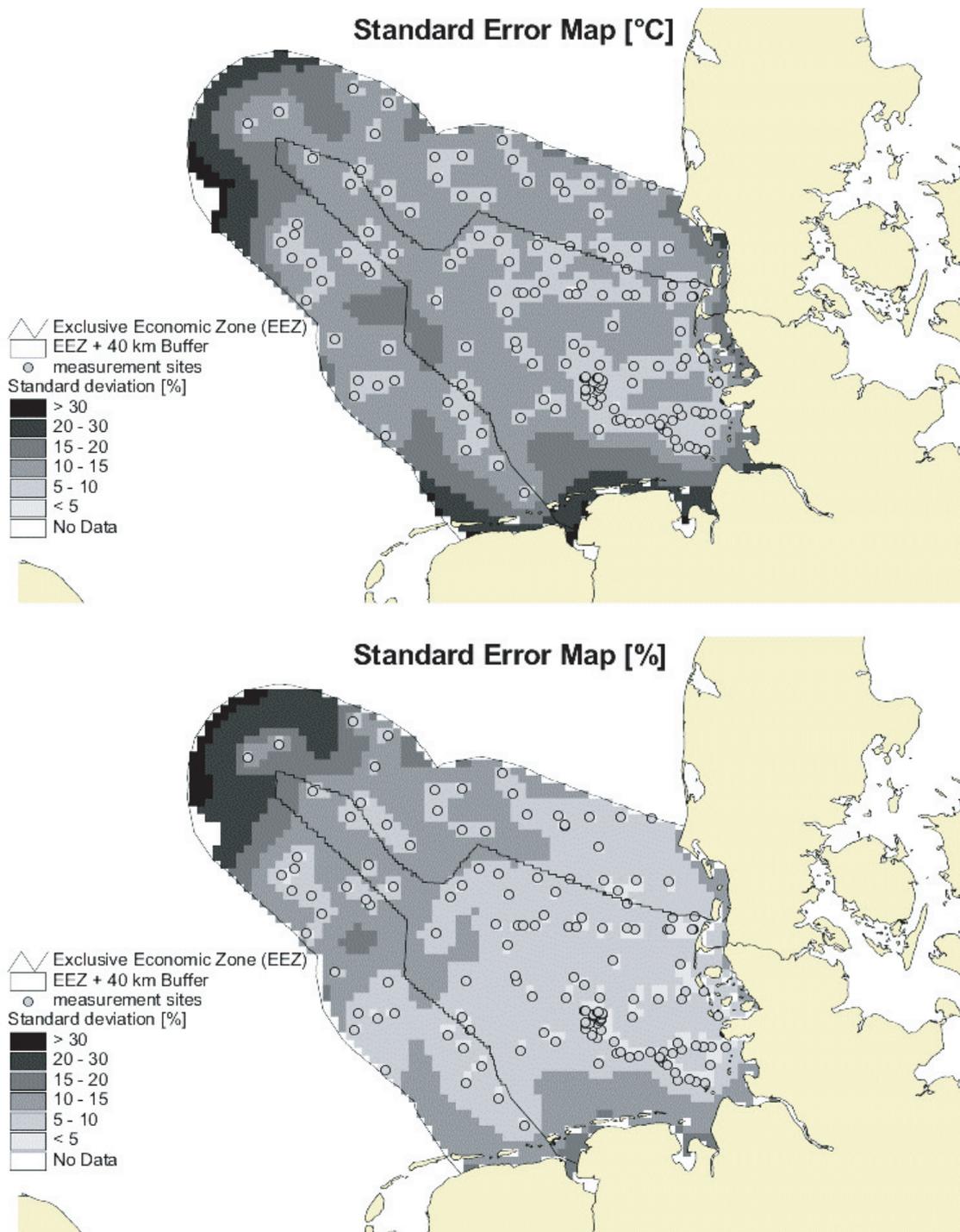
Figure 8-6: Standard error maps in [°C] and [%] for bottom water summer temperatures of the year 2000.

Spearman correlation (a = 0.01) of 0.9 can be detected. The use of kriging variance and/ or standard deviation as a local quality criterion as depicted in figure 8-6 (above) is highly disputed though (Goovaerts, 1997), since the kriging variance depends on the variogram only, and not on the measured values themselves. As exemplified in figure 8-6 (above) this can be

verified for the standard error map calculated for bottom water temperature conditions in the year 2000: The highest values lie in the greatest distance from the measurement points. One way to address this problem might be to calibrate the standard errors to the estimated values in each raster cell by setting the estimated values 100% (Fig. 8-6 below).

As was outlined in chapter, the geostatistical analysis of measurement data is only the first working step of the typology concept in the project MarGIS. The second step is to combine different geostatistically derived abiotic and biotic data layer to calculate ecological sea floor types or habitat maps. Such is currently done with help of the multivariate statistical procedure CART (Classification and Regression Trees) resulting in a habitat map for the EEZ.

9 Using decision trees to predict benthic communities within and near the German Exclusive Economic Zone (EEZ) of the North Sea

R. Pesch[1], H. Pehlke[1,] K. Jerosch[2] , W. Schroeder[1], M. Schlüter[2]

[1]Institute for Environmental Science, University of Vechta, Oldenburger Str. 97, 49377 Vechta, Germany

[2]Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

9.1 Abstract

In this article a methodological concept is introduced to predict the occurrence of benthic communities within and near the the German Exclusive Economic Zone (EEZ) of the North Sea. The approach consists of two work steps: (1) geostatistical analysis of biotic and abiotic measurement data and (2) calculation of benthic provinces by means of Classification and Regression Trees (CART) and GIS techniques. From bottom water measurements on salinity, temperature, silicate and nutrients as well as from punctual data on grain size ranges (0-20 μ, 20-63 μ, 63-2000 μ) raster maps are calculated by the use of geostatistical methods. At first the autocorrelation structure is examined and modelled with help of variogram analysis. The resulting variogram models are then used to calculate raster maps by applying ordinary point kriging. After intersecting these raster maps with punctual data on benthic communities a classication system is derived to predict the occurrence of these communities within the whole study area. This classication system is calculated from the intersected data by producing decision trees by the use of Classification and Regression Trees (CART). Since these decision trees correspond to hiarchically ordered sets of decision rules they are applied on the geostatistically estimated raster data to predict benthic habitats within and near the EEZ.

9.2 Background and objectives

The investigation of the spatial distribution of benthic organisms inhabiting the sediments on the sea floor of the North Sea has a long tradition. First attempts go back to the work of Petersen (1914), who characterised dominant benthic species, relying on bathymetrie and sediments. Since then the macrozoobenthos has been investigated in different parts of the North Sea leading to numerous publications (summarised in Glémarec et al., 1973; Kingston and Rachor, 1982). With regard to the German Bight. the survey carried out by Salzwedel et al. (1985) in October 1975 produced a detailed data set on the macrofaunal communities for

the sublitoral. The spatial distributiuon of these benthic communities were nearly affirmed in an extensive study made by Rachor and Nehmer (2003). Here data on endofaunal organisms were collected at more than 180 sites within the area of the German Bight and the bordering central North Sea. With help of multivariate statistics the abundance data was aggregated to derive eight benthic communities representative for the biological conditions at the sea floor. These communities also include information on benthic organisms that are refered to as endangered species.

The goal of this article is to calculate a prediction map for the occurrence of these benthic communities for the whole area within and near the German Exclusive Economic Zone (EEZ) of the North Sea. We refer to the resulting prediction map as a benthic habitat map able to describe the biotic and abiotic conditions at the sea floor. According to the ICES Working Group on Marine Habitat Mapping a marine habitat can be defined as "a recognisable space which can be distinguished by its abiotic characteristics and associated biological assemblage, operating at particular spatial and temporal scales." (ICES, 2005). The knowledge of such ecological units often is a prerequisite for marine planning and management needs, such as the installation of offshore wind power plants or the declaration of protection zones (Hughes, 1997; Reiniger, 1997; Lourie and Vincent, 2004). Currently, different international research groups are focussing on the derivation of such benthic habitat maps for the North Sea by using existing marine databases. In Europe the habitat types classification system of EUNIS (European Nature Information System) provides a hierarchical concept to characterise habitats in the European terrestrial and the marine environments. Our benthic habitat mapping concept depends on availible marine datasets and statistical methods as well as GIS technologies and is based on two work steps:

(1) With the help of geostatistical methods abiotic measurement data on grain size ranges as well as on salinity, nutrients, dissolved oxygen, silicate and temperature for the lower water body are spatially extrapolated in terms of raster maps.

(2) Together with site specific data on benthic organisms the calculated raster maps are used to derive a hierarchical classification system for eight benthic communities derived by Rachor and Nehmer (2003) by means of Classification and Regression Trees (CART). The resulting decision tree is applied to predict the occurrence of these benthic communities for the EEZ.

9.3 Statistical methods
9.3.1 Geostatistical methods
Originally coming from geological research applied to estimate mineral resources and reserves (Krige, 1951; Matheron, 1965; 1971), geostatistics are nowadays used in various terrestrial and marine fields of research. As far as marine research is concerned geostatistical instruments were applied by scientific disciplines like geochemistry (Montgomery et al., 1998), pollution

research (Poon et al. 2000), sedimentology (Zeiler et al., 2000), geology (Chihi et al., 2000) and biology (Harbitz and Lindstrøm, 2001; Jelinski et al., 2002). The latter exhibits a multitude of examples of geostatistical applications serving the interests of various fisheries (Petitgas, 1997; Lembo et al., 1999; 2000; Rivoirard and Wieland, 2001).

Geostatistical procedures are based on the theory of regionalised variables, first introduced in the doctoral thesis of Matheron (1965): "Les variable régionalisées et leur estimation". A regionalised variable *z(x)* is understood as a variable whose respective values depend on the location *x*. The changes of the values of this variable in space are "... neither completely statistically detectable nor deterministically recordable by mathematics in terms of exact formulae" (Akin and Siemens, 1988, pp. 27 therein). In geostatistics the realisations of $z(x_1, x_2, x_3, ... x_n)$ are therefore seen as purely random in a way that they result from a set of random variables $Z(x_1, x_2, x_3, ... x_n)$. Each of these random variables follows a probability distribution with a mean μ and a variance σ^2. By disregarding this random component the spatial variability of the occurring values may depend on an underlying spatial process referred to as spatial autocorrelation. Spatial autocorrelation expresses the notion that objects that are close to each other tend to be more alike than if they were farther apart. Such spatial relationships may be examined by means of variogram analysis. Variogram analysis consists of two work steps: calculation of an experimental semivariogram and fitting a mathematical variogram model to the experimental semivariogram. The variogram model serves as a key parameter for the kriging estimation.

Following Matheron (1963) the semivariogram can be estimated from any empirical dataset by computing half of the mean squared differences of the measured values of pairs of points *(γ)* within given distances *(h)*:

$$\gamma(h) = \frac{1}{2\,N(h)} \sum_{i=1}^{N} (Z(x_i) - Z(x_i+h))^2 \qquad [1]$$

where *N(h)*= number of sample pairs separated by distance *h*.

There are two ways to calculate such an experimental semivariogram: (1) By determining semivariances for defined distance intervals and radial angle sectors, and (2) by using so-called variogram maps. The variogram analyses reported in this article were based on the second approach which is therefore referred to shortly. Four work steps can be distinguished in modelling the spatial autocorrelation structure with the aid of variogram maps (Fig. 9-1):

(1) Forming direction vectors also referred to as lag vectors from all point pairs of the considered point distribution;

(2) Connection of the origins of all *lag vectors*;

(3) Overlaying the result of (2) with a grid of defined mesh size and assigning semivariogram values for each cell respectively *bin* according to equation [3] (The semivariogram values

are weighted according to a Kernel function depending on the distance of the ending of each lag vector to the centre of each cell (Johnston et al., 2001, pp. 254 therein) → resulting in a *variogram map*;

(4) Assigning the calculated semivariances to a coordinate system defined by separation distance (abscissa) and semivariance (ordinate) → resulting in the *experimental variogram.*
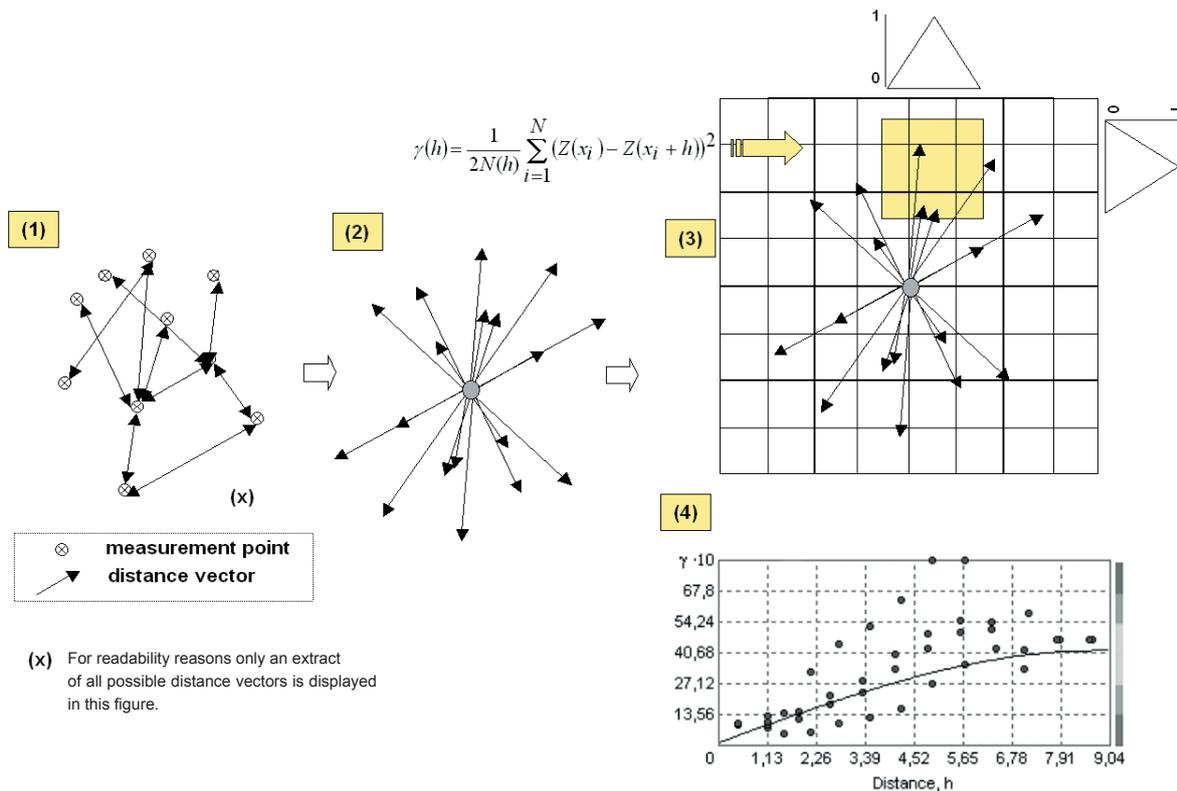
Figure 9-1: Methodology of variogram analyses by means of variogram maps (following Johnston et al., 2001).

Note that variogram maps are point-symmetrical relative to their origin because all lag vectors have a counterpart in the opposite direction (Fig. 9-1). Variogram maps automatically give hints about so-called anisotropies in the data field. Anisotropies are properties of the underlying spatial process where spatial dependence not only changes with distance but also with direction. Examples of such processes can be found, for example, in ecology where ecological patterns strongly depend upon topography.

It is necessary to adapt a defined variogram model to the experimental variogram in order to perform kriging. This can be achieved by means of mathematical models fitted to the experimental variogram in terms of a least-squares regression line. The software Geostatistical

Analyst distinguishes eleven such models (Johnston et al., 2001), from which above all the spherical and the exponential models are used most frequently. For detailed mathematical descriptions, please refer to Akin and Siemens (1988), Olea (1999), Webster and Oliver (2000) as well as Johnston et al. (2001). The variogram model may be described with respect to three parameters: sill, range, and nugget-effect. The range equals the maximum separation distance within which a distinct increase of semivariogram values can be observed. This is indicative of spatial autocorrelation. The sill corresponds to the semivariance assigned to the range. If anisotropies can be detected, both sill and range will vary with respect to direction. Small-scale variabilities or measurement errors may lead to high semivariances at nearby locations. The variogram model refers to this in terms of the so-called nugget effect, where the variogram model cuts the ordinate above the origin. A pure nugget effect indicates a complete lack of spatial autocorrelation.

Cross-validation can be used to choose the optimal variogram model. For this purpose, each measurement value is extracted from the dataset and estimated by means of the selected variogram model and the kriging method to be applied. By subtracting the measured value from the estimated value an estimation error can be calculated resulting in an error distribution for the whole dataset. Various key parameters can be calculated from the distribution to characterise the quality of the chosen variogram model. Examples of such parameters are the mean error (ME), root mean square standardised error (RMSSE) and the root mean square error (RMSE) (Johnston et al. 2000). The ME equals the average value of the cross-validation errors and describes the tendency of the variogram model to over- / underestimate the measurement values. At best, ME should equal 0. The RMSSE corresponds to the ratio of (experimentally derived) mean squared cross-validation errors and the (theoretically derived) kriging variances. Ideally, the RMSSE should equal 1. The RMSE equals the average of the root of the mean squared deviation between measured and predicted values. The RMSE should be as small as possible.

By applying a variogram model to the available information of the spatial variation pertinent to a certain property kriging can be used to make predictions for blocks (block kriging) or points (point kriging). The term kriging was invented by Matheron (1963) in order to recognise the contribution of the mining engineer D.G. Krige to improve the estimation of metal concentrations in geological deposits (Webster and Oliver, 2000). A large number of kriging options are available depending on the assumptions the user makes, (e.g. simple, ordinary or universal kriging). These methods all have in common that they minimise the estimation variance and rely on weighted averaging of the measured values within a chosen searching or kriging window. The results presented in this article were calculated by the use of ordinary point kriging. Upon applying ordinary kriging it is assumed that the mean values of all random variables μ are constant at all locations, although unknown. Ordinary kriging is a linear kriging

method in which all measurement values within a defined search window are averaged with respect to the distance of the respective measurement point to the point to be predicted. The weighting factors are calculated so they add up to one, so that the estimation error by average equals zero and the estimation variance is minimised (Akin and Siemens, 1988). Whether or not anistropies are detected in the variogram analyses determines whether this kriging window is circular (no anisotropies) or ellipsoidal (different ranges of spatial autocorrelation for different directions).

9.3.2 Classification and regression trees (CART)

In this article CART is introduced as a statistical procedure to derive a classification system or decision tree, respectively, for the occurence of benthic communities in the area of the German EEZ of the North Sea. CART is applied in various scientific disciplines to uncovers hidden structures in complex data matrices and to predict the characteristics of a chosen target variable by a set of meaningful predictor variables (Breiman et al., 1984). In psychological and medical research, for example, decision trees are applied to understand and predict human characteristics, behaviours and even deseases (Swan et al., 2004; Arentze and Timmermans, 2005; Rosenfeld et al., 2005). In environmental sciences application examples can be found in environmental monitoring (Ryan, 1995; Schröder and Schmidt, 2003), global change biology (Thuiller, 2003), applied forestry (Fin et al., 2003; Lawrence and Labus, 2003) and meteorology (Walmsey et al., 2001). In marine biology Norcross et al. (1999) used CART to classify nearshore flatfish habitats in Alaska's waters. Huettmann and Diamond (2001) predicted and modelled the distribution of seabirds in the Canadian North Atlantic by applying CART.

CART handles both categorical and metric data without data transformation.and produces decision trees to display class memberships by recursively partitioning a heterogeneous data set into subsets (also called classes, groups, nodes) by means of a series of binary splits. The aim is to create subsets that improve in terms of homogenity according to the features of the target variable. How each node is split into two subnodes is determined with the help of the predictor variables. Whether the target variable is of metric, ordinal or nominal scale dignity different impurity measures exist. The Gini index is commonly used when the target variable is categorial, allthough other options exist (entropy, twoing index) (Steinberg and Colla, 1995):

$$g(a) = 1 - \sum^{i} P(i \mid a)^2 \qquad\qquad [2]$$

In equation [2] $g(a)$ represents the gini measure of a given node a and is determined by the substracting the sum of all squared probabilities of all features i of the target variable from 1. The Gini index reaches 0 if all cases in the node belong to the same category (optimal homogenity) and reaches its maximum when all cases are evenly distributed within the node.

To choose the optimal binary split the maximum reduction of inhomogeneity is calculated over all possible splits by using the following equation (Yisehac and Webb, 1999):

$$\Delta i(s,t) = i(t) - p_L \, [\, i(t_L)] \; - p_R \, [\, i(t_R)] \qquad\qquad [3]$$

Here $\Delta i(s,t)$ stands for the decrease in impurity or the improvement at a particlular split s of node t. The terms p_L and p_R are the proportions of the cases at node t that go into the left and right subnode and $I(t_L)$ and $I(t_R)$ are the impurities of the left and the right subnode.

The growing of decision trees are processed until a maximum tree is reached depending on user specified restrictions, e.g. insufficient number of cases in a produced node or until further splitting is impossible (only one case or identical cases in the node). Smaller trees can be produced by pruning the maximum tree either automatically or interactively by expert judgements. In this way all decision trees result in so called endnodes each of which are the product of a hierarchical sequence of decision rules. By deviding all misclassified cases through the total number of cases a prediction quality measure may be expressed in terms of a misclassification rate or risk estimate.

9.4 Methodology and procedures

9.4.1 Data basis and preparation

The data used to predict benthic communies was collected within the framework of the project MarGIS financially supported by the German Federal Ministry of Education and Research. Along with thematic surface data (e.g. sediment data maps) most of this data consist of biotic (e.g. data on benthic organisms) and abiotic measurements of the lower water column. The datasets were provided by several national and international databases and projects and integrated in a relational Database Management System (rDBMS). Approximately 235,000 abiotic measurement data sets have been collected from the International Council for the Exploration of the Sea (ICES), the Marine Environmental Data Base (MUDAB) of the German Federal Maritime and Hydrographic Office (BSH) and the German Federal Environmental Agency (UBA) as well as the Institute of Marine Research (IfM), University of Hamburg for the North and the Baltic Sea.

For the calculations performed here data on eight benthic communities derived in the summer of 2000 by Rachor and Nehmer (2003) (Fig. 9-2) as well as abiotic measurement data on salinity, temperature, silicate nitrate, amonium and on grain size ranges (0-20 µ, 20-63 µ, 63-2000 µ) was taken from the MarGIS database for the study area. The latter was extended to a 40 km buffer around the EEZ in order to improve the geostatistical estimations for the outer parts of the EEZ (→ chapter 9.4.2). Since the abiotic data was to serve as the predicting variables in the CART analysis described in chapter 9.4.3, the data was aggregated over a six
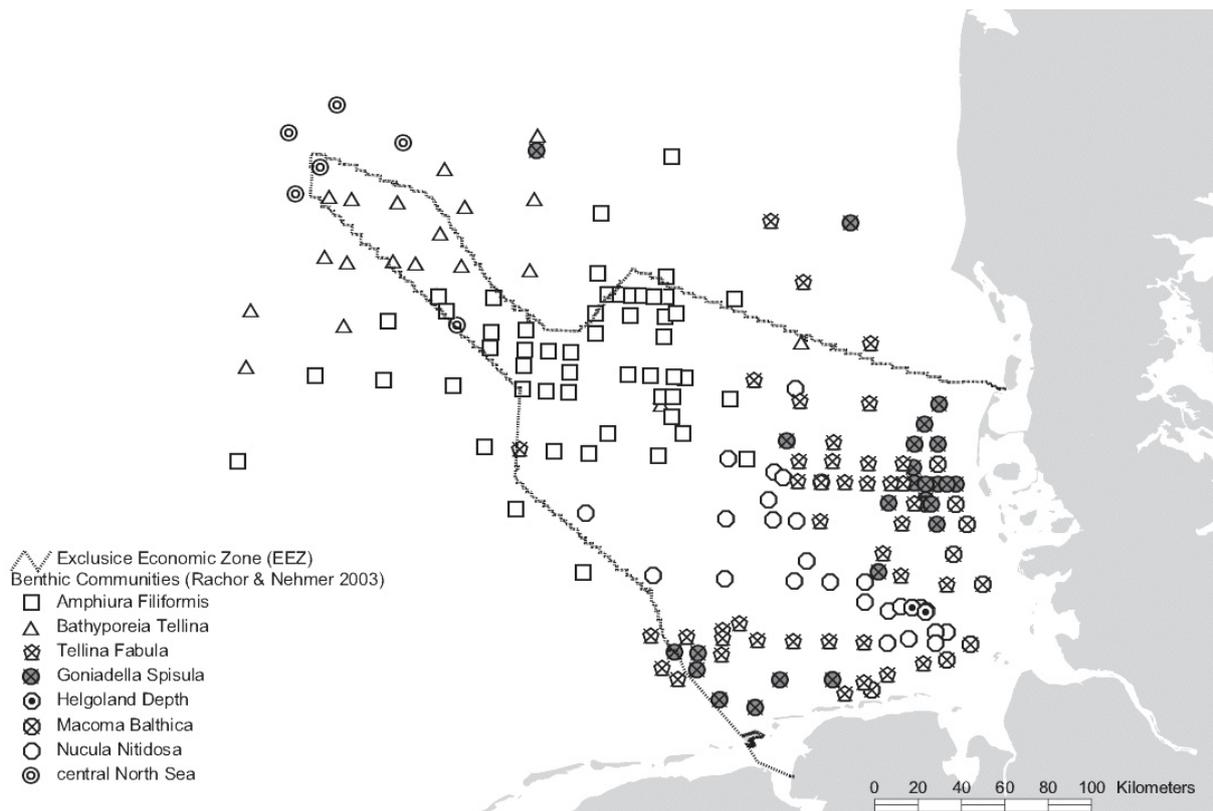
Figure 9-2: Benthic communities within the German Bight and bordering central North Sea (Rachor and Nehmer, 2003).

Table 9-1: Data basis for the prediction of benthic communities within the EEZ plus a 40 km buffer.

Parameter	Months	Sample size	Mean nearest neighbor distance [m]
Ammonium	January - March	256	4838
	July - September	197	6638
Nitrate	January - March	198	6046
	July - September	239	4076
Dissolved oxygen	July - September	232	5925
Phosphate	January - March	292	4526
	July - September	247	4270
Salinity	January - March	907	1351
	April - June	670	2200
	July - September	919	2040
	October - December	644	1199
Silicate	January - March	304	3745
	July - September	223	4993
Temperature	January - March	913	1384
	April - June	668	2191
	July - September	921	2030
	October - December	645	1184
Grain size range 0 - 20 µ		674	1454
Grain size range 20 - 63 µ		662	1451
Grain size range 63 - 2000 µ		668	1443

131

year period from 1995 to 2000 according to four monthly time intervals: January to March; April to June; July to September and October to December. By integrating these datasets in a GIS program the spatial distribution of the site specific data was checked with respect to its spatial distribution. Only those datasets were considered for further statistical analyses that covered the whole area of interest. These are summarised in table 9-1.

In addition to the abiotic parameters listed in table 9-1 an already existing map on bathymetry was included in the calculation described in chapter 9.4.3. This map was made availible by the Alfred Wegener Institute, Bremerhaven.

9.4.2 Geostatistical analysis of abiotic measurement data

The extension 'Geostatistical Analyst' from ArcGIS 9.0 was used to calculate raster data on bottom water measurements on salinity, temperature, silicate, dissolved oxygen, phosphate and nitrate as well as on grain size ranges (0-20 µ, 20-63 µ, 63-2000 µ). The calculations were performed for a six year period from 1995 to 2000 according to the four monthly time intervals carried out in chapter 9.4.1. The Universal Transverse Mercator (UTM) coordinate system (Zone 32) was used as a spatial reference system.

According to Webster and Oliver (2000) the mean distance of each measurement site to its nearest neighbour was set as a starting point for the lag size (Tab. 9-1). Whether or not the autocorrelation structure could be detected in the variogram window a multitude of this distance was applied. The number of lags allocated to each lag size was set so that the distance of significant autocorrelation (range) became clearly visible in the variogram window. In a next step we tested various variogram models, including different types of models, sills, ranges, and nugget effects. Finally, if the semivariances displayed on the variogram map indicated anisotropies in the data field, different ranges for different directions (to account for anisotropies) were compared with each other. All variogram parameters (range, sill, nugget-effect, type of variogram model, anisotropies) were chosen with respect to the ME, the RMSSE and the RMSE derived from the results of cross-validation (→ chapter 9.3.1).

Ordinary point kriging was applied to use the information of the variogram model for surface estimations. The grid cell size was set with respect to the average mean distance of neighbouring point pairs over all parameters for all time intervals, here 3150 times 3150 meters. The kriging window, which includes the measurement values to estimate a certain point, was adjusted to the range of the variogram model. Since anisotropies were detected in each year under study, the search windows had both major and minor axes (Johnston et al. 2001), hence corresponding ellipses. A four-sector neighbourhood was defined to avoid directional bias. A maximum of five points was accounted for in each sector to estimate a certain point.

Two other key parameters, primarily suited for empirical scientists to assess the quality of the estimations performed, were calculated from the cross-validation results: The correlation
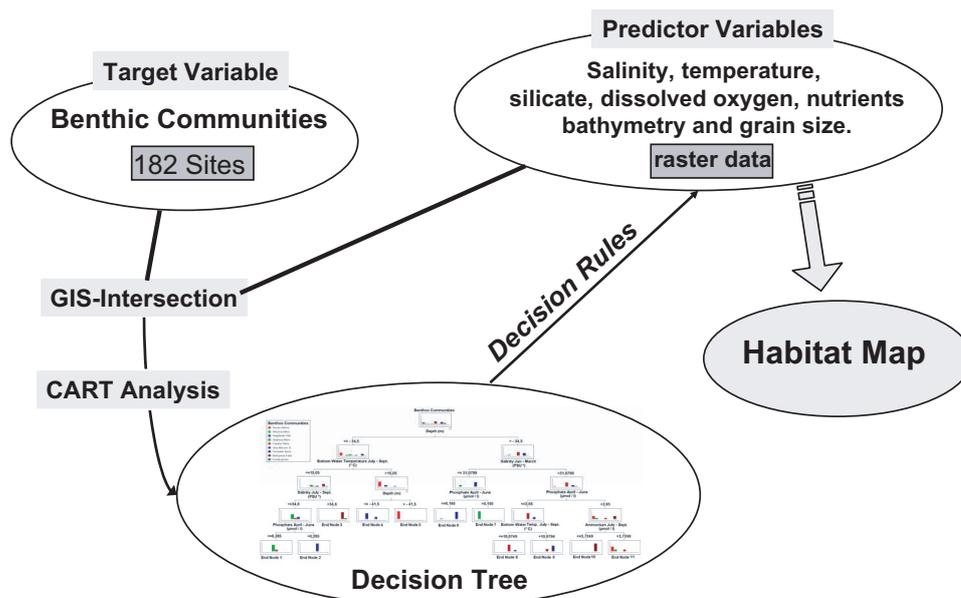
Figure 9-3: Predictive benthic habitat mapping by CART.

coefficient after Spearman (CS) and the Mean Percentile Error (MPE): In case of an ideal correlation the CS-value should be equal to 1, if no such correlation exists CS tends towards 0. By setting the measured value to 100% the cross-validation error may be quantified relative to the measured value. MPE is the median value of all relative cross-validation errors in the dataset.

9.4.3 Predictive benthic habitat mapping by CART

Our predictive habitat mapping concept relies on the derivation of CART decision models for the occurrence of benthic communities availible at 182 sites within and near the German EEZ (Fig. 9-2). The benthic communities were statistically derived by Rachor and Nehmer (2003) from abundance data on benthic epifauna organisms collected within the study area. They distinguish eight different communities: *Amphiura filiformis* (53 sites), *Tellina fabula* (40 sites), *Goniadella spisula* (27 sites), *Nucula nitidosa* (24 sites), *Bathyporeia tellina* (20 sites), *Macoma balthica* (8 sites), benthic organisms of the central North Sea (5 sites) as well as of the Helgoland depth (4 sites). The methodology to spatially predict the occurrence of these eight benthic communities for the entire study area is summarised in figure 9-3. The CART-calculations were performed with the SPSS-module 'Answer Tree'.

At first, all geostatistically estimated abiotic raster data and the bathymetry map (chapter 9.4.2) were intersected with the punctual data on benthic communities, resulting in a table build up of 182 rows and 24 columns representing X-/Y-coordinates, the eight benthic communities as well as the 20 abiotic parameters listed in table 9-1 and the bathymetry. By defining the benthic communities as the target variable and the abiotic parameters as the predictor variables a

Table 9-2: Key parameters from variogram analysis.

Parameter	Months	N-S-Ratio	A-Ratio	Direction
Ammonium	January - March	0.29	0.59	16.7
	July - September	0.31	0.61	56
Nitrate	January - March	0.13	0.39	0.1
	July - September	0.48	0.43	58.2
Dissolved oxygen	July - September	0.31	0.69	51.8
Phosphate	January - March	0.32	0.47	356.5
	July - September	0.23	0.53	40.2
Salinity	January - March	0.16	0.56	340.6
	April - June	0.02	0.73	339.4
	July - September	0.05	0.60	356.2
	October - December	0.25	0.65	327.5
Silicate	January - March	0.18	0.43	352.2
	July - September	0.39	0.74	40.3
Temperature	January - March	0.57	0.73	341.6
	April - June	0.27	0.91	16.2
	July - September	0.19	0.61	55
	October - December	0.52	0.48	93.1
Grain size range 0 - 20 µ		0.17	0.91	341.7
Grain size range 20 - 63 µ		0.22	0.73	23.4
Grain size range 63 - 2000 µ		0.06	0.74	18.6

decision tree was calculated by applying CART. The tree was grown in a way that the minimum number of cases per endnode did not exceed four. This corresponds to the number of sites belonging to the smallest benthic community, here the community "Helgoland depth" (n=4). Additionally the binary splitting was not continued when the improvement of homeginity lay below 0.001. Finally, the resulting tree was pruned back so that each of the resulting endnodes at best were dominated by one of the eight communities. Since the decision tree corresponds to a hiarchically ordered set of decision rules these rules were written into an SQL-statement and applied on the raster data availible for the entire study area. In this way a prediction map was generated with as many benthic habitat classes as there were endnodes in the underlying decision tree. The benthic prediction map at last was intersected with the prediting variables to statistically describe all habitat classes with respect to the abiotic conditions.

9.5 Results

9.5.1 Variogram analyses and surface estimations

The results of the geostatistical analyses are only carried out in parts. For a detailed description of the results of both variogram analysis and kriging maps please refer to Pehlke (2005).

The results of variogram analysis depict distinct autocorrelations patterns for all abiotic parameters and time intervals (Tab. 9-2). Almost all calculated spherical variogram models

show low nugget-sill ratios (n-s-ratio in table 9-2), ranging from 0.02 (salinity [PSU]: April - June) to 0.57 (temperature [°C]: January - March). To 90% the nugget-sill ratios lie below 0.5 which is indicative for low small-scale variabilities and strong autocorrelations of the measurement values. With the help of variogram maps anisotropies could furthermore be detected resulting in searching ellipses instead of circles in the succeeding kriging interpolations. The form of these ellipses may be described by means of the anisotropy-ratio (a-ratio in table 9-2) which is the proportion of the minor axis of the ellipse to its major axis. A low anisotropy-ratio is therefore equivalent to a narrow searching ellipse resulting in parallel spatial structures in the kriging maps. The direction of the major axes of the searching ellipses of all 20 parameters mirror different parts of the coastline of the German, Danish and Dutch coastal parts of the study area. These reach from 327.5° for the salinity conditions from October to December (NNE direction) to 93.1° for the temperature conditions in the same time interval (E direction). All variogram models were used to produce surface maps for each parameter and time interval. An example of the results of such an ordinary kriging map is given in figure 9-4 by the temperature conditions at the bottom of the sea floor of the study area from July to September. The spherical variogram model depict a strong autocorrelation of the measurement values with a nugget-sill-ratio of 0.19 [µmol/l]². The anisotpopy ratio is 0.61 meaning that the length of the minor axis of the searching ellipse is 61% of the length of the major axis. The direction of the
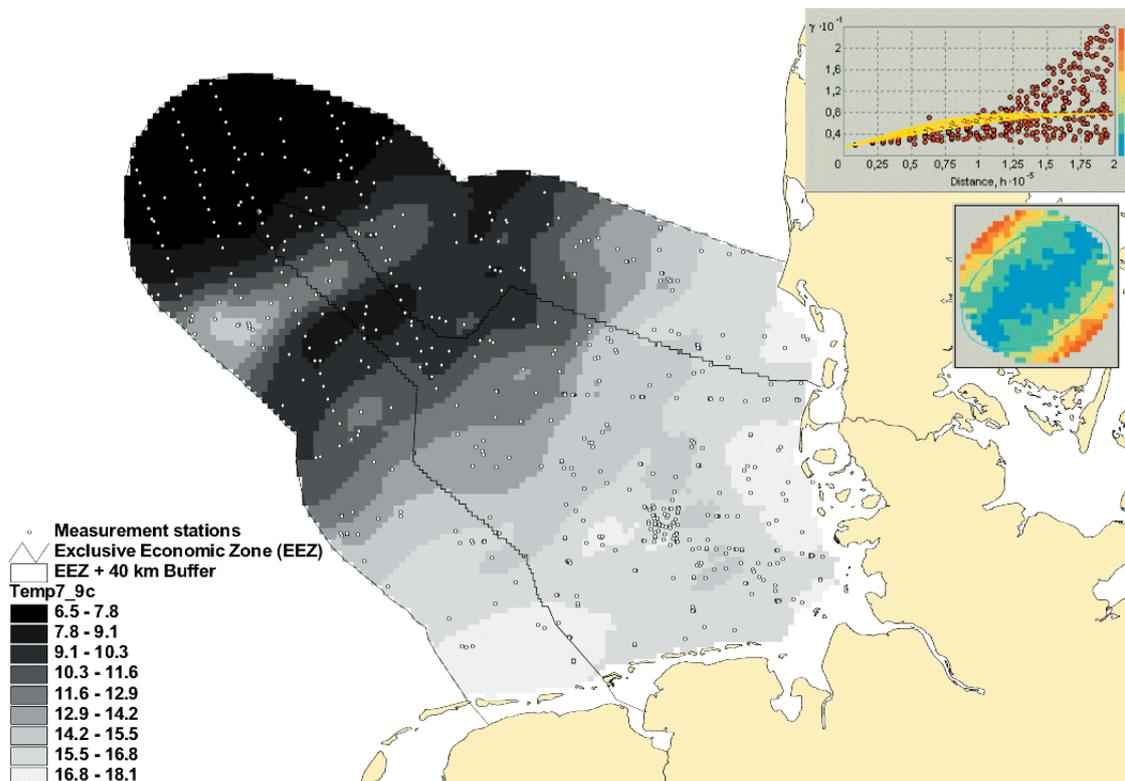

Figure 9-4: Temperature conditions at the bottom of the sea floor of the German EEZ from July to September.

latter points to 55° (NE) explaining the spatial structure of the estimated raster values.

To address the quality of the surface estimations performed key parameters were extracted from the results of cross-validation. These are summarised in table 9-4, which demonstrates the ME, RMSE, MPE as well as the Cs values. Both ME and RMSSE indicate neither underestimation or overestimation and therefore no bias in the surface estimations: ME shows that the average cross-validation errors equal almost zero in all cases. RMSSE equals almost 1 for all parameters and monthly time intervals indicating that variances calculated from the cross-validation errors by average equal the theoretical kriging variances. MPE can be observed to be low for the grain size range 63-2000 µ as well as for all temperature estimations. With acception to the monthly time interval from January to October this also holds true for the salinity estimations. Highest MPE's can be found for the grain size range 20-63 µ (71.7%), bottom water ammonium for January to Marc (33.8%) as well as silicate for July to September (32.8%). Except for temperature for October to December all coefficients of correlation after Spearman lie above 0.5. In 45% of all cases Cs lies above 0.8 indicating high degrees of associations between the measured and estimated values. Cs is permanently high for bottom water salinity for all monthly time intervals.

9.5.2 Benthic habitats within and near the EEZ

The results of the CART analysis resulted in a decision tree grown in nine binary splits leading to 10 endnodes or classes, respectively. Figure 9-5 depicts the nodes of the decision tree in terms of histograms where each bar is representative for one of the eight communities. By following the dendrogram from up to down it can be observed that the portion of each benthic communities
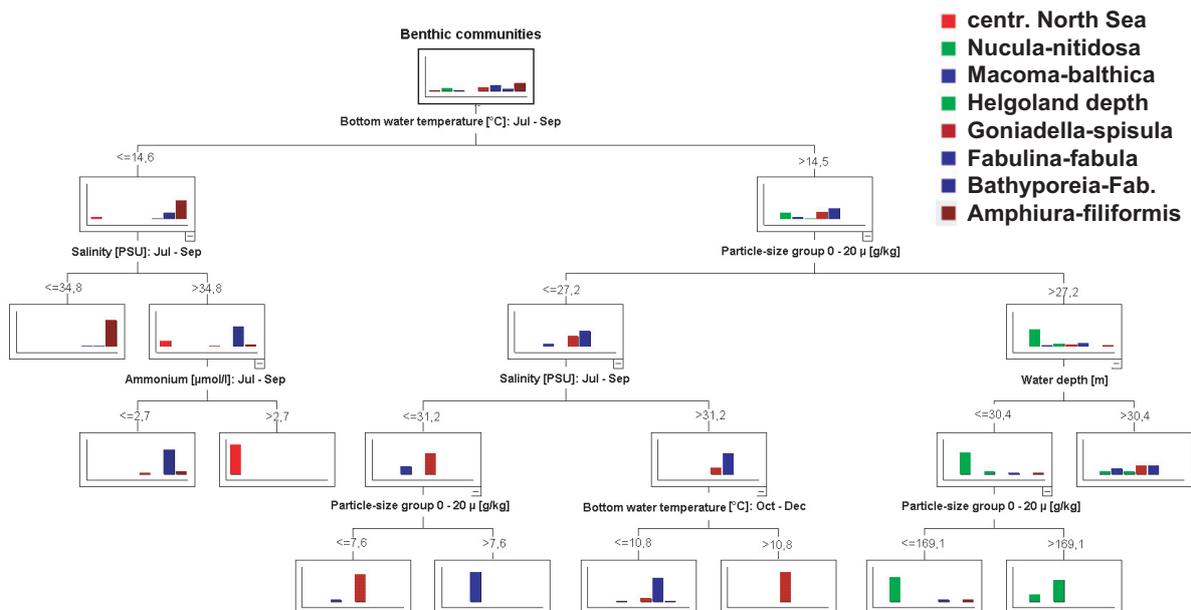

Figure 9-5: Decision tree for the occurence of eight benthic communities derived by Rachor and Nehmer (2003).

Table 9-3: Quality of estimation by means of cross-validation.

Parameter	Months	ME	RMSE	MPE	K
Ammonium	January - March	0.00	0.94	33.8	0.78
	July - September	0.00	1.00	31.9	0.88
Nitrate	January - March	0.00	0.98	22.7	0.67
	July - September	0.00	0.98	22.7	0.53
Dissolved oxygen	July - September	0.00	0.98	22.7	0.88
Phosphate	January - March	0.01	1.10	31.5	0.63
	July - September	0.00	0.99	14.7	0.74
Salinity	January - March	0.00	1.05	27.4	0.77
	April - June	0.00	1.01	0.6	0.91
	July - September	0.02	1.06	0.4	0.95
	October - December	0.00	1.02	0.2	0.97
Silicate	January - March	-0.03	0.98	19.3	0.87
	July - September	-0.04	1.08	32.8	0.56
Temperature	January - March	0.00	1.01	10.5	0.60
	April - June	0.04	1.06	9.3	0.67
	July - September	0.01	0.98	4.5	0.93
	October - December	-0.02	1.04	6.7	0.42
Grain size range 0 - 20 µ		-0.50	0.99	30.22	0.81
Grain size range 20 - 63 µ		-0.95	0.99	71.71	0.70
Grain size range 63 - 2000 µ		1.66	1.04	1.62	0.86

increases stepwise This leads to nine endnodes in which one of the eight communities is dominant (portion >75%). Each community is represented once except for Goniadella Spisula that can be found in both endnode 33 (93%) and endnode 36 (100%). Endnode 32, consisting of 10 sites, is evenly represented by five communities. The nine binary splits chosen to recursively partition the 182 sites were chosen according to six predictor variables: bottom water temperature [°C] July - September (once) and October - December (once), bottom water salinity [PSU] July - September (twice), bottom water ammonium [µmol/l] July - September, water depth [m] (once) as well particle size group 0 - 20 µ [g/kg] (three times).

The goodness of the decision tree to predict the eight benthic communities may be described with the help of the missclassification matrix depicted in table 9-4. Here for each community both succesfully classified and misclassified cases are listed. It can be seen that, except for Macoma Balthica, all other communities are succesfully classified with at least 75%. The smallest misclassification rate can be found for *Nucula Nitidosa* (12.5%) as well as *Amphiura Filiformis* (7.55%). By deviding all misclassified cases through the total number of cases a overall misclassification rate or risk estimate of merely 16% can be calculated.

Since each endnode is determined by decision rules, the tree can be applied to predict benthic communities at sites where such information is not available. By doing this for all geostatistically estimated raster cells within the EEZ and its bordering areas, a predictive habitat map was

Table 9-4: Misclassification matrix for the prediction of the eight benthic communities.

		Category							
		cNS	NN	MB	HD	GS	TF	BT	AF
	centr. North Sea (cNS)	5	0	0	0	0	0	0	0
	Nucula nitidosa (NN)	0	21	0	0	0	2	0	2
	Macoma balthica (MB)	0	0	4	0	0	0	0	0
Predicted	Helgoland Depth (HD)	0	1	0	3	0	0	0	0
	Goniadella spisula (GS)	0	1	3	1	21	3	0	0
	Tellina fabula (TF)	0	0	1	0	5	33	1	0
	Bathyporeia tellina (BT)	0	0	0	0	1	0	17	2
	Amphiura filiformis (AF)	1	1	0	0	0	2	2	49
Total		6	24	8	4	27	40	20	53
Misclassified		**1**	**3**	**4**	**1**	**6**	**7**	**3**	**4**
Misclassified %		**16.67**	**12.50**	**50.00**	**25.00**	**22.22**	**17.50**	**15.00**	**7.55**

calculated with respect to the occurrence of benthic communities derived by Rachor and Nehmer (2003). This a map is depicted in figure 9-6. Each of the 10 endnodes described in figure 9-5 corresponds to one of the 10 spatial units illustrated in figure 9-6. Each of these spatial units may therefore be described with respect to the possibility of the occurrence of one of the eight communities. This possibility of the occurrence of each community can be derived from its percentage in the corresponding endnode.

9.6 Discussion

The habitat map described in chapter 9.5.2 is the product of the application of statistical methods on marine measurement data. These data were taken from international marine data bases consisting of biotic and abiotic measurement data collected by approved international research institutes like the Alfred Wegener Institute, Bremerhaven. Therefore the quality of the measurement data used for the predictive benthic habitat mapping can be assumed to be assured.

As was carried out in chapter 9.5 the first methodological step to calculate the predictive benthic habitat map illustrated in figure 9-6 consisted of the geostatistical analysis of abiotic measurement data. Except for the sediment data the surface estimations were caclulated for four monthly time intervals within the year aggregated over a six year period from 1995-2000. On the one hand this had to done because otherwise a sufficient coverage of measurement values could not have been ensured to perform the geostatistical analysis. Secondly a the calculated raster maps were to be representative for a sufficient long time period. Kröncke and Bergfeld (2001) suggest that such a time period should at least cover three years. Such an approach might be seen as problematic because the variations of the measurement values are assumed to be to great to be aggregated over such a long time interval. Table 9-5 lists the variations of the measurement values at places were sampling was performed more than once
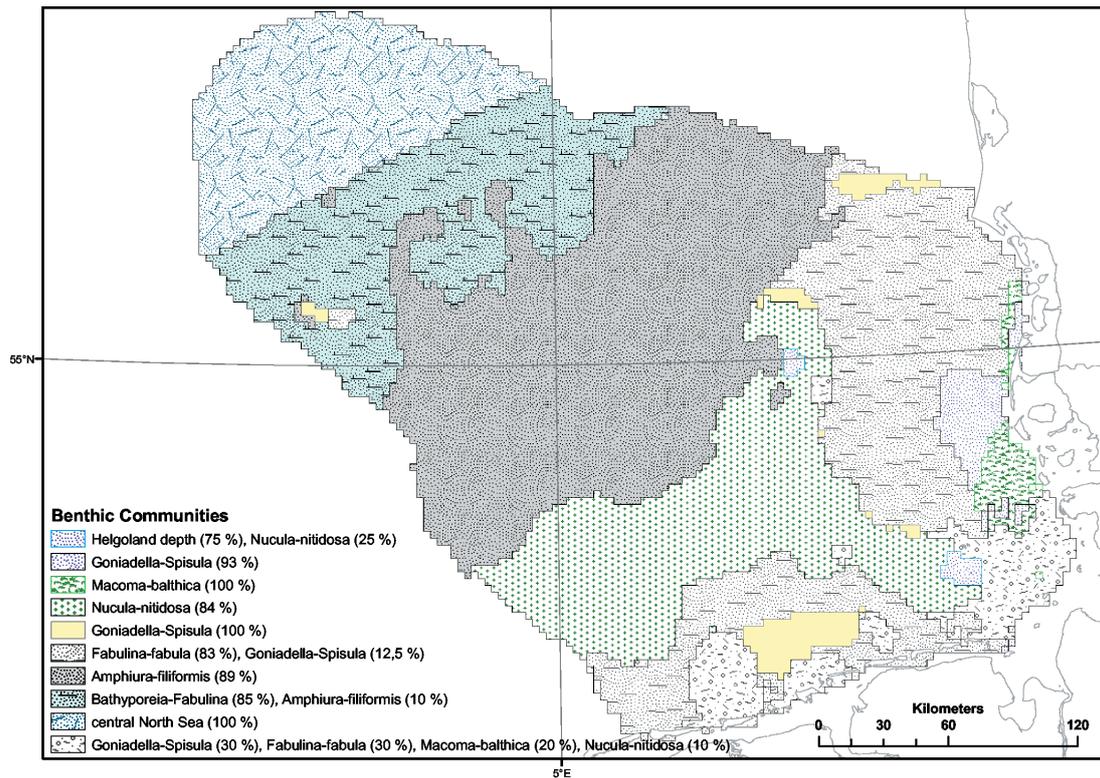
Figure 9-6: Predictive benthic habitat map for the German EEZ of the North Sea and bordering area.

between 1995 and 2000. Whereas high mean coefficients of variation can be observed for all nutrients incl. dissolved oxygen this is not the case for the parameters of interest in terms of the decision tree in figure 9-5. Only bottom water ammonium between July - September shows high mean coefficients of variation (38.7% from multiple measurements at 34 stations). The bottom water temperature between July - September shows mean coefficients of 3.26% calculated from 241 measurement stations where samples were taken up to 21 times within the six year period. For the same parameter measured between October and December the mean coefficient is 9.79% (number of measurement sites = 66). The bottom water salinity values seem to be stable throughout the investigation period. All mean coefficients amount to less than 2%.

Disregarding the variations depicted in table 9-5 the variogram analyses of all parameter and time intervals listed in table 9-1 resulted in variogram models indicating distinct autocorrelation structures. The variogram models were therefore used to calculate surface estimations. These were then set as the predictor variables to build a classification tree for the eight benthic communities derived by Rachor and Nehmer (2003). The resulting decision tree depicted in figure 9-5 describes a realistic classification system for benthic habitats because the selection of predictor variables including bottom water temperature, salinity, ammonium as well as water depth and grain size mirrors the state of scientific knowledge. Peterson (1914), Remane (1940),

Thorson (1957), Jones (1950), Duineveld et al. (1991), Künitzer et al. (1992), Salzwedel (1985) and Rachor and Nehmer (2003) refer to the bathymetry and sediment as the dominant abiotic parameters to characterise benthic habitat complexes. Künitzer et al (1992) furthermore claim bottom water temperature and the supply of nutrients to be responible for the spatial distribution of the faunistic communities. According to OSPAR (2000) changes of the benthic populations correlates among other things with changes of the degree of eutrophication. The meaning of bottom water salinity for the abundance of epibenthic organisms was proven in an investigation made by Callaway et al. (2002).

9.7 Acknowledgements

Section D: Synthesis

„Nur durch die Zusammenfassung
aller Geowissenschaften dürfen wir hoffen,
die Wahrheit zu ermitteln."

Alfred Wegener, 1880-1930
deutscher Polarforscher

10 Synthesis

This chapter outlines the main results of the previous chapters, with special emphasis on the questions and objectives mentioned in chapter 1.4. It further summarises the major outstanding problems and possible guideline for further projects.

10.1 Key results

In this cumulative thesis seafloor provinces are defined for the Håkon Mosby Mud Volcano (HMMV) using georeferenced video mosaics images, bathymetric, morphological and temperature data to identify biogeochemical habitats of chemoautotrophic communities and a suggested source location of ascending mud.

For the German Exclusive Economical Zone (EEZ) of the North Sea provinces are identified based on bathymetric, sedimentological, nutrient, biological, and hydrological investigations. For both regions GIS-supported analysis was applied as geostatistical interpolation methods, trend surface analyses or overlay techniques.

The Håkon Mosby Mud Volcano

All the results concerning the HMMV presented in this thesis are mostly conducted on the basis of georeferenced video mosaics. Due to their equipment with geographical coordinates they offer a multiplicity of GIS-based analysis possibilities, but due to the large amount of images they implicate also the challenge to develop analysis methods and an automatic content-based analysis system.

• **Definition of categorical biogeochemical habitats and a suggested source location of ascending mud at the HMMV**

Chemoautotrophic communities as *Beggiatoa* and pogonophoran tube worms are linked to the methane cycle using a product of methane consumption, the sulphide. Therefore, they provide an indicator mechanism for anaerobe oxidation of methane (Boetius et al., 2000) at active submarine methane bearing sites. A high resolution microbathymetry map and georeferenced video mosaics generated by means of the Remotely Operated Vehicle *Victor6000* of the Håkon Mosby Mud Volcano was integrated to understand how the spatial distribution of the chemoautotrophic communities and their relation to temporal mud flow sequences.

The GIS-supported visual inspection of the video mosaics yields in the spatial distribution of chemoautotrophic communities implemented into a classification scheme. The performed slope and trend surface analyses of the microbathymetric data provide detailed and general pattern about slope angles or directions. The GIS builds the platform to calculate areas of the morphological units, uncovered and covered regions by chemoautotrophic communities, and to measure features at the seafloor on a m^2 basis at the HMMV for the first time.

A classification scheme was developed indicating the region of highest methane release (HCR) (740,000 m^2), on whose basis the characteristics of 2310 analysed georeferenced video mosaics, covering 6.2% of the HCR, were stored within a GIS geodatabase. After application of the geostatistic interpolation using indicator kriging the predicted surfaces of the individual geochemical habitats were computed and assigned to a certain biogeochemical habitat (BGH 0-4). A polynomial trend surface computation based on microbathymetric data and observations of highest heat flow (Kaul and Heesemann, 2004) enabled to identify the suggested source location of freshly ascending mud within the uncovered centre of the Håkon Mosby Mud Volcano. A temporal dependence of the settlement structure by bacteria could be derived. This suggests that present mud flow ascend close to the northern edge of the flat unit of the HMMV moving to the south-south-west.

The HCR consists of two morphological units: the flat centre (26% of the HCR) and the hummocky periphery (74% of the HCR). Approximately 16% of the HCR located in the flat centre of the mud volcano is nearly void of any benthic community and represents the BGH 0, a location of methane discharge into the hydrography. A 5% surface is densely inhabited by *Beggiatoa*, which is to be found in the south-eastern central part (BGH 1). The hummocky outer part is colonised dominantly by pogonophoran tubes and only sporadically by *Beggiatoa*. It was subdivided into differentiated regions forming a transient area (BGH 2 and 3) (22.6% of the HCR) until the *Beggiatoa* are completely replaced by the tube worms (BGH 4).

- **Estimate quality of the interpolated surfaces at the HMMV**

The geostatistical methodology and GIS overlay technique applied in chapter 4 is regarded in detail within chapter 5. The basic approach of probabilistic geostatistical modelling, such as indicator kriging (IK) is to turn any unsampled value z(x) into a random variable Z(x). A probability distribution function of Z(x) characterises the uncertainty about the unknown value z(x) (Richmond, 2002). The results depend on the spatial distribution of the data (Davis, 2002). The probability distribution thus could be estimated by analysing the results of the function (cross-validation). The IK models in this case study were sensitive to the selection of mean probability values of occurrence.

Interpolation procedure represents a gain in information of non-samples areas, however, with the restriction that the results always remain a degree of uncertainty; therefore interpolation cannot replace real measured values. This uncertainty can be made more transparent by the presented approach identifying the weak points but also the strengths of interpolated maps.

Polygons resulting from the georeferenced images build the data basis for geostatistically interpolated mono-parametric surface maps. At first IK is applied to calculate raster maps from the centroids of the polygons. All category maps belonging to one parameter (bacteria, tube worms or uncovered mud) are combined concerning their occurrence at the HMMV (Fig. 10-1).
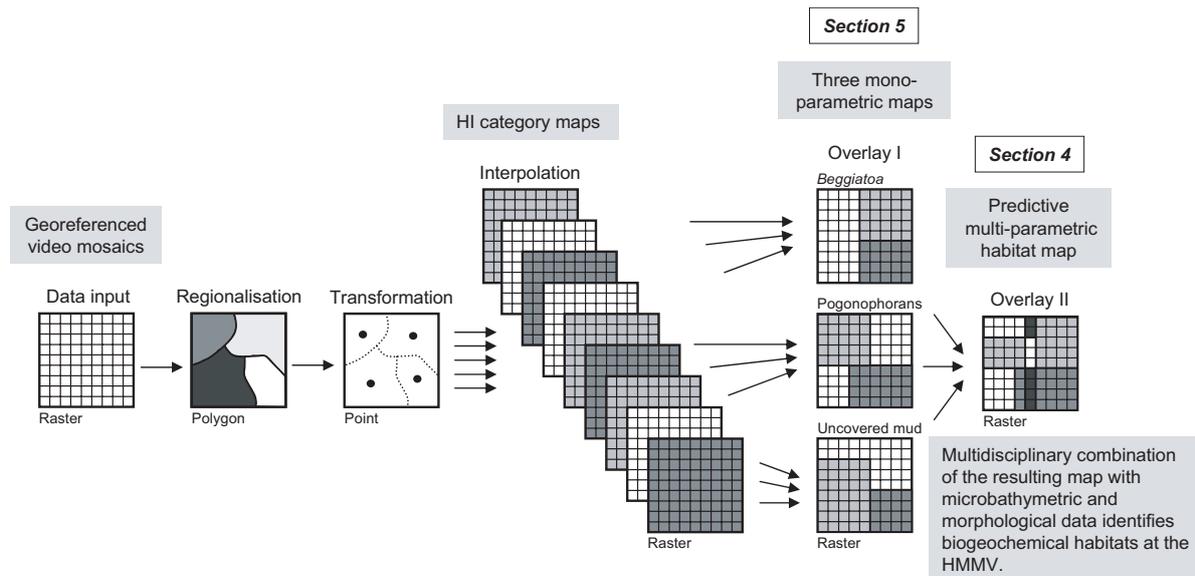
Figure 10-1: Schematic representation: identification of seafloor provinces using video mosaics from a geoinformatical view.

The proven provinces are assessed by their estimate quality by variogram analysis and cross-validation for each category. The resulting estimate quality for the individual parameter category is then expressed on the basis of statistic mean values. Apart from the evaluation of the IK method itself the results, mono-parametric habitat maps, are proven in a further step using a validation dataset consisting further visual video mosaic inspection. This validation procedure highlights the concordances and differences between real measured data and predicted surfaces (Fig. 5-8).

Image characteristics were extracted and transformed successfully into surface maps. The investigation method bears a comparison with an evaluation procedure by cross-validation and validation dataset in principle and is recommended despite of marginal deviations.

• **Developing a system of fully automatical recognition and quantification of bacteria mats on georeferenced video mosaics at the HMMV**

The software IBU is the a development of image processing techniques for the automatic content-based analysis of georeferenced Håkon Mosby video mosaics to detect and qualify features (or objects) such as bacterial mats, pogonophoran tube worms, mud structures, or small holes in the seafloor. Thereby a generic framework for the analysis of georeferenced images with useful basic functionality, e.g. shape file export, access to image metadata, coordinate transform or basic rendering of maps using the OpenMap (tm) toolkit (http://www.openmap. org), and basic image processing functionality was developed. There is also a component for manual annotation based on a generic classification scheme. The results of the module for the recognition of bacterial mats are presented in chapter 6. The analysis is mainly based on

classic image processing techniques using colour, texture and shape features.

In a first step *data*, *non data* and *redundant data* (overlapping mosaic segments) are differentiated on the mosaics using a Grassfire-like transform (Ballard and Brown, 1982). In the second step the data areas are pre-segmented into disjunctive homogeneous regions by a watershed transformation (Beucher, 1991; Roerdink and Meijster, 2000). Applying relaxation labelling (Kittler and Illingworth, 1986), then the assignment of these regions as *bacterial* or *non-bacterial* on the basis of spatial correlation and defined contrast thresholds is realised (comparing the grey values of neighbour regions). Difficulties arise particularly due to different illuminating of the seafloor and the heterogeneity of the bacteria appearance. Experiments have shown that due to inhomogeneous lighting the application of a fixed global threshold leads to unsatisfactory results. So we calculate thresholds individually for every pixel (or a small neighbourhood in order to enhance the performance of the algorithm). The idea of the algorithm is mainly based on two observations: bacteria regions typically have brighter colours than the surrounding seafloor and the variance between the intensities of the red, green and blue channel (in the RGB colour model) tends to be smaller in regions covered by bacterial mats in comparison to the surrounding seafloor. The results of the algorithm were compared to 2840 visually analysed mosaics and performed similar results which are proved satisfactorily by common used statistic characteristic numbers.

The IBU tends to produce an underestimation of the bacterial coverage, mainly due to very shallow bacterial coverage or bad illumination in certain image regions. Misdetections are mostly as a result of overexposure and reflections and light structures or objects similar to bacterial mats.

The North Sea

• **MarGIS data management and analysis system for marine geodata**

The project MarGIS intends the combination of GIS, research data, and advanced statistical techniques to characterise and identify distinct provinces at the seafloor of the North Sea with regard to the similarity of ecological characteristics. Such an ecological regionalisation often is a prerequisite for marine planning and management needs, such as installation of off-shore wind parks or the declaration of protection zones (Hughes, 1997; Moog et al., 2004; Reiniger, 1997). The eco-regionalisation approach used in MarGIS consists of two main working steps: (i) by applying geostatistical methods such as variogram analysis and kriging, surface maps are calculated from measurement data. (ii) Multivariate statistics like Classification and Regression Trees (CART) and GIS techniques are then used to calculate sea floor provinces from the kriging grid maps. Since the beginning of the project, kriging maps on temperature, salinity and nutrients (phosphate, nitrate, and ammonium) were calculated and used to derive a habitat map for the EEZ.

The data used to calculate these surface maps were taken from the database established since the beginning of the project in October 2002. Additional surface data are on sediments and on biotic and abiotic measurements. The data sets were provided by several national and international databases and projects and were integrated in a relational Database Management System (DBMS). Table 8-1 gives an overview of all abiotic parameters on which measurement data has been integrated in the DBMS until now: Nearby 235,000 abiotic data sets have been collected from the International Council for the Exploration of the Sea (ICES), the Marine Environmental Data Base (MUDAB) of the Federal Maritime and Hydrographic Office and the Federal Environmental Agency and the Institute of Marine Research, University of Hamburg (IfM).

A MarGIS geodatabase model is developed and is implemented on a client-server DBMS which allows integration of maps (vector and raster data), raw data, georeferenced images (geotiffs), and metadata. Central part of the data model are tables containing information about sampling sites and tracks studied during research cruises or obtained from the literature or reports. These tables are attributed as sites of measurements, locations of time series, or starting/end points of line features as tracks lines, dives, or net hauls. Furthermore, data derived by multilevel sampling, in situ sensors, and obtained along track lines and areas are stored. In addition to raw data aggregated information as thematic maps on sedimentology or bathymetry were integrated into the relational database management system.

For aggregated information as geological or geochemical maps, which were georeferenced, digitised, and integrated into the geodatabase, the data source, geographic projection or the applied geodetic datum are stored as metadata. The metadata are stored as XML document in the database. As considered subsequently, metadata are assigned to each information layer and are distributed in conjunction with geodata by an Internet Map Server.

• **Calculation of seafloor provinces for the German Exclusive Economic Zone (EEZ) by means of the multivariate statistical method CART adapted from on geostatistically produced surface maps**

Raster maps were calculated from the abiotic measurement data by applying geostatistical methods. Amongst others originating from applied geological research to estimate mineral resources and reserves (Krige, 1951), geostatistics is nowadays being used in different terrestrial and marine research fields (Goovaerts, 1997; Lembo et al., 2000; Petitgas, 1996). By using variogram maps, anisotropies can be accounted for in the kriging estimations. Concerning the spatial temperature distribution the variogram map indicates that strong anisotropies characterise the spatial autocorrelation pattern (Fig. 8-4). The semi-variogram values in NE-direction (44.6°) reach their sill at a distance of 140 km. In NW-direction the maximum autocorrelation range exhibits 80 km, resulting in an anisotropy ratio of 57%. Figure 8-4 furthermore depicts the result of the ordinary kriging interpolations. Except for a small

area within the inner German Bight with values below 13 °C, a continuous decrease of the temperature values can be observed from the coastal areas to the outer part of the EEZ. Further such analyses were performed for other time intervals and parameters by using different kriging techniques (indicator kriging, co-kriging). All maps were described with respect to the global quality of estimation by means of cross-validation.

Together with data on benthic communities taken from Rachor (2002) the geostatistically estimated abiotic grid data were used to calculate a habitat map for the EEZ by means CART. CART uncovers the relationship between a chosen dependent variable and a series of predictor variables by dendrogram. These trees display a hierarchical system of decision rules that allows for classifying objects (e.g. grids) according to the features of the predictor variables. CART calculates classes that are homogeneous with respect to the features of the dependent variable. For the calculations performed here, a decision tree was computed from site specific data on nine benthic communities (dependent variable) and geostatistically estimated grid data on grain size as well as on temperature, salinity, silicate, ammonium, nitrate, phosphate and dissolved oxygen aggregated over a five year period (predictor variables). On the left branch of the resulting decision tree in figure 9-5 can be seen that from up to down the portion of one of the nine benthic communities increases stepwise. The end nodes only contain one or two communities. Since each end node is determined by certain decision rules, the tree can be applied to predict benthic communities at sites where such information is not available. By doing this for all geostatistically estimated raster cells within the EEZ including a 40 km buffer, a habitat map can be calculated with respect to the occurrence of benthic communities derived by Rachor and Nehmenr (2002).

10.2 Outstanding problems and future perspective

The Håkon Mosby Mud Volcano

Natural seeps (hydrocarbons from deep-seated reservoirs, hydrothermal sources, shallow gas), are linked to biotopes/geotopes and pollution. An understanding of this link will be essential for documenting the natural state of the seabed and for management of the ocean areas in relation to biological prospecting and understanding of ecosystems in a geological context. The study at the HMMV serves as a baseline study for further investigation at that place, but probably also methodologically for other mud volcanoes or seeps with comparable biogeochemical processes and population. Whether the system is really transferable, is to be proved at other sites populated with chemosynthetic communities. For instance the Milano Mud Volcano of the Central Mediterranean Ridge (Huguen et al., 2005) or an asphalt volcano as the Chapopote Asphalt Volcano in the Campeche Knolls, Gulf of Mexico, (González et al., 2005; MacDonald et al., 2004) are potential study areas for similar evaluation procedures as conducted at the HMMV.

With the view on further investigations applying autonomous or remotely underwater vehicles and because of the huge amount of recorded data during each expedition further processing of IBU modules is recommended to be developed, for example for the automatical detection of pogonophorans and other features (small holes, structures in the mud, ...).

Additional work to improve the recent knowledge about the methane budget releasing the HMMV. Therefore, is not possible to correlate the biogeochemical habitat areas to a certain methane flux into the bottom water yet. Methane budgets of the microbial consumption could be correlated to provinces not until they are not well-established. Additional work at the HMMV is required for the investigation of the temporary development of the distribution pattern of the chemoautotrophic communities.

For future research at the HMMV a multivariate statistical analysis is proposed. Supplementary surface maps concerning methane release, small holes, water currents, sediment temperature at the sea floor but also in deeper sediment layers, gas hydrate and fluid occurrence, and the distribution of fish or clams could be produced resulting from accompanying studies. In conjunction with the already existing parameters (*Beggiatoa*, pogonophorans, uncovered mud, microbathymetry/morphology and slope, provinces could be calculated fundamentally performing CART analysis improving the understanding of the interrelationships between geochemistry and microbiology pattern.

The North Sea

A unification and combination of existing data by GIS and database management systems is an important step to a better and comprehensive view of the marine ecological system. Today is particularly important to investigate and represent the ecological needs using "hard" data intending environmental-protecting purposes or at least environmental-compatible editions (e.g. for building offshore wind parks). Usually scientists are required to provide evidence for possible ecological adverse effects, and/or disadvantages, before these aspects have an effect on the decision of building a project or imposing sustainable compensatory sanction. Habitat maps are suitable for the visualisation of provable environmental changes and are therefore required tools for regional planning of wind parks in the North Sea.

The data acquisition during the MarGIS project reveals a lack of in important chemical (e.g. dissolved CO_2 and oxygen) and sedimentological key parameters, even if a large time period would be accepted. The data availability is particularly bad e.g. regarding the parameter *dissolved oxygen* with that saisonal dependence must be considered. In figure 10-2 the available sample sites from October until December 1992-2002 are given as black dots, the samples of the years 1982-1991 as dark greyish dots. Compared to that, physical parameters (e.g. temperature, salinity, and bathymetry), and nutrients (nitrate, nitrite, ammonium, and phosphate) are distributed well, but they are not available spatially. Therefore, an advanced
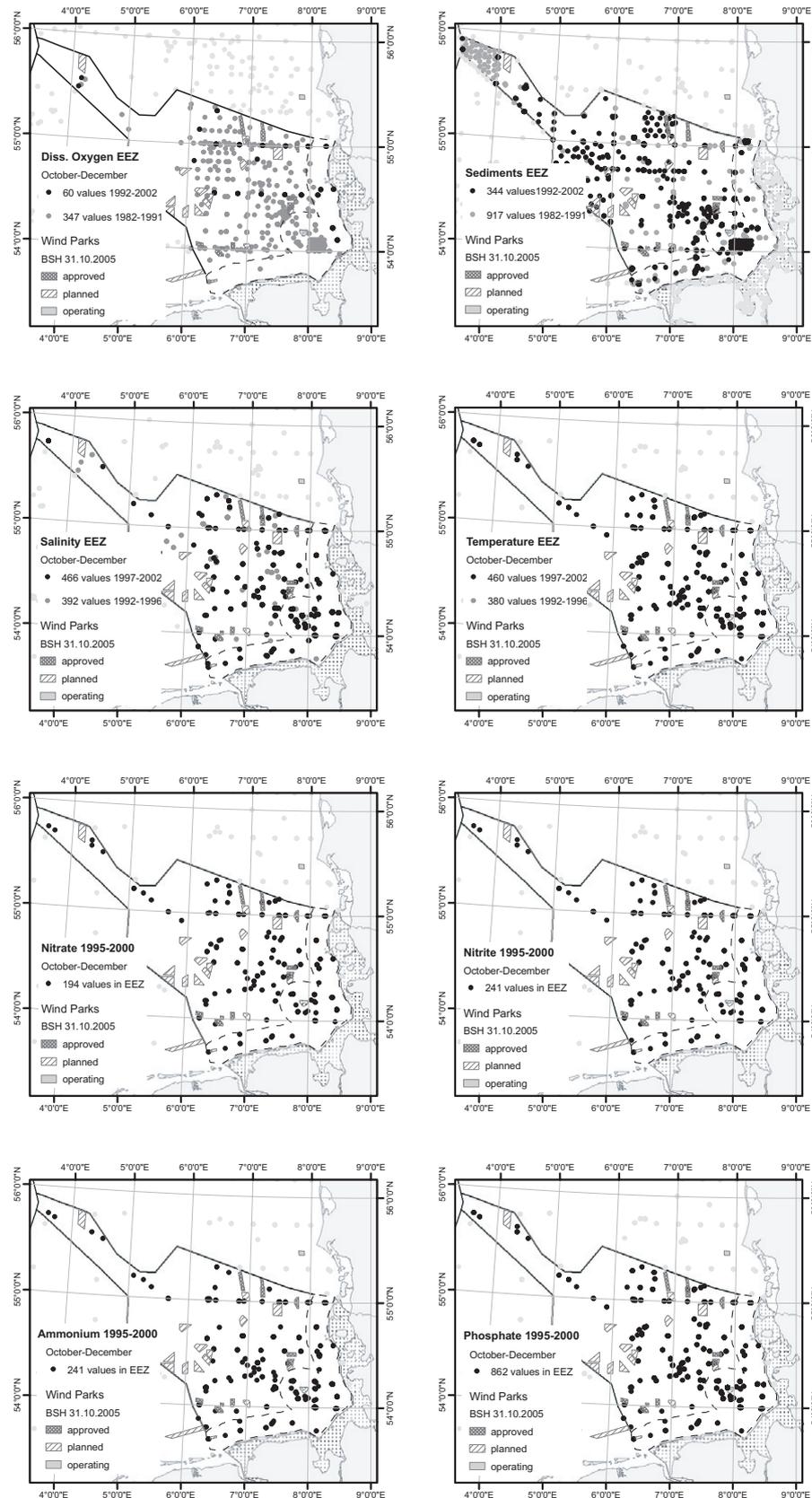
Figure 10-2: Data availability within the EEZ during 1982 and 2002 - selected key parameters and time periods. Nutrients, salinity and temperature are distributed densely compared to sediment samples or dissolved oxygen, particularly.

IMS-system is needed which provides an up-to-date status of data, which supports user-defined queries and simple GIS geoprocessing tools (e.g. overlay, cut, deterministic interpolation methods as IDW, ...), and guarantees transparent meta information (time period, interpolation method, data provider,...) of surface maps. Such a geodatabase management system would be useful for archival data storage as e.g. the national database MUDAB or the information system PANGAEA part of the Word Data Centre.

The sedimentological characteristics of the sea floor are one of the most important factors of the distribution of benthic communities. The presented study reveals a considerable need for the updating of a sediment map covering the entire North Sea. This was one of the main problems, because of a lack of up-to-date and detailed analysed sediment samples concerning the sediment composition and grain sizes covering the entire North Sea (Fig. 10-2). For this purpose, it is recommended to gain additional measured data but also to develop a method to combine point data existing analogue maps into a new sediment map using advanced kriging methods (e.g. fuzzy kriging). Furthermore, in 2005 the methodological procedure of identification of seafloor provinces using geostatistical and multivariate statistical methods was presented on an international workshop of the ICES Working Group on Marine Habitat Mapping and was supported from these experts for a similar execution North Sea wide.

Because auf the extensive anthropogenic use of the North Sea it is recommended to integrate further information layer such as organic and inorganic pollutants in the sediments and flat fish. The compiled extensive North Sea dataset could also be used for a development of fix sampling stations for each parameter. Depending on statistics of its data variance in time and place an optimisation of an observational network could be provided.

Section E:
Acknowledgements and references

11 Acknowledgements

12 References

Akin, H., Siemens, H., 1988. Praktische Geostatistik, Springer-Verlag, Berlin-Heidelberg, 304 pp.

Albrecht, H., Schmolke, S.R., 2003. Belastung der Nordsee mit anorganischen Schadstoffen. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.): Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Allais, A.-G., Borgetto, M., Opderbecke, J., Pessel, N., Rigaud V., 2004. Seabed video mosaicing with MATISSE: a technical overview and cruise results. Proceedings of 14th International Offshore and Polar Engineering Conference, ISOPE-2004, Toulon, France, May 23-28, 2004, 2: 417-421.

Arentze, T., Timmermanns, H., 2005. An analysis of context and constraints-dependent shopping behaviour using qualitative decision principles. Urban Studies, 42(3): 435-448.

Atkinson, P. M., Lloyd, C. D., 1998. Mapping precipitation in Switzerland with ordinary and indicator kriging. Journal of Geographic Information and Decision Analysis, 2(2): 65-76.

Bartelme, N., 2005. Geoinformatik: Modelle, Strukturen, Funktionen. Springer-Verlag, Berlin-Heidelberg, 4th Edition, 454 pp.

Ballard, D. H., Brown, C. M., 1982. Computer Vision. Prentice-Hall Inc., New Jersey, 523 pp.

Becker, G., 2003. Physikalische Beschreibung der Nordsee. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.). Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Beucher, S., 1992. The watershed transformation applied to image segmentation. *In*: 10th Pfefferkorn Conference on Signal and Image Processing in Microscopy and Microanalysis, September 16-19, 1991, Cambridge, UK, Suppl. 6: 299–314.

Boetius, A., Ravenschlag, K., Schubert, C.J., Rickert, D., Widdel, F., Gieseke, A., Amann, R., Jørgensen, B. B., Witte, U., Pfannkuche, O., 2000. A marine microbial consortium apparently mediating anaerobic oxidation of methane. Nature, 407: 623-626.

Boetius, A., Jørgensen, B. B., Amann, R., Henriet, J.P., Hinrichs, K.U., Lochte, K., MacGregor, B.J., Voordouw, G., 2002. Microbial systems in sedimentary environments of continental margins. *In:* Wefer, G., Billett, D., Hebbeln, D., Jørgensen, B. B., Schlüter, M., Van Weering, T. (Eds), Ocean Margin Systems, Springer-Verlag, Berlin-Heidelberg, p. 479-495.

Boetius, A., Beier, V., Niemann, H., Müller, I., Heinrich, F., Feseker, T., 2004. Geomicrobiology of sediments and bottom waters of the Håkon Mosby Mud Volcano. *In*: Klages, M., Thiede, J., Foucher, J.-P. (Eds.), The Expedition ARK XIX/3 of the Research Vessel "Polarstern" in 2003, Reports on Polar and Marine Research, 488: 190-199.

Böttcher, U., Teich, P., 2003. SQL: Grundlagen und Datenbankdesign. Herdt-Verlag, Nackenheim, 1st Edition, 190 pp.

Bohrmann, G., Torres, M., in press. Gas hydrates in marine sediments. *In*: Schulz, H., Zabel, M. (Eds.), Marine Geochemistry. Springer-Verlag, Berlin-Heidelberg, p. 481–512.

Bonham-Carter, G. F., 1996. Geographic Information Systems for Geoscientists: Modelling with GIS. Computer Methods in Geoscience. Pergamon, Ontario, 414 pp.

Borgetto, M., Rigaud, V., Lots, J. F., 2003. Lighting correction for underwater mosaicking enhancement. Proceedings of the 16th International Conference on Vision Interface, VI, Halifax.

Breiman, L., Freidman, J. H., Olshen, R. A., Stone, C. J., 1984. Classification and Regression Trees, The Wadsworth statistics/probability series, 368 pp.

Brenk, V., 2003. Verschmutzung der Nord- und Ostsee durch die Schifffahrt. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.): Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Brockmann, U., Topcu, D. H., 2002. Nutrient Atlas of the Central and Northern North Sea. Research Report 200 25 211, UBA-Texte Nr. 26/2002.

Brockmann, U., Lenhart, H., Schlünzen, K. H., Topcu, D., 2003. Nährstoffe und Eutrophierung. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.): Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Bundesamt für Bauwesen und Raumordnung, 2004. Marine Raumordnung. Heft 7/8, Bonn.

Busch, M., 2005. Die EUNIS-Habitatklassifikation – Eine Grundlage für die Kartierung mariner Meeresbodenhabitate in der deutschen Ausschließlichen Wirtschaftszone der Nordsee? Diploma thesis, Hochschule Vechta, Germany, 214 pp.

Callaway, R., Alsvåg, J., De Boois, I., Cotter, J., Ford, A., Hinz, H., Jennings, S., Kröncke, I., Lancaster, J., Piet, G., Prince, P., Ehrich, S., 2002. Diversity and community structure of epibenthic invertebrates and fish in the North Sea. ICES Journal of Marine Science, 59: 1199-1214.

Chihi, H., Galli, A., Ravenne, C., Tesson, M., De Marsily, G., 2000. Estimating the depth of stratigraphic units from marine seismic profiles using nonstationary geostatistics. Natural Resources Research 9(1): 77-95.

Clark, I., 1982. Practical Geostatistics. Elsevier Applied Science. 129 pp.

Cleverdon, C. W., Mills, J., Keen, M., 1966. Factors determining the performance of indexing systems, Volume I: Design - Volume II: Test Results. ASLIB Cranfield Project, Cranfield.

Commission of the European Communities, 2005. Proposal for a Directive of the European Parliament and of the Council - Establishing a Framework for Community Action in the Field of Marine Environmental Policy (Marine Strategy Directive) [SEC(2005) 1290], 31pp.

Conan, M., 2000. Environmentalism in Landscape Architecture. Dumbarton Oaks Colloquium on the History of Landscape Architecture 22.

Connor, D. W., Allen J. H., Golding N., Howell K. L., Liebknecht L. M., Northen K. O., Reker J. B., 2004a. The Marine Habitat Classification for Britain and Ireland Version 04.05. JNCC. - Peterborough.

Connor, D. W., Allen J. H., Golding N., Howell K. L., Liebknecht L. M., Northen K. O., Reker J. B., 2004b. The Marine Habitat Classification for Britain and Ireland Version 04.05 - Sublittoral Sediment Section. JNCC. - Peterborough.

Connor, D. W., Allen J. H., Golding N., Howell K. L., Liebknecht L. M., Northen K. O., Reker J. B., 2004c. The Marine Habitat Classification for Britain and Ireland Version 04.05 - Littoral Rock Section. JNCC. - Peterborough.

Connor, D. W., Allen J. H., Golding N., Howell K. L., Liebknecht L.M ., Northen K.O ., Reker J. B., 2004d. The Marine Habitat Classification for Britain and Ireland Version 04.05 - Littoral Sediment Section. JNCC. - Peterborough.

Connor, D. W., Allen J. H., Golding N., Howell K. L., Liebknecht L. M., Northen K.O ., Reker J. B., 2004e. The Marine Habitat Classification for Britain and Ireland Version 04.05 - Infralittoral Rock Section. JNCC. - Peterborough.

Connor, D. W., Allen J. H., Golding N., Howell K. L., Liebknecht L. M., Northen K. O., Reker J. B., 2004f. The Marine Habitat Classification for Britain and Ireland Version 04.05 - Circalittoral Rock Section. JNCC. - Peterborough.

Crane, K., Vogt, P., Sundvor, E., Shor, A., Reed, T., 1995. SeaMARC II investigations of the northern Norwegian-Greenland Sea. *In*: Crane, K., Solheim, A. (Eds.), Seafloor Atlas of the Northern Norwegian-Greenland Sea. Norsk Polarinstitutt Meddelelser, Oslo, 137: 132-140.

Damm, E., Budéus, G., 2003. Fate of vent-derived methane in seawater above the Håkon Mosby Mud Volcano (Norwegian Sea). Marine Chemistry, 82: 1-11.

Davies, E. R., 1990. Machine Vision: Theory, Algorithms and Practicalities. Academic Press, 576 pp.

Davies, C. E., Moss, D., Hill, M.-O., 2004. EUNIS Habitat Classification Revised 2004. Report to the European Topic Centre on Nature Protection and Biodiversity, European Environment Agency, Paris, Kopenhagen, 307 pp.

Davis, J.C., 2002. Statistics and data analysis in geology. John Wiley and Sons, New York, 3rd Edition, 638 pp.

De Beer, D., Sauter, E., Niemann, H., Witte, U., Schlüter, M., Boetius, A., in press. In situ fluxes and zonation of microbial activity in surface sediments of the Håkon Mosby Mud Volcano. Limnology and Oceanography.

Dimitrov, L. I., 2002. Mud volcanoes - the most important pathway for degassing deeply buried sediments. Earth-Science Reviews, 59: 49–76.

Duineveld, G. C. A., Kuenitzer, A., Nierman, U., De Wilde, P. A., Gray, J. S., 1991. The macrobenthos of the North Sea. Netherlands Journal of Sea Research, 28(1-2): 53-65.

Edy, C., Bisquay, H., Foucher, J. P., Opderbecke, J., Simeoni, P., Allais, A. G., Beyer, A., Jerosch, K., R. Rathlau, R., 2004. Microbathymetry of the Håkon Mosby mud volcano off Northern Norway: results of a ROV-born multibeam survey. European Geosciences Union, Geophysical Research Abstracts, Vol. 6, 04619, SRef-ID: 1607-7962/gra/EGU04-A-04619.

Egorov, A., Crane, K., Vogt, P., Rozhkov, A., 1999. Gas hydrates that outcrop on the sea floor: stability models. Geo-Marine Letters, 19: 68-75.

Eldholm, O., Sundvor, E., Vogt, P. R., Hjelstuen, B. O., Crane, K., Nilsen, A. K., Gladczenko, T. P., 1999. SW Barents Sea continental margin heat flow and Håkon Mosby Mud Volcano. Geo-Marine Letters, 19: 29-37.

ESRI Shapefile Technical Description. URL http://www.esri.com/

Fan, Z., Larsen, D. R., Shifley, S. R., Thompson, F. R., 2003. Estimating cavity tree abundance by stand age and basal area. Forest Ecology and Management, 179: 231-242.

Figge, K., 1981. Sedimentverteilung in der Deutschen Bucht. In: Karte Nr. 2900, Deutsches Hydrographisches Institut, Hamburg.

Fleischer, P., Orsi, T. H., Richardson, M. D., Anderson, A. L., 2001. Distribution of free gas in marine sediments: a global overview. Geo-Marine Letters, 21: 103-122.

Flemming, B., Bartholomä, A., 2003. Sedimentation und Erosion an der Nordseeküste. In: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.), Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Folk, R. L., 1954. A system for classifying marine sediments. Journal of Geology, 62: 344-359.

Folk, R. L., 1974: The petrology of sedimentary rocks. Hemphill Publishing Co., Austin, Texas, 182 pp.

Gätje, Ch., 2003. Tourismus und Erholung im Wattenmeer. In: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.), Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Gebruk, A. V., Krylova, E. M., Lein, A. Y., Vinogradov, G. M., Anderson, E., Pimenov, N. V., Cherkashev, G. A., Crane, K., 2003: Methane seep community of the Håkon Mosby Mud Volcano (the Norwegian Sea): composition and trophic aspects. Sarsia, 88: 394-403.

Ginsburg, G. D., Milkov, A. V., Soloviev, V. A., Egorov, A. V., Cherkashev, G. A., Vogt, P. R., Crane, K., Lorenson, T. D., Khutorskoy M. D., 1999. Gas hydrate accumulation at the Håkon Mosby Mud Volcano. Geo-Marine Letters, 19: 57-67.

Gonzáles-Gurriarán, E., Freire, J., Fernandez, L., 1993. Geostatistical analysis of spatial distribution of *Liocarcinus depurator*, *Macropipus tuberculatus* and *Polybius henslowii* (Crustacea: Brachyura) over the Galician continental shelf (NW Spain). Marine Biology, 115: 453-461.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 496 pp.

Green, D. R., King, S. D., 2003. Coastal and Marine Geo-Information Systems, Kluwer Academics Publishers, Dordrecht, 580 pp.

Greene, H.G., Stakes, D.S., Orange, D.L. Barry, P., Robinson, B.H., 1993. Application of a ROV in geologic

mapping of Monterey Bay, California, USA, Proc. Am. Acad. Underw. Sci.; 13th Annual Scientific Diving Symposium, Pacific Grove, CA, pp. 17-32.

Greene, H. G., Yoklavich, M. M., Starr, R. M., O'Connell, V. M., Wakefield, W. W., Sullivan, D. E., McRea, J. E., Cailliet, G. M., 1999. A classification scheme for deep seafloor habitats. Oceanologica Acta, 22(6): 663-678.

Greinert, J., 1999. Rezente Submarine Mineralbildungen: Abbild geochemischer Prozesse an aktiven Fluidaustrittsstellen im Aleuten- und Cascadia-Akkretionskomplex. GEOMAR Report Nr. 87, 196 pp.

Harbitz, A., Lindstrom, U., 2001. Stochastic spatial analysis of marine resources with application to minke whales (*Balaenoptera acutorostrata*) foraging: A synoptic case study from the southern Barents Sea. Sarsia, 86: 485-501.

Hinrichs, K. U., Boetius, A., 2002. The anaerobic oxidation of methane: new insights in microbial ecology and biogeochemistry. *In*: Wefer, G., Billett, D., Hebbeln, D., Jørgensen, B. B., Schlüter, M., Van Weering, T. (Eds), Ocean Margin Systems, Springer-Verlag, Berlin-Heidelberg, p. 457-477.

Hjelstuen, B. O., Eldholm, O., Faleide, J. I., Vogt, P. R., 1999. Regional setting of Håkon Mosby Mud Volcano, SW Barents Sea margin. Geo-Marine Letters, 19: 22-28.

Hooker, S. B., Rees, N. W., Aiken, J., 2000. An objective methodology for identifying oceanic provinces. Progress in Oceanography, 45: 313-338.

Hovland, M., 2002. On the self-sealing nature of marine seeps. Continental Shelf Research, 22: 2387-2394.

Hovland, M., MacDonald, I.R., Rueslåtten, H., Johnsen, H.K., Naehr, T., Bohrmann, G., 2005. Chapopote asphalt volcano may have been generarted by supercritical water. EOS, 86(42): 397, 402.

Huettmann, F., Diamond, A. W., 2001. Seabird colony locations and environmental determination of seabird distribution: a spatially explicit breeding seabird model for the Northwest Atlantic. Ecological Modelling, 141(1): 261-298.

Hughes, R. M., Heiskary, S. A., Matthews, W. J., Yoder, C. O., 1994. Use of ecoregions in biological monitoring. In: Loeb, S. L., Spacie A., (Eds.), Biological Monitoring of Aquatic Systems. Lewis Publishers, Boca Raton, LA, p. 125-151.

Huguen, C., Mascle, J., Chaumillon, E., Kopf, A. J., Woodside, J., Zitter, T., 2004. Structural setting and tectonic control of mud volcanoes from the Central Mediterranean Ridge (Eastern Mediterranean). Marine Geology, 209(1-4): 245-263.

Huguen, C., Mascle, J., Woodside, J., Zitter, T., Foucher, J.P., 2005: Mud volcanoes and mud domes of the Central Mediterranean Ridge: Near-bottom and *in situ* observations. Deep-Sea Research I, 52: 1011-1931.

ICES (International Council for the Exploration of the Sea), 2005. Report of the Working Group on Habitat Mapping (WGMHM), ICES, April 5-8,2005, Bremerhaven, Germany, CM 2005/E:05, REF. ACE, B, 91 pp.

Isaaks, E. H., Srivastava, R. M., 1992. An Introduction to Applied Geostatistics. Oxford University Press 561 pp.

Ivanov, M. K., Limonov, A. F., Van Weering, T., 1996. Comparative characteristics of the Black Sea and Mediterranean Ridge mud volcanoes. Marine Geology, 132: 253-271.

Jahnke, R. A., 1996. The global ocean flux of particulate organic carbon: Areal distribution and magnitude. Global Biogeochemical Cycles, 10: 71-88.

Jarass, H. D., 2002. Naturschutz in der Ausschließlichen Wirtschaftszone, Nomos Verlagsgesellschaft, Baden-Baden, 70 pp.

Jelinski, D. E., Krueger, C.C., Duffus, D.A., 2002. Geostatistical analyses of interaction between killer whales (*Orcinus orca*) and recreational whale-watching boats. Applied Geography (22): 393-411.

Jerosch, K., Schlüter, M., Pesch, R., Schöder, W., Köberle, A., Vetter, L., 2005a. MarGIS Marine Geo-Information-System for Visualisation and Typology of Marine Geodata. Congress Proceedings, Informatics for Environmental Protection, September 7 - 9, 2005, Brno, p. 165-166.

Jerosch, K., Lüdtke, A., Schlüter, M., Ioannidis, G., 2005b. Image analysis methods, geo-statistics and GIS techniques to derive spatial budgets for benthic communities at mud volcanoes. General Assembly of the European Geosciences Union (EUG), 24 - 29 April, 2005., Vienna.

Jerosch, K., Schlüter, M., Foucher, J., Allais, A., Klages, M., Edy, C., subm-a. Spatial distribution of benthic communities affecting the methane concentration at Håkon Mosby Mud Volcano. Marine Geology.

Jerosch, K., Schlüter, M., Pesch, R., subm-b. Indicator kriging applied for video mosaics: a case study of habitat mapping at the deep-sea submarine Håkon Mosby Mud Volcano. Ecological Informatics.

Jerosch, K., Lüdtke, A., Schlüter, M., Ioannidis, G. T., subm-c. Automatic content-based analysis of georeferenced image data: detection of *Beggiatoa* mats in seafloor video mosaics from the Håkon Mosby Mud Volcano. Computers and Geosciences.

Johnston, K., Ver Hoef, J. M., Krivoruchko, K., Lucas, N., 2001. Using ArcGIS Geostatistical Analyst. ESRI, Redlands, USA, 40 pp.

Jones, L. A., Hiscock, K. Connor, D. W., 2001. Marine habitat reviews. A summary of ecological requirements and sensitivity characteristics for the conservation and management of marine SACs. Peterborough, Joint Nature Conservation Committee. (UK Marine SACs Project report).

Jørgensen, N. O.,1992. Methane-derived carbonate cementation of marine sediments from the Kattegat, Denmark: Geochemical and geological evidence. Marine Geology, 103(1-3): 1-13.

Jouffroy, J., Opderbecke, J., 2004. Underwater vehicle trajectory estimation using contracting PDE-based observers. American Control Conference (ACC 2004), Boston.

Journel A. G., Huijbregts, C. J., 1978. Mining Geostatistics. Academic Press, London, 600 pp.

Kaul, N., Heesemann, B., 2004. Heat probe measurements down to 3 m depth. *In*: Klages, M., Thiede, J., Foucher, J.-P. (Eds.), The Expedition ARK XIX/3 of the Research Vessel "Polarstern" in 2003, Reports on Polar and Marine Research, 488: 164-169.

Kelly, A., Powell, D., Riggs, R. A., 2005. Predicting potential natural vegetation in an interior northwest landscape using classification tree modeling and a GIS. Western Journal of Applied Forestry, 20 (2): 117-127.

Kittler, J., Illingworth, J., 1986. Relaxation labelling algorithms - a review. Image and Vision Computing 3(4): 206–216.

Klages, M., Thiede, J., Foucher, J.-P., 2004. The Expedition ARK XIX/3 of the Research Vessel "Polarstern" in 2003, Reports on Polar and Marine Research, 488, 346 pp.

Kohl, B., Roberts, H. H., 1994. Fossil foraminifera from four active mud volcanoes in the Gulf of Mexico. Geo-Marine Letters, 14: 126-134.

Kopf, A. J., 2002. Significance of mud volcanism, Reviews of Geophysics, 40(2), 1005, doi: 10.1029/ 2000RG000093.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand, Journal of the Chemical, Metallurgical and Mining Society of South Africa, 52(6): 119-139.

Kröncke, I., Bergfeld, C., 2001. Synthesis and new conception of North Sea Research (SYCON), Working Group 10: Review of the current knowledge on North Sea benthos, Berichte aus dem Zentrum für Meeres- und Klimaforschung, University of Hamburg, Reihe Z, 12: 138 pp.

Kuenitzer, A., Basford, D., Craeymeersch, J. A., Dewarumez, J. M., Dörjes, J., Duineveld, G. C. A., Eleftheriou, A., Heip, C., Herman, P., 1992. The benthic infauna of the North Sea: Species distribution and assemblages. ICES Journal of Marine Science, 49(2): 127-143.

Kvenvolden, K. A., Lorenson,T. D. 2001. The global occurrence of natural gas hydrate. *In*: Paull, C. K., Dillon, W. D. (Eds.), Natural Gas Hydrates: Occurrence, Distribution, and Detection . Geophysical

Monograph,124: 3-18.

Lang, T., Dethlefsen, V., Von Westernhagen, H., 2003. Fischkrankheiten und embryonale Missbildungen. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.), Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Lawrence R., Labus M., 2003. Early detection of Douglas-Fir beetle infestation with subcanopy resolution hyperspectral imagery. Western Journal of Applied Forestry, 18(3): 202-206.

Lee A.J., Ramster J.W., 1981. Atlas of the Seas around the British Isles. Ministry of Agriculture, Fisheries and Food of the United Kingdom. sections 1.00-5.04, 50 sheets.

Lein, A., Vogt, P. R., Crane, K., Egorov, A., Ivanov, M., 1999. Chemical and isotopic evidence for the nature of the fluid in CH_4-containing sediments of the Håkon Mosby Mud Volcano. Geo-Marine Letters, 19: 76-83.

Lembo, G., Silecchia, T., Carbonara, P., Contegiacomo ,M., Spedicato, M.T., 1999. Localization of nursery areas of *Parapenaeus longirostris* (Lucas 1946) in the central-southern Tyrrhenian Sea by geostatistics. Crustacea 73: 39-51.

Lembo, G., Silecchia, T., Carbonara, P., Spedicato, M.T., 2000. Nursery areas of *Merluccius merluccius* in the Italian Seas and in the east side of the Adriatic Sea. Biologia Marina Mediterranea, 7(3): 98-116.

Levin, L. A., Ziebis, W., Mendoza, G. F., Growney, V. A., Tryon, M. D., Brown, K. M., Mahn, C., Gieskes, J. M., Rathburn, A. E., 2003. Spatial heterogeneity of macrofauna at northern California methane seeps: influence of sulphide concentration and fluid flow. Marine Ecology Progress Series 265: 123-139.

Limonov, A. F., Woodside, J. M., Cita, M. B., Ivanov, M. K., 1996. The Mediterranean Ridge and related mud diapirism: a background. Marine Geolology, 132: 7-19.

Longhurst, A., 1995. Seasonal cycles of pelagic production and consumption. Progress in Oceanography, 36: 77-167.

Longhurst, A. 1998. Longhurst Areas: Ecological Geography of the Sea. Academic Press, 398 pp.

Longhurst, A., Sathyendranath, S., Platt, T., Caverhill, C., 1995. An estimate of global primary production in the ocean from satellite radiometer data. Journal of Plankton Research, 17: 1245-1271.

Lozán, J. L., 2003. Nicht nachhaltige Nutzung der Nordsee durch Fischerei. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.): Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

MacDonald, I. R., Bohrmann, G., Escobar, E., Abegg, F., Blanchon, P., Blinova, V., Brückmann, W., Drews, M., Eisenhauer, A., Han, X., Heeschen, K., Meier, F., Mortera, C., Naehr, T., Orcutt, B., Bernard, B., Brooks, J., De Faragó, M., 2004. Asphalt volcarism and chemosysthetic life in the Campeche Knolls, Gulf of Mexico. Sience, 304: 999-1002.

Matheron, G., 1963. Principles of geostatistics. Economic Geology, 58: 1246-1266.

Matheron, G., 1965. Les variables régionalisées et leur estimation. Masson et Cie., Paris, 305 pp.

Matheron, G., 1971. The theory of regionalized variables and its application. Cahiers du Centre de Morphologie Mathématique de Fontainebleau, no. 5. Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.

Maynou, F., 1998. The application of geostatistics in mapping and assessment of demersal resources. *Nephrops norvegicus* (L.) in the north-western Mediterranean: a case study. Scientia Marina, 62(1): 117-133.

McCoy, J., 2004. Geoprocessing in ArcGIS. Redlands, 370 pp.

Mienert, J., Posewang, J., 1999. Evidence of shallow- and deep-water gas hydrate destabilizations in North Atlantic polar continental margin sediments. Geo-Marine Letters, 19: 143-149.

Milkov, A. V., Vogt, P. R., Cherkashev, G., 1999. Sea-floor terrains of Håkon Mosby Mud Volcano as

surveyed by deep-tow video and still photography. Geo-Marine Letters, 19: 38-47.

Milkov, A.V., 2000. Worldwide distribution of submarine mud volcanoes and associated gas hydrates. Marine Geology, 167: 29-42.

Milkov, A. V., Sassen, R., Apanasovich, T. V., Dadashev, F. G., 2003. Global gas flux from mud volcanoes: A significant source of fossil methane in the athmosphere and the ocean. Geophysical Research Letters, 30(2): 9/1-9/4.

Milkov, A. V., Vogt, P. R., Crane K., Lein, A. Y., Sassen, R., Cherkashev, G. A., 2004. Geological, geochemical, and microbial processes at the hydrate-bearing Håkon Mosby mud volcano: A review. Chemical Geology, 205: 347-366.

Moog, O., Schmidt-Kloiber, T., Ofenböck, T., Gerritsen, J., 2004. Does the ecoregion approach support the typological demands of the EU Water Framework Directive? Hydrobiologia, 516(1): 21-33.

Norcross B. L., Blanchard A., Holladay, B. A., 1999. Comparison of models for defining nearshore flatfish nursery areas in Alaskan waters. Fisheries Oceanography, 8(1): 50-67.

Odobez, J. M., Bouthémy, P., 1994. Robust multi-resolution estimation of parametric motion models applied to complex scenes. IRISA Intern Report, 788 pp.

Olea, R. A., 1999. Geostatistics for Engineers and Earth Scientists. Kluwer Academic Publishers, Boston, Dordrecht, London, 328 pp.

OSPAR Commission, 2000. Quality status report 2000: Region II - Greater North Sea, OSPAR Commission, London, UK, 136 pp.

Pehlke, H., 2005: Prädiktive Habitatkartierung für die Ausschließliche Wirtschaftszone (AWZ) der Nordsee. Diploma thesis, Hochschule Vechta, Germany, 282 pp.

Perencsika, A., Woo, S., Booth, B., Crosier, S., Clark, J., MacDonald, A., 1999. Building a Geodatabase. ESRI Press, Redlands, 390 pp.

Pesch, R., Jerosch, K., in prep-a. Combining geostatistical methods and GIS to estimate temperature maps of the seafloor in the German Exclusive Economic Zone (EEZ) of the North Sea. International Journal of Geographical Information Science.

Pesch, R., Pehlke, H., Jerosch, K., Schroeder, W., Schlüter, M., in prep-b. Using decision trees to predict benthic communities within and near the German Exclusive Economic Zone (EEZ) of the North Sea. Environmental Monitoring and Assessment.

Pessel, N., Opderbecke, J., Aldon, M. J., 2003. An experimental study of robust self-calibration method for a single camera. Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, ISPA, Rome.

Petitgas, P., 1996. Geostatistics and their applications to fisheries survey data. *In*: Moksness, E., (Ed.): Computers in Fisheries Research. Chapman & Hall, London, p. 113-141.

Petitgas, P., 1997. Sole egg distributions in space and time characterised by a geostatistical model and its estimation variance. ICES Journal of Marine Science, 54: 213-225.

Petschick, R., Kuhn, G., Gingele, F., 1996. Clay mineral distribution in surface sediments of the South Atlantic: sources, transport, and relation to oceanography. Marine Geology, 130: 203-229.

Pfeil, B., 2005. Vergleich unterschiedlicher WebGIS-Produkte am Beispiel pflanzensoziologischer Untersuchungen des Fjellbirkenwaldes in Südnorwegen. Möglichkeiten der Publikation und Verbreitung wissenschaftlicher Arbeit. Diploma thesis, University of Bremen, Germany, 146 pp.

Pimenov, N., Savvichev, A., Rusanov, I., Lein, A., Egorov, A., Gebruk, A., Moskalev, L., Vogt, P., 1999. Microbial processes of carbon cycle as the base of food chain of Håkon Mosby Mud Volcano benthic community. Geo-Marine Letters, 19: 89-96.

Pimenov, N. V., Savvichev, A. S., Rusanov, I. I., Lein, A. Y., Ivanov, M. V., 2000. Microbiological processes of the carbon and sulphur cycles at cold methane seeps of the North Atlantic. Microbiology, 69: 709-720.

Poon, K.-F., Wong, R.W.-H., Lam, M. H.-W., Yeung, H.-Y., Chiu,T. K.-T., 2000. Geostatistical modelling of the spatial distribution of sewage pollution in coastal sediments. Water Research 34(1): 99-108.

Pratje, O., 1951. Die Fortsetzung der Endmoränen am Boden der Nordsee. Zeitschrift der Deutschen Geologischen Gesellschaft, 103: 75-77.

QRS, 2000. Quality Status Report, Region II Greater North Sea. OSPAR Commission, London, 136 pp.

Rachor, E., Nehmer, P. 2003. Erfassung und Bewertung ökologisch wertvoller Lebensräume in der Nordsee. Abschlußbericht für das F & E-Vorhaben FKZ 899 85 310 (Bundesamt für Naturschutz). Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven.

Reiniger, C., 1997. Bioregional planning and ecosystem protection. *In*: Thompson, G.F.; Steiner, F. (Eds.) Ecological Design and Planning. John Wiley & Sons, New York, p. 185-200.

Remane, A., 1940. Einführung in die zoologische Ökologie der Nord- und Ostsee. *In*: Grimpe, G., Wagler, E. (Eds.), Tierwelt der Nord- und Ostsee: Ia, Akademische Verlagsgesellschaft, Leipzig, 238 pp.

Riek, R.; Wolff, B., 1997. Repräsentanz des europäischen 16x16 km-Erhebungsnetzes für Aussagen zum deutschen Waldbodenzustand. Arbeitsbericht des Instituts für Forstökologie und Walderfassung 97/2, Eberswalde.

Richmond, A., 2005. An alternative implementation of indicator kriging. Computers and Geosciences, 28: 555-565.

Rivoirard, J., Simmonds, E. J., Foote, K., Fernandes P., Bez, N., 2000. Geostatistics for estimating fish abundance. Blackwell Science, UK, 216 pp.

Rivoirard, J., Wieland, K., 2001. Correcting for the effect of daylight in abundance estimation of juvenile haddock (*Melanogrammus aeglefinus*) in the North Sea: An application of kriging with external drift. ICES Journal of Marine Science, 58: 1272-1285.

Roerdink, J. B., Meijster, A., 2000. The watershed transform: Definitions, algorithms and parallelization strategies. Fundamenta Informaticae, 41(1-2): 187-228.

Rosenfeld, B., Lewis, C. H., 2005. Assessing violence risk in stalking cases: A regression tree. Law and Human Behavior, 29(3): 343-357.

Rufino, M. M., Maynou, F., Abelló, P., Gil de Sola, L., Yule, A. B., 2005. The effect of methodological options on geostatistical modelling of animal distribution: A case study with *Liocarcinus depurator* (Crustacea: Brachyura) trawl survey data. Fisheries Research, 76: 252-265.

Ryan, W. F., 1995. Forecasting severe ozone episodes in the Baltimore metropolitan area. Atmospheric Environment, 29(17): 2387-2398.

Saetre, M., Becker, G., 1990. Review of physical oceanography of the North Sea. Neth. Journal of Sea Research, 26(2-4): 161-238.

Sahling, H., Rickert, D., Lee, R. W., Linke, P., Suess, E., 2002. Macrofaunal community structure and sulfide flux at gas hydrate deposits from the Cascadia convergent margin, NE Pacific. Marine Ecology Progress Series, 231: 121-138.

Sahling, H., Galkin, S. V., Salyuk, A., Greinert, J., Foerstel, H., Piepenburg, D., Suess, E., 2003. Depth-related structure and ecological significance of cold-seep communities – a case study from the Sea of Okhotsk. Deep-Sea Research I, 50: 1391-1409.

Sahling, H., Wallmann, K., Dählmann, A., Schmaljohann, R., Petersen, S., 2005. The physicochemical habitat of *Sclerolinum sp.* at Hook Ridge hydrothermal vent, Bransfield Strait, Antarctica. Limnology and Oceanography, 50(2): 598-606.

Salzwedel, H., Rachor, E., Gerdes, D., 1985. Benthic macrofauna communities in the German Bight. Veröffentlichungen des Institutes für Meeresforschung in Bremerhaven, 20: 199-267.

Sauter, E. J., Muyakshin, S. I., Charlou, J.-L., Schlüter, M., Boetius, A., Jerosch, K., Damm, E., Foucher, J.-P., Klages, M., in press. Methane discharge from a deep-sea submarine mud volcano into the upper water column by gas hydrate-coated methane bubbles. Earth and Planetary Science Letters.

Schlüter, M., 1996. Einführung in geomathematische Verfahren und deren Programmierung. Enke Verlag, Stuttgart, 326 pp.

Schlüter, M., Rutgers van der Loeff, M. M., Holby, O., Kuhn, G., 1998. Silica cycle in surface sediments of the South Atlantic. Deep-Sea Research I, 45: 1085-1109.

Schlüter, M., Sauter, E. J., Schäfer, A., Ritzau, W., 2000. Spatial budget of organic carbon flux to the seafloor of the northern North Atlantic (60°N - 80°N). Global Biogeochemical Cycles, 14(1): 329-340.

Schönwiese, C.D., 2000. Praktische Statistik für Meteorologen und Geowissenschaftler. Gebrüder Bornträger, Berlin, Stuttgart, 298 pp.

Schröder, W., Schmidt, G., 2003. Medienübergreifende Umweltbeobachtung in Baden-Württemberg. Ergebnisse eines Modellprojekts. *In*: Landesanstalt für Umweltschutz (Eds.), Medienübergreifende Umweltbeobachtung. Stand und Perspektiven. Karlsruhe, p. 39-60.

Seiter, K., Hensen, C., Schröter, J., Zabel, M., 2004. Organic carbon content in surface sediments – defining regional provinces. Deep-sea Research I, 51: 201-2026.

Shi, J., Tomasi, C., 1994. Good features to track. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR94), Seattle, WA, p. 593-600.

Sibuet, M., Olu, K., 1998. Biogeography, biodiversity and fluid dependence of deep-sea cold-seep communities at active and passive margins. Deep-Sea Res. II, 45: 517-567.

Sibuet, M., Olu-Le Roy, K., 2002. Cold seep communities on continental margins: structure and quantitative distribution relative to geological and fluid venting patterns. *In*: Wefer, G., Billett, D., Hebbeln, D., Jørgensen, B. B., Schlüter, M., Van Weering, T. (Eds.), Ocean Margin Systems, Springer-Verlag, Berlin-Heidelberg, p. 235-251.

Smirnov, R. V., 2000. Two new species of pogonophoran from the Arctic mud volcano off northwestern Norway. Sarsia, 85: 141-150.

Soltwedel, T., Portnova, D., Kolar, I., Mokievsky, V., Schewe, I., 2005. The small-sized benthic biota of the Håkon Mosby Mud Volcano (SW Barents Sea slope). Journal of Marine Systems, 55: 271-290.

Söntgerath, K., 2003. Gewinnung von Rohstoffen. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.): Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Steinberg, D., Colla, P., 1995. CART: Tree-structured Non-Parametric Data Analysis. Salford Systems, San Diego, CA.

Streif, H. 2003. Die Nordsee im Wandel - vom Eiszeitalter bis zu Neuzeit. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.), Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Streit, U., 2004. Werkzeuge zur numerischen Modellierung. Das Interpolationsverfahren Kriging. Online lecture, Institute for Geoinformatics, University of Münster, Germany (online at: http://ifgivor.uni-muenster.de/vorlesungen/Num_Modellierung/index.html).

Swan, G. E., Javitz, H. S., Jack, L. M., Curry, S. J., McAfee, T., 2004. Heterogeneity in 12-month outcome among female and male smokers. Addiction, 99(2): 237-250.

Taniguchi, M., Burnett, W. C., Cable, J. E., Turner, J. E., 2002. Investigation of submarine groundwater discharge. Hydrological Processes, 16: 2115-2129.

Thorson, G. E., 1957. Bottom communities (sublittoral or shallow shelf). Geological Society of America Memoirs, 67(1): 461-534.

Thuiller, W., 2003. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology, 9(10): 1353-1362.

Tiedemann, A., 2003. Windenergieparke im Meer. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.): Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz.

Wissenschaftliche Auswertungen, 447 pp.

Van Westen, C. J., Rengers, N., Terlien, M. T. J., Soeters, R., 1997. Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. International Journal of Earth Sciences, 86: 404-414.

Vincent, A. G., Jouffroy, J., Pessel, N., Opderbecke, J., Borgetto, M., Rigaud, V., 2003. Real-time georeferenced video mosaicing with the MATISSE system. Proceedings of the Oceans 2003 Marine Technology and Ocean Science Conference, MTS/IEEE OCEANS'03, San Diego, USA, September 22-26, 2003, 4: 2319-2324.

Vogt, P. R., Cherkashev, G., Ginsburg, G., Ivanov, G., Milkov, A., Crane, K., Lein, A., Sundvor, E., Pimenov, N., Egorov, A., 1997. Håkon Mosby Mud Volcano provides unusual example of venting. EOS 78(549): 556-557.

Vogt, P. R., Gardner, J., Crane, K., 1999. The Norwegian-Barents-Svalbard (NBS) continental margin: Introducing a natural laboratory of mass wasting, hydrates and ascent of sediment pore water and methane. Geo-Marine Letters, 19: 2-21.

Von Rohr, G., 2003. Die Entwicklung der Universalhäfen zwischen Hamburg und Antwerpen. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.): Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Wackernagel, H., 1998. Multivariate Geostatistics: an introduction with applications. Springer-Verlag, Berlin, Heidelberg, 312 pp.

Walmsley, J. L., Barthelmie, R. J., Burrows, W. R., 2001. The statistical prediction of offshore winds from land-based data for wind-energy applications. Boundary-Layer Meteorology, 101(3): 409-433.

Webster R., Oliver M. A., 2001. Geostatistics for environmental scientists. John Wiley and Sons. Ltd, Chichester, New York, 286 pp.

Weigel, S., 2003. Belastung der Nordsee mit organischen Schadstoffen. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.): Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Wentworth, C.K., 1922. A scale of grade and class terms for clastic sediments. Journal of Geology, 30: 377-392.

Whiticar, M., 1996. Isotope tracking of microbial methane formation and oxidation. Mitteilungen Internationale Vereinigung für theoretische und angewandte Limnologie, 25: 39-54.

Wirtz K., Schuchardt, B., 2003. Auswirkungen von Rohrleitungen und Stromkabeln. *In*: Lozán, J. L., Rachor, E., Reise, K., Sündermann, J., Von Westernhagen, H. (Eds.): Warnsignale aus Nordsee & Wattenmeer: Eine aktuelle Umweltbilanz. Wissenschaftliche Auswertungen, 447 pp.

Woodhouse, S., Lovett, A., Dolman, P., Fuller, R., 2000. Using a GIS to select priority areas for conservation. Computers, Environment and Urban Systems, 24: 79-93.

Wright, D., Bartlett, D., 2000. Marine and Coastal Geographical Information System, Taylor & Francis, London, 360 pp.

Yisehac, Y., Webb, P., 1999. Classification and Regression Trees, Cart: A user manual for identifying indicators of vulnerability to famine and chronic food insecurity. Microcomputers in Policy Research 3, International Food Policy Research Institute, Washington, DC, 50 pp.

Yoklavich, M. M., 1997. Applications of side-scan-sonar and in-situ submersible survey techniques to marine fisheries habitat research. *In*: Boehlert, G. W., Schuhmacher, J. D. (Eds.): Changing Oceans and Changing Fisheries: Environmental Data for Fisheries Research and Management, NOAA Technical Memorandum, p. 140-141.

Zabel, M., Hensen, C., Schlüter, M., 2000. Back to the ocean cycles: Benthic fluxes and their distribution patterns. *In*: Schulz, H. D., Zabel, M., (Eds.), Marine Geochemistry. Springer-Verlag, Berlin-Heidelberg, 455 pp.