NGU REPORT
2015.043

An assessment of scale, sampling effort and confidence for maps based on visual and acoustic data in MAREANO

GEOLOGY FOR SOCIETY
SINCE 1858

GEOLOGICAL
SURVEY OF
NORWAY
- NGU -

NORGES GEOLOGISKE UNDERSØKELSE
- NGU -

REPORT

Geological Survey of Norway
P.O.Box 6315 Sluppen
NO-7491 TRONDHEIM
Tel.: 47 73 90 40 00

Report no.: 2015.043
ISSN: 0800-3416 (print)
ISSN: 2387-3515 (online)
Grading: Open

Title:
An assessment of scale, sampling effort and confidence for maps based on visual and acoustic data in MAREANO

Authors:
Terje Thorsnes, Thijs Christiaan van Son, Margaret F.J. Dolan, Genoveva Gonzalez-Mirelis, Nicole Baeten, Pål Buhl-Mortensen, Lilja Rún Bjarnadóttir, Hanne Hodnesdal, Valérie Bellec

Client:
MAREANO

County:

Commune:

Map-sheet name (M=1:250.000)

Map-sheet no. and -name (M=1:50.000)

Deposit name and grid-reference:

Number of pages: 97 **Price (NOK):** 280
Map enclosures:

Fieldwork carried out:

Date of report:
20.10.2015

Project no.:
311720

Person responsible:
Reidulv Bøe
Reidulv Bøe

Summary:

MAREANO has mapped Norwegian waters for ten years, and produced several map series at different scales addressing bathymetry, geology, biology and pollution. There has been an increasing demand to document the confidence of these maps, and to evaluate the effectiveness of the present survey strategy in relation to these map products. One aspect of this study has been to examine whether the future survey strategy should be modified in order to optimise the cost/benefit of the programme and produce the best possible map products.

This report provides an overview of current survey strategy and methods; a confidence assessment of selected maps (namely, sediment grain size, biotopes, vulnerable habitats); an overview and discussion of different sampling strategies with recommendations; and an overview of the user needs in national management and industry. As part of the study, we have attempted to quantify the natural environmental variation of the seabed, and developed a tool for doing this. This work is presented in full in a separate report.

The study has provided answers to many of the questions raised at the start of the work, but has raised even more questions that should be adressed, in order to optimise future MAREANO survey and sampling strategy. MAREANO is a unique programme, addressing multiple objectives. Mapping the seabed requires substantial resources, and it is important that the strategy and methods employed are optimal with respect to programme objectives, in order to maximise the cost-benefit ratio. The recommendations for future work which are outlined at the end of this report therefore represent an important part of this report. Further optimisation of the programme will require resources for testing and development.

Keywords: Scale	Sampling effort	Confidence
Video	Grain size	Biotope
Vulnerable habitats	GRTS	Survey strategy

CONTENTS

1. INTRODUCTION

The ultimate goal for the MAREANO programme is to provide knowledge for ecosystem based management, including data needed for marine spatial planning. In order to do this, the programme collects acoustic and visual data, and physical samples from the seabed. In the period since MAREANO started in 2005 to the end of 2014, the programme has collected 160 000 square kilometres of multibeam echosounder data, and mapped a similar area with respect to geology, biology and chemistry. The data are analysed, and results reported primarily through maps, supported by reports and scientific publications on the programme website (www.mareano.no). Key map products include full coverage bathymetric maps, geological maps, biotope maps, vulnerable biotopes/habitats maps, and point maps for biology and chemistry.

The goal of this study is to evaluate whether the current survey design strategy and sampling effort are optimal for the products MAREANO makes (sampling effort is defined as the number of stations with visual observations and/or physical sampling per unit area). Several key questions emerge in relation to this evaluation. Are there different needs for mapping geology versus biotopes? If that is the case, how can we accommodate both needs in a cost-effective manner? Can we define a certain sampling effort for a certain scale, or should the sampling effort be dynamic and respond to expected physical (ocean currents, geology) and biological complexity? What is the optimal sampling effort for mapping at different scales? What levels of confidence are associated with given combinations of nature complexity, sampling effort and scale? What is the optimal way of collecting visual data (e.g. along lines or in squares)? And what is the optimal length of video lines and segments used for further analysis? Shall sediment map production in the future be more based on modelling, rather than expert interpretation? How can we integrate the expertise of the scientists in a modelling based map production? Is the approach to biotope classification and modelling optimal, or are there advances to be made? During the course of this study, many of these questions have proved to be more complex and demanding to solve than was realised at the start. Whilst we have been able to answer some of the questions, perhaps more importantly, as a result of this study we are now in a much better position to ask the right questions, and have a clearer idea of how to answer them through future studies.

At the start of the programme, the ambitions for sampling effort were high, and the aim was to collect 30 video stations and 7 physical stations per 1000 square kilometres. The planned sampling effort was reduced to 15 video and 3 physical stations in 2009. Since 2010, the planned sampling effort has been relatively constant at 10 video and 2 physical stations per

1000 square kilometres for the surveyed areas between Finnmark and Møre. For the southeastern part of the Norwegian Barents Sea, the planned sampling effort was reduced to 5 video stations and 1 physical station per 1000 square kilometres.

In this report, we focus on a selection of full coverage maps (sediment grain size, biotope, vulnerable habitats) which are based mainly on visual data (video) in conjunction with acoustic data. As far as possible we evaluate the confidence of the maps currently produced and also give recommendations for future survey design strategy. Results based on single point observations for biology and chemistry (physical samples) are beyond the scope of this report but it is suggested these be examined in a separate report in the near future.

The questions raised in this study require complex analytical methods in order to be answered, and it has been challenging to give answers and numbers to quantify many of our findings. Nevertheless this report provides important documentation and a basis for improvement of MAREANO and we have defined several outstanding questions that will need to be addressed by follow-up studies. The process of this evaluation has also shed light on good and weaker parts of the MAREANO work flow, all the way from survey planning to implementation of the results in the national management.

2. SURVEY STRATEGY AND METHODS - CURRENT STATUS

The areas to be surveyed with acoustic and visual tools, and physically sampled, are chosen by the MAREANO programme management (programme group, steering group). This is usually concluded the year before the surveys, following long-term planning spanning over several years.

2.1 Acoustic data

2.1.1 Multibeam echosounders

Data on bathymetry, backscatter and the water column are collected by multibeam echosounders mounted on surface vessels (figure 1).

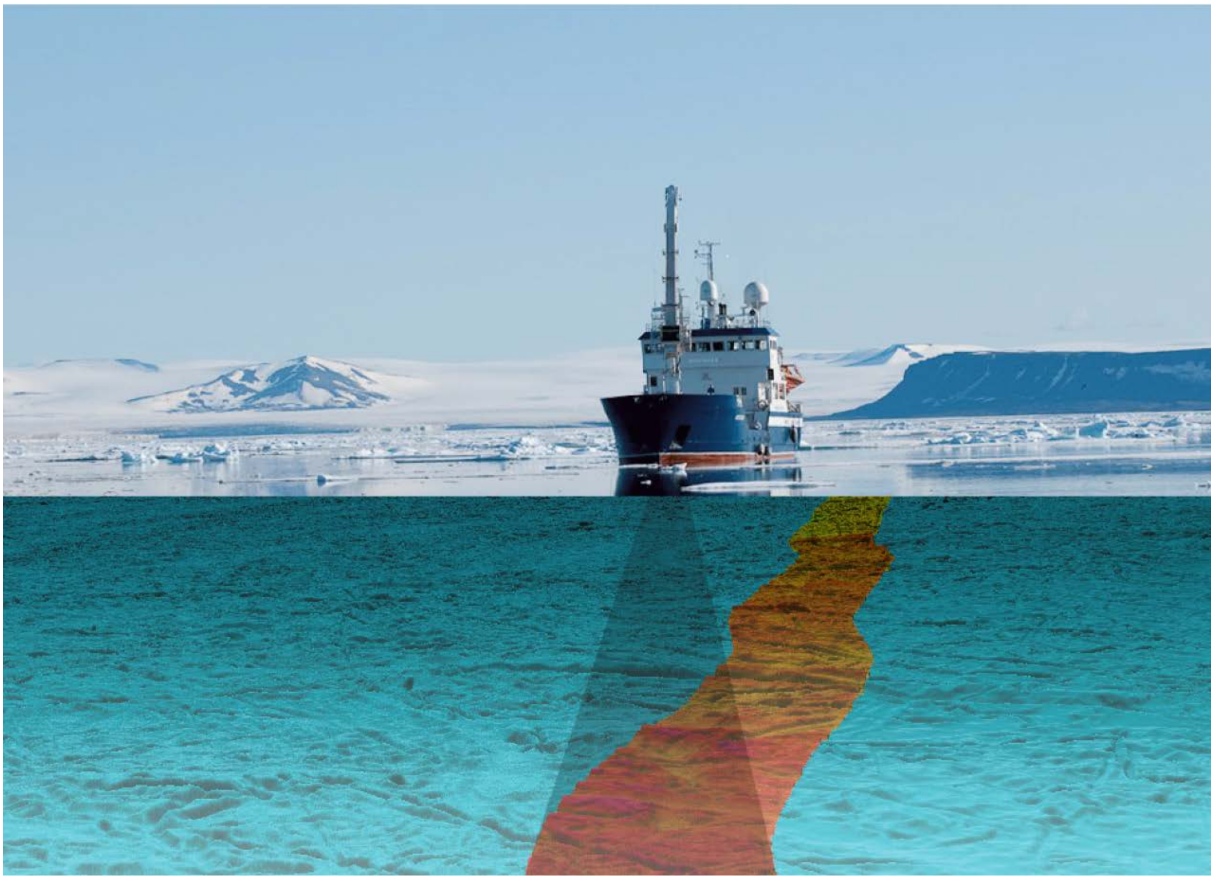


Figure 1. Illustration of surveying using multibeam echosounder.

Bathymetry data are used to produce digital terrain models (DTM) of 5, 25, and 50-metre resolution (figure 2). In some areas it may be possible to produce DTMs with even higher resolution depending on the depth, multibeam echosounder used and the resulting sounding density. Inside the territorial boundary, the resolution is limited to 50 metres in compliance with Norwegian military classification regulations. It is also possible to collect bathymetric data using autonomous underwater vehicles (AUVs). This can give data with grid sizes less than 10 cm at any water depth, because the AUV is moving close to the seabed. This has not yet been done in MAREANO but AUV trials will be conducted in October 2015 including acquisition of bathymetry data.

Bathymetry data give important information on the general seabed topography, as well as all seabed features (like iceberg ploughmarks, coral reefs, pockmarks, sand waves, boulders etc.). Both the surveying and processing must be done in a careful manner to preserve all the seabed information, yet facilitate removal of any faulty soundings. Data artefacts can occur; it is important that seabed features are not camouflaged by artefacts, or that artefacts appear as false seabed features. Despite cleaning, some motion-related artefacts may persist in otherwise clean data, especially when the surveys have been conducted in poor sea conditions.

Echosounder measurements also provide reflectivity (backscatter) data, which are essential for interpreting seabed sediments. Backscatter provide a first indication of the hardness of the seabed and thus the sediment type, but represent only an acoustic proxy to sediment properties. Backscatter must therefore be ground truthed – i.e. verified by visual and/or physical sampling. Together with bathymetry and ground truthing information, acoustic backscatter data provide the basis for interpretation of sediment distribution. Backscatter can also be useful for prediction of seabed habitats (biotopes). However, it is worth noting that there are several challenges associated with the use of backscatter data. They are usually not normalised, meaning that absolute reflectivities cannot be compared from one area to another. Also, backscatter for a given sediment varies if echosounders with different frequencies are used. The backscatter is texture dependant, meaning that an uneven seabed may give a different response from an even seabed. Finally, the modes and settings of echosounders vary, being an additional source of variations not related to sediment variability. For these reasons, extra care must be taken when interpreting reflectivity data (see also section. 2.2.1). Successful use of backscatter is dependant on sufficient ground-truthing and qualified expert interpretation.

Since 2010, water column data have also been acquired using multibeam echosounders. These are reflections of the sonic pulse as it travels through the water, before it reaches the seabed. Such data can be used to detect objects or anomalies in the water column, for example to detect gas bubbles rising from the seabed.

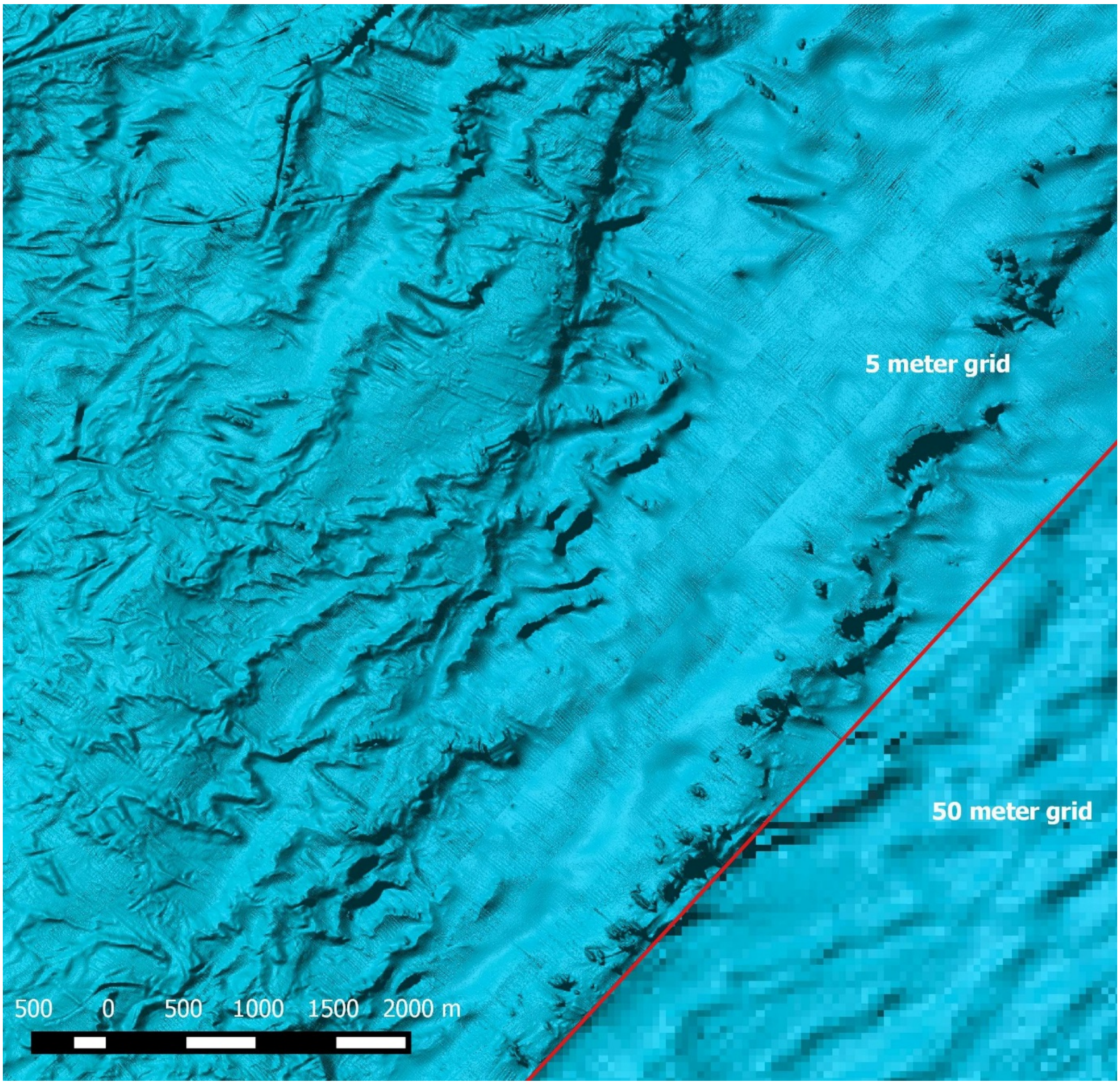


Figure 2. Digital terrain models from Sularevet with 50 metres resolution (down to the right) within the territorial border, and 5 metres resolution (up to the left) outside the territorial border. Coral reefs, moraines and iceberg ploughmarks can be identified using 5 metres resolution, but not using 50 m.

If bathymetric data of adequate quality and resolution exist, these are used by MAREANO. Sources of existing data are the Norwegian Mapping Authority (Norwegian Hydrographic Service, NHS), the Norwegian Defence Research Establishment (FFI), the Norwegian Petroleum Directorate, Olex AS, the petroleum industry, universities, research institutes and other bodies. Olex data (figure 3) is a compilation of mostly single beam echosounder data acquired by working vessels using the Olex vessel navigation system. Use of Olex for sediment and biotope mapping has been evaluated (Elvenes et al. 2012), showing a considerable potential, with some important limitations. Data provided by Olex to MAREANO have been used at Mørebanken. The coverage varies, and in areas with good coverage (important fishing areas), it is possible to produce a DTM with a resolution of 50 metres. Olex data do normally not have backscatter or water column data, nor do some of the older multibeam surveys.

Since 2006, most bathymetric surveys, in water depths shallower than 1200 metres, have been carried out using the EM710 echosounder manufactured by Kongsberg Maritime. EM2040 has been used for some areas, but this echosounder has higher frequency and is best suited for areas shallower than 400-500 metres. EM710 and EM2040 provide a sounding density of approximately 0.9 metres at a water depth of 100 metres, 1.8 metres at 200 metres depth, and so on. This enables use of high resolution (5 metre grid) terrain models (figure 2), which can be used to interpret small seabed features (coral reefs/bioclastic mounds, sandwaves etc.). In areas of more than 1000 metres water depth, it is mainly the EM300 echosounder that is used, which provides coarser resolution. In 2011, EM302 was also used in deep water areas, and effectively provides the same number of measurements as the EM710, although the sounding density is less because of the water depth (12 metres density at a water depth of 1500 metres, and 24 metres at 3000 metres). Whilst sounding density is the main concern related to bathymetric data, frequency is important for mapping seabed properties. Variations in frequency present an additional challenge when interpreting backscatter across neighbouring datasets, especially when ground truthing data only document the uppermost seabed (see section 2.2.1 for further information).

Multibeam echosounders are normally set up to measure up to 60 degrees on either side of the vertical. In this case, a boat line covers a belt on the seabed, which is about 3.5 times as wide as the water depth. In other words, if the water depth is 100 metres, a 350-metre wide belt is covered, in 200 metres of water the belt is 700 metres wide, and so on. This is a general rule of thumb and may vary with multibeam system and data acquisition setup.

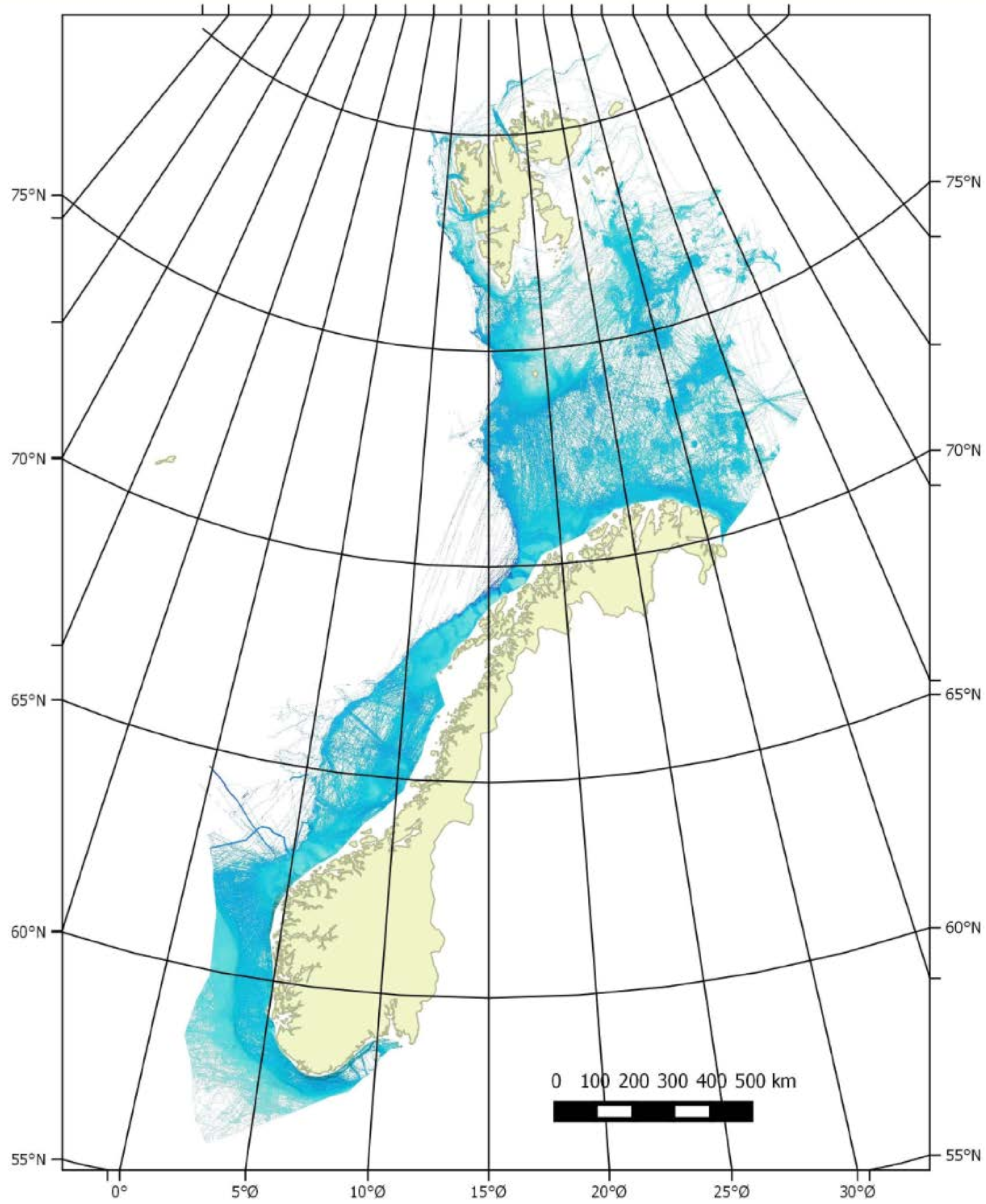


Figure 3: Coverage of Olex data in Norwegian sea areas (2010).

Various instruments are used to measure the vessel's position, motion (course, speed, heave, roll and pitch) and the velocity of sound in the water, both close to the sonar transducer and through the water column. The interaction between the various sensors is important, and the installation of the equipment must be accurately calibrated with reference to the vessel's coordinate system. For details on this and other aspects of multibeam mapping, see the technical specification for bathymetric mapping in MAREANO (NHS 2010). Note that these specifications currently include limited description of backscatter standards which have been introduced by MAREANO in order to assure that the best possible backscatter data are acquired. These standards should be re-evaluated in accordance with the recently published Guidelines and Recommendations for Backscatter measurements by seafloor mapping sonars (Lurton and Lamarche, 2015) as soon as possible.

Weather limitations also need to be taken into account. Large waves result in poor quality bathymetric data, partly because air bubbles created under the vessel interfere with the echosounder measurements. To some extent this can be compensated for by sailing survey lines along the prevailing wave direction, rather than against it, or by reducing the aperture of the echosounder and planning a greater overlap between survey lines. Such measures will result in slower progress. At some stage there will be too much noise, and data acquisition must be suspended. At what sea state this will occur depends on, among other things, the size of the vessel, the design of the hull and the location of the echosounder.

Data processing is another important aspect contributing to the quality of the data. To obtain the best possible geographical positioning, correction data are downloaded from GPS satellites following the field work, and these are used to refine the positioning data. The recorded bathymetric data are filtered to remove noise and checked for gaps. This involves a certain amount of manual work and is therefore a time-consuming process. Data are not smoothed, as this may camouflage seabed structures. Backscatter data are also processed using software that makes the necessary geometric corrections. Typically backscatter data are gridded to the same or higher resolution than the bathymetry data for the same area.

Hydrographic surveying is expensive. The price depends on the water depth, with the price per square kilometre being highest in shallow water areas. Some of the hydrographic surveys conducted for MAREANO have been carried out using government owned vessels: Hydrograf (NHS), H.U. Sverdrup (FFI) and G.O. Sars (IMR). However, in general, hydrographic services have been purchased on the international market. About 40% of the funding allocated to MAREANO has been used to finance multibeam mapping. This includes hydrographic surveying (field work), data processing (filtering of noise), administration of bathymetric data and the production and distribution of products and services. Management of backscatter and water column data are covered by the same budget. Processing of backscatter and water column data is done by NGU on a separate budget.

2.1.2 High resolution seismic data

High resolution seismic data are collected to complement multibeam data, permit characterization of sediment stratigraphy and sediment types, and for studies of geological processes on the seabed and in the upper sediment column. Lines with high resolution seismic data should ideally cover a wide range of geological settings, in order to provide information on the geological processes which have formed the seabed. This would give additional information for optimal location of visual and physical sampling. In practice, the majority of high resolution seismic lines acquired by MAREANO are collected on the sampling cruises, in transit between sampling stations, or on transits to and from ports. This still gives valuable information, but is not ideal.

G.O. Sars, which is the main vessel used by MAREANO, is equipped with a hull mounted TOPAS PS 18 (TOpographic PArametric Sonar). The TOPAS data are used, in combination with other seabed data, for mapping seabed sediment distribution and linking biological and geological processes. TOPAS PS 18 is designed for sub-bottom profiling with very high spatial resolution in water depths from less than 20 metres to full ocean depth. The low frequency signal is generated in the water column by non-linear interaction between two high frequency signals (centered symmetrically around 18 kHz). Similarly, a sum frequency signal is also generated. However, only the low frequency signal (0.5-6 kHz) is used for sub-bottom profiling. The parametric sources have the advantage of generating a low frequency, narrow (4.5 degrees for TOPAS PS 18) signal beam with no distinct side lobe structure.

The system can operate with various signal waveforms for optimal performance: Typically, Ricker pulses are used for very high resolution work, Chirp pulses are used for deep water, high penetration work and CW pulses are used for narrow band, frequency sensitive work. Penetration performance depends on sediment characteristics, water depth, transmitted signature, noise level etc. According to Kongsberg Defence Systems, penetration of 200 metres has been achieved in water depths of more than 3000 metres with a sediment layer resolution of typically 20 cm or better.

MAREANO has experienced that TOPAS PS 18 can penetrate 100-150 metres into soft, fine-grained sediments and that the resolution is better than 0.5 metre in Chirp-mode. The penetration is less in sand, in coarse-grained sediments, or where the bottom is hard (i.e. till). The resolution increases towards shallower water due to the decreasing area of the signal footprint. In areas with irregular seafloor topography or steep slopes the resolution decreases because of larger footprint. Where the seabed slopes more than 4-5 degrees, the data cannot be used. Reasonably good data are obtained with up to 4-5 metres wave height when the survey vessel runs in the direction of the weather.

The primary frequency pulses around 20 kHz or its harmonics can interfere with other acoustic systems onboard, and the systems should be synchronized so that high quality data can be acquired both during transit and planned surveys. Raw data from the TOPAS are stored as TOPAS-files and later processed and converted to segy-files for seismic interpretation utilizing commercial software.

High resolution seismic data are crucial for reliable mapping of seabed sediments and bottom types. TOPAS PS 18 is installed also onboard H.U. Sverdrup (FFI). Other bottom penetrating sonar/pinger systems onboard of other ships can be used for collection of high resolution seismic data. Data quality may vary depending on weather condition, type of system, processing etc. It is also possible to collect very high resolution bottom penetrating sonar data utilizing autonomous underwater vehicle (AUV). For further details on high resolution seismic data, see Bøe et al. (2010).

It is possible to use high resolution seismic systems together with multibeam echosounders, but they must be carefully synchronized with the echosounder, so that they do not disturb each other. This will reduce the ping rate of the echosounder (i.e. how often it can measure), but high resolution does not need to be measured on every line. Every fourth line should be adequate in order to meet the objectives of providing relevant information to support the interpretation of seabed substrates.

2.2 Visual data for geology and biology

2.2.1 Sampling strategy

Sampling stations for geological and biological sampling have been selected primarily on the basis of multibeam data. Since 2015, oceanographic data have also been taken into account. Stations are placed to ensure representative sampling across the survey area, based on the allotted number of samples per square kilometres set in the annual budget (typically 10 video stations per 1000 square kilometres).

The sampling stations acquire data for multiple objectives:

- Ground truthing of multibeam backscatter data for sediment interpretation
- Visual documentation of biotopes (expected to vary with physical environment)
- Visual documentation of vulnerable habitats (Norwegian Environment Agency and OSPAR - Convention for the Protection of the Marine Environment of the North-East Atlantic)
- Video and sample material for assessment of biodiversity and production
- Sample material for geological and geochemical analysis
- Verification of morphological features identified by multibeam bathymetry data, particularly those with ecological relevance, e.g. coral reefs

An order of priority for positioning sampling stations with respect to each of these objectives has not yet been set by MAREANO. Selection of sampling stations in MAREANO has largely been guided by expert judgement in an attempt to obtain representative samples of all major habitats in the study area and a good geographic spread spanning depth intervals of known biological relevance (e.g. water mass boundaries). The station planning process is typically as follows: A preliminary assessment of the geology and geomorphology of an area is conducted by NGU scientists based on multibeam bathymetry data from NHS. Multibeam backscatter data are processed by NGU and assessed as a first pass indication of surficial sediment type. After combining bathymetry and backscatter data, sampling stations are planned to obtain representative sampling of the major geomorphic features and to ground-truth the different acoustic signatures in the multibeam backscatter. Multibeam data for a bio-geo survey area typically come from several mapping surveys that have used different survey vessels/contractors/multibeam systems (figure 4). Sampling must ensure, as far as possible, that backscatter levels from each of the individual surveys are ground truthed so that the data can be used to interpret sediment distribution across the entire area.

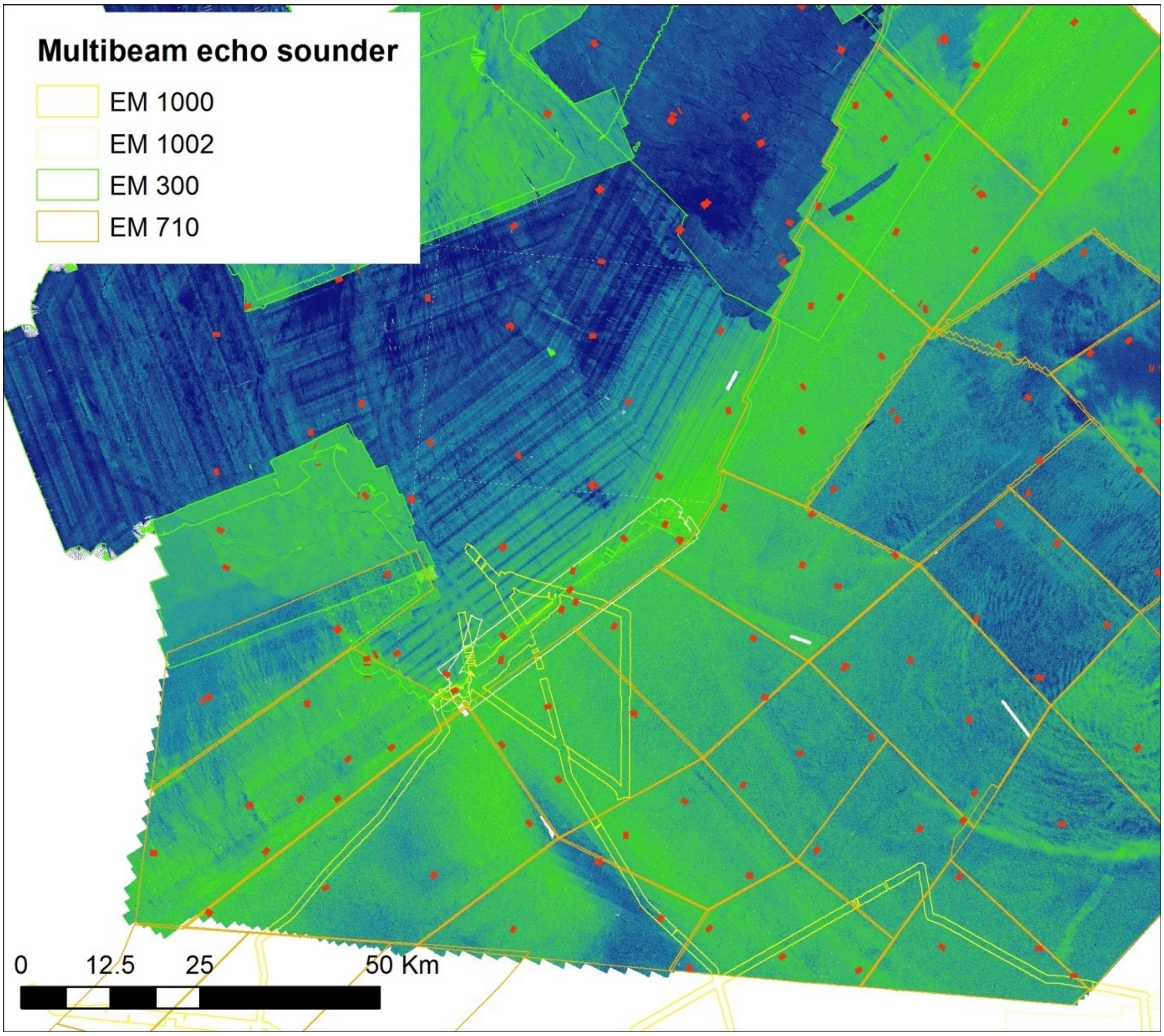


Figure 4: Example of backscatter data from Nordland VI. Data are from 4 different echosounders - EM1000, EM1002 (95kHz), EM710 (70-100), EM300, (30kHz), which have different penetration depths. Backscatter dB from each survey are shown on the same colour ramp highlighting differences in the values from low (blue) to high (green). MAREANO video stations are shown in red.

In addition to a visual assessment, these multibeam data are also subjected to a coarse unsupervised classification which splits the survey area into areas with similar physical characteristics based on bathymetry, backscatter, and terrain attributes (e.g. slope, curvature etc.) at both local (100s metres) and broader scales (kilometres). To date the ISOCLUSTER algorithm with maximum likelihood classification in ArcGIS has been used to provide this initial, broad-scale assessment of physically distinct areas, which are likely to reflect different seabed habitats in terms of the bottom type and structure, and the biological communities living there.

This ISOCLUSTER layer (figure 5) is used to provide a more objective overview of the study area and sufficient number of samples from each major class is ensured by expert judgements. Other data/knowledge from the study area is also considered at the cruise planning stage. Oceanographic data has been taken into account since 2015.

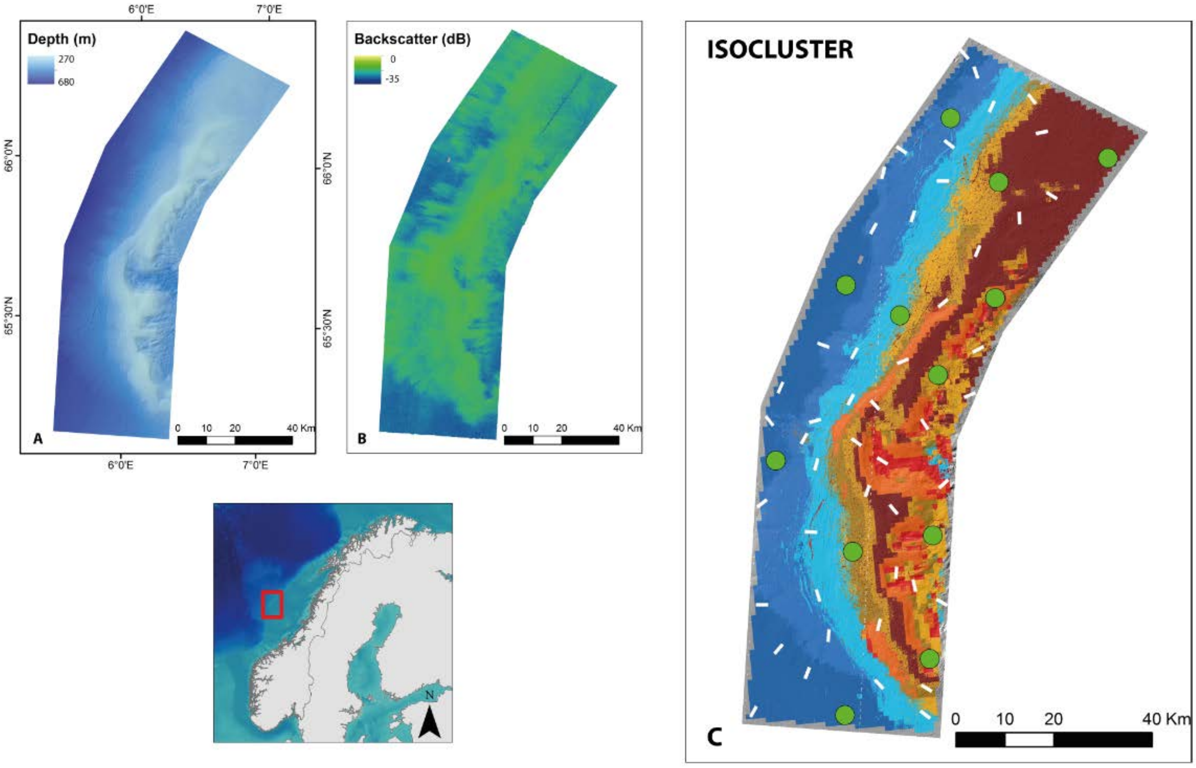


Figure 5. Example data used for planning MAREANO survey stations with location map at Skjoldryggen in the Norwegian Sea, surveyed by MAREANO in 2013. (A) Multibeam bathymetry shown as colour shaded relief. (B) Multibeam backscatter map indicating variation in sediment properties. (C) ISOCLUSTER map based on unsupervised classification of bathymetry, backscatter and derived quantitative terrain variables. The colours indicate physically different areas based on these variables — similar colours indicate areas with similar physical characteristics while contrasting colours indicate physically different areas. Planned video lines indicated with white lines, green circles indicate full sampling stations.

In due time before each cruise, a more formal cruise planning meeting is held to review suggested survey plans and harmonise biological, geological and chemical sampling objectives. This meeting is also open to members of the MAREANO programme group, representing different stakeholder interests. The survey area is discussed using a live GIS presentation of all available data, and any suggested sampling stations from NGU and IMR are discussed and optimized. In addition to planning visual stations, a selection of the video stations are chosen as stations where also physical sampling is conducted. These stations are considered to be representative of the study area, and also suitable for the sampling gear, to minimize damage to the seabed/sampling gear.

The cruise planning meeting marks a formal acceptance of the sampling design for a forthcoming cruise. Any onboard revisions of this plan (e.g. due to bad weather, equipment problems, fishing gear in sampling area) are made by the cruise leader (IMR) in consultation with NGU scientists on board, involving MAREANO project management onshore as necessary.

During some cruises, MAREANO has been able to acquire additional samples due to particularly favourable weather/sea conditions, other cruises have not managed to acquire all the planned samples. Whether or not these missed samples are acquired by later cruises has been assessed by the MAREANO project managers etc. in planning the next MAREANO activities.

2.2.2 Methods

Individual video survey lines were first selected to be 1000 metres long, but after the initial MAREANO sampling cruise in 2006, results were assessed and the distance was reduced to 700 m, based on cumulative curves for the number of observed taxa along individual transects (Buhl-Mortensen et al. 2015). Each transect starts and ends with the video platform standing on the seabed, enabling close-ups and visual scanning within an area of approximately 6 m^2. Between start and end-points, the video platform is towed behind the survey vessel at a speed of approximately 0.7 knots and controlled with the winch to provide a distance of around 1.5 metres above the seabed.

2.2.3 Technical description of visual platforms

The survey performance and the technical specifications used for visual inspection cover all requirements described by the European standard for “Visual seabed surveys using remotely operated and/or towed observation gear for collection of environmental data” (CEN 2012).

The seabed and its epifauna are documented by means of underwater video using the video platforms Campod and Chimaera (since 2014). Campod is a tripod equipped with a high definition colour video camera (Sony HDC-3200) for inspection purposes, and an analog CCD video camera for navigation. It also has four parallel laser pointers (10 cm apart arranged in a square), for scaling of the imagery, and an altimeter to measure the height above the seabed. This video platform was built in 2005 based on the Canadian Campod (Gordon et al. 2000). With the high survey activity of MAREANO, the pressure on the equipment is also high, and production of a new platform was needed after seven years. During planning for a new video platform, experience with Campod was used to make modifications (see Fig. 6).

The new platform, Chimaera (named after the Latin name of the rat tail fish) has two wide runners instead of three legs like Campod and is also equipped with additional cameras for safety and navigation purposes, directed backwards and upwards, monitoring the cable. The main camera on Chimaera is a Sony HDC-P1camera equipped with a Canon HJ17 EX7.6B IASE lens. The rest of the instrumentation is similar to Campod. Chimaera also has two rudders to prevent side-way drifting. Geo-positioning of observations and of the track of the video platforms is provided by a hydroacoustic positioning system (Simrad HIPAP and Eiva Navipac software) with a transponder mounted on the platform. This system provides positions accurate to about 2% of the water depth.

2.2.4 Video resolution and formats, data storage volumes

The image resolution of the video images is 192030080 pixels, and images are sampled at a rate of 25 images per second. Video is recorded in a slightly compressed High-definition format (HD-DVCpro), and stored in Quick time format. The video files are written to file in real time during the survey. For this purpose a 3TB raid is used. The files on the raid are later transferred to external hard discs. One hour of video recording occupies about 60 GB of disk space, and one video transect results in approximately 45 minutes of video footage

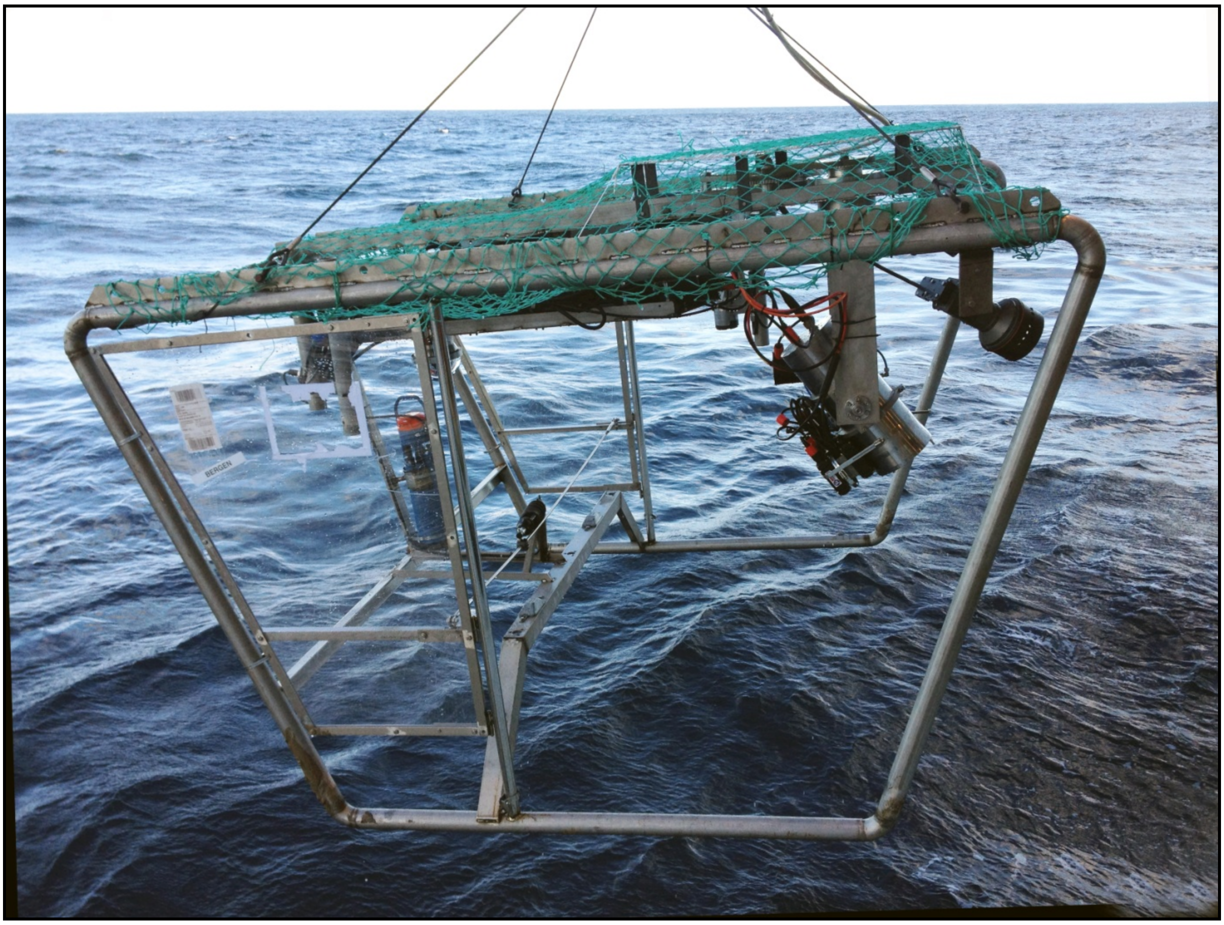


Figure 6. The video platform Chimaera has four video cameras. The main camera has High-Definition standard whereas the three additional cameras have Standard Definition. In addition, the video platform has current meters (short range ADCP) and sensors for temperature, salinity, and optical backscatter.

2.2.5 Field recording and sub-sampling

Real-time registration of observations of the seabed substrates and fauna are made along video lines. Bottom types and organisms are identified and recorded in the field using the event-logging software ‘Campod Logger’ developed at the Institute of Marine Research. The bottom types are classified into one of the following classes: mud, sandy mud, sand, gravelly mud, gravelly sand, sandy gravel, gravel, boulders, bedrock, and coral reef, modified from the Folk grain size scale (Folk, 1954), provided in a list appearing in a drop-down menu in the logging program. More substrate classes have been added to the list of standard bottom types as MAREANO has gained more experience from field observation.

Since registration of all occurrences of organisms is not feasible in the field, the registration is carried out as described below. Navigational data from transponder and HIPAP (Date, UTC time, and positions) and depth are recorded automatically at ten seconds intervals. Each transect is divided into five sequences: two locations (start and end of transect) with detailed inspection while the video platform was parked on the seabed, and three consecutive sequences between 150 and 250 metres long when the video platform is towed between the start and end location. Each observed taxon is recorded only the first time that it occurs within each of the five sequences. The identified bottom type is recorded automatically at the same intervals until a change is observed. Then, the bottom type is manually changed to the new bottom type. Separate logs are recorded for biological and geological observations.

2.2.6 Post-cruise handling and analysis of video data

Video records are analyzed in detail with respect to biology after the cruises using the software VideoNavigator (developed at the Institute of Marine Research). The analysis provides quantitative species data for samples consisting of video sequences of desired length. After initial analyses (ordination of biological and environmental video-data) of video sequences of different lengths (50, 200, and 1000 m), it was concluded that 200 metres was the best compromise between needed spatial detail and time consumption and computing capacity. This conclusion was based on comparisons of similarities between the different ordination plots. The adequacy of this size has been recently confirmed through additional statistical analyses of spatial patterns of megafauna (see later in this report). The size of the areas for the sequences is calculated based on distance travelled (from navigation data) and average field width. The field width is estimated from the ratio between measurements of the distance between two laser points on the video screen, and the width of the screen.

All organisms are identified to the lowest possible taxon and counted, or quantified as a percentage of seabed coverage following the method described by Mortensen and Buhl-Mortensen (2005). Lebensspuren, burrows, and encounters with lost fishing gear and litter are also counted. Abundance data for solitary organisms is standardized to number of individuals observed per 100 m^2. The relative composition of bottom substrate classes (mud, sandy mud, sand, pebble, cobble, boulder, bedrock, consolidated mud, coral rubble, dead *Lophelia*, and live *Lophelia*) is estimated based on the average of estimates of at least three still images within the video sequence. The percentage cover of these classes is estimated subjectively at a scale of 5% intervals.

Analysis of video data for the purposes of geological mapping is mainly limited to quality control of the field logs and examination of features of special scientific interest. Video data and extracted still images are analysed by the geologist as required along each video line and the logged bottom type checked and aligned with the SOSI sediment grain size classes used in mapping (Bøe et al. 2010) in a GIS environment. Video observations are related to the multibeam backscatter data, bedforms and geological setting as well as physical samples, where available, in order to make the best possible interpretation of seabed sediment type at the relevant map scale for the area in question.

3. MAP SCALES, SAMPLING EFFORT AND CONFIDENCE

3.1 Introduction - confidence

A map is one way to provide a spatial representation of natural features. Maps depict different features, depending on the target users. In MAREANO, some examples of map products are bathymetry maps, sediment grain-size maps, biotope maps and vulnerable habitat maps. All the data presented on the maps rely upon real data acquired in the field, but the degree of interpretation and modelling used to turn these data into maps differs between products. The bathymetry maps are based on raw bathymetric data which is subsequently cleaned, and compiled into digital terrain models using suitable algorithms. Although there is some manual checking of data no interpretation is involved. The production of sediment grain-size maps involves a considerable degree of expert interpretation, integrating bathymetry terrain features, backscatter, high-resolution seismic data, visual observations and physical sampling. The biotope and vulnerable habitats maps are based on predictive modelling, which integrate expert biological classification of visual data with full coverage environmental predictor variables derived from multibeam data, geological maps, and oceanography data where available.

Maps may be broad scale (typically 1: 1.000.000 – 5.000.000) or fine scale (typically 1:20.000 - 1:100.000). One centimetre on a 1:5.000.000 map equals a distance of 50 kilometres, while one centimetre on a 1:100.000 map equals a distance of 1 kilometre. Fine scale maps will allow a large number of classes to be represented and include detailed delineation of boundaries, while broad scale maps will normally be more generalized with fewer and broader classes.

The confidence of maps can be considered as the probability that a map is correct for any given point - e.g. if video data are recorded from a seabed area depicted as "sand" on the geological map, what is the probability that the images show a seabed dominated by sand? There are several ways of calculating confidence, depending on the objects which are mapped, and the data which are available. In most cases, confidence will be tested using additional data which are not included in the interpretation and/or modelling. The confidence of maps will rely upon a number of conditions. These conditions can mainly be divided into four groups: 1) *Data collection quality*; 2) *Ground truthing quality*; 3) *Map interpretation quality*; and 4) *Map accuracy* (figure7).

In MAREANO, the ***data collection quality*** is related to the quality of bathymetry and backscatter surveys as well as video lines and physical samples. For example, there may be issues related to acquisition settings and/or sea conditions introducing noise or inconsistencies in the bathymetry and/or backscatter data. Insufficient balancing of the backscatter signal along one line, shifts of bathymetry acquisition mode (long versus short pulse), or level differences between surveys all contribute to lower consistency of backscatter data which results in lower overall data quality. This may have a negative effect in interpreted and modelled map products.

Ground truthing quality is related to the quality of the video observations (e.g. sufficient lightning, image resolution, focus, correct distance over seabed etc.) and physical samples. It is also very much dependant on the sampling strategy used for planning stations.

Map interpretation quality is related to the map production process, for example whether a sediment grain size map is produced manually by expert interpretation or by a more supervised, semi-automated modelling approach, or even fully-automated classification. Manual interpretation will introduce interpreter-bias (be subjective) and is very difficult to reproduce while semi- and fully-automated approaches may yield complex polygons and may be confused by systematic artefacts in data collection.

Finally, the ***map accuracy*** informs us about the accuracy of the final product, which is influenced by map resolution and scale, and number of mapped classes. Normally, a manually interpreted map does not leave any data for accuracy assessment. It is however possible to set aside some stations for testing. An interpretation based on modelling opens up for a much wider range of accuracy assessment tools. Such tools include cross-validation techniques which allow for successive partioning of the data into training and testing data and applying multiple combinations of these data to model and evaluate performance. However, it is important to keep in mind that a greater availability of accuracy assessment tools does not automatically mean that maps based on modelling are more accurate than maps based on manual interpretation.

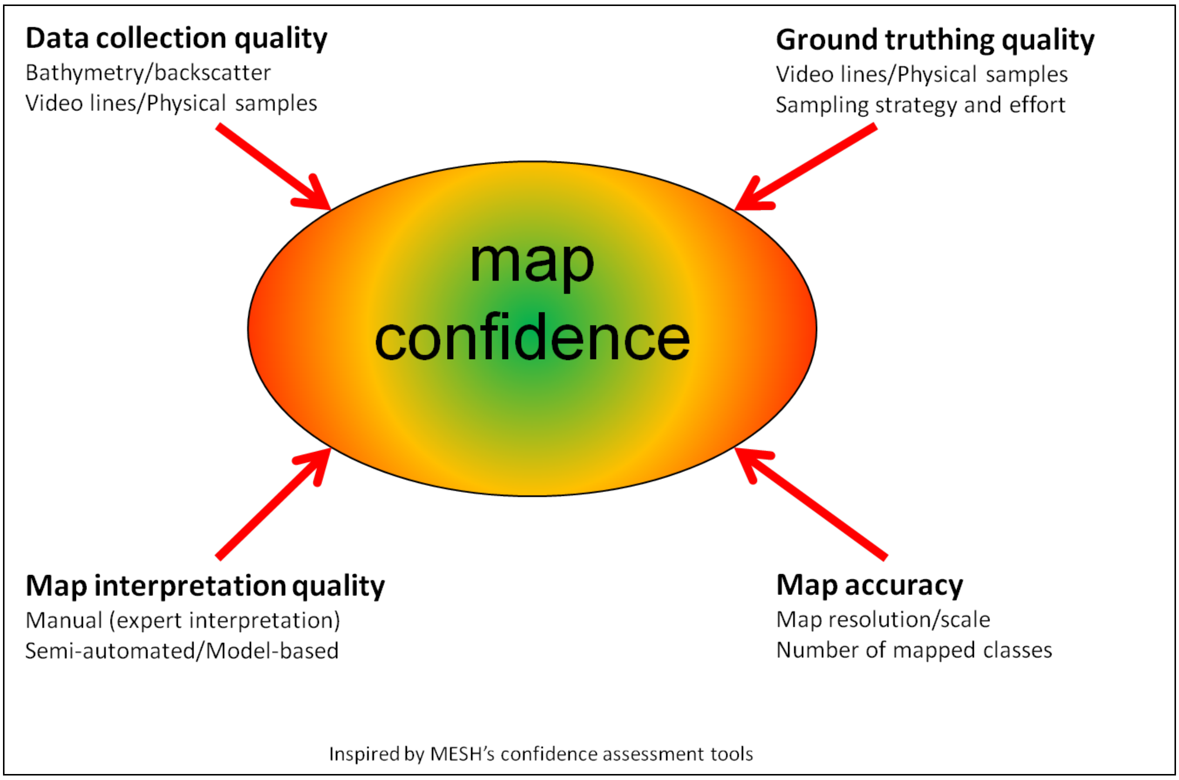


Figure 7. The four main groups of conditions that drives the confidence in a map.

Whichever method is used, the map accuracy depends on the quality of data collection, ground truthing, and map interpretation. Therefore, a complete analysis of confidence should include an evaluation at each of the four groups, not only in the accuracy assessment step. The MESH project provides an example of how this can be implemented (http://www.emodnet-seabedhabitats.eu/default.aspx?page=1693). In this report, we focus on *map accuracy assessment* as the main proxy for map confidence.

Proper sampling strategies and survey design are a very important basis for creating good map products and models. A good sampling strategy should ensure that the environmental variability in the area is covered in a satisfactory manner. The environmental variability is illustrated in figure 8, showing two areas of similar size (2000 square kilometres). The first area is west of Senja, forming part of the Specially Valuable Area (SVO) defined in the management plan for the Barents Sea and the areas outside Lofoten. This area is complex, with a great deal of environmental variability. It covers water depths from less than 100 metres to more than 2500 metres, and includes fishing banks, the Andfjorden glacial trough, the Bleiksdjupet Canyon, and a steep continental slope where many underwater landslides have occurred.

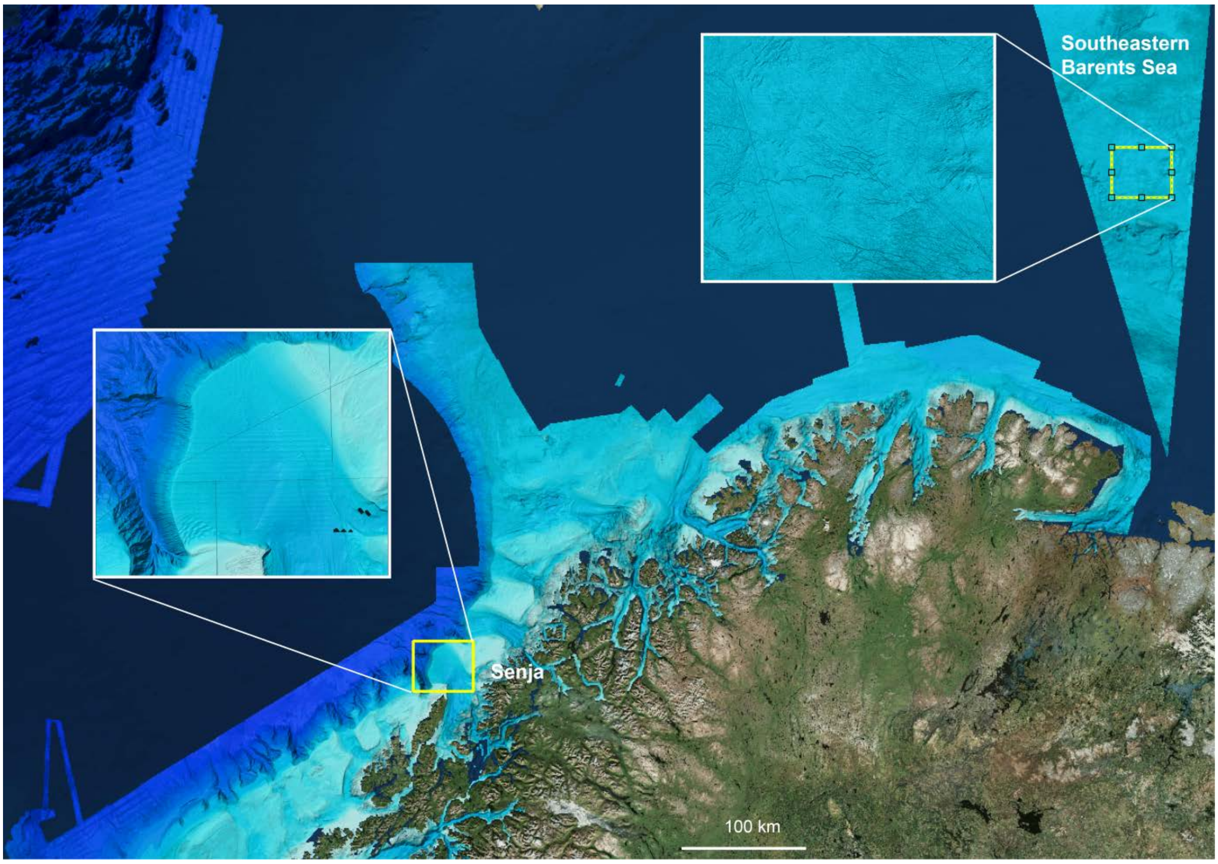


Figure 8. Two areas (2000 square kilometres) with different environmental variability. The area west of Senja (left) is a complex area with high variability, while the area in the southeastern Barents Sea is a rather homogeneous area with low environmental variability.

Large oceanographic gradients are present here, with both Atlantic water and coastal water interacting dynamically. Based on this high environmental complexity, we can expect high geodiversity and biodiversity in this area.

The other area is located within the southeastern parts of the Norwegian Barents Sea. This area is rather homogeneous, with low environmental variability, at least on broader scales. The water depths are much more constant – ranging between 250 and 300 metres, and the seabed is rather even, despite a myriad of small-scale geomorphic features such as moraines, iceberg ploughmarks and pockmarks. Only Atlantic Water is present here, and broad scale oceanographic data show far less variation than in the area west of Senja. The expected broad scale geodiversity and biodiversity are therefore less here. Nevertheless, at finer scales, we find evidence of complex geological structures and processes some of which are relevant to 1:100.000 mapping, others which are on even finer scales. If these two example areas are mapped with the same sampling effort (i.e. sampling stations per area unit), we would expect that the map confidence (e.g. in a 1:100.000 sediment grain-size map) will be lower in the complex area west of Senja, because we can expect more classes, with each class covering smaller areas (patches). Or we can consider the differences in environmental complexity the other way around – i.e. we need to have a higher sampling effort in the complex area than in the rather homogeneous area, in order to achieve the same level of confidence. As part of a related study we have examined this physical environmental variability in more detail, and proposed a method for how this can be quantified with suggested use in guiding sampling effort (figure 9). This is described in a separate report (van Son et al, 2015).

The map confidence will also be influenced by the method chosen for locating the sampling stations. For example – positioning the stations evenly will give a different result from locating the stations in a tight cluster in a small part of the survey area. Slightly different strategies have been used for the MAREANO sampling cruises over time, and this is partly reflected in actual spatial distribution of the sampling stations (figure 10). Different methods for positioning sampling stations have been studied, including Generalised Random-Tessellation Stratified sampling (GRTS) (Stevens and Olsen, 2004) which takes into account important factors such as randomness, independence (spatial balance), and environmental variability. Different sampling strategies and their pros and cons are discussed more in detail in sections 4.2 and 4.3.

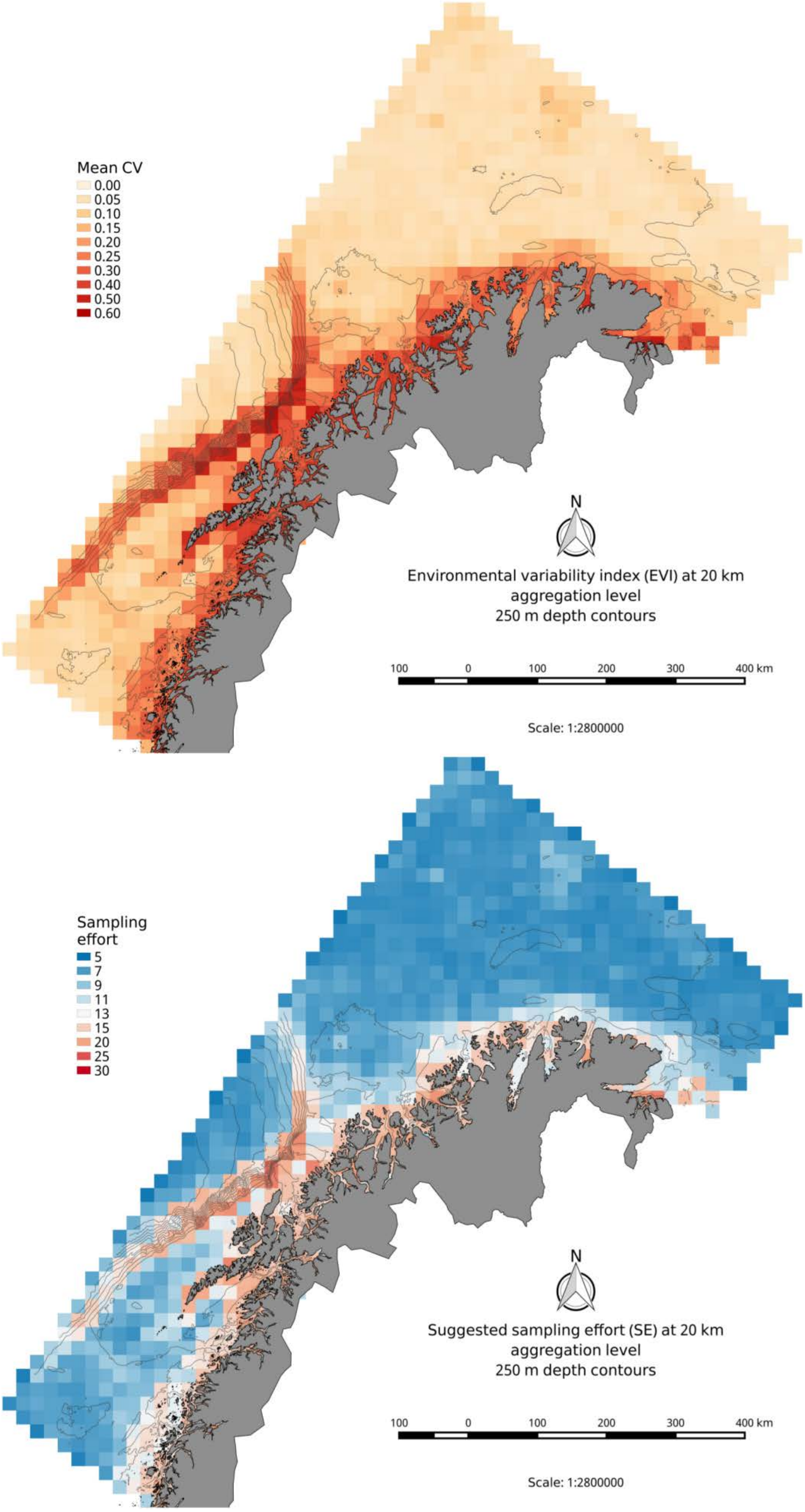


Figure 9. Environmental variability index (EVI, Top) and the corresponding suggested sampling effort (Bottom) when the environmental variability is aggregated and evaluated within an area of 20 km.

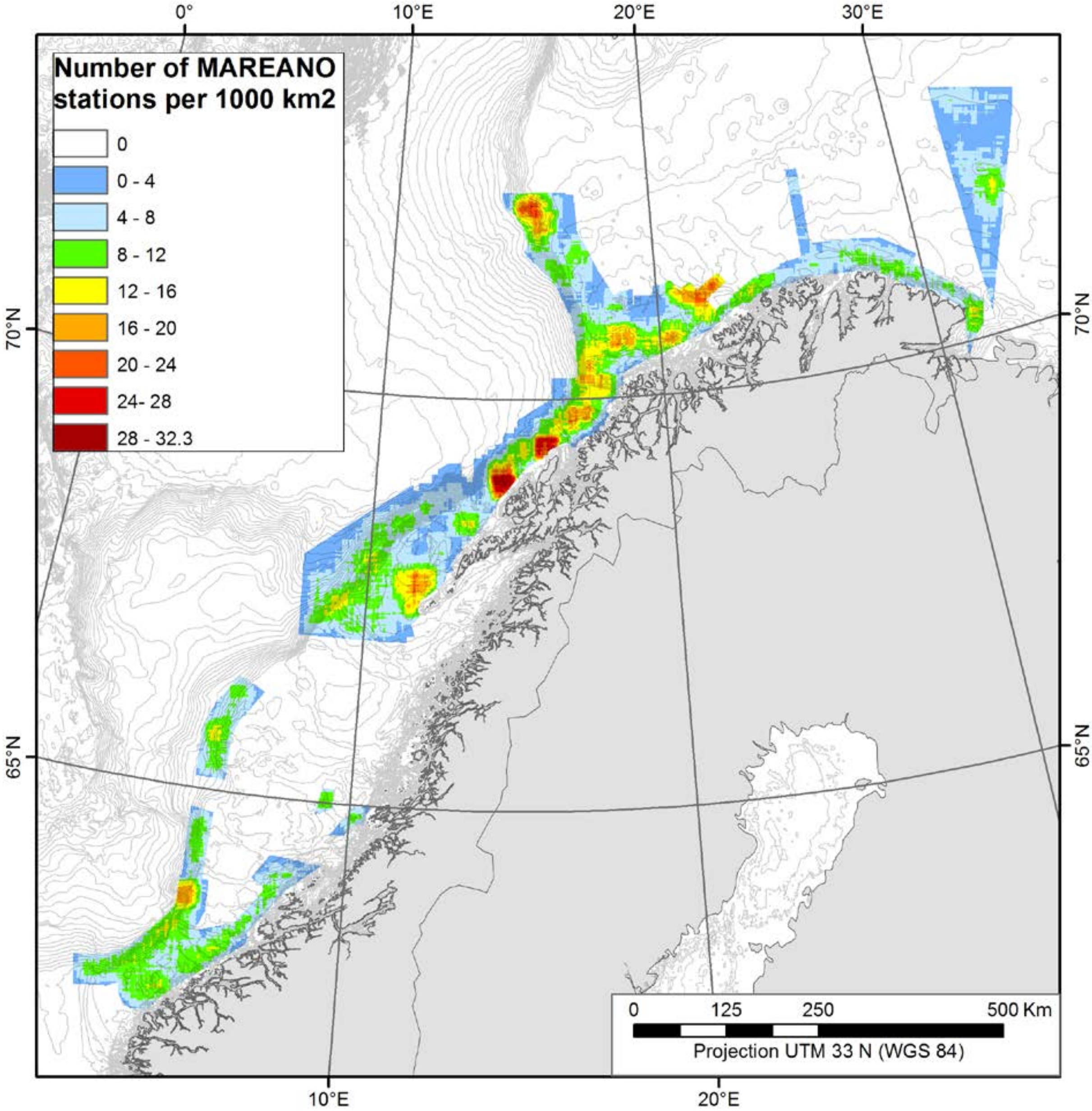


Figure 10. Actual sampling effort in the areas sampled by MAREANO in 2006-2014.

3.2 Sediment grain size maps

Sediment grain size maps produced by MAREANO are based on multibeam bathymetry and backscatter, using visual and physical sampling for ground truthing. Other environmental data such as oceanography are used if available.

3.2.1 Map interpretation based on different sampling efforts

ArcGIS from ESRI (www.esri.com) is the GIS interface used for interpretation and compilation of geological maps. The expert-based interpretation is based on integration of bathymetry (terrain indices, shaded relief imagery, backscatter), high resolution seismic data, observations from visual platforms and physical samples. Bathymetric data are received from NHS in formats compatible with Fledermaus. The bathymetry is used as a stand-alone product, or derived products are generated. Terrain indices maps are used to quantify terrain variation in an objective manner. Shaded relief maps are generated in order to highlight structures and features which provide information regarding the sediment type or processes (active or past). Backscatter data are processed using Fledermaus GeoCoder. The backscatter data are classified according to reflectivity, together with expert judgement of texture. Subbottom data, usually TOPAS PS 18 data from G.O. Sars are used to constrain boundaries between different sediment types. Visual information - normally from ten 700 metre long video lines pr. 1000 km^2 together with visual and granulometric data from physical samples (grab, box corer or multicorer) are used for ground truthing. Existing data from external sources, such as universities or site investigations performed by oil companies are integrated if available and of sufficient quality.

Digitizing of geological boundaries and features is done at different scales, depending on the resolution of the bathymetry, and the density of ground truthing data. The most detailed maps are produced when full coverage multibeam echosunder bathymetry data, backscatter, and around 10 video lines pr. 1000 square kilometres are available. In this case, the final maps are presented at a scale of 1:100.000. The length of the smallest object digitised is at least 50 metres, but the width may be less, as in the case of e.g. iceberg ploughmarks. Mounds with bioclastic sediments are mapped even if they are smaller than this - down to dimensions like 40 x 40 metres (for more information - see Bellec et al. 2014).

Less detailed digitisation is done in areas where only regional bathymetry (grid size typically 50 metres) without backscatter, combined with some multibeam data, and/or reduced density of ground truthing data is available. In this case, the digitising is done in order for the final maps to be presented at a scale of 1:250.000. The length of the smallest object digitised is at least 125 metres, but the width may be less. This has been the case for the maps produced at Mørebankene and in the southeastern Barents Sea. Interpreted geological data are stored in a database made accessible for users through several map services (Bøe et al. 2010), and thematic maps are published on the MAREANO web site (www.mareano.no).

Normally all stations (on average 10 per 1000 square kilometres) are used for map integration. This impairs testing of confidence, because additional sampling stations (i.e., not forming part of the interpretation) are needed to estimate confidence. In order that some testing could be conducted for the purposes of this study, we selected an area at Tromsøflaket (figure 11), which was sampled in 2006 at a higher density (exceeding 20 stations per 1000 square kilometres) for use in re-interpretation and confidence testing. The chosen area and distribution of sampling stations is shown in figure 11. Note that the video lines recorded here are longer than today's standard, with an average length of 1000 metres.

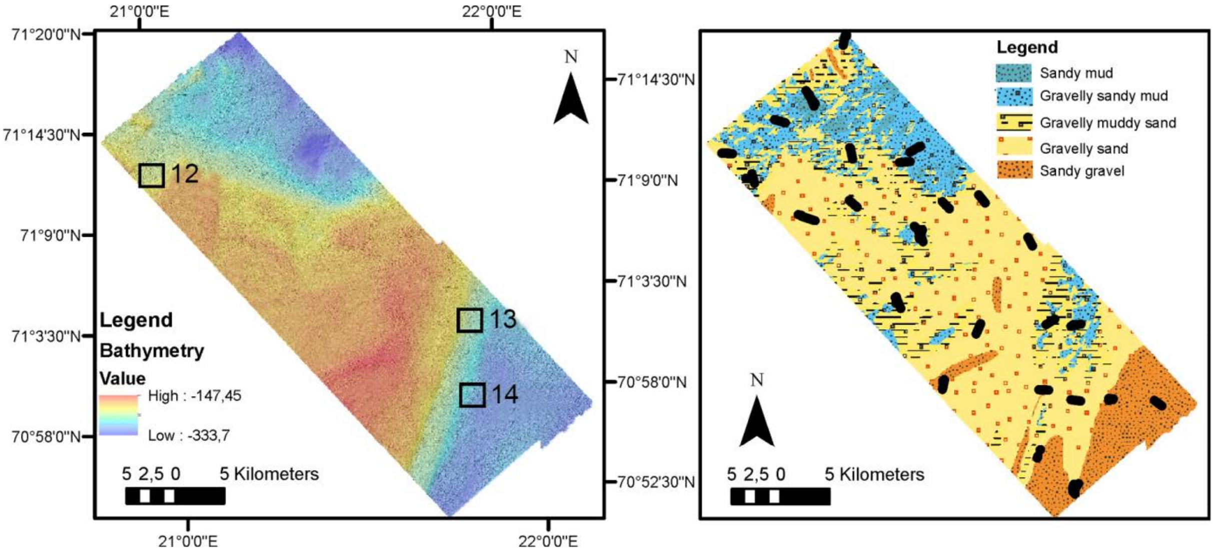


Figure 11. Bathymetric map for the area used for testing at Tromsøflaket , showing areas in figures 12,13 and 14 (left) and the original interpretation from 2007 (right), with all video stations available for interpretation superimposed.

The plan was to make interpreted maps using 5, 10 and 15 stations, respectively, and set aside a sufficient number of stations in order to allow for testing of confidence, following these steps:

1. Pseudo cruise planning – select stations for each sampling density to be tested from existing stations.
2. Pseudo cruise with video logging (geologist views all video data and logs with Campod logger)
3. Video lines were standardised to 700 m for comparability with present video lengths. Package up all available data in 5, 10, 15 samples worth packages for geologist to interpret.
4. Geologist interprets sediments
5. Testing and evaluation of maps.

This plan was challenged by several factors, described below:

- It was not possible to link positions and the actual video for all lines. Some video lines had poor quality or other technical problems meaning that 700 metres worth of video data was not available. This reduced the maximum number of available stations for interpretation (excluding those set aside for confidence testing) to 13 stations. It was therefore decided to make interpreted maps using 5 and 13 stations per 1000 square kilometres.
- The backscatter data contains a relatively high level of nadir noise (higher values recorded at the centre of the beam) and other artefacts that persist even after processing (figure 12). Note that these backscatter data were acquired using a now outdated multibeam echosounder (EM1002) and that data acquisition and processing procedures have since been improved.
- The seabed is heavily indented by iceberg ploughmarks, which pose a strong control on the distribution of seabed sediments. This results in sediment patches which are smaller than the lower limit for objects to be digitised (smallest mappable object), and consequently that a sediment class which is shown on the map may contain smaller patches of a different sediment class (figure 13)

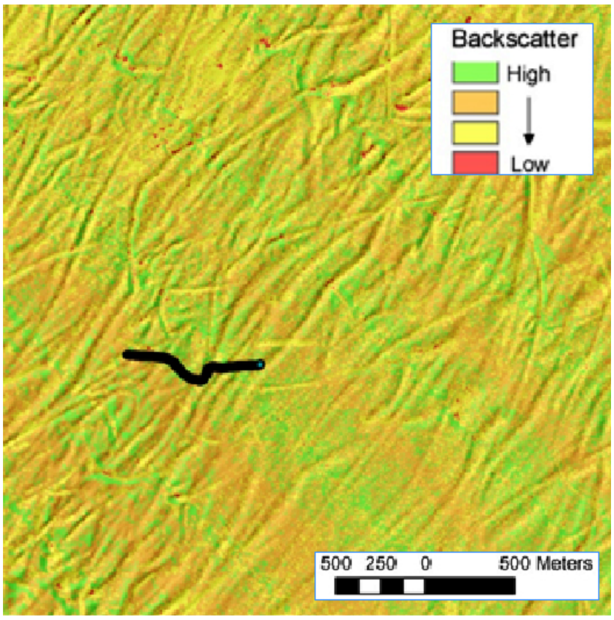


Figure 12. Backscatter showing influence of nadir noise over a shaded relief image, giving challenges for sediment interpretation. Black line shows video line.

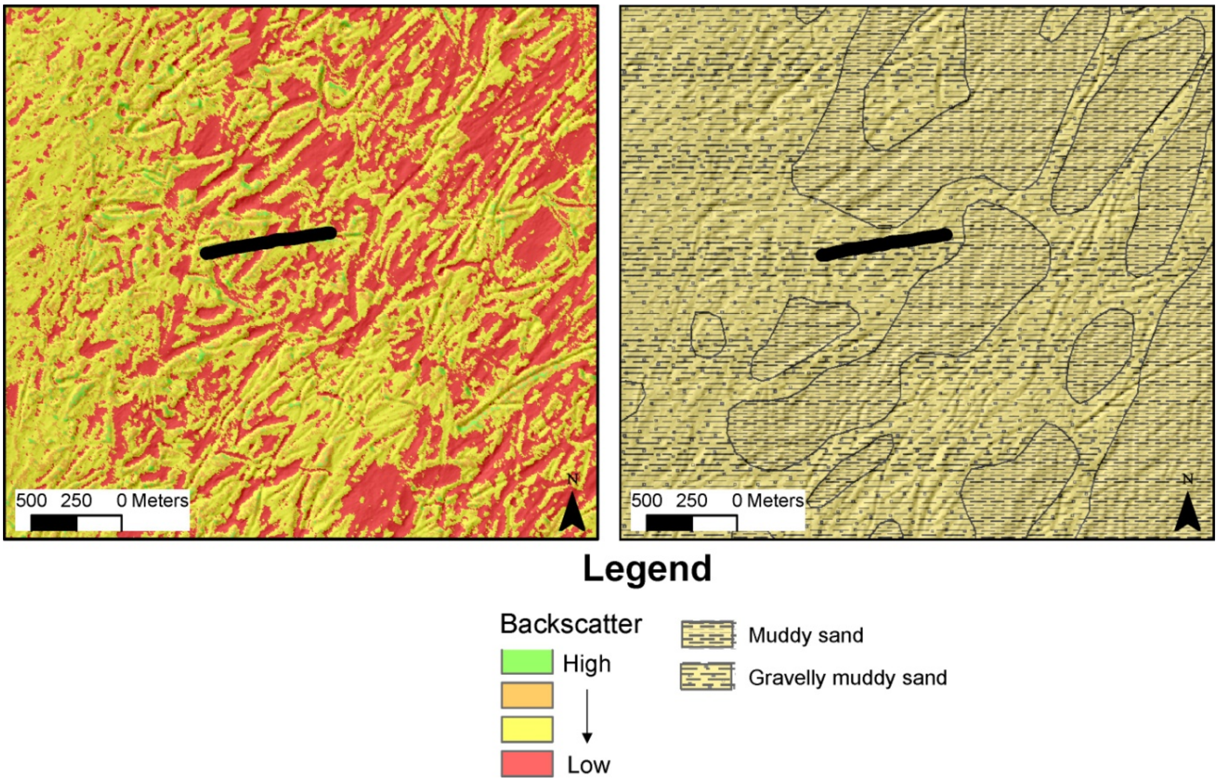


Figure 13. Backscatter over shaded relief image (left) and interpreted sediment map (right), with video line (black). The size of the sediment patches is smaller than the smallest mappable object. This gives challenges for both classification and confidence testing.

3.2.2 Confidence testing

The confidence of the maps based on 5 and 13 stations, respectively, were tested by performing a map accuracy assessment. Five video lines were set aside for accuracy assessment. Each video line was divided into three segments (approximately 235 m long). The length of the segments was chosen to match with the smallest mappable objects (250m). For each segment, in each video line, the dominant sediment grain size class was estimated. This dominant class was then positioned at the midpoint in the corresponding segment and then tested against the interpreted class. The analysis yielded an overall accuracy of 53% for 5 stations per 1000 square kilometres (Table 1) and an overall accuracy of 73% for 13 stations per 1000 square kilometres (Table 2) The test data set only contained two of the four sediment classes and as such did not provide a completely satisfactory test of the accuracy of the interpreted sediment maps.

Table 1. Error matrix for the accuracy assessment of the interpretation of Tromsøflaket based on a sampling effort of 5 stations per 1000 square kilometres. The overall accuracy is 53%, since 8 (numbers in bold) of 15 test points were correctly classified.

	Predicted class (interpreted)				
Observed class (test)	80	120	200	215	Row sums
80	**3**	0	0	0	*3*
120	0	**5**	7	0	*12*
200	0	0	**0**	0	*0*
215	0	0	0	**0**	*0*
Col sums	*3*	*5*	*7*	*0*	***15***

It is also important to note that we tested the midpoint in a line segment against the corresponding class of the interpreted map. In certain cases, this may create a mismatch between the dominating sediment class in the segment and the assigned class in the interpreted map. This is particularly prone to happen in areas with long, elongated interpreted polygons.

Table 2. Error matrix for the accuracy assessment of the interpretation of Tromsøflaket based on a sampling effort of 13 stations per 1000 square kilometres. The overall accuracy is 73%, since 11 (numbers in bold) of 15 test points were correctly classified.

	Predicted class (interpreted)				
Observed class (test)	80	120	185	200	Row sums
80	**3**	0	0	0	*3*
120	1	**8**	3	0	*12*
185	0	0	**0**	0	*0*
200	0	0	0	**0**	*0*
Col sums	*4*	*8*	*3*	*0*	***15***

3.2.3 Challenges for confidence testing

We tested the interpretation in a line segment against the value of a midpoint on the interpreted map (figure 14). If the midpoint of the segment (highlighted in blue in figure 14) falls exactly on a minority within the segment (in this case segment C), this can give result that is not representative of the rest of the dominant class.

The aim of this study was not to compare the newly interpreted results with those for the published sediment map from this area. However an important factor to note, and which gives rise to the dissimilarities observed in figure 15 is that the sediment classification scheme has developed considerably since 2007. For example in 2007, gravelly sand and gravelly muddy sand were interpreted together in the same class (gravelly sand) with additional text added in the “legend” (gravelly sand with a variable content of mud). The number of classes was limited compared to the present classification scheme. Based on the experience from several years of video observation, sediment classification and map production, new routines and classes have been implemented. The differences in the finer fractions are very small, from gravelly mud and gravelly sandy mud to muddy sand. The intermediate class used is gravelly sand in one and gravelly muddy sand in the other map,

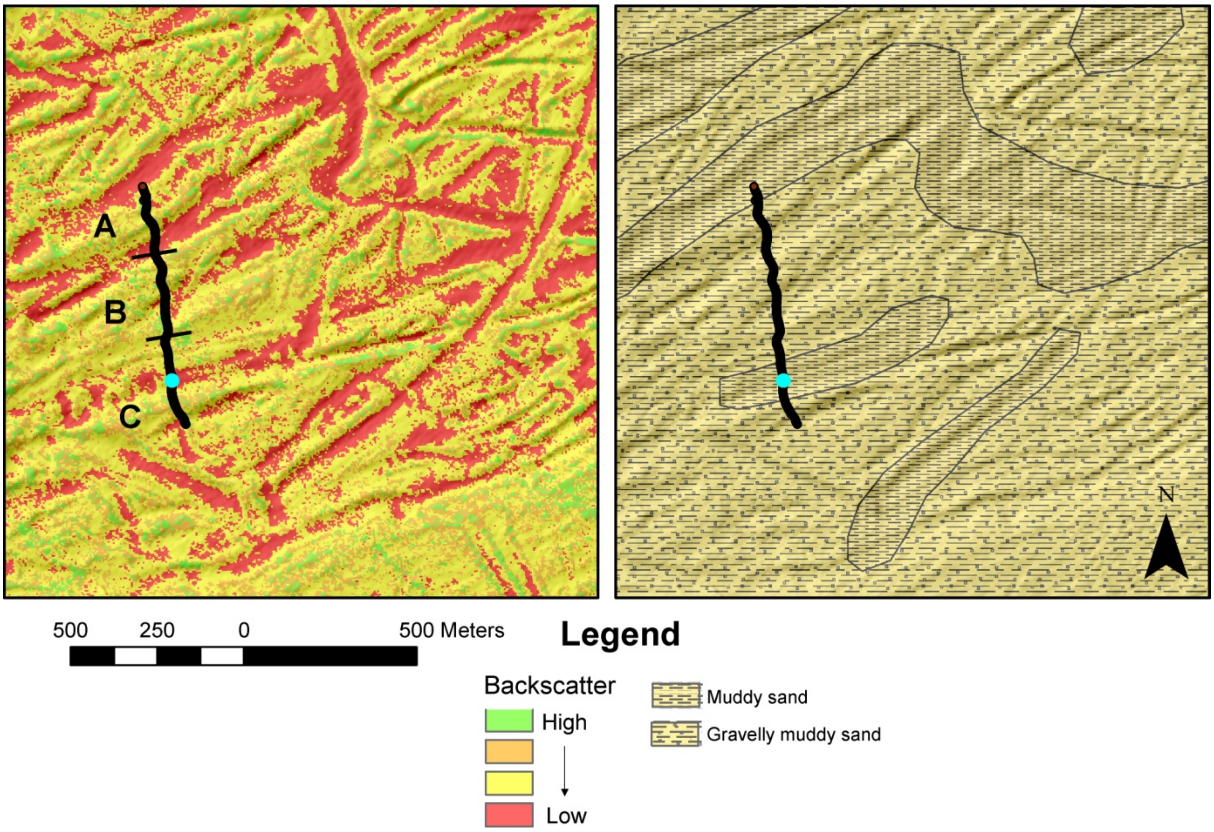


Figure 14. Backscatter over shaded relief image (left) and interpreted sediment map (right), with video line (black).

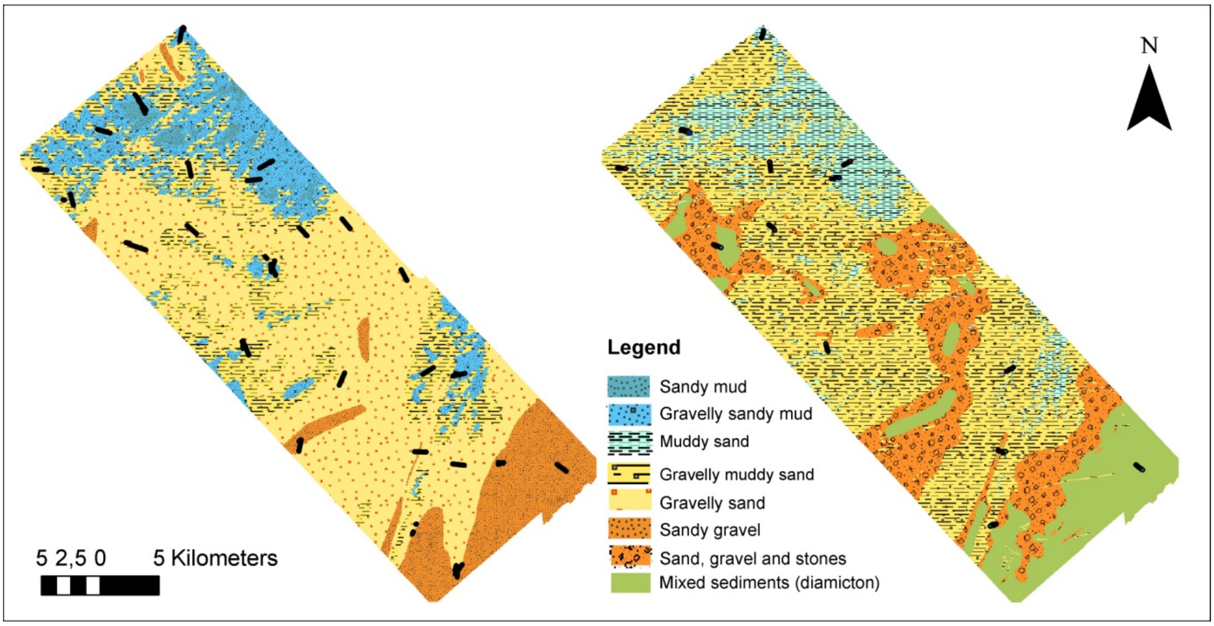


Figure 15. Dissimilarities between the original map interpretation from 2007 (left) and the new classification from 2015 (right).

which is also comparable. The difference is however large if we look at the coarser classes, because these are the classes affected by the new map production routines and classification.

The interpretation using 5 and 13 stations gave somewhat different maps (figure 16). This is particularly significant for the coarse-grained sediments, while the fine-grained sediments have a pretty similar distribution. This is because it was found that 5 stations do not include enough documentation of all sediment types, too little to base a comprehensive classification on. The five lines covered the finer classes found in the study area, but were insufficient to characterise the coarser classes. More lines were necessary to fully characterise all classes present in the study area, providing sufficient visual material for determining boundaries between classes and ground truthing backscatter data.

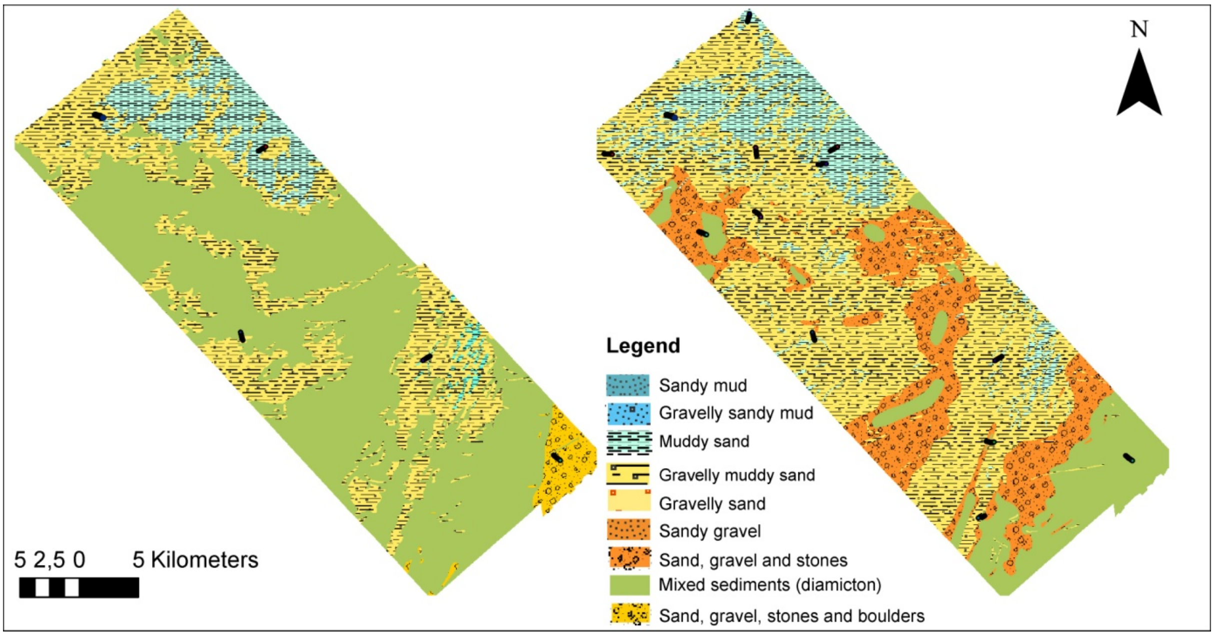


Figure 16. New interpretation, based on 5 stations (left) and 13 stations (right) pr. 1000 square kilometres, respectively. Note that the interpreted distribution of the fine-grained sediments is basically similar, while the interpreted distribution of coarse-grained sediments (yellow, orange and green) differ significantly.

From a mapping point of view, 5 stations were too little to adequately base a classification on for an area of this size and complexity. This is in agreement with EVI analyses (van Son et al. 2015) indicating that the sampling effort should be between 9 and 14 stations pr. 1000 square kilometres for this area. With only 5 stations for reference, not all the backscatter intensity classes were ground truthed and it was hard to know which classes should be used, and how much they occurred in the area. Increasing to 13 stations gave a much better representation of the area, the classes present and their distribution. Adding more stations (from 5 to 13 stations pr. 1000 km^2) for ground truthing resulted in a change of classes for the coarse-grained sediments (figure 16). The classification based on 13 stations pr. 1000 square kilometres was considerably more time consuming, but gave a map representation which is significantly better than the map based on 5 stations. This observation supports the EVI analysis which suggests that, because of the complexity, between 9 and 14 stations pr. 10000 km^2 are needed for ground truthing in this area.

3.2.4 Discussion about future developments

Manual interpretation of sediment grain size maps is time-consuming and MAREANO may benefit from adopting alternative and potentially more efficient ways of producing this map product. Methods, such as object-based image analysis (OBIA), which has been widely used in satellite image classification is now being implemented more and more in classification of sediment grain size (e.g. Lucieer and Lamarche, 2011). The National Oceanography Centre in the UK has recently released a first version of a plug-in to ArcMap 10.x (Le Bas et al. 2015). This plug-in allows for segmenting backscatter and other variables related to sediment grain size into objects using k-means on variables scaled between 0 and 255 (download at www.codemap.eu/outputs).

Also, linking backscatter strength to sediment classes is challenging, particularly when few stations are available for ground truthing. One way to at least partially overcome this may be to take into account more features, such as backscatter image texture and properties based on the backscatter signal such as those obtained by ARA (Angular Range Analysis, see e.g. Fonseca et al. 2009), rather than just the intensity (dB level). It may also be beneficial to implement semi-automatic classification to reduce the manual input required and increase objectivity and standardisation as well as offering more opportunities for confidence testing. Another approach is to make spatial predictions based on sediment compositional data based on physical samples rather than class. This can be achieved using machine learning methods (Stephens and Diesing 2014). MAREANO does not currently collect detailed sediment compositional data at a spatial density necessary to achieve this.

3.3 Biotope maps

To produce full coverage maps of the distribution of biotopes, as required by management, predictive modelling techniques are used. These models use information on the characteristics and distribution of biological communities (based on visual documentation at MAREANO stations) and combine it with physical characteristics of the seabed identified by terrain analysis and geological interpretation. Since 2014, when oceanographic data first became available to MAREANO, these data have also been incorporated in biotope modelling.

MAREANO has refined the methods used for spatial prediction of biotopes over the years since the first map for Tromsøflaket was produced in 2008, testing various methods for classification and modelling as well as improving methods for map validation. Currently prediction of biotope distribution is performed using maximum entropy distribution modelling (Maxent: Phillips et al. 2004). The workflow for biotope modelling is summarized in figure 17 (see also Mortensen et al. 2009, Dolan et al. 2009, Elvenes et al. 2014)

3.3.1 Biotope mapping data and workflow

MAREANO biotope mapping utilizes all video data acquired by MAREANO with station density as shown in figure 10, as the basis for identification of faunal groups indicating biotope classes. The taxonomic nomenclature is revised on a regular basis in order to provide comparable data sets from different cruises. Results from video lines with unreliable navigational data are not included in analyses of the relationship between faunal composition and spatially explicit predictors. Data from physical sampling are not directly used in the production of biotope mapping.

Following detailed video analysis and documentation of fauna the first step in biotope mapping is the identification of biotope classes. Detrended correspondence analysis (DCA) (Hill, 1973) using the software PC-Ord (McCune and Mefford, 2006) is the ordination method currently used to identify groups of samples (video sequences of 100-200 metres length) with high similarity of fauna composition. DCA is an unconstrained ordination that is suitable for conducting indirect gradient analysis. DCA results facilitate the identification of groups of samples with similar species composition. The correlation of the environmental variables in relation to the various axes ordination result can also be analysed for identification of complex gradients. By plotting samples on the first 2 (or 3) DCA axes and then clustering of these points through expert judgement (2008-2012) or statistical algorithms (k-means since 2012) the samples are split into biotope classes with similar faunal composition (figure 18).

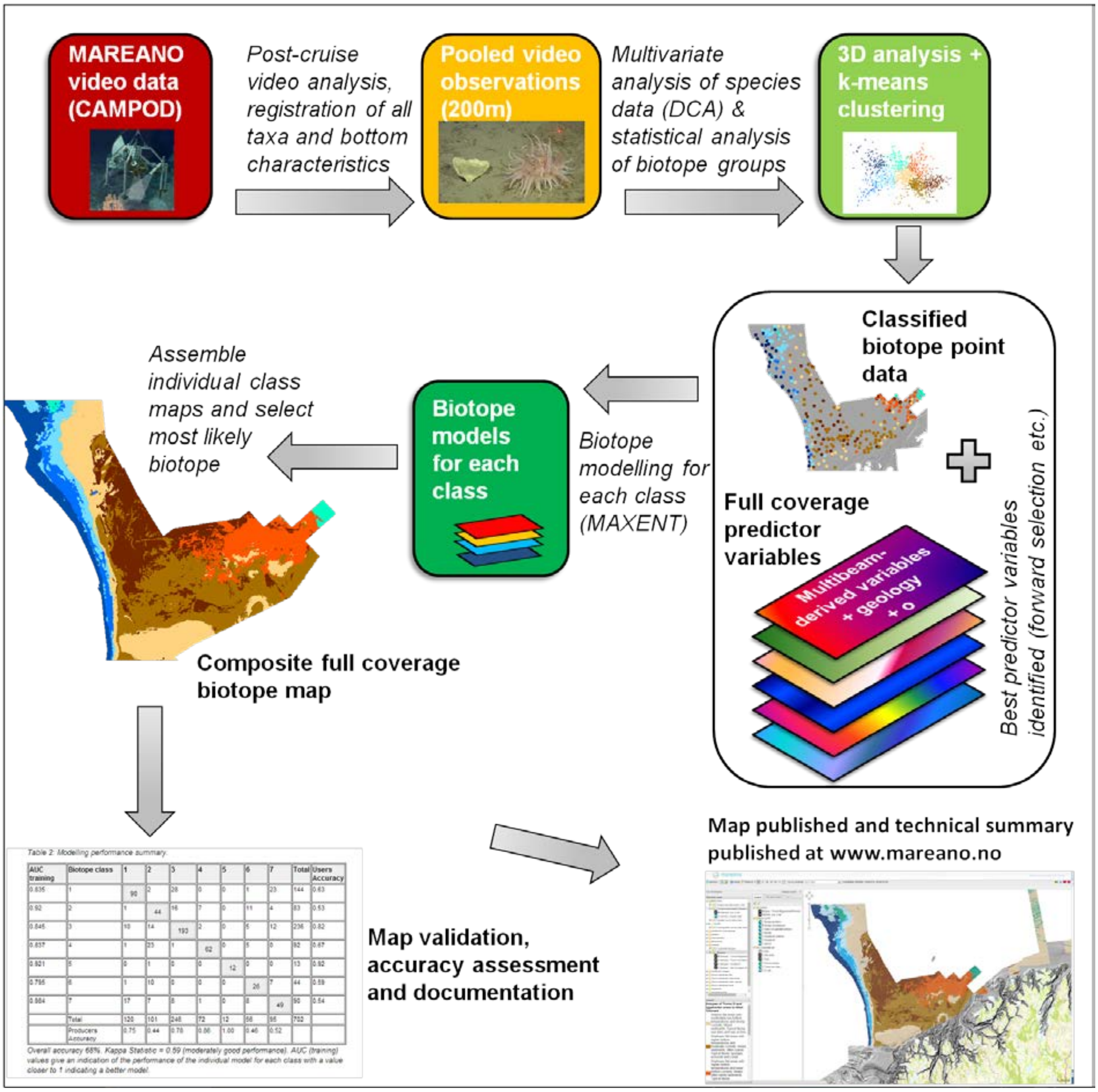


Table 2: Modelling performance summary.

AUC training	Biotope class	1	2	3	4	5	6	7	Total	Users Accuracy
0.835	1	90	2	28	0	0	1	23	144	0.63
0.92	2	1	44	16	7	0	11	4	83	0.53
0.845	3	10	14	193	2	0	5	12	236	0.82
0.857	4	1	23	1	62	0	5	0	92	0.67
0.921	5	0	1	0	0	12	0	0	13	0.92
0.795	6	1	10	0	0	0	26	7	44	0.59
0.864	7	17	7	8	1	0	8	49	90	0.54
	Total	120	101	246	72	12	56	95	752	
	Producers Accuracy	0.75	0.44	0.78	0.86	1.00	0.46	0.52		

Overall accuracy 68%. Kappa Statistic = 0.59 (moderately good performance). AUC (training) values give an indication of the performance of the individual model for each class with a value closer to 1 indicating a better model.

Figure 17: Summary of MAREANO biotope mapping workflow.

By re-plotting these samples in geographic space using GIS we are able to see the geographic distribution of the samples now classified by biotope, and examine how the spatial distribution of biotopes varies in relation to environmental variables (figure 19). Moreover, this step allows extraction of physical characteristics for each sample (biotope) point from the full coverage quantitative terrain variables (e.g. slope, curvature etc.), as well as sediment grain-size class, landscape type and oceanographic variables (temperature, salinity, bottom currents). Together these provide a large number of physical seabed

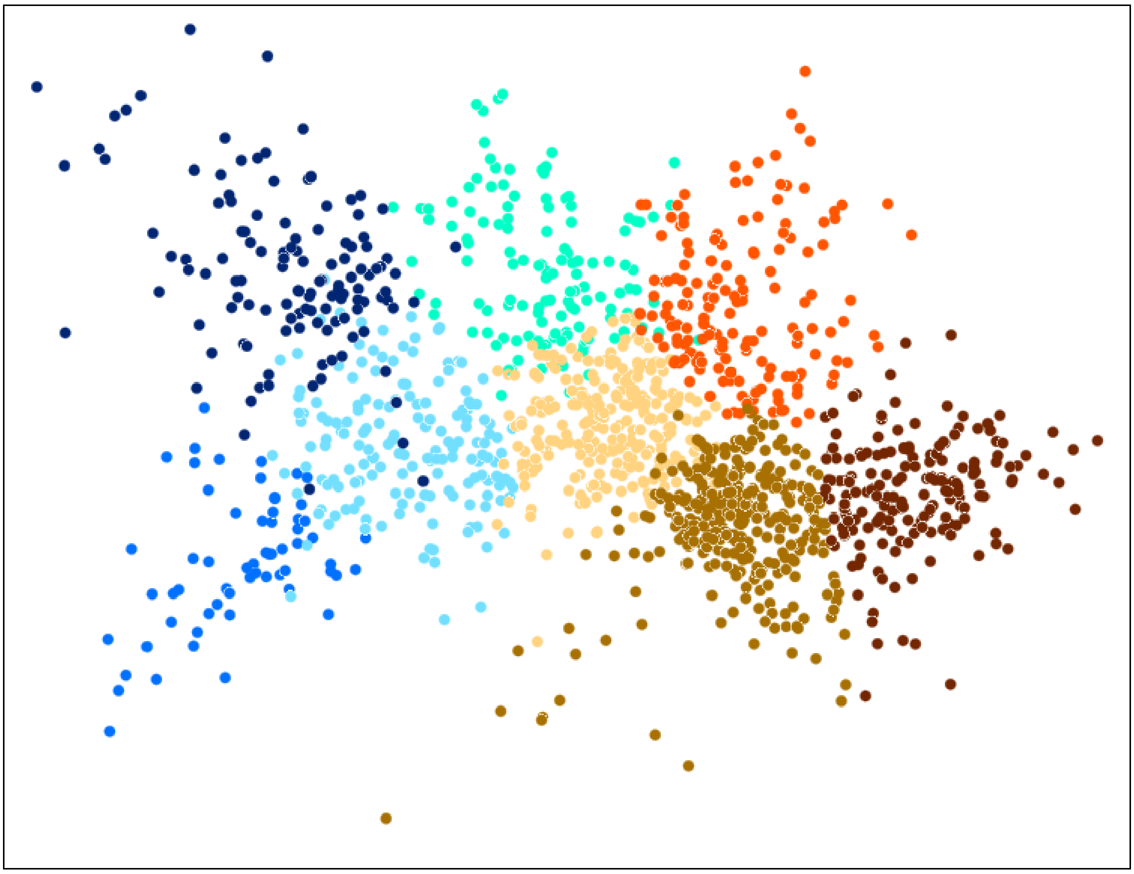


Figure 18: Example of samples plotted in multivariate space on DCA axes 1, 2 and 3. 3D visualization from ArcScene - points are rotated to show best separation of the 8 biotope classes determined by k-means classification, each of which is shown with a unique colour (see also figures 19 and 22).

descriptors which can potentially serve as predictor variables which allow the model to predict from point observations to a full coverage map.

To avoid using predictors that are strongly collinear, or that are proxies for the same structuring processes, and to avoid overfitting of the models, we use forward selection with Monte Carlo permutation using CANOCO for Windows 4.52 (ter Braak and Smilauer, 2002). This is done to select the most suitable (continuous) predictor variables from all those available, as well as examining correlations graphically. Internal routines in Maxent are used to assess the importance of categorical variables (sediment, landscape) and Maxent and Random Forest Models (Breiman, 2001) are also used as a check on the ranking of variable importance suggested by forward selection.

Although biotopes are defined first on the basis of their faunal composition, each biotope is characterized by different substrate, depths, terrain characteristics and oceanographic

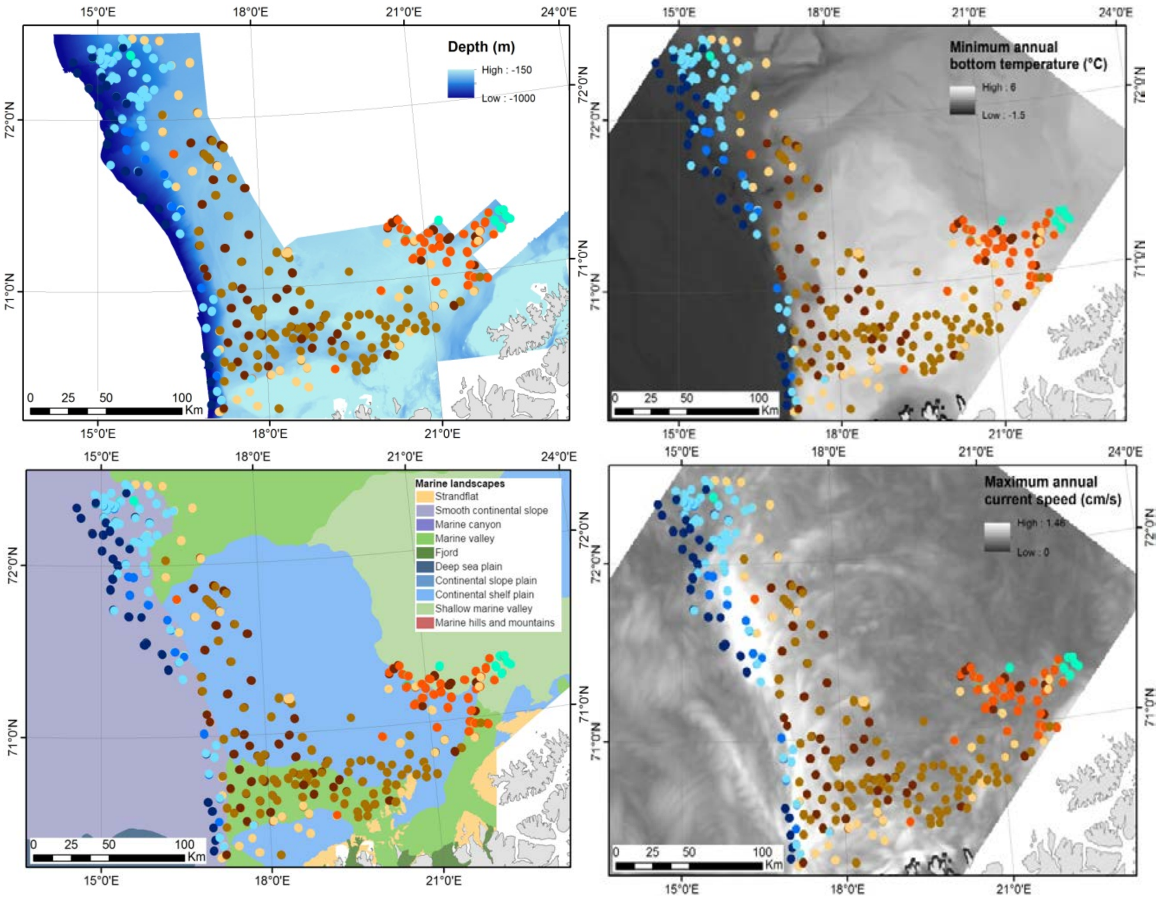


Figure 19: Examples of how the distribution of biotope classified samples (model response variable) varies across the Troms III-Eggakanten-Tromsøflaket mapping area in relation to potential model predictor variables reflecting different properties of the physical environment (a - upper left) bathymetry, (b - upper right) minimum annual bottom temperature, (c - lower left) marine landscape, (d - lower right) Maximum annual current speed. Sample points are coloured by biotope class (see figure 22) – note not all points are visible at this map scale.

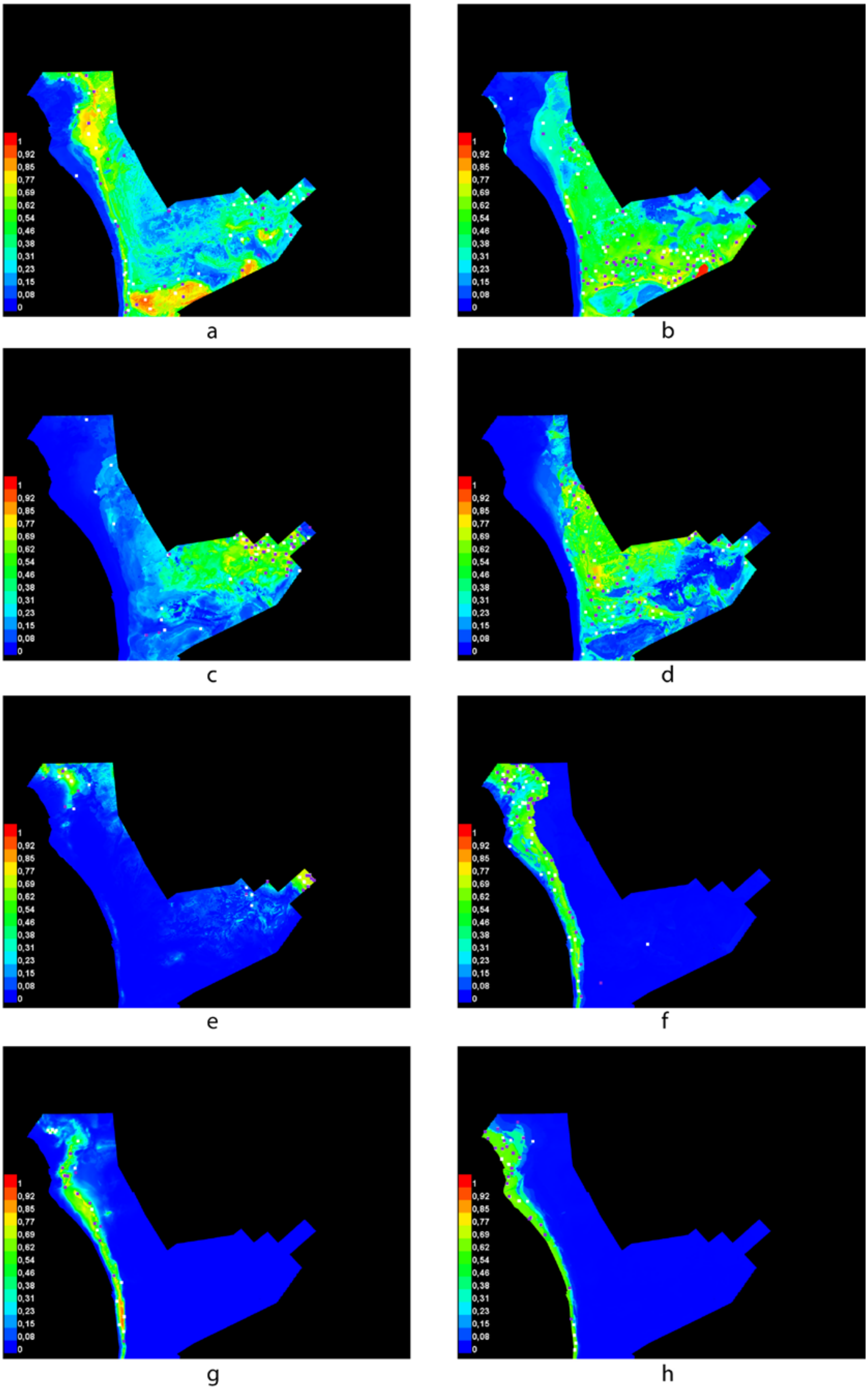
conditions. These characteristics, based on the most important predictor variables and typical taxa, are summarised in the map legend and a more detailed description given in the technical summary document that is published on www.mareano.no as supporting information to each biotope map.

Modelling and prediction is conducted using the software program Maxent (Phillips et al., 2004), which implements the maximum entropy principle to predict biotope distribution based on presence-only point data and full-coverage environmental predictor variables. The Maxent method is one that performs well in comparison with other modelling approaches and which has gained widespread use in terrestrial and increasingly in marine habitat modelling applications (see Elvenes et al. 2014 for further background and examples).

Modelling and prediction typically involves production of several alternative models and selecting the best based on model performance and expert validation. Although Maxent remains the primary modelling tool at the time of writing since 2013 parallel modelling using Random Forests has been conducted as a cross check on the Maxent-based map result (and as part of an evalution and development of alternative methods). Maxent modelling requires that one model and prediction is generated for each biotope class (figure 20). Each map is a probability distribution map (0 to 1) indicating the likelihood of finding that class at any location across the study area. All models are produced at a resolution of 50 m pixel size which provides an adequate level of detail for regional mapping while allowing reasonably large areas to be modelled at once. Therefore, all predictor variables are resampled to 50 m for use in modelling including those originating from finer scale data (e.g. 5 metres bathymetry), coarser data (e.g. 800 metres oceanography), and categorical data (e.g. sediment grain size).

Validation statistics are generated for each map in the form of a Receiver Operator Characteristics (ROC) curve (figure 21), indicating how well the models based on training and test data (typically 25%) compare with random prediction (AUC = 0.5). These AUC statistics are included in the MAREANO technical summary for each area in tables similar to Table 1 where values close to 1 indicate good performance. Typical AUC values for MAREANO models to date are 0.75 – 0.95 although this value varies between classes and mapping areas. Whilst the AUC value gives an indication of model performance its importance should not be overplayed as it has a number of shortcomings (Lobo et al. 2008).

Since the required map product for MAREANO is a single biotope map showing the overall distribution of all biotopes, not one map for each biotope, a composite map which combines the results of all the individual models also needs to be produced. This step is performed in ArcGIS Spatial Analyst using the probability scores from each individual map to determine the most likely biotope occurring at any given location (figure 22).



a b c d e f g h

Figure 20: Individual biotope maps for the 8 biotope classes in the Troms III – Eggakanten – Tromsøflaket mapping area (see figure 22). Maps are shown on a rainbow colour scale with values 0 to 1 where blue indicates low probability and red indicates the highest probability.

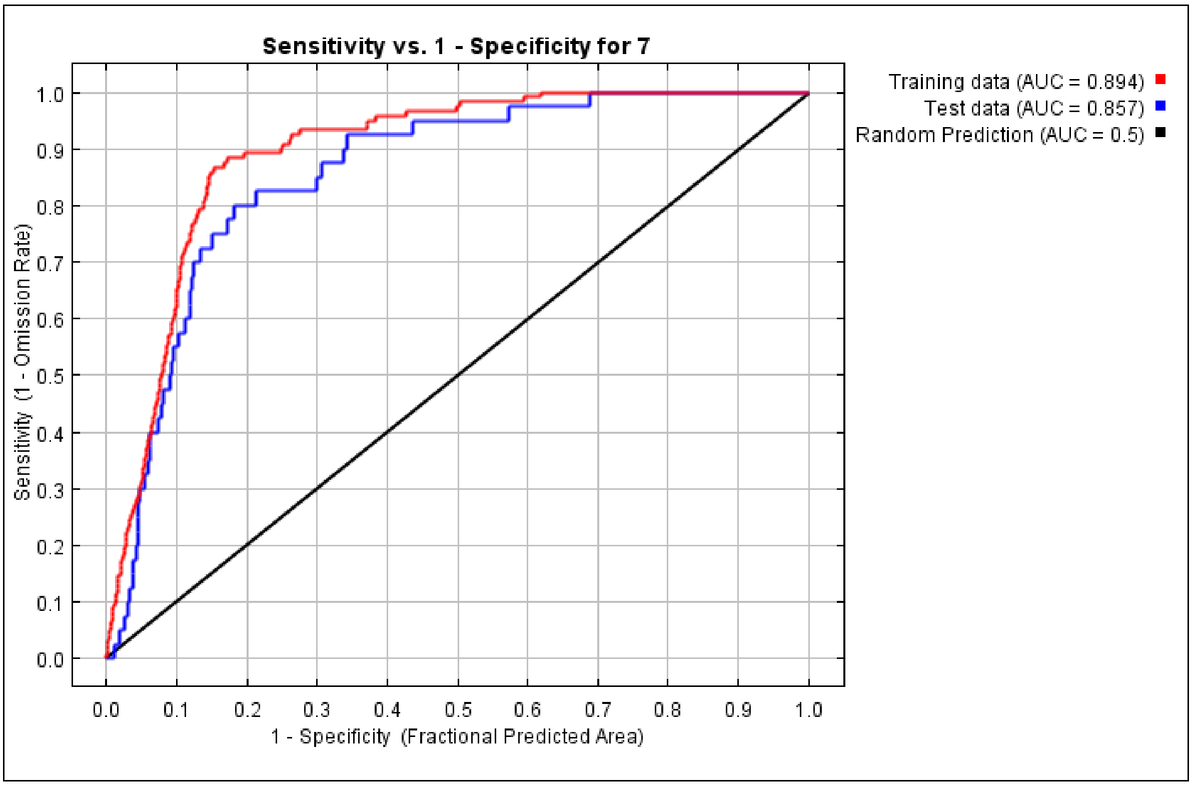


Figure 21: Example ROC curve generated by Maxent for Class 1 of the Troms III – Eggakanten-Tromsøflaket biotope map.

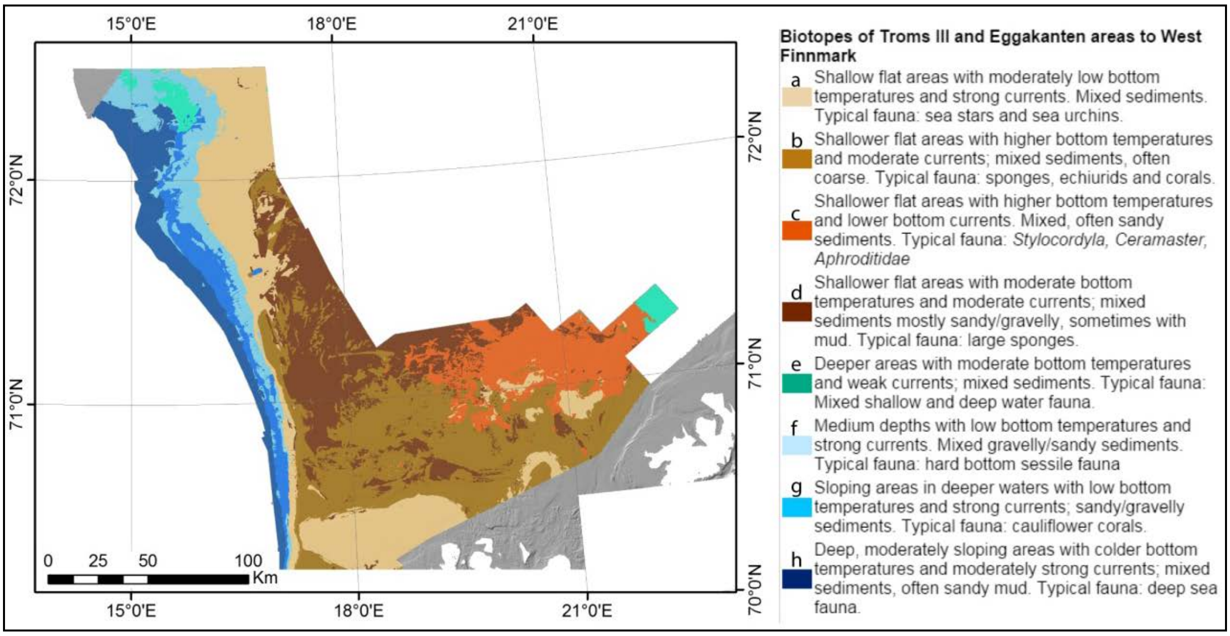


Figure 22: Example Biotope map for Troms III – Eggakanten - Tromsøflaket. See figure 20 for individual biotope maps with probabilities.

Until 2012, this map was tested using only an overall accuracy figure indicating the percentage of input points that are predicted correctly with respect to all available classified biotope points by the final composite map. Since 2013, an additional check on model performance has been introduced where the user's and the producer's accuracy provide a summary of performance across biotope classes and summarised in the Kappa Statistic (Table 3). This additional test on map performance was first reported by MAREANO in Elvenes et al. (2014). Since MAREANO biotope maps for all areas have either been published for the first time or updated since that time, confusion matrices and Kappa Statistics are available for all areas (except west Finnmark, where only low resolution oceanography data were available).

The producer's accuracy refers to the probability that a certain biotope observed on the seabed is classified as such by the model, while the user's accuracy refers to the probability that a pixel with a certain biotope class value in the modelled biotope map really is this class. The Kappa statistic (K), calculated using these accuracy values, provides a measure of overall performance assessing the degree to which the biotope map and point data agree over and above that which could be expected by chance alone. According to the interpretation scale of Altman (1991), which was adopted by Lucieer et al. (2013) for benthic habitat mapping, the values of the Kappa statistic can be interpreted as: $K < 0.2$ poor, $0.2 < K \leq 0.4$ fair, $0.4 < K \leq 0.6$ moderate, $0.6 < K \leq 0.8$ good, $0.8 < K \leq 1.0$ very good.

Although the confusion matrices and Kappa Statistic give a good level of information on map accuracy they do not fully address the more challenging issue of assigning a map confidence rating. As we see from figure 20, many of the biotope classes typically have very different distributions, however there are typically a few classes where the probability of more than one biotope occurring at a given location is more similar. It is in these areas that our confidence in the most likely biotope is less, even if accuracy assessments indicate that the map performs well. This issue is highlighted by the current biotope modelling approach combing results from single models for each species, but it is an underlying issue even if other modelling and prediction methods are used which take a more direct route from input data to a composite map.

Table 3: Confusion matrix table and AUC summaries for 8 biotope classes (Troms III-Eggakanten-Tromsøflaket).

AUC training	*0.909*	*0.833*	*0.951*	*0.953*	*0.848*	*0.883*	*0.894*	*0.95*		
AUC test	*0.869*	*0.803*	*0.936*	*0.956*	*0.763*	*0.82*	*0.857*	*0.93*		
BIOTOPE CLASS	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Total**	**User’s accuracy**
1	*127*	1	3	18	9	1	0	7	166	0.7651
2	2	*234*	0	0	24	32	4	1	594	0.8939
3	5	0	*59*	1	0	0	8	1	74	0.7973
4	14	0	0	*46*	2	0	0	0	62	0.7419
5	8	21	7	1	*74*	3	20	1	135	0.5481
6	1	45	0	1	5	*126*	9	0	187	0.6738
7	0	12	7	0	8	12	*122*	0	161	0.7578
8	6	0	5	2	0	0	0	*102*	115	0.8869
TOTAL	163	610	81	69	122	174	163	112	**1197**	
Producer’s accuracy	0.779	0.8705	0.7284	0.6667	0.6065	0.7241	0.7485	0.9107		
Kappa statistic	***0.737043***									
Overall accuracy	***74.35%***									

It is important to note that the performance of all these model assessment statistics will depend on the number of biotope classes. Models with fewer classes will generally give better statistics as there are more training (and test) points per class. However, fewer classes means a lower level of information content in the resulting map. The pragmatic solution is therefore to find a reasonable trade off between level of detail and an acceptable accuracy measure.

A related issue currently under investigation as part of the MAREANO biotope methods development study is the sample length extracted from video lines. To date a sample length of 200 metres has been standard for biotope mapping since an initial test on cumulative species richness at Tromsøflaket (figure 23). This choice of length is currently being evaluated across several representative mapping areas in order to test if this length remains the most suitable choice for future biotope modelling. Rank correlograms of the data from Nordland VI have provided further evidence in favour of this length (see below, and section 4.3.3).

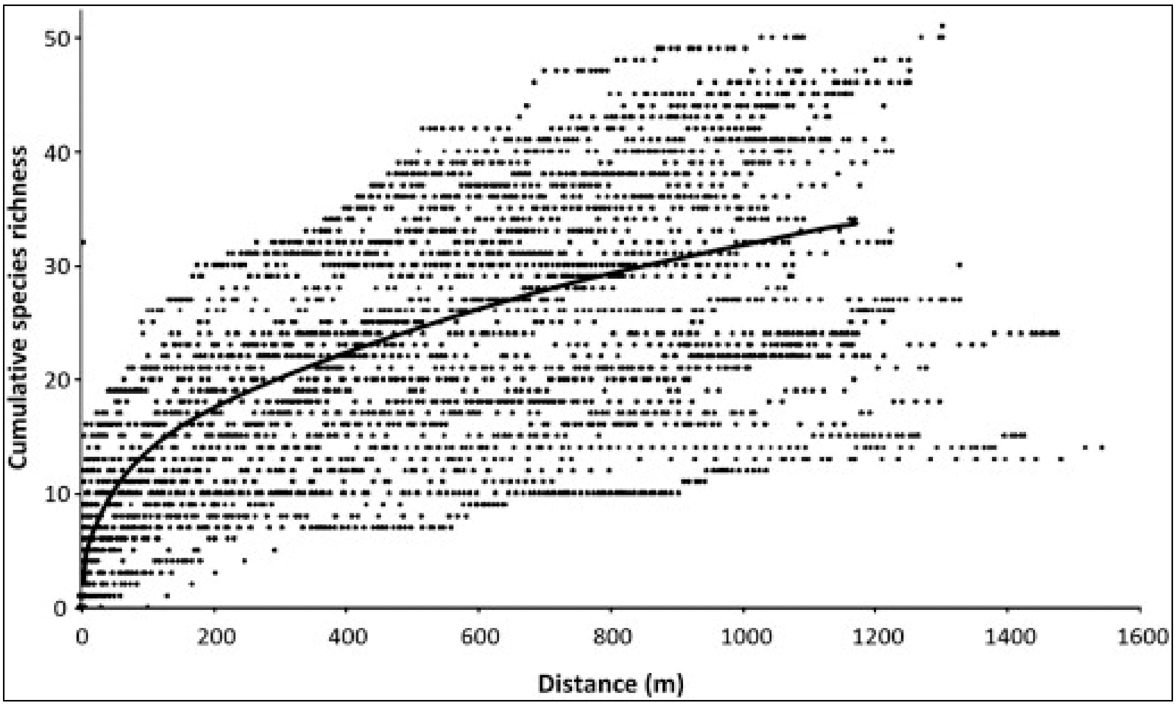
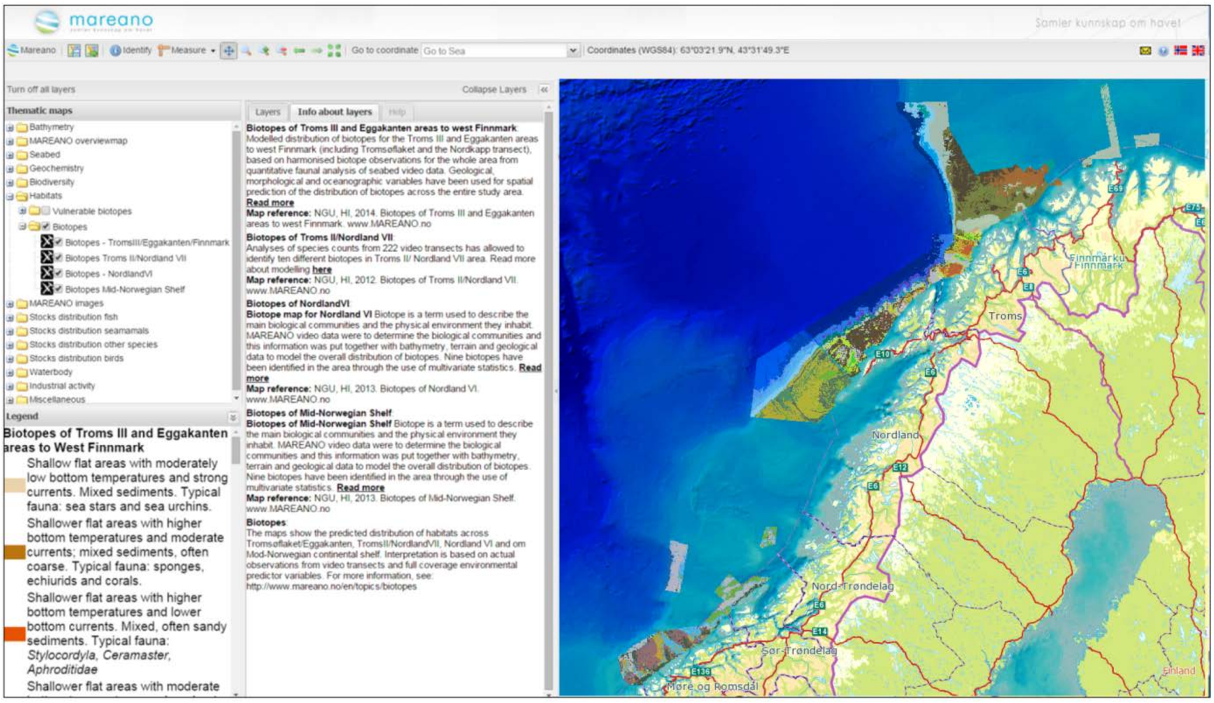


Figure 23. Cumulative number of taxa versus distance along 48 video lines recorded at Tromsøflaket. The solid line indicates the mean cumulative number of taxa for the same distances for all 48 video lines. From Buhl-Mortensen et al. (2015).

The method for definition of biotope classes has been changed in recent years from expert-based to automated (k-means) providing greater transparency and repeatability to the classification, however, the current approach still requires that biotope classes are determined from ordination (DCA) results specific to each mapping area. This means that maps for different MAREANO mapping areas are not harmonised (figure 24) and leads to changes in class definitions over time as mapping area boundaries extend and some areas e.g. Tromsøflaket are remodelled. With no pre-defined biotope (habitat) classification to work according to and an initial lack of knowledge about Norwegian offshore fauna MAREANO's first ten years have been largely a discovery phase for biotope mapping. Within this framework evolving biotope maps have generally proved acceptable and useful for management and other users. However, now that MAREANO mapping is well established and knowledge of the benthic fauna of this area much advanced thanks to MAREANO the demand for harmonised and stable biotope maps is growing.



Biotopes of Troms III and Eggakanten areas to west Finnmark:
Modelled distribution of biotopes for the Troms III and Eggakanten areas to west Finnmark (including Tromsøflaket and the Nordkapp transect), based on harmonised biotope observations for the whole area from quantitative faunal analysis of seabed video data. Geological, morphological and oceanographic variables have been used for spatial prediction of the distribution of biotopes across the entire study area. **Read more**
Map reference: NGU, HI, 2014. Biotopes of Troms III and Eggakanten areas to west Finnmark. www.MAREANO.no

Biotopes of Troms II/Nordland VII:
Analyses of species counts from 222 video transects has allowed to identify ten different biotopes in Troms II/ Nordland VII area. Read more about modelling **here**
Map reference: NGU, HI, 2012. Biotopes of Troms II/Nordland VII. www.MAREANO.no

Biotopes of NordlandVI:
Biotope map for Nordland VI Biotope is a term used to describe the main biological communities and the physical environment they inhabit. MAREANO video data were to determine the biological communities and this information was put together with bathymetry, terrain and geological data to model the overall distribution of biotopes. Nine biotopes have been identified in the area through the use of multivariate statistics. **Read more**
Map reference: NGU, HI, 2013. Biotopes of Nordland VI. www.MAREANO.no

Biotopes of Mid-Norwegian Shelf:
Biotopes of Mid-Norwegian Shelf Biotope is a term used to describe the main biological communities and the physical environment they inhabit. MAREANO video data were to determine the biological communities and this information was put together with bathymetry, terrain and geological data to model the overall distribution of biotopes. Nine biotopes have been identified in the area through the use of multivariate statistics. **Read more**
Map reference: NGU, HI, 2013. Biotopes of Mid-Norwegian Shelf. www.MAREANO.no

Biotopes:
The maps show the predicted distribution of habitats across Tromsøflaket/Eggakanten, TromsII/NordlandVII, Nordland VI and om Mod-Norwegian continental shelf. Interpretation is based on actual observations from video transects and full coverage environmental predictor variables. For more information, see: http://www.mareano.no/en/topics/biotopes

Biotopes of Troms III and Eggakanten areas to West Finnmark

▨	Shallow flat areas with moderately low bottom temperatures and strong currents. Mixed sediments. Typical fauna: sea stars and sea urchins.
▨	Shallower flat areas with higher bottom temperatures and moderate currents; mixed sediments, often coarse. Typical fauna: sponges, echiurids and corals.
▨	Shallower flat areas with higher bottom temperatures and lower bottom currents. Mixed, often sandy sediments. Typical fauna: *Stylocordyla, Ceramaster, Aphroditidae*
	Shallower flat areas with moderate

Figure 24: All MAREANO biotope maps currently published on www.mareano.no. Biotope maps have been separately produced for 4 areas: Troms III and Eggakanten to west Finnmark, Troms II and Nordland VII, Nordland VI, Mid-Norwegian Shelf. Some of these maps are already updated from previous versions.

The classes of video samples based on similarities in species composition are challenging to compare between areas modelled separately. As new areas are surveyed and new video samples are added to the database, biotope classes need to be revised. Although tests of this between MAREANO survey areas have proved useful with only small changes of clear classes identified by the prior separate ordinations this area-by-area based analysis strategy does not seem sustainable especially in light of the growing demand for a harmonised map. Tests to date indicate, however, that the solution to harmonisation is not simply a matter of working with an ordination that pools all available biological data to date. Important faunal groups can be obscured by such methods, and the species turnover may be too great for the ordination methods to handle when operating across several biogeographic regions. Alternatives to DCA such as Non-metric multidimensional scaling (NMDS) (Kruskal 1964 a, b) are also under investigation, either as ways to obtain more relevant information for identification of biotope classes and/or as parallel ordination (van Son and Halvorsen 2014) to validate the DCA results, increasing the overall confidence.

MAREANO biotope mapping has now been conducted across several biogeographic regions from the mid-Norwegian shelf to the Barents Sea. It is important to remember that when the geographic area increases, the “biological signals” resulting from regional “climatic” gradients become increasingly important. Across larger biogeographic regions oceanographic gradients become less or varyingly correlated with predictor variables derived from the terrain, that may have served as an adequate proxy within a certain ‘climatic setting’. Across areas spanning thousands of square kilometres spatial prediction based on depth, terrain and sediment classes will have less explanatory ability and the model will be poor. To address this problem, the best solution is to include full-coverage information about the seabed “climate” as predictor variables (e.g. temperature, salinity and currents, including estimates for their variability, maximum and minimum values). The first biotope maps incorporating oceanographic data were published in 2014 (Troms III, Eggakanten and west Finnmark) incorporating data from related projects and IMRs database. These data have demonstrated that oceanographic data will be an important input to future MAREANO biotope modelling as well as prerequisite for the production of “harmonised” biotope maps across larger areas of the MAREANO mapping area. Oceanographic data from the NorKyst 800 model (Albretsen et al. 2011) have been provided to MAREANO in 2015 and will be used in future biotope modelling within the area covered by these data (figure 25).

Another practical consideration that will need to be addressed when modelling over large areas is that of model resolution. A raster size of 50 metres offers a good compromise between detail and processing capability for limited mapping areas however in predicting across wider areas this may need to be revised to a coarser resolution, or the mapping area subdivided into areas of a size suitable for predictions to be run at 50 m.

Biotope maps are just one component of the overarching theme of habitat mapping. Another potential, and possibly parallel, route towards harmonised habitat maps for MAREANO mapping areas is through use of the newly updated Norwegian Nature Types classification descriptive system (Halvorsen et al. 2015).

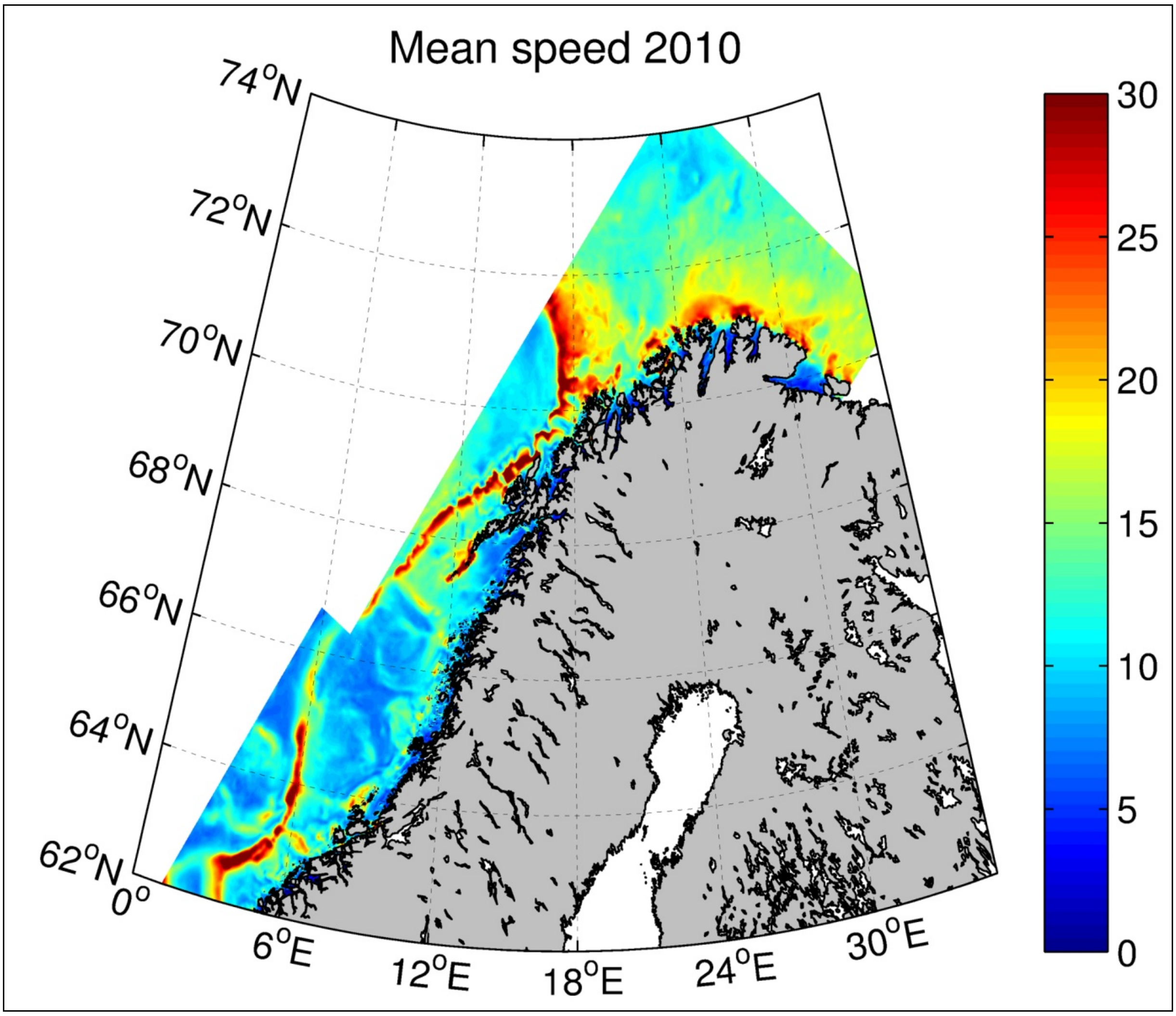


Figure 25: Example of oceanographic data from the NorKyst 800 model. Bottom current speed calculated from the lowest layer in the model given in cm/s.

3.3.2 NiN 2.0 - towards nature type mapping

‘Natur i Norge’ (NiN) – Nature Types in Norway version 2.0 was released April 15, 2015. NiN is now regarded by management bodies as the standard way of classifying and describing nature in Norway. MAREANO has had a close working relationship with other institutes leading to the development of NiN v.2.0.The natural progression of this collaboration is to work towards production of NiN-based nature type maps that make full use of MAREANO data, especially since MAREANO is financed directly through the national budget and produces map-related information that is used in marine management.

A move towards NiN-based mapping potentially removes the need to classify biotopes from scratch each time a new area is being mapped. In contrast to the methods currently used by

MAREANO for biotope classification, NiN operates on non-metric multidimensional scaling (NMDS) ordinations on theoretical and generalised lists of species and how they respond to and are structured by environmental complex-gradients. This way, NiN is a fixed, although adjustable, system that will classify species-compositions into known nature types. The delineation of these types is based on so-called ecological distance units. A nature type (grunntype) is bounded by one ecological distance unit along both axes (represented by a box in Table 4). These units correspond to a certain range along major complex-gradients (represented by NMDS axes) scaled in half-change units, which again correspond to a certain amount of change in species composition.

Thus, a NiN-classification will be based on a combination of (a) looking at where observations fall into the combination of major complex-gradients (which defines the nature type, 'grunntype') and (b) the position in ordination space. The advantage of this is that classifications based on NiN should always be the same, irrespective of the amount of biological data and the extent of the study area that are used for the classification. Furthermore, while biotope maps have faced problems with regard to naming conventions, NiN will always yield the same and recognizable name for areas of similar environmental conditions giving information with greater stability for management and other users.

In addition to providing a basis for a more stable system of nature type classification the NiN approach, using generalised species lists, overcomes another problem. There will always be quite a bit of noise in species data collected in the field that will tend to distort the ordination result. Such distortions will propagate into biotope modelling and introduce errors. Further, by basing the class boundaries on half-change units directly related to species turnover we remove the need to use a non-deterministic method such as k-means to split the biotope classes into a more arbitrary number of classes not directly linked to the species turnover.

Table 4: Example of NiN classification for M4 Eufotisk marin sedimentbunn (translated from and modified after Halvorsen et al. 2015) showing the concept of types (so-called 'grunntyper') varying along two major complex gradients. Each box represents a 'grunntype', and each box represents a standardised change in species composition along both axes. For example, it is expected that grunntype S1-g corresponds to a certain species composition that is characterised by little resistance to erosion and contain little fine-grained sediment particles.

S3–F– Mud content	¤ High mud content	unconsolid. mud with high water content	unconsolid. mud	S1·h coarse silt	S1·h fine and intermediate silt	S1·i clay	consolidated clay		
	bc Intermediate mud content			coarse sand with intermediate mud content	fine sand with intermediate mud content S3–S·c	gravel and rocks with intermediate mud content S3–S·d			
	0a Low mud content		S1·g fine and intermediate sand	S1·f coarse sand (and very fine gravel)	S1·e fine and intermediate gravel	S1·d coarse gravel S3–S·a S3–S·b	S1·c rocks	S1·b boulders	S1·a bedrock
M4 Explanation To S3-Diagram		0 very low	a low	b quite low	c intermediate	d quite high	e high	f very high	¤ bedrock
		S3–E – Resistance to erosion							

At the time of writing we see that the NiN system is promising, however we note that the publication of NiN 2.0 only presents the theory behind NiN. There is considerable work to be done in taking this theory into practical use in classification of nature and mapping, especially in the marine realm. MAREANO has gathered a wealth of biological and environmental information over the past decade but a dedicated effort will be required in collaboration with NiN to make full use of these data to produce generalised species list for marine nature types. This work should be prioritised as soon as possible. Until this is done any NiN-based mapping will rely on theory alone and therefore be based largely on physical environmental data while underusing MAREANO's rich biological data resource. As long as this is the case a separate biotope map is likely to be required in order to present biological results from MAREANO.

3.4 Vulnerable habitat map

For this exercise we used visual data from video lines conducted off Nordland and Troms between 2006 and 2009. The area selection was dictated by management needs and is approximately 63 000 square kilometres. The total number of video lines was 388. The species occurrence data was pooled at a 200-m scale. Each transect yielded 3-5 point-locality samples, and the total number of samples (*n*) was 1709. For additional details, see Gonzalez-Mirelis and Buhl-Mortensen (2015).

3.4.1 Habitat selection

We used three commonly-used criteria to prioritize marine benthic habitats: being dominated by long-lived species, being unique in Norway, or simply, being included in the OSPAR list of Threatened and/or Declining Habitats (OSPAR 2008). They are all considered highly vulnerable to physical or mechanical damage by e.g. demersal fishing gear, and some host significant biological diversity. These classes are defined on the basis of both species composition and dominant substrate type, and are summarized in Table 5. These initial classes were refined through multivariate analyses of biological data similar to those conducted prior to biotope modelling. Because so much variation was found in Deep Sea Sponge Aggregations this class was split into three sub-classes according to substrate type and presence of glass sponges (class Hexactinellida). The final number of modelled classes was seven (Hard bottom coral garden; Soft bottom sponge aggregations; Seapen and burrowing megafauna; Umbellula stands; Glass sponge communities; Hard bottom sponge aggregations; Soft bottom coral gardens).

3.4.2 Modelling and model performance

For each habitat class, we created a spatially-explicit dataset of the total abundance of all indicator species pooled together (total density of organisms/colonies of indicator species) for each video line subsample (referred to simply as “samples”), as response variable. Samples where none of the qualifying species were found for each habitat were used as absence data. We reserved 677 samples for model evaluation. These were selected randomly, with the additional condition that they be as geographically isolated as possible to avoid inflating model performance measures on account of possible spatial autocorrelation.

Table 5. Habitat and biotope classes considered of special concern in offshore, Northern Norway.

Class	Indicator species	Criterion
Umbellula Stands	*Umbellula encrinus*	Long-lived species
Radicipes Meadows	*Radicipes* cf. *gracilis*	Red-listed in Norway
Deep Sea Sponge Aggregations	Fam. Axinellidae, Geodiidae, Mycaliidae, Darwinellidae, Ancorinidae; class Hexactinellida	OSPAR
Seapen and Burrowing Megafauna Communities	*Funiculina quadrangularis*, *Kophobelemnon stelliferum*, *Pennatula* sp. (and other pennatulaceans), *Virgularia* spp.	OSPAR
Coral Gardens (particularly, Hard Bottom Coral Gardens)	*Paragorgia arborea*, *Paramuricea placomus*, *Primnoa resedaeformis*	OSPAR

The predictor data was derived from geophysical data collected through hydroacoustic remote sensing techniques (multibeam echosounder), following a similar procedure as for biotope modelling. We used depth, slope, topographic complexity (surface area), and landscape type. Layers on sedimentary environment, and dominant grain size were generated from backscatter data, aided by expert interpretation and ground-truthed by video data (for details see Bøe et al. 2010 and this report).

We used Conditional Inference Forests (CIF, Hothorn et al.2006) to predict habitat suitability. This method has been applied to modelling of vegetation types (Czúcz et al.2011) and benthic biotopes (Gonzalez-Mirelis & Lindegarth 2012). As a machine-learning method, it is data-driven rather than model-driven, and it is based on recursive partitioning. Recursive partitioning-based methods have become popular owing to their ability to solve classification problems where data are multidimensional, explanatory variables are correlated, and relationships are non-linear (Strobl et al. 2008), as is the case in ecology.

For each habitat we (1) fit a CIF to the training data (*n*=1032); (2) used this model to predict the probability of presence (of any of the species in the group) for every observation in the evaluation dataset; and (3) checked the observed presences and absences in the evaluation dataset against the probabilities returned by the model to compute a measure of model performance (see below). We then approximately repeated this procedure to generate spatial

predictions, except that at this time we used the whole dataset (*n*=1709), and we set the model to predict values in the same scale as the response in the data (density).
Accuracy of the models was measured by the AUC (area under curve) of the ROC (receiver operating characteristic, Fielding and Bell, 1997). We used the following system for classifying the accuracy of each model: 0.9-1 = excellent, 0.8-0.9 = good, 0.7-0.8 = fair, 0.6-0.7 = poor, 0.5-0.6 = fail. Modelling results are summarized in Table 6, and except for two classes, they were all good or above.

The best performing models were those for *Radicipes* Meadows, and *Umbellula* Stands with AUC values of 0.99, and 0.85. These were conventional Species Distribution Models and their high performance is a result (among other factors) of models dealing with a single species and thus a uniform response (Zimmermann and Kienast 1999). The high accuracy of the *Radicipes* model can also be partly due to the low prevalence of this taxon (0.1%, Table 6), which makes it easy to predict its absence. However, it is in fact known that this biotope is not present outside the Bjørnøya slide area, as the data extends well beyond that locality. The limited distribution of this species and the fact that where it occurs it forms extensive, dense meadows makes this species easy to model and the model particularly useful to define the boundaries of the distribution range of this red-listed biotope (Lindgaard and Henriksen 2011).

Habitat-forming organisms can cover the whole surface of an area (and fill up a large volume over the seabed) at this scale (hundreds of metres) leaving only small, randomly-distributed gaps, much like vegetation does. A situation where the scale of analysis is smaller than the size of the patch of the feature of interest is known to be the optimal scenario to detect patterns from observational data (Fortin and Dale 2005). Therefore, the predictions from the Hard Bottom Coral Gardens model can aid both in understanding the factors shaping the patterns of distribution, and in selecting target areas for further investigation and ultimately for protection of hard bottom gorgonian corals and Coral Gardens.

Demosponge aggregations, both on soft bottom (ostur) and hard bottom were also adequately predicted by models. Ostur is predicted (and known) to occur in massive, dense aggregations at Eggakanten and Tromsøflaket area. All other models were less reliable and probably habitat definitions or sampling scales will need to be revised.

Table 6. Modelling results, including model performance as Area Under curve (AUC), prevalence of the qualifying species in the dataset, maximum observed density, and variables that ranked first and second as explanatory variables

Habitat name	**AUC**	**Prevalence (%)**	**Maximum observed density (n/100m^2)**	**Top-two explanatory variables**
Umbellula Stands	0.85 (good)	0.3	6.4	Surface area Depth
Radicipes Meadow	0.99 (excellent)	0.1	239.8	Landscape type Sedimentary environment
Hard bottom Demosponges	0.81 (good)	44.4	127.4	Depth Landscape type
Soft bottom Demosponges	0.80 (good)	37.7	258.3	Landscape type Grain size
Other Deep Sea Sponges, including Glass Sponges	0.77 (fair)	9.0	30.6	Sedimentary environment Landscape type
Seapens and Burrowing Megafauna	0.73 (fair)	14.0	52.0	Landscape type Grain size
Hard Bottom Coral Gardens	0.85 (good)	4.0	44.4	Landscape type Surface area

3.4.3 Creating the final distribution map and calculating additional accuracy statistics

Plotting the predictions from the model in geographic space results in what is known as a “heat map”, such as those in figure 20. Management authorities, however, are often more interested in crisp boundaries, which in turn allow implementation of marine spatial planning. Therefore, we decided to convert heat maps into binary maps. To achieve this, a decision on an appropriate threshold value needs to be made. This effectively means that only pixels whose predicted density of indicator species is above that value will be classified as “present” for that habitat type. Additionally, to minimize overlap between the different habitats at any given location only the habitat of maximum predicted density is ascribed to each pixel. Lastly, we applied a boundary clean algorithm which removed single or very small groups of pixels as well as very small gaps inside polygons. The result is shown in figure 26. and Table 3, and more details about some of these steps are found below.

3.4.4 Density thresholds

We attempted to generate threshold values by splitting up all samples (within each habitat type) into density groups; fit a separate model for each subset (i.e., first using only low-density samples, then intermediate-density samples, etc.), and plot model performance against density, expecting an increase in performance at a given density value. The assumption behind this approach was that where the density of the species of interest is low the species should be responding mostly to chance events and any patterns should be harder to detect. This effectively is similar to modelling the distribution of a species outside of its range. Unfortunately, for most habitats the number of samples was not enough and this could only be implemented for the Seapens case. All other thresholds were picked arbitrarily, only in reference to the range of observed values (see maximum observed values in Table 6) and how much each threshold value affected the total area predicted.

The resulting polygon layers can be checked for accuracy. To compute the confusion matrix we used the same threshold to classify both the observed values and spatial predictions. In other words, samples where the observed total density was above the threshold and which intersected the relevant polygon were considered true positives, and those where the observed density was below the threshold and did not overlap with the corresponding polygon were true negatives.

Table 7. Density thresholds and resulting accuracy values for each map layer.

Habitat name	**Threshold used (n/100m^2)**	**Percent correctly classified (%)**	**Kappa**
Umbellula Stands	0.06	95	0.45
Radicipes Meadow	4.4	100	0.73
Hard bottom + Demosponges	5	87	0.44
Soft bottom + Demosponges	6	92	0.57
Other Deep Sea Sponges, including Glass Sponges	1	96	0.40
Seapens and Burrowing Megafauna	2	95	0.50
Hard Bottom Coral Gardens	0.26	98	0.54

In relation to OSPAR habitats, some authors have said that the density of the qualifying species needs to be at least 10 times the background density (Rogers and Gianni 2009), but the problem remains as how to measure the background density. OSPAR has made attempts to make the density of colonies or individuals a criterion in the definition of habitats (e.g. for Coral Gardens, Christiansen 2010). This approach has not succeeded owing to difficulties in adequately quantifying the variation of density across space (Bullimore et al. 2013, ICES 2007) but should such a figure be available it would be straightforward to use it to classify the spatial predictions from a model.

This is an issue where data like that collected in MAREANO can be of great help, as video data, particularly when pooled at small scales (10s of metres, or even single video frames) is ideally suited to explore variation of density across space. Furthermore, a promising avenue is the use of geostatistics, whether to create interpolated surfaces or to simulate distributions.

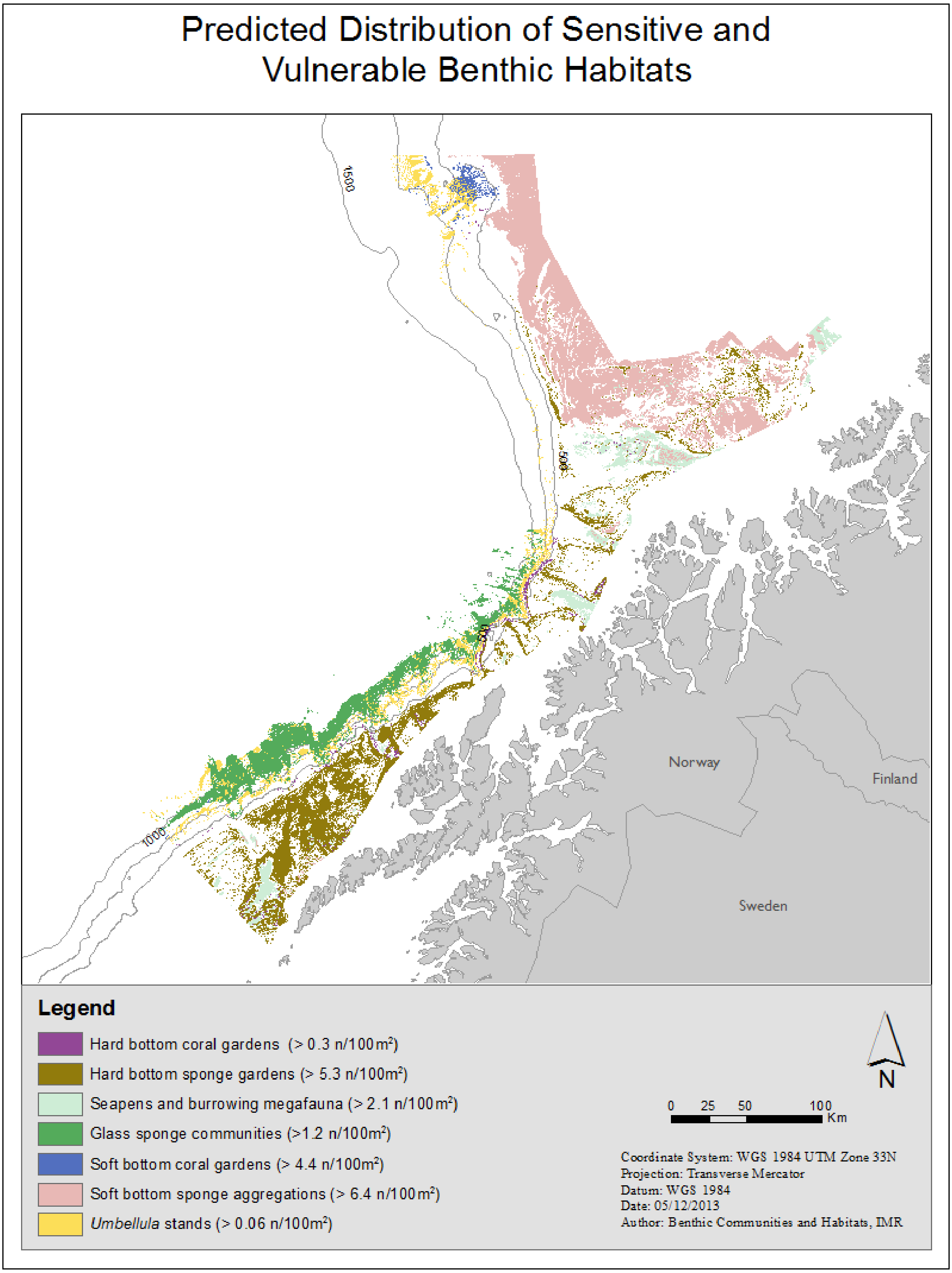


Figure 26. Predicted distribution of vulnerable habitat types, as defined by MAREANO based on prior knowledge and OSPAR definitions. The distributions were generated by means of spatially-explicit modelling of sets of qualifying species (indicator species).

3.5 Standardised procedures for interpretation, modelling and map representation

There is a need for methods to document data manipulations, spatial analysis, and modelling taking place in MAREANO. The focus on web-based, rather than report-based map delivery has also contributed, at least in part, to the fact that documentation of many products and routines (e.g. metadata) is variable. Although ArcGIS remains the common platform for data integration and much analysis, lately, R is becoming more and more used for many of the numerical-based analyses from cruise planning to data analysis. The increased use of R has opened up opportunities for better documentation, standardisation, and reproducibility of the MAREANO's deliverables. The use of Rmarkdown documents (created in Rstudio) is a particularly promising way to combine documentation, code and illustrations (see Appendices of van Son et al. 2015). Another interesting option could be to use IPython (http://ipython.org/notebook.html) for this purpose. A transition to these methods will require that MAREANO staff have sufficient opportunity to become familiar with these tools and adjust their workflows accordingly. It is clear, however, that different map products will have differing levels of suitability for these approaches.

Metadata is essential to disambiguate data and enable data reuse. Data provenance, one kind of metadata, pertains to the derivation history of a data product starting from its original sources, and providing evidence of how, what, and who was involved in creating it, as well as when it happened. These are two aspects that show room for improvement in most MAREANO data and map products.

The biotope workflow (figure 17) is one that is largely repeated each time MAREANO maps a new area. There is potential for automating at least parts of this workflow so that it is reproducible. Automation could increase the efficiency of map production, as well as allow for testing of new settings and optimization of methods so that outputs can be adapted and improved as necessary. Typically the modelling process is an iterative procedure with several models and prediction results being produced and evaluated on the basis of both performance statistics and expert judgement. Documentation of data quality is important, as is the ability to track any sources of error through to the final product, ensuring that complete provenance metadata can be provided to accompany the resulting map.

We examine the potential for automation with reference to the biotope maps (Section 3.3.) although many of the points are relevant to other map products, particularly those produced by numerical rather than expert-based analysis (at present the biotope and vulnerable habitat maps). The workflow can be broken down into blocks, each producing an output which is a requirement to run the next block. An example outline of this process follows, where data sources would include the video data pooled at 200 metres (in the form of a species-by-site matrix), the predictor variables, and a database of biotope classes.

Block 1: Multivariate analyses
Input datasets: (1) A species-by-site matrix for the 200 m-long subsamples, (2) a list of existing biotope classes with typifying species (to check against, and continually update)
Input parameters: number of clusters (can be implemented as a slider)
Outputs: (1) updated list of classes with typifying species, (2) samples with assigned classes

Block 2: Spatial overlay, variable extraction, and model selection
Inputs: (1) samples with assigned classes, (2) predictor variables
Outputs: list of important variables ("best model")

Block 3: Modelling, spatial predictions, and map algebra
Inputs: (1) list of important variables, (2) samples with assigned classes, (3) predictor variables
Outputs: composite map of biotopes

A tool or platform for sharing and collaboratively running the code would need to be selected, and the workflow implemented.This type of approach could offer benefits to MAREANO, including reproducibility, ease of map updating including reference to data sources, plus a well-documented workflow (e.g., which species are being removed? Which samples are being removed? etc). If this type of automated workflow could be realised it would allow MAREANO to generate data products accompanied by complete metadata and provenance information, as increasingly required by big data science projects. To achieve this level of automation would, however, require a substantial programming effort.

The development of any such system would be sufficiently flexible that the methodology can evolve and improve over time. Techniques, particularly for automated classification and modelling, are in continuous development among the scientific community and MAREANO needs to remain sufficiently adaptable in its methodologies to be able to take up improved methods into the standard workflow, once they are sufficiently mature for use on large datasets. It is important to note that flexible automation would facilitate documentation of the workflow which is vital, not only for final documentation, but in the development and testing phase as products are being fine-tuned.

Biotope mapping, in particular is likely to see changes over the coming years as harmonization of biotopes between mapping areas and/or adoption of NiN as the basis for nature type mapping are addressed (section 3.3). The need for complete documentation remains paramount and flexible automation may offer significant benefits. Automation of routine tasks (e.g. handling of cruise data) may also be a particularly beneficial starting point for automating parts of the MAREANO workflow.

4. MAPPING AND SAMPLING STRATEGIES

4.1 MAREANO survey strategy

The main survey strategy for MAREANO is full area-coverage bathymetry/backscatter/water column data, with visual and physical sampling at a pre-defined density (10 and 2 stations per 1000 square kilometers), respectively. However, specific data needs and other circumstances have led to different survey strategies being employed in certain areas. The so-called "Nordkapp transekt" is a 15 kilometres wide and 100 kilometres long transect going at a right angle from the Finnmark coast (figure 27). These data were included in the biotope model for Troms III-Eggakanten-west Finnmark, but gave poorer results than the model for the area west of Tromsøflaket, mostly due to the fact that only coarse (25 km) oceanography data were available for this area at the time modelling was conducted, as compared with 800 m oceanography data for the remaining area. See http://mareano.no/en/topics/biotopes/biotopes_troms3_to_westfinnmark.

Predicted biotope distribution in this area is also challenged by the shape of the area which makes it difficult to capture east-west variations in biotope distribution. The resulting map for the Nordkapp transect contains just 2 biotope classes.

The Mørebanken SVO (specially valuable area) (figure 28) is another area where full area coverage of multibeam echosounder has not been used. In this area, multibeam data collected by the FFI from the coastal zone (outside 4 nautical miles) were combined with MAREANO multibeam data in a 20 kilometres wide corridor from the coast covering the Breisund area, and multibeam data from MAREANO and the industry covering the outer parts of the Mørebanken SVO. For the remaining areas, single beam echosounder data from Olex (www.olex.no) were used for the production of sediment and biotope maps. The sampling effort was 10 visual and 2 physical stations per 1000 square kilometres.

In the Barents Sea, a third approach is used. The Programme group has underlined the need for acquiring data from a transect crossing several important climatic zones. The original proposal was to establish two 10 kilometre wide and c. 500 - 600 kilometre long transects extending from the Nordkapp transect northwards (figure 29a). Following discussions, this has now been modified to a series of square areas each being c. 1000 square kilometres, connected by one or a few multibeam echosounder lines (figure 29b).

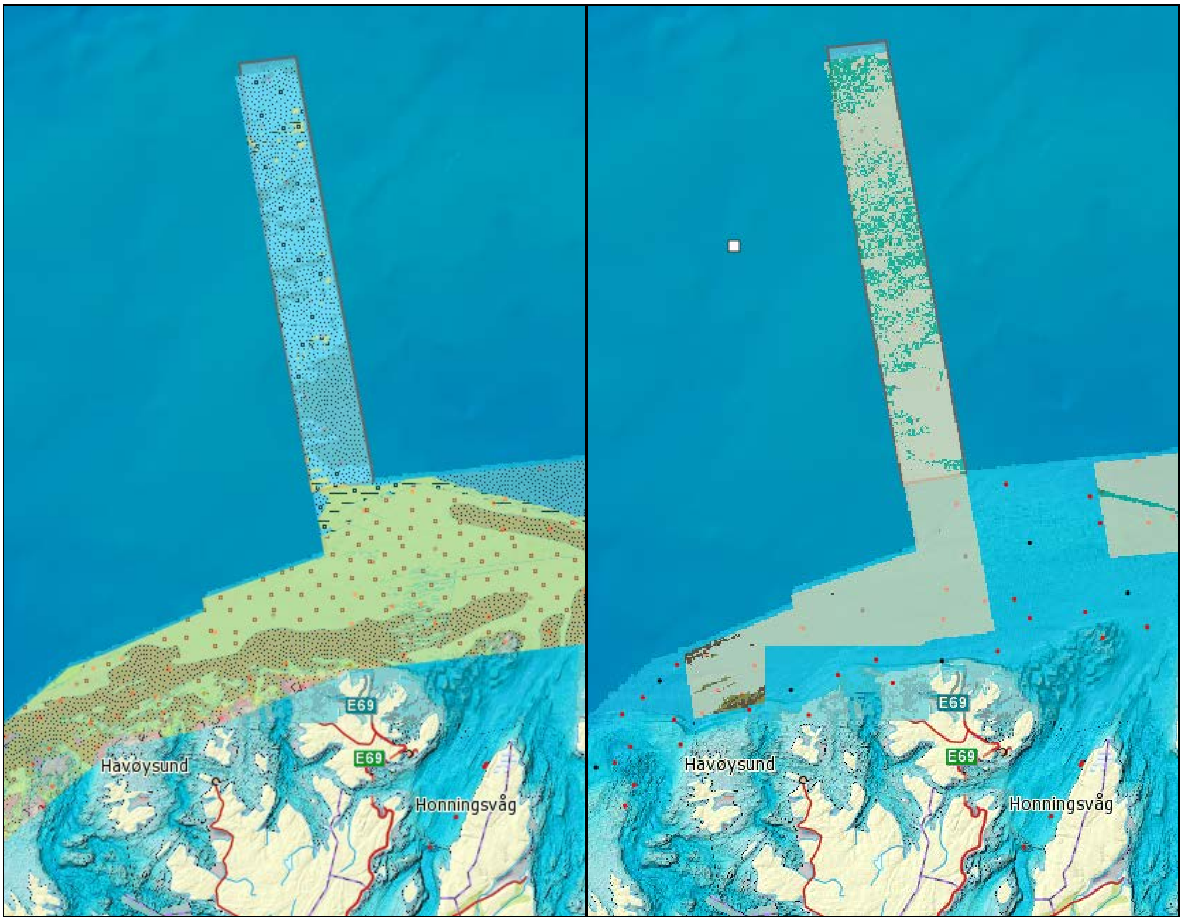


Figure 27. The Nordkapp transect extends in total 150 kilometres from the Finnmark coast, and is 15 kilometres wide. Left - sediment grain size. Right - biotope map.

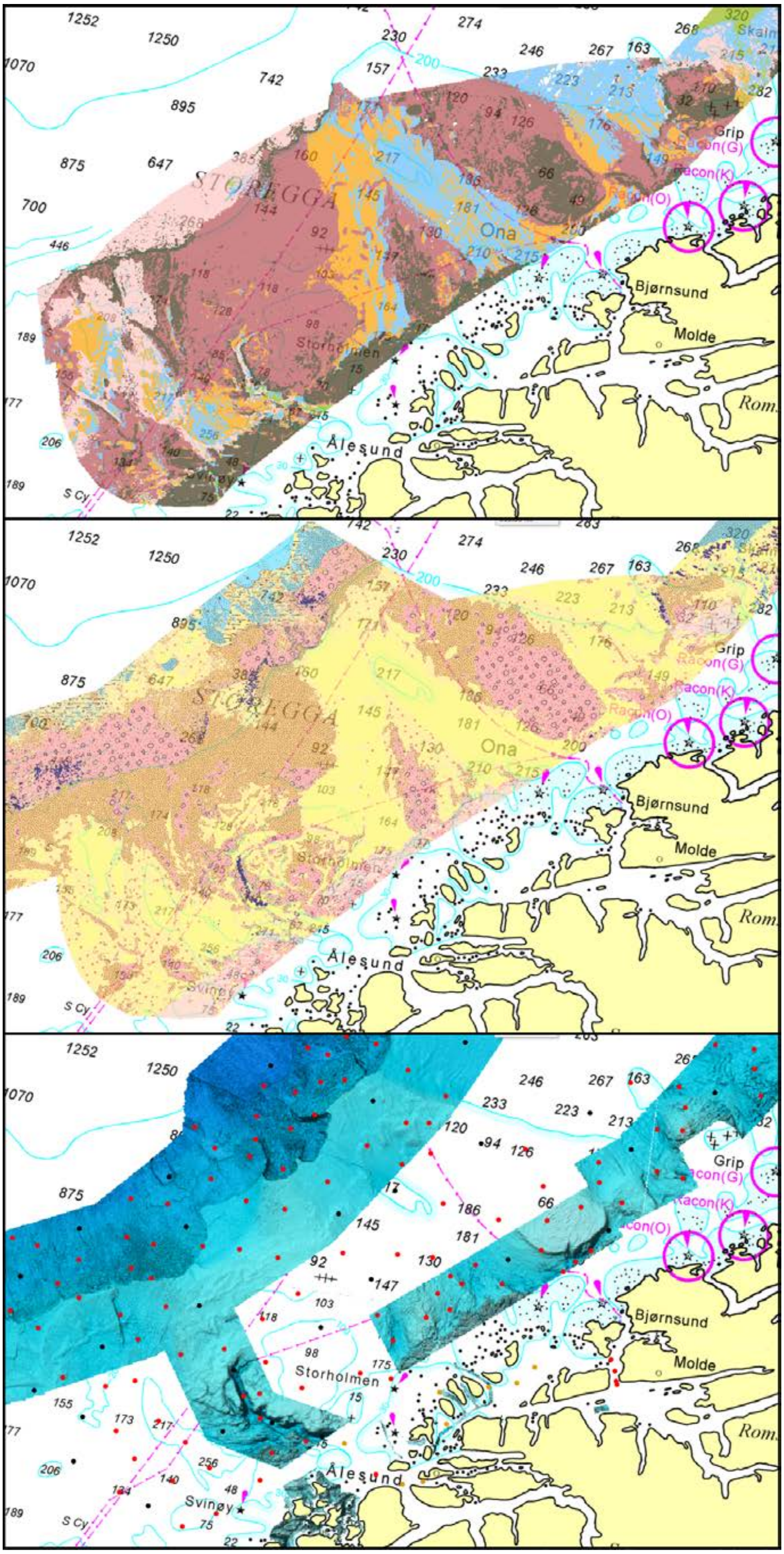


Figure 28. The Mørebanken SVO biotope (upper panel) and sediment map (middle panel) is based on only partly multibeam echosounder coverage (lower panel). For the remaining areas, single beam echosounder data from Olex was used.

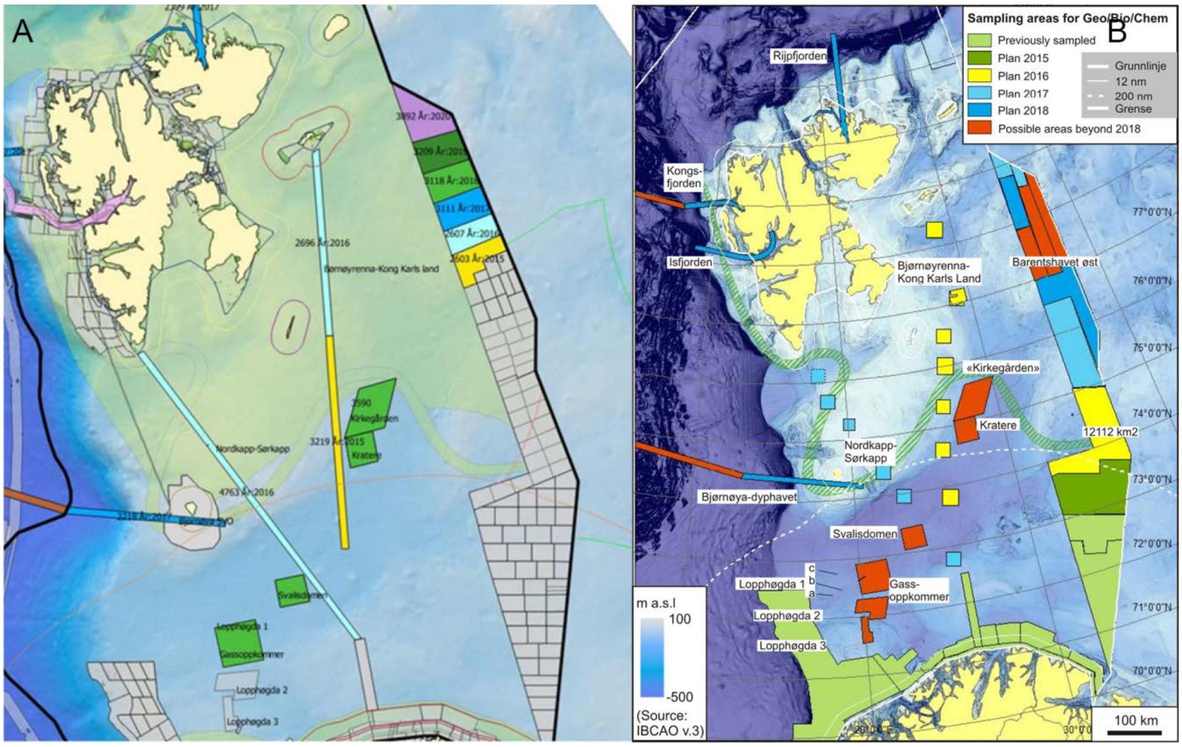


Figure 29. Proposed transects in the Barents Sea, addressing climate related aspects. A - original proposal with continuous, 10 km wide transects. B - modified proposal, with 1000 square kilometres boxes linked by multibeam lines.

4.2 General overview of sampling strategies

Sampling is performed to provide an estimate of parameters of interest, as such the general goal of sampling in a particular area is to obtain representative samples of the parameter(s) of interest (e.g. the abundance of a species). Despite the clear importance of the method used to locate samples, few studies have actually studied the implications of different approaches (Chap 4 in Murray et al. 2002, but see Albert et al. 2010). The aim of a future survey design in MAREANO should be to sample in such a way that as much as possible of the environmental space as defined by important, structuring environmental variables of the study area is sampled.

MAREANO samples both biological and geological data and the needs may not always be overlapping. To further complicate the matters, for modelling purposes, MAREANO should strive to sample data that account for environmental variability, that is independent among stations, and that is random. In certain cases, geologists and biologists may want to target certain areas of particular interest that may have been overlooked by a sampling strategy. To solve this issue, it is proposed to set aside a certain percentage (e.g., 20%) of the pool of stations for so-called 'targeted stations'. These targeted stations will allow for positioning stations in areas of particular interest, but will also introduce bias into the modelling process if used in modelling of for example biotopes or nature types.

The Environmental Variability Index (EVI; van Son et al. 2015) is intended to recommend the total sampling effort to go into a study area (figure 8). The suggested number of stations will then be the pool of stations to position optimally using the most appropriate sampling strategy for the purpose. A sampling strategy covering spatial and environmental variability will optimise the likelihood of obtaining representative samples.

Below follows a list of strategies for sampling marine seabed habitats, starting with Expert judgement. This is the survey design strategy MAREANO is currently based on, guided by unsupervised classification based on bathymetry, backscatter and oceanography, if available. Each sampling strategy is evaluated in general terms for its ability to account for environmental variability, independence among samples, and whether it is random or not.

4.2.1 Expert judgement

Accounting for environment: low to high
Independence: low to high
Random: no

The way this is implemented in MAREANO has been thoroughly described in section 2.2. Expert judgement can potentially provide sampling designs that are spatially balanced and that take environmental variability into account in an appropriate manner. However, expert judgement is not optimal as a basis for sampling strategy because it will usually ignore important gradients that are yet unknown and unidentified (Albert et al. 2010). The biggest drawback however is the high level of subjectivity involved. The lack of randomness in the positioning of stations leads to lack of independence among samples. Without randomness and independence, standard statistical inference and statements about confidence intervals and confidence level are invalid. However, in certain cases where the environmental data is of such poor and/or variable quality that the positioning of stations cannot adequately be guided by measures of environmental variability, then expert judgement or regular grid sampling (see below) may be the only viable options for cruise-planning. In many cases it is likely that some degree of expert judgement will also be necessary in order to identify and sample features of particular geological/biological/management interest which are inevitably overlooked by automated methods.

4.2.2 Transects

Accounting for environment: intermediate
Independence: low
Random: no

Transects represent long lines positioned to account for a large fraction of the range of values for a *single*, putatively important environmental gradient. This is a classical sampling method that is often used in fieldwork. In the marine realm, especially among marine benthic biologists, depth is the most commonly used environmental gradient forming the basis for a transect. With regard to the overall environment, transects have the potential to cover both geological and biological variation to a certain degree, and this can be achieved with relatively little effort. The main drawbacks with transects are that they are positioned subjectively (i.e., lack of randomness) and that they have very low interspersion between stations (i.e., lack of both spatial balance and independence between samples).

4.2.3 Regular grid sampling

Accounting for environment: low
Independence: low - high (depending on randomness and the interspersion between the samples)
Random: no (yes, if starting point of the grid is randomly defined)

Regular grid sampling is often regarded as a good sampling strategy for distribution modelling (Hirzel and Guisan 2002), but it actually has several drawbacks. The starting point of a grid is normally subjectively selected, thus there is lack of randomness among stations. A random start of the grid can deal with this issue. However, depending on the distance between stations within the grid, the stations may or may not be independent. Regular grid sampling does not take into account environmental information with the consequence that it is not possible to allocate weights to environmental strata and is therefore likely to fail to sample certain, rare combinations of environmental conditions. The whole area will receive the same sampling effort, regardless of differences in environmental variability within the area. Another disadvantage is that depending on the distance between stations, regular grid sampling is not a very cost-effective strategy. It will tend to sample more samples than necessary. In certain cases with bad quality data (artefacts), regular grid sampling and/or expert judgement may be the best alternatives.

4.2.4 Random sampling

Accounting for environment: low
Independence: low to intermediate
Random: yes

Ordinary random sampling places samples completely randomly within the study area. On the positive side, the strategy is random. However, random sampling tend to form clusters that negatively affect the spatial balance between samples. Furthermore, just as regular grid sampling, random sampling does not take into account environmental information, however the potential consequences of the two strategies are different. While regular grid sampling allocates the same sampling effort regardless of differences in environmental variability, ordinary random sampling may, in a worst case scenario, locate the highest sampling effort in the most environmental homogeneous parts of the study area.

4.2.5 Random stratified sampling

Accounting for environment: high
Independence: intermediate
Random: yes

Random stratified sampling is an extension of ordinary random sampling accounting for environmental variability. By combining environmental variables into strata, random stratified sampling has the potential to cover a very large fraction of the major environmental variation existing in the study area, given that the environmental variables have high predictive power in relation to the natural phenomena under study. Data from random stratified sampling designs is suitable for statistical modelling purposes. However, the problem with clustered samples is still an issue. Because of the stratification, the problem with spatial balance and independence is less of an issue than for ordinary random sampling.

4.2.6 Generalised Random-Tessellation Stratified (GRTS) spatially balanced sampling

Accounting for environment: high
Independence: intermediate - high
Random: yes

The Generalised Random-Tessellation Stratified (GRTS) sampling is very flexible and can be implemented in a wide range of different ways for point, linear and areal features (Stevens and Olsen 2004). Just as for random stratified sampling, GRTS is a stratified design and has all the same potential advantages compared to non-stratified sampling strategies. However, while random stratified sampling faces problems with clustered samples and spatial balance, GRTS accounts for this by using a restricted hierarchical randomisation process that provides spatially well-balanced sampling designs. The positioning of the samples themselves, is a completely objective process. However, as for all stratified sampling strategies, some subjective decisions need to be made, e.g., when deciding on the number of strata. In addition, GRTS depends on a probability inclusion layer that is crucial for dividing the samples among the different strata.

Another sampling strategy that tries to optimise the spatial balance is Balanced Acceptance Sampling (BAS; Robertson et al.2013). BAS can be implemented in R using the BalancedSampling package (Grafström 2014). So far, this sampling strategy has not been evaluated in MAREANO, but it will be in the near future.

4.2.7 Discussion of sampling strategies

The above overview of sampling strategies demonstrates their pros and cons related to important factors such as independence, how much they account for environmental variability and their level of randomness. Their relation to these factors and to each other can be summarised in a space spanned by independence and environment accounted for (figure 30).

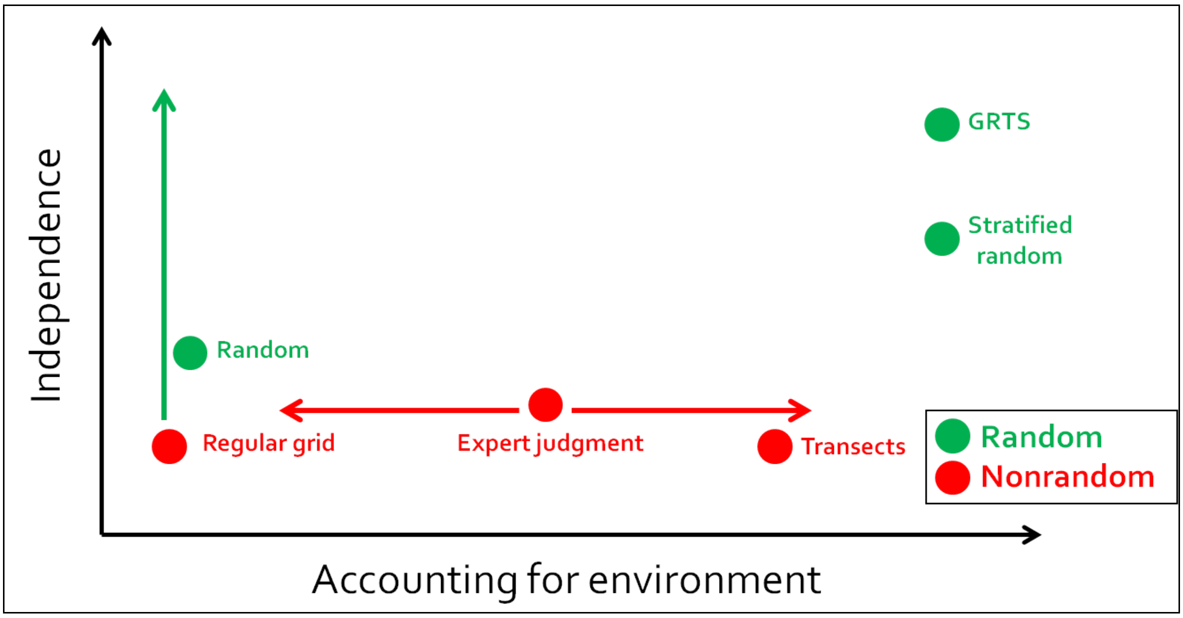


Figure 30. The location of sampling strategies in relation to their independence between samples and how much they account for environmental variability. Whether positioning of stations are random or not is also indicated. The red lines for Expert judgement indicate that its position depends on its implementation. The green line for Regular grid indicate that when the starting point of the grid is assigned randomly it becomes a random strategy and that the independence can vary depending on the interspersion between the stations.

Independence, accounting for environmental variability and level of randomness are the three most crucial factors in relation to the appropriateness of sampling strategies for mapping based on modelling. For geological mapping which is currently based on expert judgement supported by environmental indices, there may be a greater need for targeted stations. This is particularly important where the backscatter data set is noisy, and within areas where the backscatter data comes from different surveys with different surveys or settings. Spatial balance is important because it helps to avoid the issue of spatial autocorrelation, which can have adverse effects of the independence of samples (Legendre 1993; Dormann et al. 2007). Consideration of environmental variability can guide and scale the sampling effort in strata in a way that both optimises the cost-benefit trade-off and allows for improved statistical models. Randomness is also essential in order to analyse the data statistically. Investigator bias occurs when stations are targeted. When targeting stations, the assumption of independence of errors may be severely violated (Chap 4 in Murray et al. 2002).

4.3 A potential future sampling strategy for MAREANO

4.3.1 Generalised Random-Tesselation Stratified (GRTS)

No sampling strategy is perfect, but GRTS has a feature list that makes it very suitable for many types of sampling frameworks. The restricted hierarchical randomisation process, put into action by quadrant recursive partitioning of strata (Stevens and Olsen 2004), is the main feature that really makes it stand out compared to other good sampling strategies such as random stratified sampling. This allows for sampling strategies that are well balanced spatially, avoiding clumping of stations typically found by conventional random sampling, and hence also reducing and to a large degree avoiding problems with spatial autocorrelation. Stratification is of course an important feature of GRTS, but another, maybe equally important feature that is shared with other stratified sampling strategies, is the ability to optimally allocate stations based on knowledge about differences in environmental variability between strata. Optimal allocation analysis (e.g., Clements et al. 2010) provides a semi-objective method for determining the number of samples within predetermined strata/classes. The theory behind optimal allocation does not offer any guidance on how to place the samples within the strata, but this is handled in GRTS by the restricted hierarchical randomisation process.

Another attractive feature of GRTS is that it allows for dynamically adding new samples when non-target or inaccessible points are discovered. Another option to deal with non-target stations in GRTS is to ask it to return a certain amount of oversamples that are ready-to-use stations that can be used when facing the challenge of inaccessible points in the field. Both dynamically adding and oversampling is possible without sacrificing the overall spatial balance of the sampling design.

4.3.2 GRTS Implementation

There is a dedicated package in R, 'spsurvey' (Kincaid and Olsen 2015) for implementation of GRTS designs. In 'spsurvey' the GRTS implementation can take a shapefile containing environmental strata as input and it has means to weigh and optimally allocate samples to strata according to a wide range of different needs (both scientific and management). An example of such implementation is shown in figure 31.

The proper stratification of the environment is crucial for a successful implementation of

GRTS. After environmental variables have been selected carefully according to their importance for describing patterns of natural diversity, there exist at least two ways the environment can be stratified. It can be either stratified by conventional stratification or by application of the ISOCLUSTER algorithm (Ball and Hall 1965; Richards 1986) in ArcGIS. The k-means algorithm (MacQueen 1967) seems to produce very similar classes (i.e., strata) as the ISOCLUSTER algorithm. A more thorough comparison of these two methods will be conducted in the near future.

By conventional stratification, each environmental variable is divided into a number of sub-layers where each sub-layer covers a part of the total range of possible values for that layer. Setting the boundaries for these sub-layers is perhaps the most subjective part of any stratified sampling design. There are many possible ways for how to set these boundaries, but the two most important ways is probably to either 1) set boundaries based on biological and/or geological knowledge, this means that sub-layers will cover unequal ranges of values; or 2) set the range of the boundaries equal for all sub-layers.

At first, the first option for setting the boundaries may intuitively seem the best. However, if the aim of the sampling is to cover as much as possible of the environmental variation in the area, the latter option is the most successful. This is because it makes sure that every potential value for a given environmental variable is given equal opportunity not only of being sampled, but also of being combined with sub-layers of other environmental variables.

The ISOCLUSTER algorithm in ArcGIS as well as the kmeans() function in R use an iterative optimisation clustering procedure, also called the migrating means technique. Because the number of optimal classes is unknown, it is advised to assign a conservatively high number. At the initiation, each class is assigned a random mean (or cluster centre) in the multidimensional space defined by all the environmental variables submitted to the tool. The algorithm then for each iteration computes the minimum Euclidean distance when assigning each cell to a cluster, and the mean is recalculated. The algorithm proceeds until the migration of cells between clusters falls below a threshold value. The two algorithms strive to minimise the variance within classes and maximise the variance between classes, thus producing classes hat are environmentally homogeneous.

There are some pros and cons associated with use of either conventional stratification or the ISOCLUSTER/k-means algorithm. One advantage of conventional stratification is the ease of providing meaningful names to each stratum (e.g., deep-rugged-hard). This is not so straight forward in strata resulting from ISOCLUSTER/k-means although extraction of ranges of values for each environmental variable for each strata could potentially provide a similar naming convention as for conventional stratification.

An advantage of ISOCLUSTER/k-means is that it seems to provide smoother and less noisy results than conventional stratification. When the environmental variation is high, conventional stratification can be very noisy with many small polygons that potentially can be assigned stations. This is also related to the level of smoothing that should be applied to any stratification process after the initial stratification has taken place. A wide array of single cells and small groups of cells creates a lot of noise that needs to be removed by smoothing. The level of smoothing applied is a somewhat subjective decision to make, although the amount of smoothing can be chosen to comply with certain predetermined criteria.

GRTS relatively often place stations very close to borders between two strata, this is particularly a problem for noisy stratifications. In general this would not be a problem, but in a MAREANO context this creates a challenge. Currently, video lines in MAREANO are 700 metres long, thus, since GRTS provides the centre point of each video line, 350 metres is the minimum distance needed from a border between strata. Such a minimum distance allows for the video line to take any random direction while staying within a stratum. So far this issue has been solved by oversampling each strata to such a degree that stations placed too close to the border can be replaced by oversampled stations complying with the minimum distance needed.

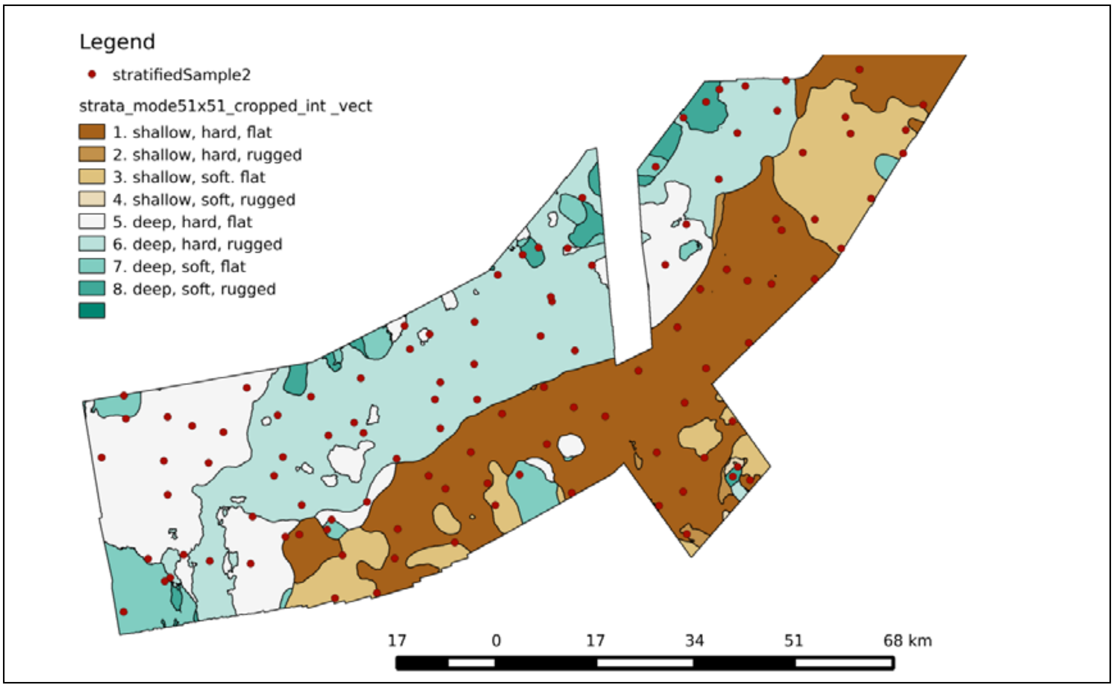


Figure 31. A simple example of a sampling design based on stratified GRTS. Bathymetry, backscatter and terrain ruggedness index have each been divided in two strata and then combined together to form a total of eight strata. Weighting has only been given to the proportion of extent for each of the eight strata. The red points represent the location of stations as GRTS would allocate them when given an input of 115 stations.

4.3.3 Optimal length for video lines

Should the length of video lines be reduced? For example 200 m long lines? Should we keep the 700 m long lines for comparability with already sampled video lines? What is more important, back-compatibility or sampling in a way that potentially can lead to better models and map products coming out of MAREANO? Today 200 m segments derived from the 700 m video line are used in biotope modelling. Will the 200 m segments be too long in areas with high spatial variability where communities change rapidly? We want as much of the variation in our samples as possible to be contained between, and as little as possible within observation units. This will yield homogeneous samples for biotope modelling as well as provide adequate ground truthing of backscatter data for geological mapping.

This will in many circumstances call for shorter video lines, and this may open an opportunity to obtain more, shorter samples. Longer video lines may in some cases be necessary, particularly for geological mapping. This means that most of the video lines could be a standard length of 200 metres, but that longer lines could also be accommodated as required by joining several 200 metres lines together.

4.3.4 Time-use analysis of video sampling

In order to address the question of whether a greater number of shorter samples would be more cost effective we have done some analysis on the time use related to recording of video lines and the transit time between stations using data from MAREANO's 2012 cruise in the Norwegian Sea. Currently, the recording of a video line includes two still stations, at the beginning and at the end. Each still station take on average between 11 and 12 minutes to complete while the video observation along the line takes on average about 38 minutes to complete (figure 32). The maneuvering of the ship in order to get to the start of the video line takes on average about 10 - 15 minutes, thus, in total, a video station of 700 m length needs about one hour to be completed. The transit time between stations has a median of 73 minutes and hence the median total time spent on a video observation and transit time is 133 minutes (figure 3). It's important to note that time use related to video sampling is very much dependent on weather conditions and other logistical constraints and challenges.

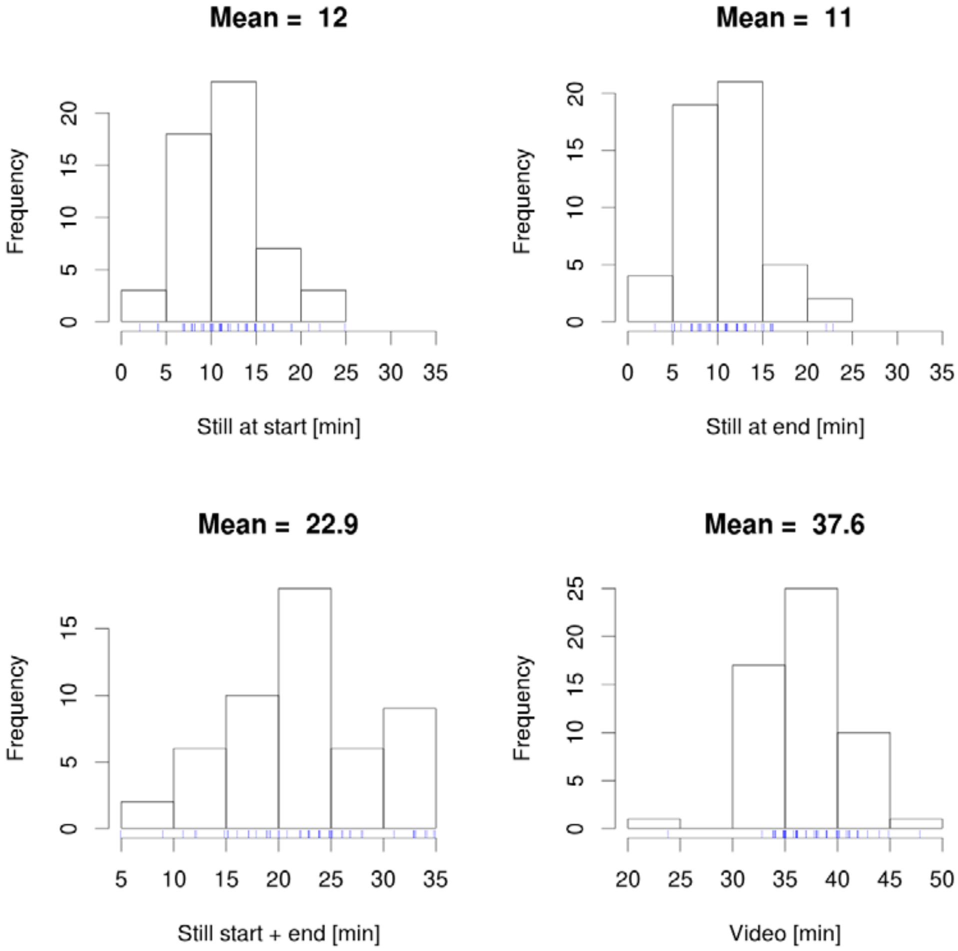


Figure 32. The time use related to video observations.

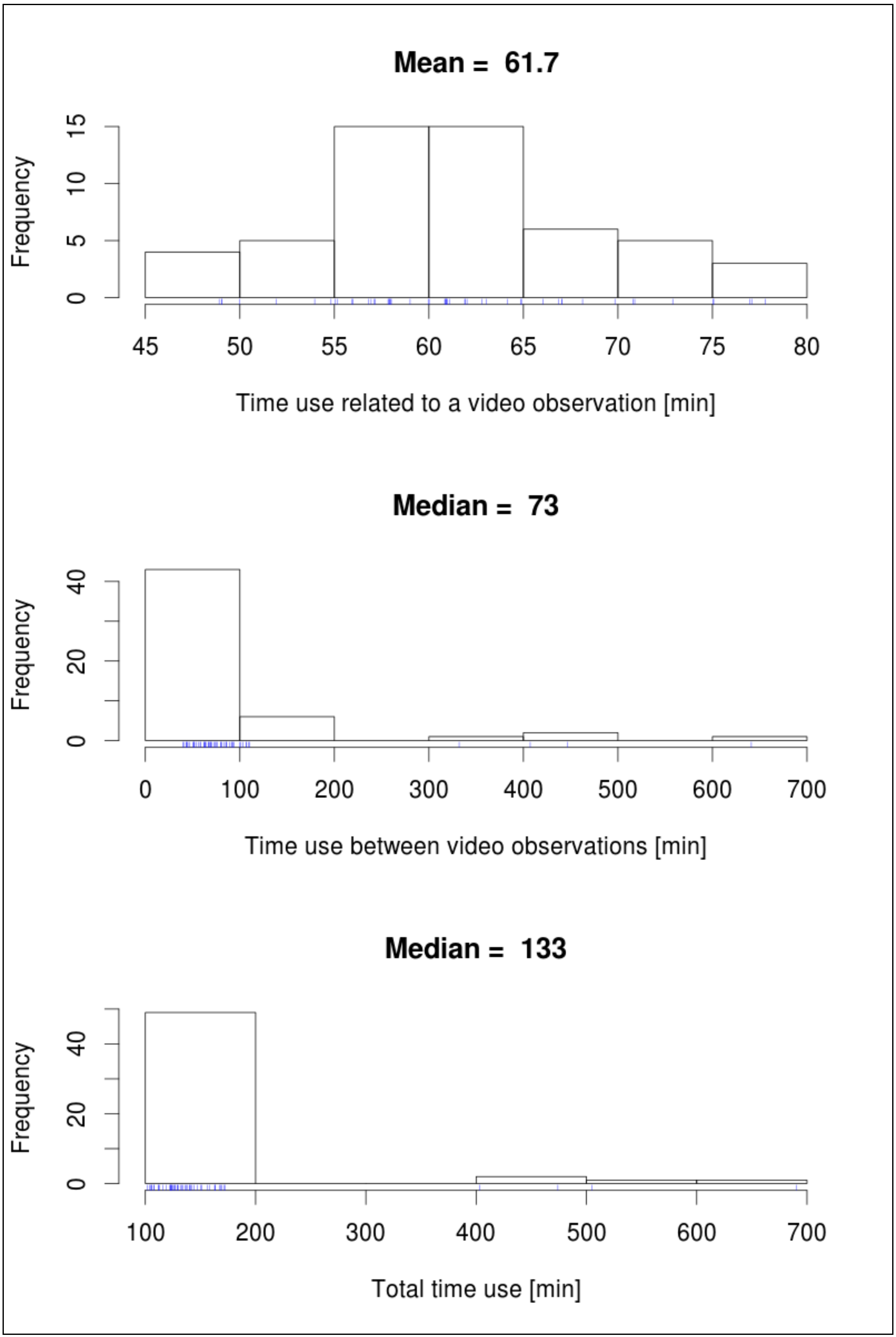


Figure 33. The time use related to a video observation (above), the transit between video stations (middle), and the total time use of video observation and transit time (bottom).

Having this information we can now analyze the time use related to different lengths of the video lines, the number of still stations per video line, and the transit time (figure 34). This analysis shows that, if the video length is reduced to 200 m and the number of still stations per video line is reduced to one, and that the transit time between stations is halved, then we can sample twice the amount of stations in the same amount of time when compared to the current strategy used in MAREANO.

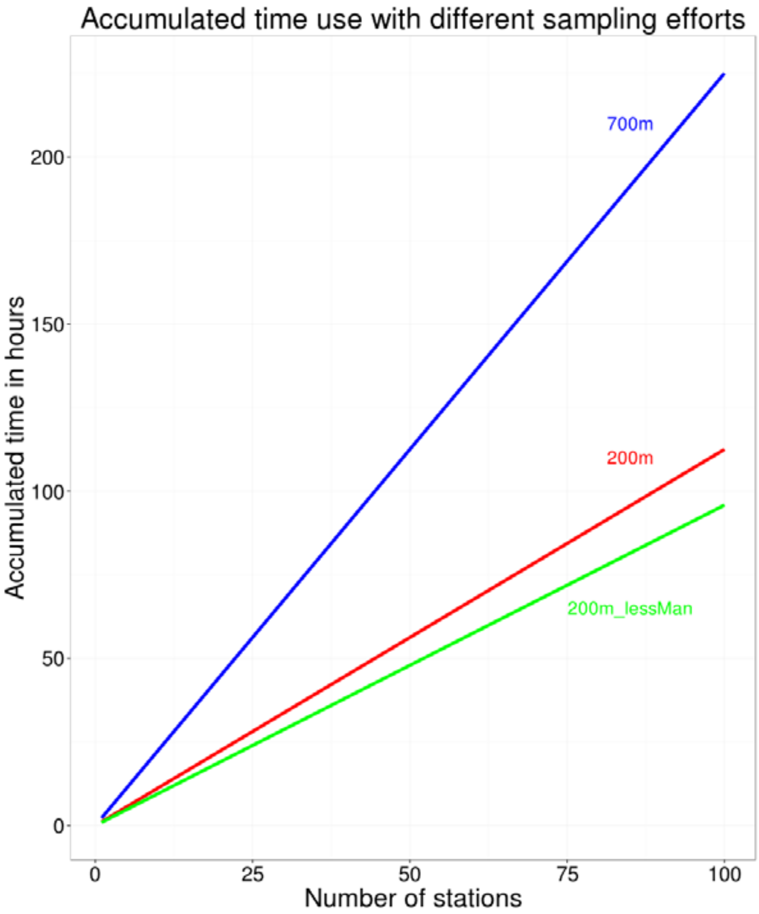


Figure 34. Accumulated time use for different sampling efforts as a function of number of stations. The blue line shows the estimated time use for the current sampling strategy in MAREANO (700 m video lines, two still stations, transit time of one hour; the red line (200 m video lines, one still stations, half transit time); green line (same as red, but with reduced maneuvering time, i.e., less precision in start point).

4.3.5 Analysis of spatial autocorrelation of megafaunal communities

In relation to optimal lengths of video observations, in addition to the time use analysis of video sampling, we have done analysis on patch sizes of megafaunal communities in Nordland VI. Apart from setting the distance between samples, appropriately matching the sampling unit size to the variables being measured (e.g. abundance of organisms) significantly enhances our ability to detect pattern from ecological observational data. O'Neill et al. (1996, 1999) suggested that the sampling unit size should be two to five times smaller than the biological patch or feature of interest. The sampling unit should be, however, large enough to contain more than one individual, but not so big that there is too much within-unit variability. When the data are not randomly distributed, a sampling unit that is too small (e.g. 1×1 in figure 35) will increase the variance and a sampling unit that is too large (e.g. 4×4) will reduce the variability. Here, the optimal sampling unit size is 2×2. Fortin and Dale (2005) recommend that a smaller sampling unit should be favoured because small units can be aggregated into larger ones without the loss of information, but the reverse is not true. It becomes apparent then that in order to determine the appropriate size of MAREANO sampling units (namely, video line length) one should first find out the average size of any biological patches. This exercise results in a sampling design which maximizes inference power from the data, and therefore yields optimal models for pattern recognition.

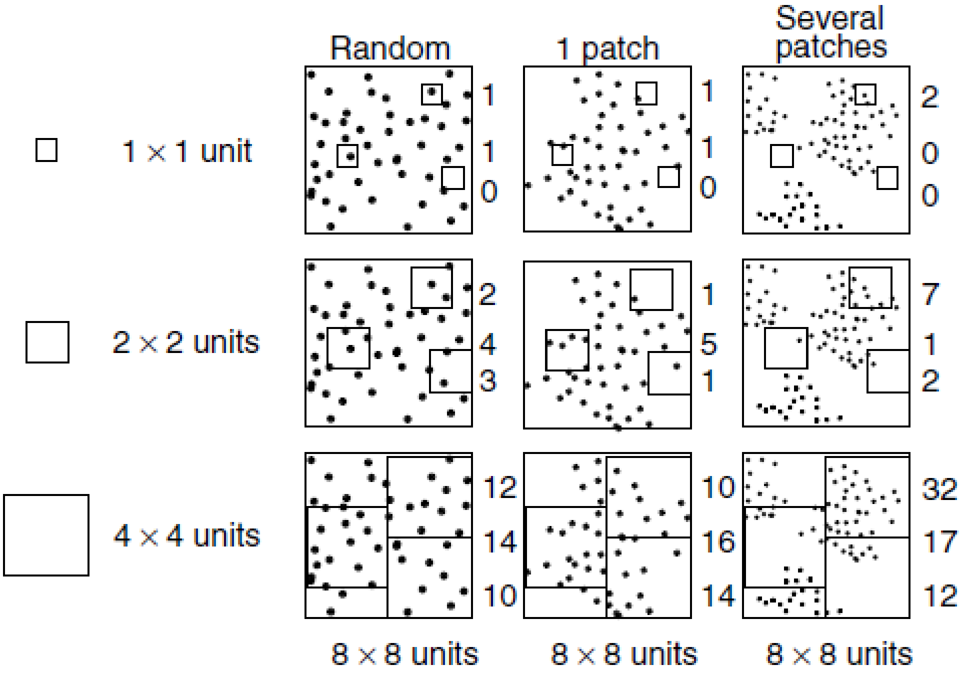


Figure 35. Effect of sampling unit size where the numbers to the right of each study area are the counts of individuals per sampling unit. Source Fortin and Dale (2005)

We re-worked the video data from Nordland VI to perform analyses aimed at answering the question of what is the average size of patches of epibenthic megafaunal communities? All video lines available from that area were split into sections that were 22 metres in length, which equals to 1 minute of footage. The reason for this size relates to the way video analysis is performed, where at least one entry needs to be made every minute. We used Mantel rank-correlograms (Oden and Sokal, 1986) to determine the relationship between biological similarity and distance, between pairs of 22 x 3.5 (the approximate width of the video line) m^2 sampling units. This method has been used before to quantify patchiness of benthic communities (e.g. Parry et al.2003, Gonzalez-Mirelis et al.2009). Through this method one can easily find the distance at which the correlation between samples (Mantel autocorrelation) becomes 0, indicating the distance at which samples become too different to be part of the same community, or patch.

When using all the data together we found that this distance was never quite reached (figure 36). Even at 700 metres distance these sampling units were positively autocorrelated; nevertheless, had samples separated by 800-1000 metres been available, it seems reasonable to assume that correlation would have become not different from 0 then. Overall scale of patchiness can reasonably be placed at 800-900 metres on average for this area. When looking at subsets of samples that had similar substrate composition patchiness was detected at distances less than 700 metres (figure37). The communities of sand-dominated habitat change more rapidly in space than others, reaching spatial independence at 500 metres.

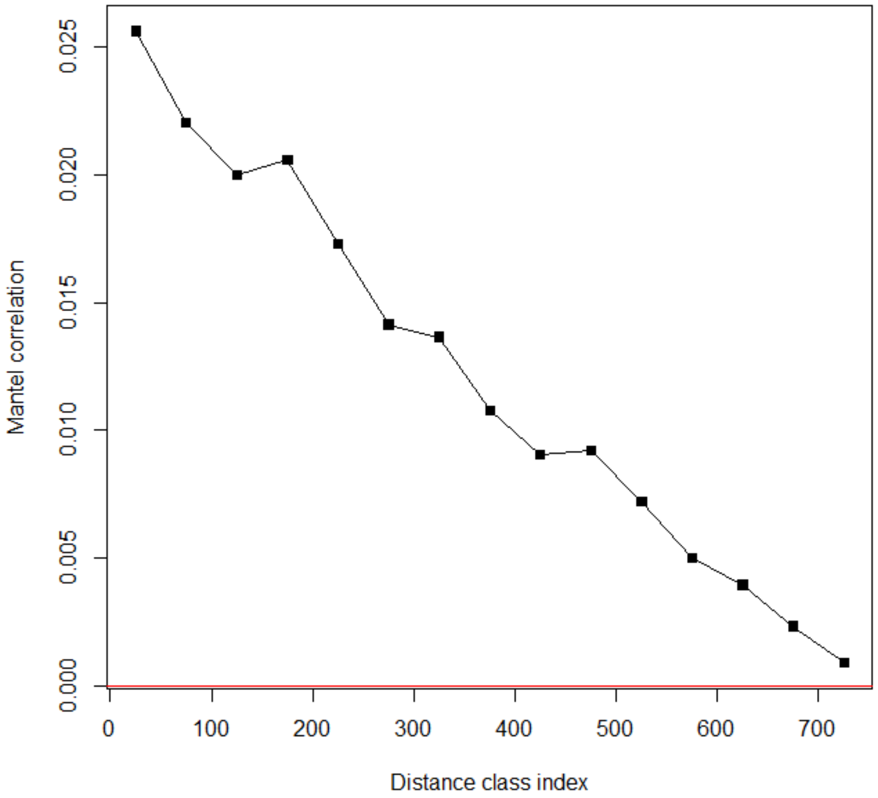


Figure 36. Mantel rank-autocorrelogram for Nordland VI using all data together. All correlation values were positive and significantly different from 0 at these scales.

In light of these findings and the relevant literature summarized above we recommend that video samples (or subsamples) whose main purpose is the development of species/community distribution models be no longer than 500 m, and preferably at least half that size (250 m). This sizes will ensure that most samples are homogeneous (i.e., from the same community type), allowing biologists to detect patterns of response of these communities to environmental variables. Note that by the same token (within-sample homogeneity) a classification of communities can be derived. This is in line with the sizes that we have used so far in MAREANO where 700m-long video liness are split into ~200m-long subsamples for modelling. The 200 metres subsample length were arrived at by means of assessing species accumulation curves (see figure 23).

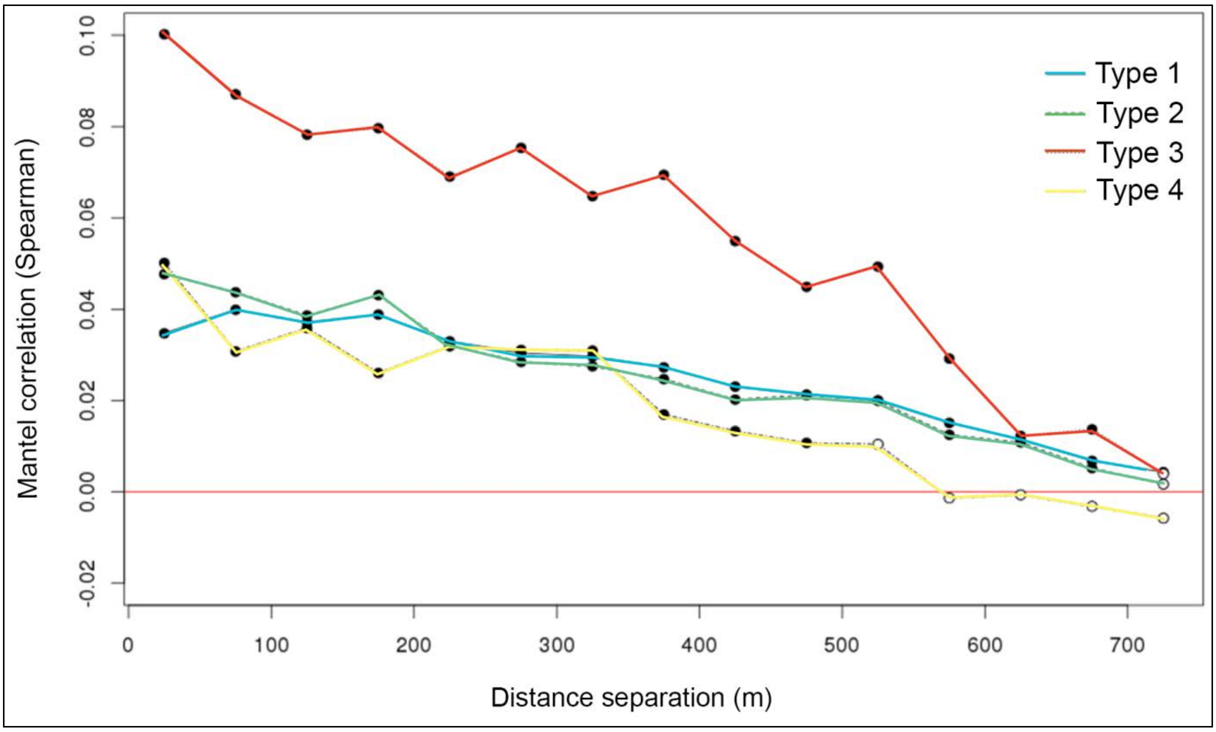


Figure 37. Mantel rank-autocorrelogram for Nordland VI. Type 1: sandy mud, Type 2: mud, Type 3: sand but mixed with pebble or cobble, Type 4: sand. Open circles represent correlation values not significantly different from 0. Solid circles represent correlation values significantly different from 0. Significance of the test was determined applying a Bonferroni correction to account for the different number of samples.

4.3.6 Discussion of video line length versus spatial autocorrelation

Analysis of spatial autocorrelation of megafaunal communities at Nordland VI indicates that about 250 metres is an adequate length of video observations. The time-use analysis for the 2012 cruise in the Norwegian Sea indicates that a doubling of the number of stations may be feasible to obtain if video lengths are reduced from 700m to 200 m, and if still stations are reduced from two to one for each video line. A doubling of stations literally means a doubling of both ground-truthing observations and video samples, which is likely to have a very positive effect on MAREANO products in terms of quality and confidence.

However, more studies conducted in different areas using different sampling strategies are necessary in order to properly assess the most appropriate length of video samples. An important contribution to this end will be to pick an area with good availability of environmental data, simulate the distribution of species, and subsequently sample the area using the different sampling strategies presented above. This will allow us to test and compare different sampling strategies and see how well they sample the simulated 'truth'.

5. END USER NEEDS

The end users of the results from MAREANO are national management, industry, research institutions and the general public. The results are also used internationally by e.g. OSPAR. This chapter is based on input from the relevant institutions and organisations, but the authors of this report are responsible for the content below.

5.1 National management

The MAREANO results have been widely used for the management plans for the Barents Sea and the Norwegian Sea. The results are also used on a regular basis for the day-to-day management, particularly in relation to the oil industry and the fishing industry.
Primary users in the national management are:

- the Norwegian Environment Agency
- the Directorate of Fisheries
- the Norwegian Petroleum Directorate

The purpose of a management plan is to provide a framework for the sustainable use of natural resources and goods derived from an ocean area, while maintaining the structure, function and productivity of the ecosystems in the area. A multi-sector basis for decision making lies at the core of the ecosystem-based approach. In the plans, several areas are identified as particularly valuable and vulnerable. For areas identified as particularly valuable and vulnerable (SVOs), special caution will be required and special considerations will apply to the assessments of standards for, and restrictions on, activities. Vulnerability is assessed with respect to specific environmental pressures such as oil pollution, climate change, fluctuation in food supply and physical damage within the plan area. Sessile animals will be vulnerable with respect to climate change, pollution and certain types of fishing operation, particularly trawling, which can have direct effects on the seabed, by damaging and disturbing benthic communities, re-suspending particles and shifting sediments.

An important part of a management plan process is also to identify knowledge gaps and to prioritize them according to a defined set of criteria. Improved seabed data (topography, sediment quality, pollutants, flora and fauna etc) achieved through systematic surveys of the sea bed within the MAREANO programme along with surveys carried out by other programmes/institutes are important with respect to this. Furthermore, benthic communities may be damaged or disturbed by trawls and other demersal gear types. Thus, better mapping of fishing activities, would also be a major step forward.

Appropriate management must include a means to evaluate the state of the ecosystem, i.e. ecological quality, at any given time. In order to do so a set of indicators, including benthic, have been identified. These, among several other indicators are being used, when managing biodiversity and productivity of an ecosystem. A follow-up system has been established for the management plans to ensure that it is up-dated as needed, e.g. in the light of new findings emerging through mapping (e.g. MAREANO), monitoring and research (e.g. MAREANO) or significant changes taking place in the environment. Permanent working groups give advice and produce reports which are used for this purpose.

Maps of biotopes and particularly vulnerable habitats (largely based on the OSPAR list of threatened and/or declining habitats in the North-East Atlantic (chapter 8) are priority 1 for the national management institutions. These maps should have a 75% confidence (see separate definition in section 3.1). The relevant scales and areas vary considerably. For certain purposes, such as e.g. management regulations of the petroleum industry with respect to the herring spawning fields at Mørebankene, a scale of 1:1.000.000 can be considered adequate. The areas involved here (herring spawning fields) are in the order of 200 - 2.000 square kilometres. At the other end of the spectrum, management regulations concerning fishing or petroleum activities in the vicinity of coral reefs may involve decisions taken at a scale of 1 km or less. This requires maps at scales of 1:100.000.

5.2 Fishing industry

It is assumed that data from the MAREANO-programme could provide the basis for more environmental friendly, resource-saving and profitable fisheries. For instance, data from MAREANO could make it possible to fish with better accuracy, making it easier to prevent damage to cold water coral reefs. It is also believed that better access to detailed seabed data could reduce bottom impact, fuel consumption and CO_2 emissions.

Fishermen will have a particular advantage of better access to maps showing bathymetry, sediment distribution, coral reefs and other vulnerable ecosystems. The data must be adapted to map plotting systems, commonly used by the fishing fleet. This has been extensively done through the a pilot project in the Astafjordene region in cooperation with

Olex (Thorsnes et al. 2013). There is work in progress to implement this for the rest of the Norwegian shelf.

In general, the fishermen need the best data sets available. Data sets containing bathymetry and backscatter (and to some degree sediment grain size) should be as detailed as possible (1-5 metres grid). This provided that the amount of data can be handled in normal chart plotter systems, like the Olex system (www.olex.no). Data sets like habitats, coral reefs (including buffer zones), breeding grounds and other vulnerable areas could have less resolution (5-30 metres grid).

5.3 Energy industry

The petroleum industry has a wide range of applications where results from MAREANO have a potential use. For engineering purposes (subsea installations, platforms, pipelines etc.), very detailed bathymetry data with a resolution of 1 x 1 metre is requested. High resolution sediment grain size maps (e.g. 1:10 000) are also requested. Priority map products with regard to environment includes the vulnerable habitats likely to be subject to management regulations. Point observations from physical sampling stations includes the species represented in OSPAR list of threatened and/or declining habitats in the North-East Atlantic (chapter 8), havmiljo.no, and the Norwegian Red and Black lists (Lindgaard and Henriksen, 2011)). The offshore wind mill industry can be expected to have similar needs, both for engineering and environmental purposes.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

1. The MAREANO survey and sampling strategy has continuously developed since the programme started. Sampling effort for visual stations has varied between less than 5 stations to more than 15 stations pr. 1000 square kilometres. MAREANO has so far produced maps at scales of 1:100.000 - 1:250.000.
2. User needs vary in spatial scale from grids on metre scale (scale 1.10.000) to maps at scales of 1:1.000.000. Confidence assessments for maps used by management are very important.
3. The current sampling strategy for MAREANO is based on expert judgement. An evaluation of several strategies concludes that a combination of targeted stations and a more objective and automated strategy should be used, for example "Generalised Random-Tesselation Stratified".
4. The study shows that the current approach with 700 metre long video lines (divided into c. 200 metre long segments for biotope mapping) should be further discussed, and alternatives should be considered.
5. A confidence assessment for a sediment grain-size map from an area with medium to high environmental variability (Tromsøflaket) showed clear differences when data from 5 and 13 stations pr. 1000 square kilometres were used. Confidence values rose from 53 to 73% when a greater number of stations were used. The assessment also highlighted methodological challenges when testing confidence within small mapping areas.
6. Confidence assessment of the biotope maps is a complex process that is an integral part of modelling and prediction. Results to date indicate good confidence (c. 75%) in areas such as Troms III-Tromsøflaket-Eggakanten. It has not been possible to test confidence using different numbers of stations within the scope of this study.
7. Confidence assessment of the vulnerable habitat maps yielded results varying from "fair" to "excellent", with most of the habitats falling into "good".
8. Manual interpretation of sediment maps in MAREANO is time-consuming and a non-repeatable process.

6.2 Recommendations

1. Establish a dedicated effort in collaboration with NiN to make full use of MAREANO biological data to produce generalised species list for marine nature types. This work should be prioritised.
2. Use EVI to guide sampling effort.
3. Investigate the effect of increasing heterogeneity of the environment and the size of the study area on the number of classes as defined by the ISOCLUSTER/k-means algorithms.
4. Initiate a separate study looking at the spatial distribution of physical stations, for geology, biology and chemistry. This work should be prioritised.
5. Evaluate semi-automated classification/modelling approaches to guide sediment map production. This work should be prioritised.
6. Adopt tools for fully documenting the production of modelled maps, including metadata and provenance of the data product. This may also include automation of parts of the modelling workflow and routine tasks.
7. Further development of an objective (semi) automated sampling strategy, including simulation studies, that meets the needs of biological and geological mapping plus management objectives.
8. Conduct simulation studies in order to test the efficiency and appropriateness of different sampling strategies as well as to address the optimal length of video observations.

7. REFERENCES

Albert, C.H, Yoccoz, N.G., Edwards, T.C., Graham, C.H., Zimmermann, N.E. and Thuiller, W. 2010. Sampling in ecology and evolution - bridging the gap between theory and practice. Ecography 33, 1028 - 1037.

Albretsen, J., Sperrevik, A.K., Staalstrøm, A., Sandvik, A.D., Vikebø, F. and Asplin, L. 2011. NorKyst-800 Report No. 1 User Manual and technical descriptions. Fisken og Havet 2/2011.

Altman, D.G. 1991. Practical Statistics for Medical Research. Chapman & Hall, London, UK. 611 pp.

Ball, G. H., and D. J. Hall. 1965. A Novel Method of Data Analysis and Pattern Classification. Menlo Park, California: Stanford Research Institute.

Breiman, L. 2001. Random forests. Machine Learning 45, 5-32.

Brunsell, N.A. and Young, C.B. 2008. Land surface response to precipitation events using MODIS and NEXRAD data. International Journal of Remote Sensing, Vol. 29, Iss. 7, 1965-1982.

Buhl-Mortensen, L., Buhl-Mortensen, P., Dolan, M. and Gonzales-Mirelis, G. 2015. Habitat mapping as a tool for conservation and sustainable use of marine resources: Some perspectives from the MAREANO Programme, Norway. Journal of Sea Research, Vol. 100, 46-61.

Bullimore, R. D.; Foster, N. L. & Howell, K. L. 2013. ‘Coral-characterized benthic assemblages of the deep Northeast Atlantic: defining “Coral Gardens” to support future habitat mapping efforts’. *ICES Journal of Marine Science* **70**, 511–522.

Bøe, R., Dolan, M., Thorsnes, T., Lepland, A., Olsen, H., Totland, O. and Elvenes, S. 2010: Standard for geological seabed mapping offshore. NGU report 2010.033, 15 pages.

CEN 2012. Water quality – Visual seabed surveys using remotely operated and/or towed observation gear for collection of environmental data. Ref. No. 1620:2 012: E, 22pp.

Christiansen S. 2010. Background Document for Coral gardens. OSPAR commission. ISBN 978-1-907390-27-2 Publication Number: 486/2010, 39 pp.

Clements, A.J., Strong, J.A., Flanagan, C. & Service, M. 2010. Objective stratification and sampling-effort allocation of ground-truthing in benthic-mapping surveys. ICES Journal of Marine Science 67, 628 - 637.

Czúcz, B.; Gálhidy, L. & Mátyás, C. 2011, 'Present and forecasted xeric climatic limits of beech and sessile oak distribution at low altitudes in Central Europe', *Annals of Forest Science* **68**, 99-108.

Dolan, M.F.J., Buhl-Mortensen, P., Thorsnes, T., Buhl-Mortensen, L., Bellec, V.K. & Bøe, R. 2009: Developing seabed nature-type maps offshore Norway: initial results from the MAREANO programme. Norwegian Journal of Geology. 89, 17-28.

Dormann, C., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. and Wilson, R.2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609 - 628.

Elvenes, S., Buhl-Mortensen, P. and Dolan, M. F. J. 2012. Evaluation of alternative bathymetry data sources for MAREANO: A comparison of Olex bathymetry and multibeam data for substrate and biotope mapping. NGU report 2012.030, 54 pp.

Elvenes, S., M. F. J. Dolan, P. Buhl-Mortensen, and V. K. Bellec, 2014, An evaluation of compiled single-beam bathymetry data as a basis for regional sediment and biotope mapping: Ices Journal of Marine Science, 71, 867-881.

Fielding, A. H. & Bell, J. F. 1997. 'A review of methods for the assessment of prediction errors in conservation presence/absence models'. *Environmental conservation* **24**, 38-49.

Folk, R.L., 1954. The distinction between grain size and mineral composition in sedimentary rock nomenclature. Journal of Geology, 62 (4), 344-359

Fonseca, L., Brown, C., Calder, B., Mayer, L. and Rzhanov, Y. 2009. Angular range analysis of acoustic themes from Stanton Banks Ireland: A link between visual interpretation and multibeam echosounder angular signatures. Applied Acoustics, 70, 1298-1304.

Fortin, M.-J. & Dale, M. 2005, Spatial Analysis: A guide for ecologists, Cambridge University Press.

Gonzalez-Mirelis, G., Bergström, P., Lundälv, T., Jonsson, L. & Lindegarth, M. 2009, Mapping the benthos: Spatial patterns of seabed-dwelling megafauna in a Swedish Fjord, as derived from opportunistic video data. Journal of Marine Biodiversity 39, 291-302.

Gonzalez-Mirelis, G. & Buhl-Mortensen, P. 2015. Modelling benthic habitats and biotopes off the coast of Norway to support spatial management. Ecological Informatics. Available online 25 June 2015

Gonzalez-Mirelis, G. & Lindegarth, M. 2012. 'Predicting the distribution of out-of-reach biotopes with decision trees in a Swedish marine protected area', *Ecological Applications* **22**, 2248-2264.

Gordon, D.C., Kenchington, E.L.R., Gilkinson, K.D., McKeown, D.L., Steeves, G., Chin-Yee, M., Vass, W.P., Bentham, K., Boudreau, P.R. 2000 Canadian imaging and sampling technology for studying marine benthic habitat and biological communities. CM 2000/T:07. ICES 2000 Annual Science Conference 27-30 September 2000, Bruges, Belgium.

Grafström, A. 2014. BalancedSampling: Balanced and spatially balanced sampling. R package version 1.4. http://www.antongrafstrom.se/balancedsampling

Halvorsen, R., Bryn, A., Erikstad, L. & Lindgaard, A. 2015. Natur i Norge - NiN. Versjon 2.0.0. Artsdatabanken, Trondheim (http://www.artsdatabanken.no/nin).

Hill, M.O. 1973. Reciprocal averaging, an eigenvector method of ordination. Journal of Ecology, 61, 237–249.

Hirzel, A. and Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling 157, 331-341.

Hothorn, T., Hornik, K. & Zeileis, A. 2006. Unbiased recursive partitioning: A conditional inference framework, Journal of Computational and Graphical Statistics **15**, 651-674.

ICES 2007. Report of the Working Group on Deep-Water Ecology (WGDEC). 26–28 February 2007. 2007. ICES Advisory Committee on Ecosystems. ICES Document CM 2007/ACE: 01.

Kincaid, T. and Olsen, A.R. 2015. spsurvey: Spatial Survey Design and Analysis. R package version 3.0. URL: http://www.epa.gov/nheerl/arm/.

Kruskal J.B. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1-27.

Kruskal J.B. 1964b. Nonmetric multidimensional scaling: a numerical method. Psychometrika 29:115-129.

LeBas, T., Scarth., A. & Bunting, P. 2015. A new Object Based Image Analysis toolbar for ArcGIS 10.x designed for combined multibeam bathymetry and backscatter interpretation. GEOHAB2015, Salvador, Brazil, May 2015.

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology 74, 1659 - 1673.

Lindgaard, A. & Henriksen, S. (eds.) 2011. *The 2011 Norwegian Red List for Ecosystems and Habitat Types*. Norwegian Biodiversity Information Centre, Trondheim.

Lobo, J.M., Jimenez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17, 145-151.

Lucieer, V., Hill, N. A., Barrett, N. S., & Nichol, S. 2013. Do marine substrates ‘look’ and ‘sound’ the same? Supervised classification of multibeam acoustic data using autonomous underwater vehicle images. Estuarine Coastal and Shelf Science, 117, 94–106.

Lucieer, V. & Lamarche, G. (2011). Unsupervised fuzzy classification and object-based image analysis of multibeam data to map deep water substrates, Cook Strait, New Zealand. Continental Shelf Research, 31(11), 1236-1247.

Lurton, X. & Lamarche, G. (eds.) 2015. Backscatter measurements by seafloor- mapping sonars. Guidelines and Recommendations. 200p.
http://geohab.org/wp- content/uploads/2014/05/BSWG- REPORT- MAY2015.pdf

MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, eds L. M. Le Cam & J. Neyman, 1, pp. 281–297. Berkeley, CA: University of California Press.

McCune, B., & Mefford, M.J. 2006.PC-ORD: Multivariate Analysis of Ecological Data. MJM Software, Gleneden Beach, Oregon, USA.

Mortensen, P.B., & Buhl-Mortensen, L. 2005. Coral habitats in The Gully, a submarine canyon off Atlantic Canada. Pp 247-277 in Freiwald, A., Roberts, J.M. (Eds.): Cold-water Corals and *Ecosystems*. Springer-Verlag Berlin Heidelberg, 1244pp

Mortensen, P. B., Dolan, M., Buhl-Mortensen, L. 2009. Prediction of benthic biotopes on a Norwegian offshore bank using a combination of multivariate analysis and GIS classification. – ICES Journal of Marine Science, 66, 2026–2032.

Murray, S.N., Ambrose, R.F., Dethier, M.N. 2002. Methods for Performing Monitoring, Impact, and Ecological Studies on Rocky shores. MMS OCS Study 2001-070. Coastal Research Center, Marine Science Institute, University of California, Santa Barbara, California. MMS Cooperative Agreement Number 14-35-0001-30761. 217 pages.

NHS 2010: Technical Specifications Seabed Mapping. MAREANO programme. http://www.mareano.no/__data/page/9162/Technical_Specifications.pdf.

Oden, N. L. & Sokal, R. R. 1986. Directional autocorrelation: an extension of spatial correlograms to two dimensions. Systematic Zoology 35, 608-617.

O’Neill, R. V., Hunsaker, C. T., Timmins, S. P., Jackson, B. L., Jones, K. B., Riitters, K. H. & Wickham, J. D. 1996. Scale problems in reporting landscape pattern at the regional scale. Landscape Ecology 11, 169-80.

O’Neill, R. V., Riitters, K. H., Wickham, J. D. & Jones, K. B. 1999. Landscape pattern metrics and regional assessment. Ecosystem Health 5, 225-33.

OSPAR 2008. OSPAR List of Threatened and/or Declining Species and Habitats. OSPAR Agreement 2008-06, pp. 1–4.

Parry, D., Kendall, M., Pilgrim, D. & Jones, M. 2003. Identification of patch structure within marine benthic landscapes using a remotely operated vehicle. Journal of Experimental Marine Biology and Ecology 285-286, 497-511.

Phillips, S. J., Dudik, M., and Schapire, R. E. 2004. A Maximum Entropy Approach to Species Distribution Modelling. In Proceedings of the 21st International Conference on Machine Learning, pp. 655–662. Banff, Alberta, Canada.

Richards, J. A. 1986. Remote Sensing Digital Image Analysis: An Introduction. Berlin: Springer–Verlag.

Robertson, B.L., Brown, J.A., McDonald, T., Jaksons, P. 2013. BAS: Balanced Acceptance Sampling of Natural Resources. Biometrics 69: 776 - 784.

Rogers, A. & Gianni, M. 2009. Implementation of UN Resolution 61/105 in the management of deep-sea fisheries on the High Seas. Provisional Report: North Atlantic status and recommendations. Technical report, International Programme on the State of the Ocean.

van Son, T.C. & Halvorsen, R. 2014. Multiple Parallel Ordinations: the importance of choice of ordination method and weighting of species abundance data. Sommerfeltia 37: 1 - 37.

van Son, T.C., Dolan, M.F.J., Gonzales-Mirelis, G., Thorsnes, T., Bjarnadóttir, L.R. & Buhl-Mortensen, P. 2015. Environmental Variability Index (EVI) - a MAREANO methods study for guidance of sampling effort. NGU report 2015.027, 54 pp.

Stephens, D. & Diesing, M. 2014. A Comparison of Supervised Classification Methods for the Prediction of Substrate Type Using Multibeam Acoustic and Legacy Grain-Size Data. PLoS ONE: 9(4):e93950.

Stevens, D.L. & Olsen, A.R. 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99, 262 - 278.

Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T. & Zeileis, A. 2008, 'Conditional variable importance for random forests', *Bioinformatics* 9: 307.

ter Braak, C. J. F., and Smilauer, P. 2002. CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (version 4.5). Microcomputer Power, Ithaca, NY, USA.

Thorsnes, T., Sandberg, J.H., Longva, O., Røyland, G., Jakobsen, P.-A. & Hestvik, O.B. 2013. Nye marine grunnkart i fiskeri-og havbruksnæringen - Fase 2. NGU-rapport 2013.037, 21 pp.

Zimmermann, N. & Kienast, F. 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. *Journal of Vegetation Science* **10**, 469-482.

8. Appendix - OSPAR list habitats

DESCRIPTION	OSPAR Regions where the habitat occurs	OSPAR Regions where such habitats are under threat and/or in decline
HABITATS		
Carbonate mounds	**I, V**	**V**[1]
Coral Gardens	**I, II, III, IV, V**	**All where they occur**
Cymodocea meadows	**IV**	**All where they occur**
Deep-sea sponge aggregations	**I, III, IV, V**	**All where they occur**
Intertidal *Mytilus edulis* beds on mixed and sandy sediments	II, III	**All where they occur**
Intertidal mudflats	**I, II, III, IV**	**All where they occur**
Littoral chalk communities	**II**	**All where they occur**
Lophelia pertusa reefs	**All**	**All where they occur**
Maerl beds	All	III
Modiolus modiolus beds	All	**All where they occur**
Oceanic ridges with hydrothermal vents/fields	**I, V**	**V**
Ostrea edulis beds	II, III, IV	All where they occur
Sabellaria spinulosa reefs	**All**	**II, III**
Seamounts	**I, IV, V**	**All where they occur**
Sea-pen and burrowing megafauna communities	I, II, III, IV	**II, III**
Zostera beds	**I, II, III, IV**	**All where they occur**

To be confirmed in the light of further survey work being undertaken by Ireland.

GEOLOGICAL
SURVEY OF
NORWAY
- NGU -

Geological Survey of Norway
PO Box 6315, Sluppen
N-7491 Trondheim, Norway

Visitor address
Leiv Eirikssons vei 39
7040 Trondheim

Tel (+ 47) 73 90 40 00
E-mail ngu@ngu.no
Web www.ngu.no/en-gb/