NGU REPORT
2015.027

Environmental Variability Index (EVI) -
a MAREANO methods study for guidance
of sampling effort

GEOLOGY FOR SOCIETY

SINCE 1858

GEOLOGICAL
SURVEY OF
NORWAY
- NGU -

NORGES GEOLOGISKE UNDERSØKELSE
· NGU ·

REPORT

Report no.: 2015.027	**ISSN: 0800-3416 (print)** **ISSN: 2387-3515 (online)**	**Grading:** Open

Title:
Environmental Variability Index (EVI) - a MAREANO methods study for guidance of sampling effort

Authors: Thijs Christiaan van Son, Margaret Dolan, Genoveva Gonzales-Mirelis, Terje Thorsnes, Lilja Rún Bjarnadóttir, Pål Buhl-Mortensen		**Client:** MAREANO	
County: Norway		**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:** 54 **Price (NOK):** 215,- **Map enclosures:**	
Fieldwork carried out:	**Date of report:** 21.04.2015	**Project no.:** 311720	**Person responsible:** Reidulv Bøe R. Bøe

Summary: MAREANO is an interdisciplinary programme mapping Norwegian offshore bathymetry, geology, biology and geochemistry. Following bathymetric mapping by the Norwegian Hydrographic Service, biological and geological sampling are undertaken by the Institute of Marine Research (IMR) and the Geological Survey of Norway (NGU) to obtain visual (video) and physical sample data that are used in conjunction with bathymetric and backscatter data for the generation of thematic maps. Currently, MAREANO samples 10 video lines per 1000 km^2 in the Norwegian Sea and 5 video lines per 1000 km^2 in the eastern part of the Norwegian Barents Sea. As the environmental and spatial variability of the Norwegian seabed has become better known, the need for a more flexible strategy has become apparent. In order to optimise MAREANO sampling effort there is a need for a sampling strategy that takes into account the level of environmental and spatial variability, i.e., increasing sampling effort in heterogeneous areas and decreasing the effort in homogeneous areas. In a response to this need for flexibility the development of an Environmental Variability Index (EVI) was initiated by MAREANO. The purpose of the EVI is to provide a basis for scaling the sampling effort to environmental and spatial variability. The EVI can be used to guide sampling effort both for long-term strategic as well as detailed cruise planning. It is designed to account for the mean environmental dispersion estimated from the coefficient of variation of the environmental layers. The EVI appears to pick up important environmental and spatial variability related to the physical environment, however, in practical terms it needs to correlate well with spatial turnover of biological and geological diversity. To test this correlation we applied linear models using either beta diversity (species turnover and nestedness from video line observations) or beta geodiversity (here measured as sediment grain size turnover and nestedness) as response variables and the EVI as the predictor. We found that both were positively correlated with the EVI. These results are promising and highlight the potential of EVI to properly scale sampling effort. Further testing in new areas as well as development of better diversity measures (especially related to beta geodiversity) is needed to prove the general validity of the EVI.

Keywords: Marine Geology	MAREANO	Biodiversity
Geological diversity	Sediment grain size	Sampling strategy
Optimal allocation	Oceanography	Modelling

CONTENTS

1. SUMMARY

MAREANO is an interdisciplinary programme mapping Norwegian offshore bathymetry, geology, biology and geochemistry. Following bathymetric mapping by the Norwegian Hydrographic Service, biological and geological sampling are undertaken by the Institute of Marine Research (IMR) and the Geological Survey of Norway (NGU) to obtain visual (video) and physical sample data that are used in conjunction with bathymetric and backscatter data for the generation of thematic maps.

Currently, MAREANO samples 10 video lines per 1000 km^2 in the Norwegian Sea and 5 video lines per 1000 km^2 in the eastern part of the Norwegian Barents Sea. The reduced sampling effort in the Barents Sea is based on the assumption of less variability, judging from the regional bathymetry. As the environmental and spatial variability of the Norwegian seabed has become better known, the need for a more flexible strategy has become apparent. In order to optimise MAREANO sampling effort there is a need for a sampling strategy that takes into account the level of environmental and spatial variability, i.e., increasing sampling effort in heterogeneous areas and decreasing the effort in homogeneous areas.

In a response to this need for flexibility the development of an Environmental Variability Index (EVI) was initiated by MAREANO. The purpose of the EVI is to provide a basis for scaling the sampling effort to environmental and spatial variability. The EVI can be used to guide sampling effort both for long-term strategic as well as detailed cruise planning. It is designed to account for the mean environmental dispersion estimated from the coefficient of variation of the environmental layers.

The EVI appears to pick up important environmental and spatial variability related to the physical environment, however, in practical terms it needs to correlate well with spatial turnover of biological and geological diversity. To test this correlation we applied linear models using either beta diversity (species turnover and nestedness from video line observations) or beta geodiversity (here measured as sediment grain size turnover and nestedness) as response variables and the EVI as the predictor. We found that both were positively correlated with the EVI. These results are promising and highlight the potential of EVI to properly scale sampling effort. Further testing in new areas as well as development of better diversity measures (especially related to beta geodiversity) is needed to prove the general validity of the EVI.

2. NORSK SAMMENDRAG (SUMMARY IN NORWEGIAN)

MAREANO er et tverrfaglig program som kartlegger batymetri, geologi, biologi og miljøtilstand på havbunnen i norske farvann. Etter at detaljerte dybdemålinger er utført av

Kartverket, gjennomføres biologisk og geologisk prøvetaking av Havforskningsinstituttet (HI) og Norges geologiske undersøkelse (NGU). HI og NGU gjør visuelle observasjoner (video) og samler inn fysiske prøver som sammen med dybdedata og reflektivitetsdata innsamlet med multistråleekkolodd brukes til å generere kart.

I dag samler MAREANO inn 10 video linjer per 1000 km^2 i Norskehavet og 5 video linjer per 1000 km^2 i den østlige norske delen av Barentshavet. Den reduserte prøvetakingsinnsatsen i Barentshavet baserer seg på antagelser om mindre miljøvariasjon, en antagelse som igjen baserer seg på inspeksjon av dybdedata i det aktuelle området. Etter hvert som den romlige variabiliteten i miljøet har blitt bedre kjent har det også vokst fram et ønske og et behov for en mer fleksibel prøvetakingsstrategi. For å optimalisere MAREANO sin prøvetakingsinnsats er det et behov for en prøvetakingsstrategi som tar hensyn til graden av variabilitet i rom og miljø. Det er derfor ønskelig å øke prøvetakingsinnsatsen i heterogene områder og redusere innsatsen tilsvarende i homogene områder.

Som en respons til dette behovet har MAREANO satt i gang utviklingen av en miljøvariabilitetsindeks (Environmental Variability Index, EVI). Miljøvariabilitetsindeksen har til hensikt å skalere prøvetakingsinnsatsen basert på romlig og miljømessig variabilitet. Målet er at indeksen kan brukes til å gi råd om prøvetakingsinnsatsen både i et langtidsperspektiv samt for detaljert toktplanlegging. Indeksen er utformet på en slik måte at den tar hensyn til den gjennomsnittlige miljøvariabiliteten estimert fra variasjonskoeffisienten til miljølagene.

Miljøvariabilitetsindeksen ser ut til å beskrive viktig variabilitet i rom og miljø, men for å ha nytteverdi, må den korrelere godt med romlige endringer i biologisk og geologisk diversitet. Vi testet denne korrelasjonen ved å bruke lineære modeller som brukte enten betadiversitet (endringer i artssammensetningen fra videolinje-observasjoner) eller betageodiversitet (endringer i dekning av sedimentkornstørrelse) som responsvariabler og miljøvariabilitetsindeksen som prediktor. Resultatene viser at både betadiversitet og betageodiversitet var positivt korrelert med indeksen. Dette er lovende resultater og understreker potensialet til miljøvariabilitetsindeksen til å skalere prøvetakingsinnsatsen på en tilfredsstillende måte. Det er nødvendig med mer testing i nye områder samt utvikling av bedre mål på diversitet (dette gjelder særlig for betageodiversitet) for å kunne si noe om den generelle gyldigheten av miljøvariabilitetsindeksen.

3. INTRODUCTION

Norway's national offshore seabed mapping programme MAREANO (www.mareano.no) is an interdisciplinary initiative mapping bathymetry, geology, biology and environmental geochemistry. Three major partners contribute to the mapping effort in MAREANO. Firstly the Norwegian Hydrographic Service (NHS) performs baseline full-coverage mapping of the bathymetry, acquiring co-registered acoustic backscatter data where multibeam echosounder surveys are conducted (and water column data since 2010). In some areas, high quality single beam data are used (Elvenes et al. 2014) together with limited coverage of multibeam data.

Building upon these baseline acoustic data, the Institute of Marine Research (IMR) and the Geological Survey of Norway (NGU) collaborate to acquire video data and physical samples of the seabed. These data are used to deliver information on the geology and biology of the seabed and assess its environmental status. Results are presented as thematic maps on www.mareano.no. By the end of 2014, MAREANO had completed multidisciplinary data acquisition over 157 585 km^2 of the Barents and Norwegian Seas, in areas prioritised by government and management bodies.

3.1 Background

This report is primarily related to the sampling (video surveys) conducted by IMR and NGU. The first of these sampling cruises was conducted in 2006 at Tromsøflaket in the southwest Barents Sea. A total of 58 video lines and 24 physical sampling stations were acquired on eastern Tromsøflaket within an area of 2057 km^2. This number of sample stations proved to be more than adequate for documenting biology and geology of the seabed in this area for the purposes of regional mapping. Over the next couple of years, sampling continued in western Tromsøflaket and areas offshore Lofoten and Vesterålen with varied sampling effort, mainly related to meteorological and technical issues. Hereinafter we use the term ‘sample effort’ to refer to the number of sampling stations (video lines unless specified) per 1000 km^2, and the term 'total sampling effort' to refer to the total number of sampling stations collected within a study area of a given size.

By 2010 MAREANO had settled on a pragmatic, and budgetary-guided sampling effort standard, based on scientific experience, of 10 sampling stations per 1000 km^2. Two of these stations are ‘physical stations’ where all sampling gears are used and both species and sediment are retrieved from the seabed for laboratory analyses, in addition to video. The remainder are video stations only. For areas where full-coverage detailed bathymetry and backscatter, and preferably bottom-penetrating data (e.g. TOPAS) are available, this level of sampling effort was considered by expert judgement to be sufficient for the production of seabed sediment maps in scale 1:100 000. Several years later this *de facto* standard of 10 stations per 1000 km^2 is still in operation (except in the eastern part of the Norwegian Barents Sea), but as MAREANO maps areas of varying topographic, geological and biological diversity, it has become increasingly apparent that a flexible sampling strategy would be more optimal.

The environmental variability of the seabed within Norwegian Sea areas has become far better understood thanks to new data. In order to optimise the effectiveness of MAREANO sampling there is need for a sampling strategy that can adequately map both heterogeneous and more homogenous areas. This is difficult to achieve with a fixed sampling effort. Expert-based recognition of this fact was reflected by a 50% reduction in the sampling effort in the eastern part of the Norwegian Barents Sea, which was deemed to have low environmental variability relative to other mapping areas. This was the first step towards an adaptive sampling strategy for MAREANO, however, it also raised the demand for a new approach based on the best available scientific knowledge.

In 2013, MAREANO initiated a methods review study to provide an internal assessment of a process aimed at evaluating the existing methodology, data acquisition methods and testing out new methods. Among the main aims of this methods review is the exploration of options for a more optimised method for guiding sampling effort in MAREANO. It is intended that development of an index for this variability will provide a quantitative formula for sampling effort that can be applied to any given mapping area. In addition to providing a more cost-effective strategy for mapping large areas, a move toward a sampling strategy based on sampling effort guided by environmental variability, has the potential to improve the quality of the deliverables from MAREANO. Increased sampling effort in environmentally heterogeneous areas can result in better biological predictions based on statistical modelling as well as improved geological maps. At the same time, decreased sampling effort in environmentally homogeneous areas is likely to have no significant negative impacts on the quality of neither geological nor biological deliverables.

This report is part of a series of reports related to the ongoing MAREANO methods review and specifically deals with sampling effort scaled to environmental and spatial variability in a certain area. Note that this work has been conducted in parallel with investigations into potential revisions of other aspects of MAREANO sampling strategy including location, number, and size of the sampling unit. These will be documented in other reports.

Here we develop an Environmental Variability Index (EVI) as a means to scale sampling effort to environmental and spatial variability. Depending on which data are used as input to the EVI, the concept described here has the potential to contribute to long-term strategic planning, and/or the first step in detailed planning of specific survey areas.

There are three main aims of the EVI:

1) To quantify the environmental and spatial variability found within any area and make it comparable to the variability found in other areas.
2) To recommend, by linear rescaling, sampling effort so that a given study area with high environmental variability (expressed as high EVI values) should see a higher sampling effort compared to a different study area with lower environmental variability (expressed as lower EVI values)
3) To achieve a better allocation of resources, both in terms of direct costs, ship time and man hours.

Marine surveys are particularly prone to sub-optimal survey designs due to the need to match the high cost of offshore surveys with multiple survey objectives, and/or re-use of data over time for purposes other than the original. Nevertheless there exist other attempts to optimise sampling effort for various types of marine surveys. For example Clements et al. (2010) applied an optimal allocation strategy to guide sampling effort when ground-truthing benthic habitats, and Harbitz et al. (1998) used a similar optimal allocation strategy to guide effort in trawl surveys. Both these studies used coefficient of variation (CV) as the measure of dispersion of environmental variables to guide and optimally allocate sampling effort, and CV is also fundamental to the EVI developed here.

Whilst there are some similarities, these other studies are generally tailored to far more specific sampling objectives than those faced by MAREANO. The challenge in developing an EVI for MAREANO lies in the fact that (a) the sampling surveys have to cover multiple objectives and (b) that a reasonable level of information on required sampling effort is needed well in advance of the mapping surveys for budgetary planning purposes and ship time applications.

MAREANO’s mandate is comprehensive, unbiased mapping of the seabed on a regional scale to inform management. This necessitates obtaining the best possible information for geology and biology through co-located samples. The data used for EVI estimation should therefore attempt to capture as much of the environmental variability affecting biology and geology as possible.

4. ENVIRONMENTAL DATA – MATCHING AVAILABILITY WITH REQUIREMENTS

The usefulness of EVI in guiding sampling effort depends on the appropriateness of the environmental data fed into the EVI formula (see 'Computation of EVI'). Hence, potential environmental data layers need to fulfil a set of requirements.

The most important criterion is the **relevance** of the environmental variable for describing and/or predicting variations and patterns of diversity in nature such as species composition, habitat type, sediment grain size, or landscape type. Any variable that is not correlated with diversity in nature at the seabed will not contribute to, and may potentially confound, the ability of EVI to provide good advice about scaled sampling effort.

Some variables are more relevant for capturing variation in biology (e.g. oceanographic variables related to seabed temperature and salinity) while others are relevant for both biological and geological diversity (e.g. seabed currents, variables describing terrain properties and variability; see Table 1). Very few, if any, variables that are relevant in an EVI context are important for describing geology only. This reflects the fact that variability in

Table 1. Overview of potentially important environmental layers and their relevance, availability and quality in relation to the estimation of the EVI.

Environmental layer	Geological relevance	Biological relevance	Availability & quality	References
Bathymetry	Depth is correlated with several environmental predictors and as such work as a proxy for many, often unknown, factors.	Depth is correlated with several environmental predictors and as such work as a proxy for many, often unknown, factors [1,2].	GEBCO/IBCAO: Good avail, intermediate/low quality MAREANO: Intermediate availability, good quality	[1] Rosenberg et al. 2000 [2] Rosenberg et al. 2002
Temperature	Within the areas considered for MAREANO sampling, temperature is not considered as important.	Marine invertebrates are so-called poikilothermic meaning that they have no mechanism for controlling temperature. Each species have their specific temperature tolerance range [3,4].	IMR(4k): Intermediate availability, intermediate quality IMR(800m): Intermediate availability, intermediate/good quality	[3] Hiscock et al. 2004 [4] Hale et al. 2011
Salinity	Within the study areas considered in this report, salinity exhibits minimal variation and is not considered as important for MAREANO sampling.	Marine invertebrates normally have no mechanism to control osmosis and can therefore be highly affected by even small changes in salinity.	IMR(4k): Intermediate availability, intermediate quality IMR(800m): Intermediate availability, intermediate/good quality	
Seabed/ocean bottom currents	The ocean currents directly affects shear stress, sedimentation rates and grain size distributions [5].	The spatial structuring of marine invertebrates are highly affected by ocean bottom currents. The ocean currents directly affects shear stress, sedimentation rates and grain size distributions [6].	IMR(4k): Intermediate availability, intermediate quality IMR(800m): Intermediate availability, intermediate/good quality	[5] Grabowski et al. 2011 [6] van Son et al. 2014
Measure of terrain variability and form	Terrain variability and form can often be used as a proxy for sediment genesis and grain size [7].	High terrain variability is often correlated with high diversity of marine benthos.	MAREANO: Intermediate availability, good quality	[7] Johnson et al. 2003
Acoustic backscatter	Very important for sediment grain size maps. Problem: Calibration/levelling between areas and surveys limits usefulusefulnessin EVI calculations [8].	Is a good predictor for the distribution of marine benthos because it says something about the relative degree of softness and hardness of the seabed.	MAREANO: Intermediate availability, good quality	[8] Stephens & Diesing 2014

geology and biology are often correlated and jointly form the basis of seabed habitat (Harris and Baker, 2012; Todd and Greene, 2007).

Data availability is another important data requirement for EVI in practical terms. Reliable estimations of EVI will be difficult to achieve unless the relevant environmental variables are readily available. The availability of the data is also critical if EVI is going to be used in a long-term sampling-planning perspective. This poses a challenge related to the point in time at which data becomes available ahead of both long-term and cruise planning. For optimal long-term planning, it is desirable to have adequate data available several years ahead of detailed cruise planning. Unfortunately, in most cases this is very difficult to achieve. For detailed cruise planning, baseline environmental data such as bathymetry and oceanographic model data should be available several months before the planned cruise is taking place. This should be possible to achieve although it may necessitate some revisions to MAREANO's current time scales for data availability and workflow. Optimal cruise planning is dependent on timely access to environmental data, and good planning will directly improve the quality of MAREANO deliverables.

Estimation of EVI is performed in the statistical computing language, R (R Core Team, 2014) and associated R packages related to spatial analyses (see Appendices for more information). Input data to EVI needs to be available as rasters covering the whole area of interest. For oceanographic data, unless derived from satellite data, this requirement poses a challenge because they are models based on and calibrated by recorded observations.

This leads to a third important data requirement, namely the **quality** of the environmental data. The environmental data need to be of such quality that they contain and properly describe real patterns and changes in values of the phenomenon they represent. This is especially crucial for variables that are represented by a model themselves. Furthermore, the quality of the environmental data is not necessarily tied to their resolution. For example, the GEBCO (2014 version; http://www.gebco.net) bathymetry was tested and compared to high-quality bathymetry data collected by MAREANO. GEBCO is compiled from IBCAO (http://www.ngdc.noaa.gov/mgg/bathymetry/arctic/arctic.html; includes MAREANO data) and a range of other bathymetric data sets that have variable data resolution. At the desired level of resolution for EVI, we found little difference between results produced using GEBCO and the MAREANO data. Overall, GEBCO data was deemed most appropriate for incorporation in estimations of EVI as it offers full coverage across all Norwegian waters, allowing use in long-term planning whereas MAREANO data are restricted to MAREANO multibeam coverage. As the GEBCO 2014 data becomes less up to date it may become necessary to integrate more recent MAREANO multibeam with this dataset.

5. ENVIRONMENTAL DATA INCLUDED IN EVI ESTIMATIONS

5.1 Bathymetry data

For our area of interest, the 30-arcsec grid GEBCO data was downloaded and converted to a projected raster grid with a resolution of about 425 m, which was subsequently resampled to 400 m (Figure 1a-b). From this data set we used the bathymetry data itself as well as vector ruggedness measure (VRM; Figure 1c; Hobson 1972, Sappington et al. 2007) and relative relief (Figure 1d; Rudberg 1968, Pike et al. 2009, Erikstad et al. 2013) that are measures of geomorphometry derived from the original bathymetry.

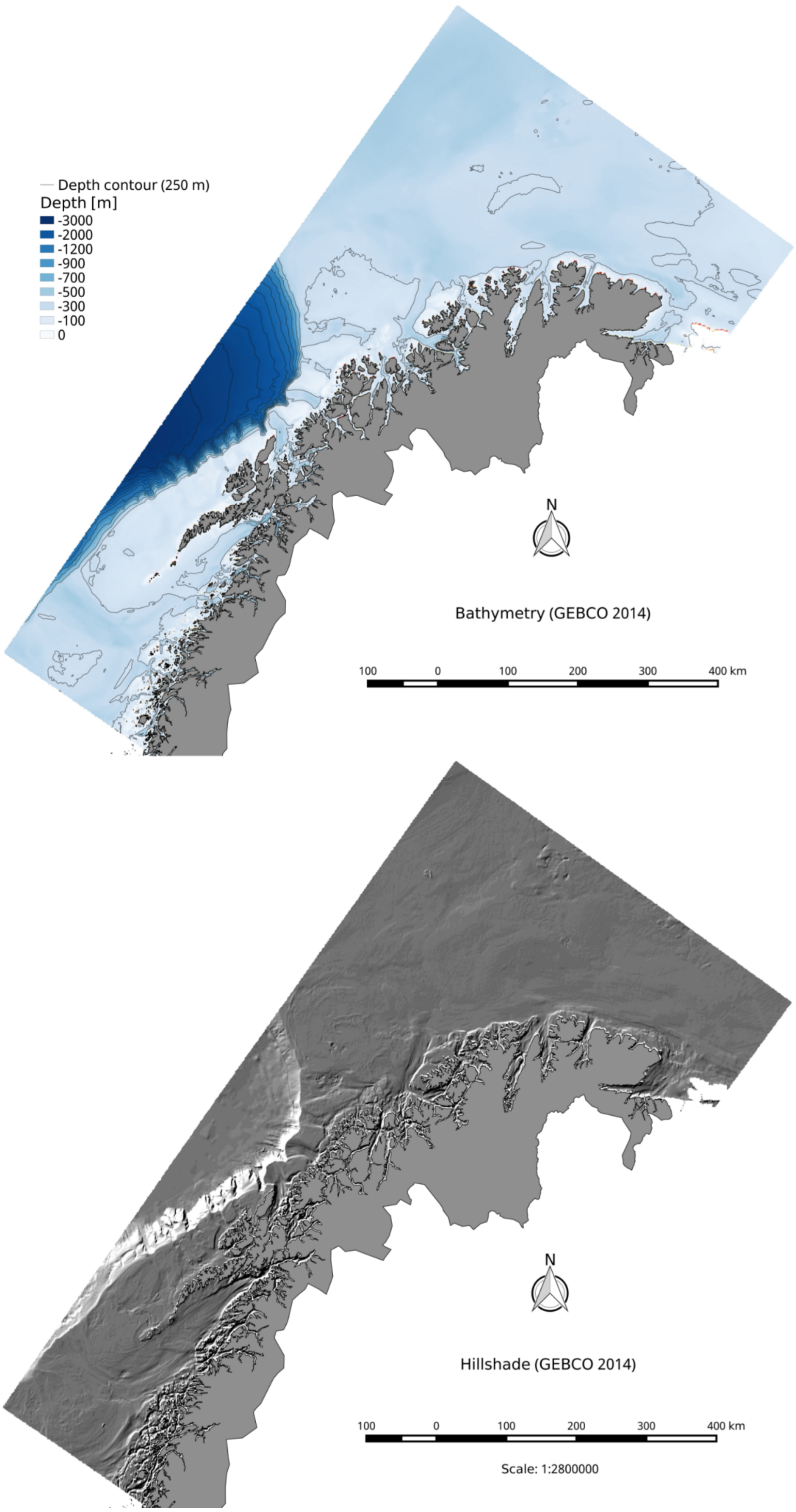
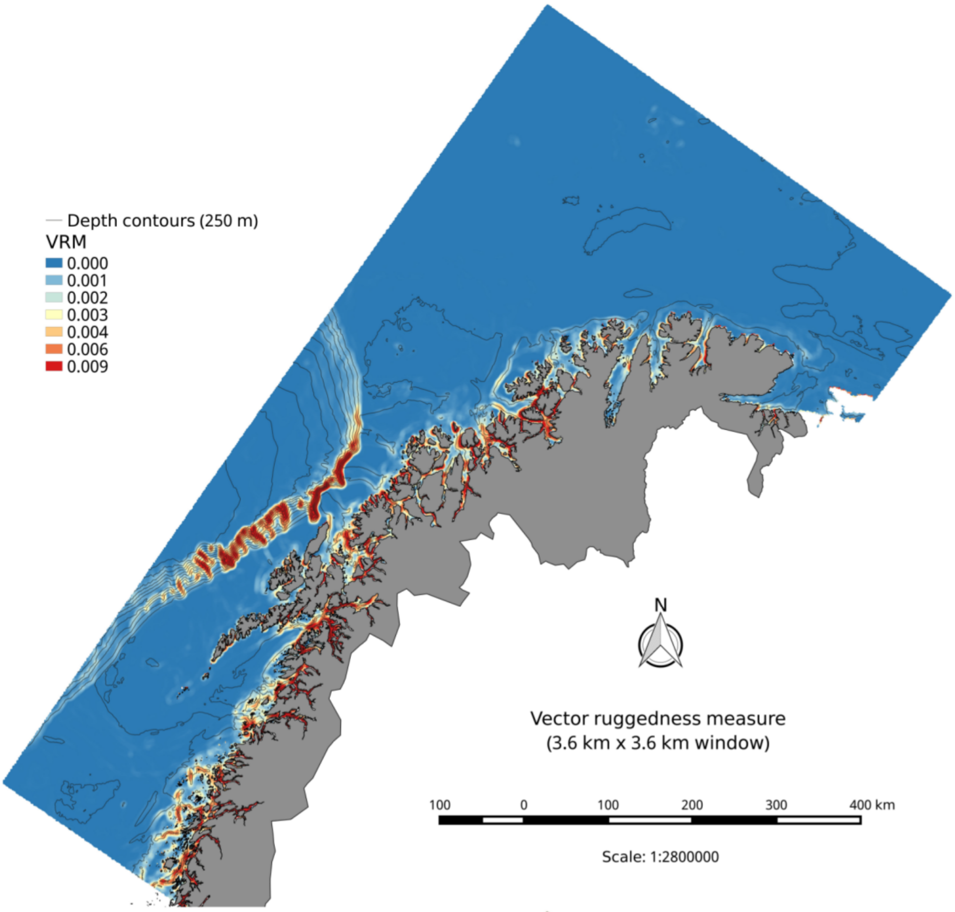


Figure 1. Raster maps of environmental layers used to in the estimation of EVI. Resolution of maps are 400 m. (a, top) Bathymetry. (b, bottom) Hillshade image.



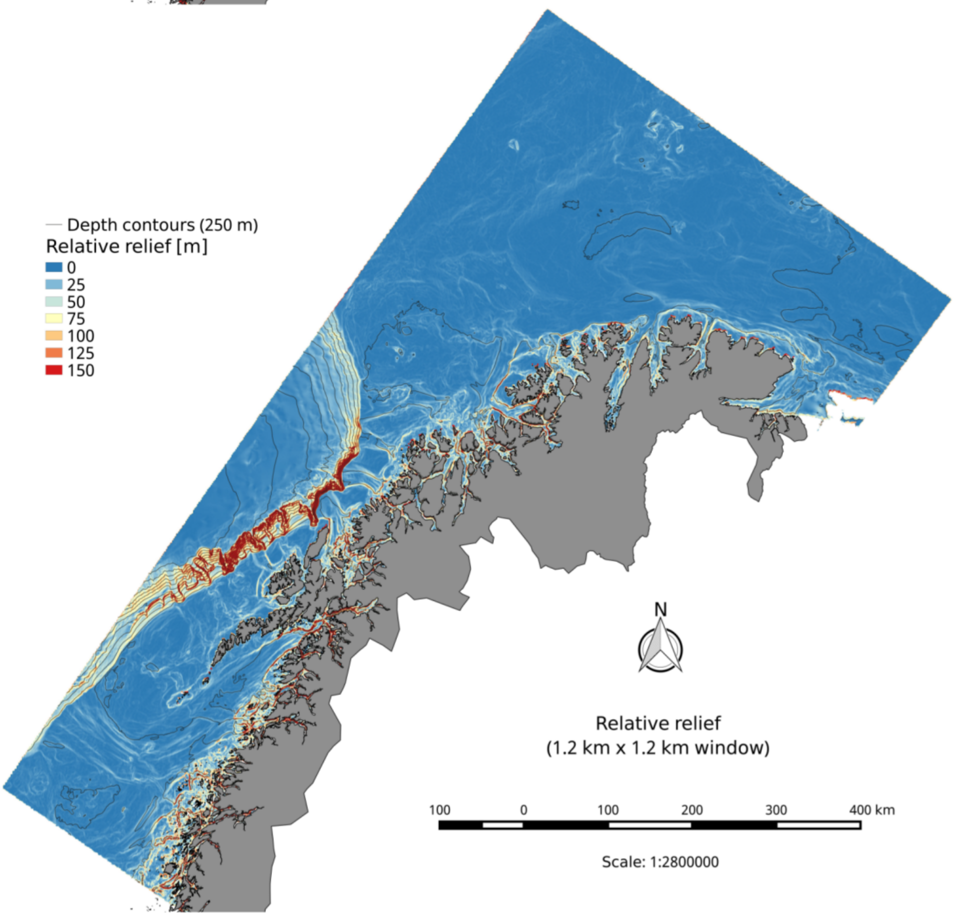
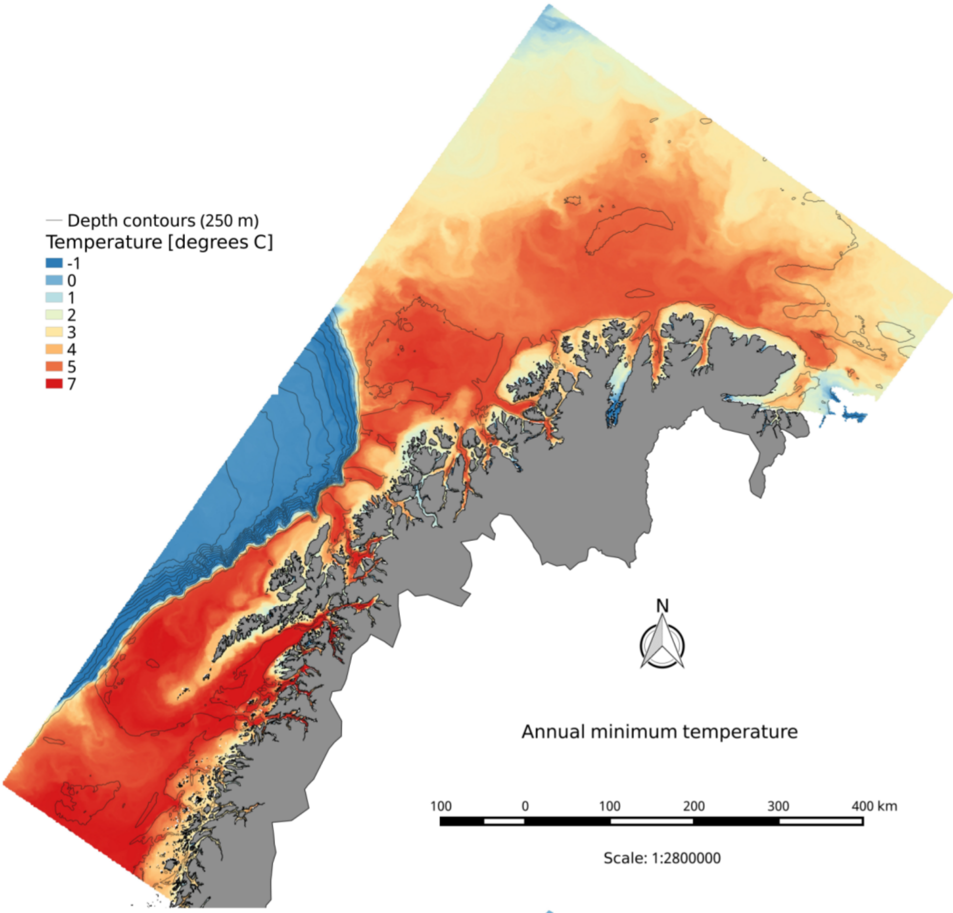


Figure 1 continued. Raster maps of environmental layers used to in the estimation of EVI. Resolution of maps are 400 m. (c, top) VRM. (d, bottom) relative relief.



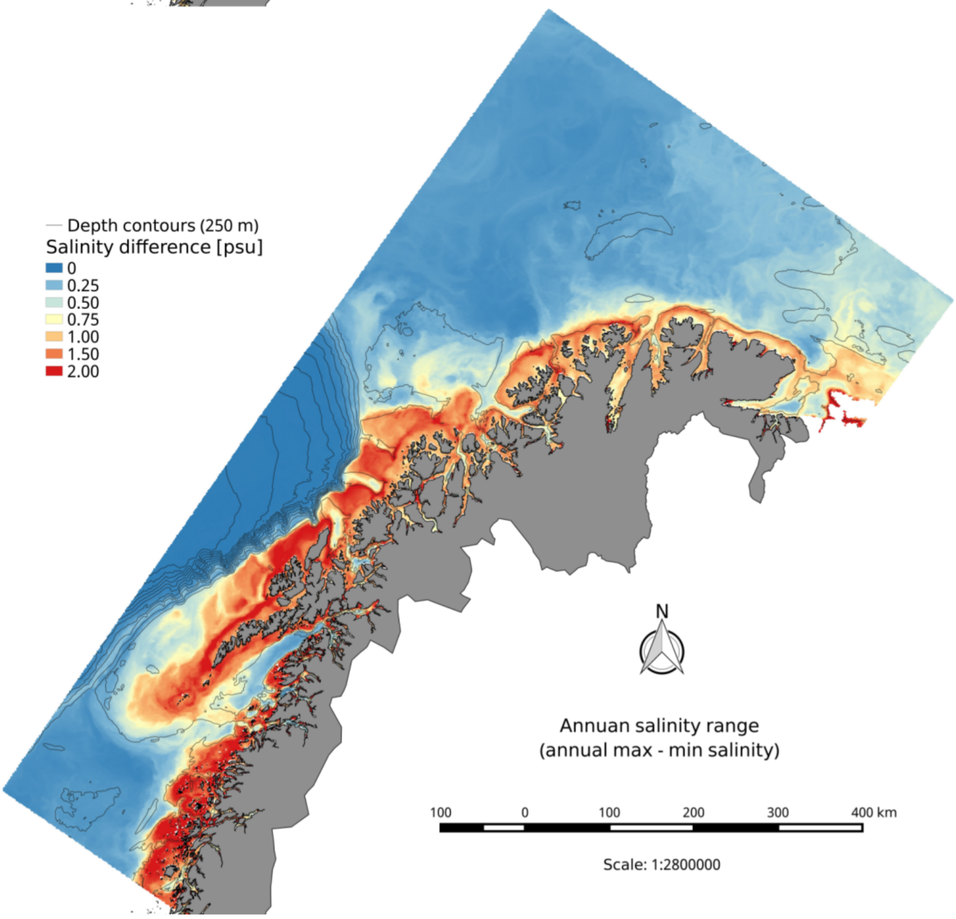


Figure 1 continued. Raster maps of environmental layers used to in the estimation of EVI. Resolution of maps are 400 m. (e, top) Minimum temperature. (f, bottom) salinity range.

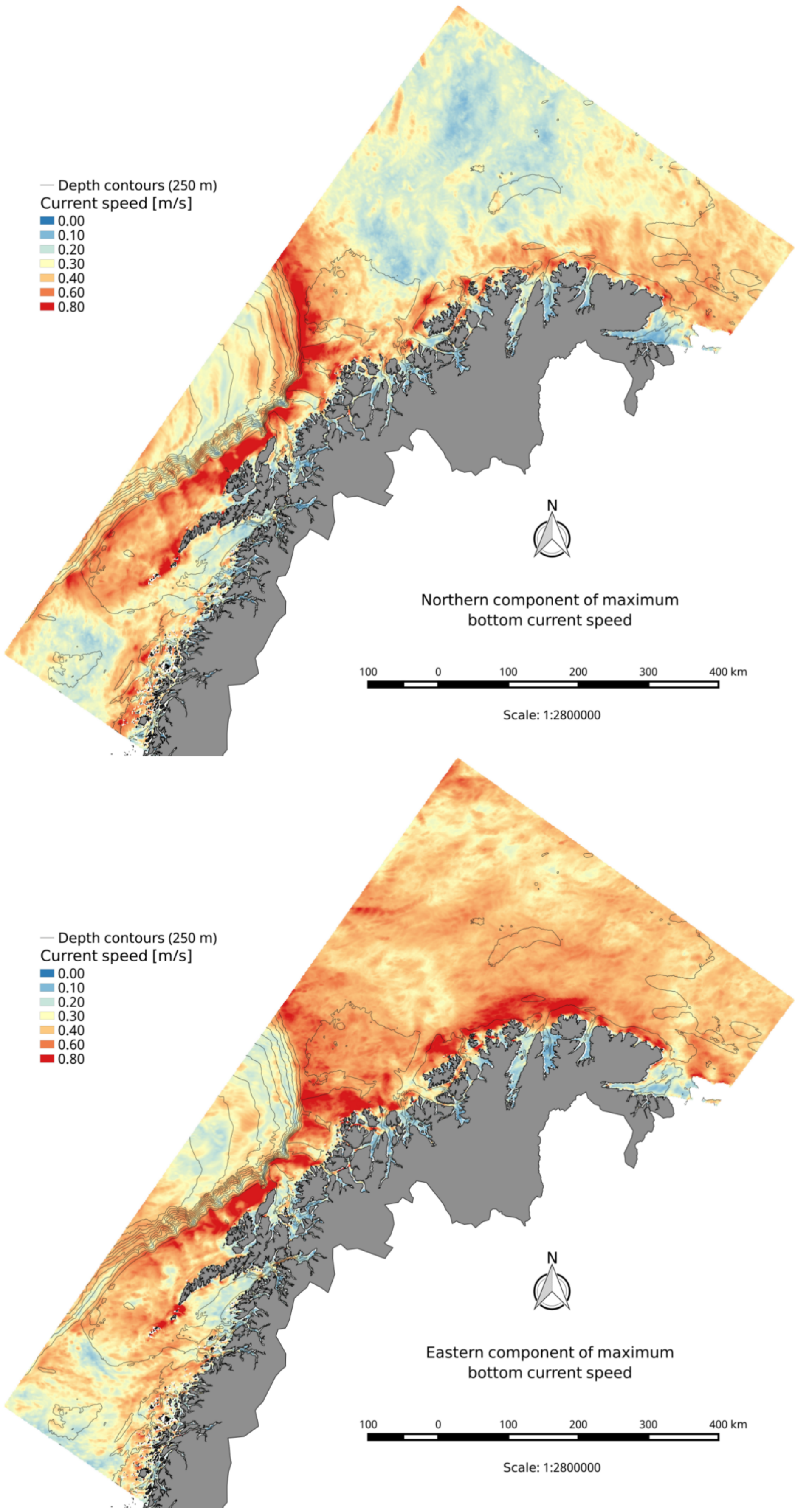


Figure 1 continued. Raster maps of environmental layers used to in the estimation of EVI. Resolution of maps are 400 m. (g, top) Northern and (h, bottom) eastern component of maximum bottom current speed.

5.2 Oceanographic data

In addition, a wide range of oceanographic variables was available for incorporation in EVI. Based on knowledge about which of these variables that may affect biology and geology, and those variables available to MAREANO at the time of this study, the following were chosen:

- annual minimum temperature (Figure 1e)
- salinity range, the annual maximum value minus the annual minimum value (Figure 1f)
- northern component of maximum current speed (Figure 1g)
- eastern component of maximum current speed (Figure 1h).

All oceanographic variables have a resolution of 800 m, but they were all resampled to 400 m to match the resolution of bathymetric variables.

6. COMPUTATION OF EVI

The computations used for the estimation of EVI in R are performed on a pixel-by-pixel basis. It is intended that the EVI should be generic in the sense that it should neither depend on which variables are used (as long as they comply with quality and availability requirements) nor on the number of variables used. Furthermore, the EVI needs to be able to combine several environmental layers into one unified index rendering a single value for each pixel.

The EVI is designed to account for the mean dispersion (variability) at each pixel of all the environmental layers used in the formula. As such it needs to incorporate a measure of dispersion such as standard deviation or variance. However, different environmental layers will operate on completely different numeric scales. Layers that happen to operate on scales with high numeric values will have larger measures of dispersion than others operating on scales with lower values, which consequently make them incomparable.
A common work-around to this problem is to calculate the coefficient of variation (CV) for each layer, where CV is defined as the standard deviation divided by the mean. The CV is very sensitive to low means, especially means below the value of 1, since these will spuriously inflate the CV estimation. Furthermore, the CV may not make sense if a layer contains both negative and positive values. Therefore, in our EVI computations, each environmental layer is linearly scaled between 5 and 260 before estimating the CV [note that we do not use the commonly used linear scale between 0 and 255 (e.g. Clements et al. 2010) because this will also result in spurious pixel means close to zero that will inflate CV values]. The CV values needs to be estimated at a certain level of aggregation of the 400 m rasters in order to capture the environmental and spatial variability. We have tested different levels of aggregation between 10 and 30 km. The CV was calculated for bathymetry (Figure 2a), VRM (Figure 2b), annual minimum temperature (Figure 2c), annual salinity range (Figure 2d),

northern and eastern component of maximum bottom current speed (Figure 2e-f), and relative relief (Figure 2g).

Taking these CV rasters into account, the formula for calculation of the EVI in each pixel at a certain level of aggregation is as follows:

$$EVI = \frac{\sum_{i=1}^{n} CV_i W_i}{\sum_{i=i}^{n} W_i}$$

where *i* is the index of the *n* environmental layers used to estimate EVI, CV_i is the coefficient of variation in each pixel of layer *i*, and W_i is the weight given to that layer. The weighting allows the user to give more weight to layers that are considered to be more important than others for describing environmental and spatial variability in biological and geological diversity.

A differentiated weighting is something that may be implemented in the future as we gain more experience in using EVI for MAREANO survey planning. By default, each layer is given weight 1. The EVI is thus standardized to the number of layers and weights included in its estimation. As such, the EVI expresses the mean environmental dispersion (variability) for each pixel and is expressed as the proportion of standard deviation to the mean.

6.1 Recommended sampling effort based on EVI

After having estimated the EVI, the next step is to use it to guide the recommended pixel-wise sampling effort. The recommended sampling effort will depend on the desired and feasible level of sampling effort. By setting a minimum and maximum sampling effort we can simply linearly rescale the sampling effort by application of a linear model. The following formula achieves this goal:

$$SE = SE_{min} + \frac{SE_{max} - SE_{min}}{EVI_{max}} \times EVI$$

where *SE* is the estimated and linearly rescaled pixel-wise sampling effort, $SE_{min/max}$ is the minimum/maximum desired sampling effort, EVI_{max} is the maximum EVI value, and EVI is the pixel-wise estimated EVI value.

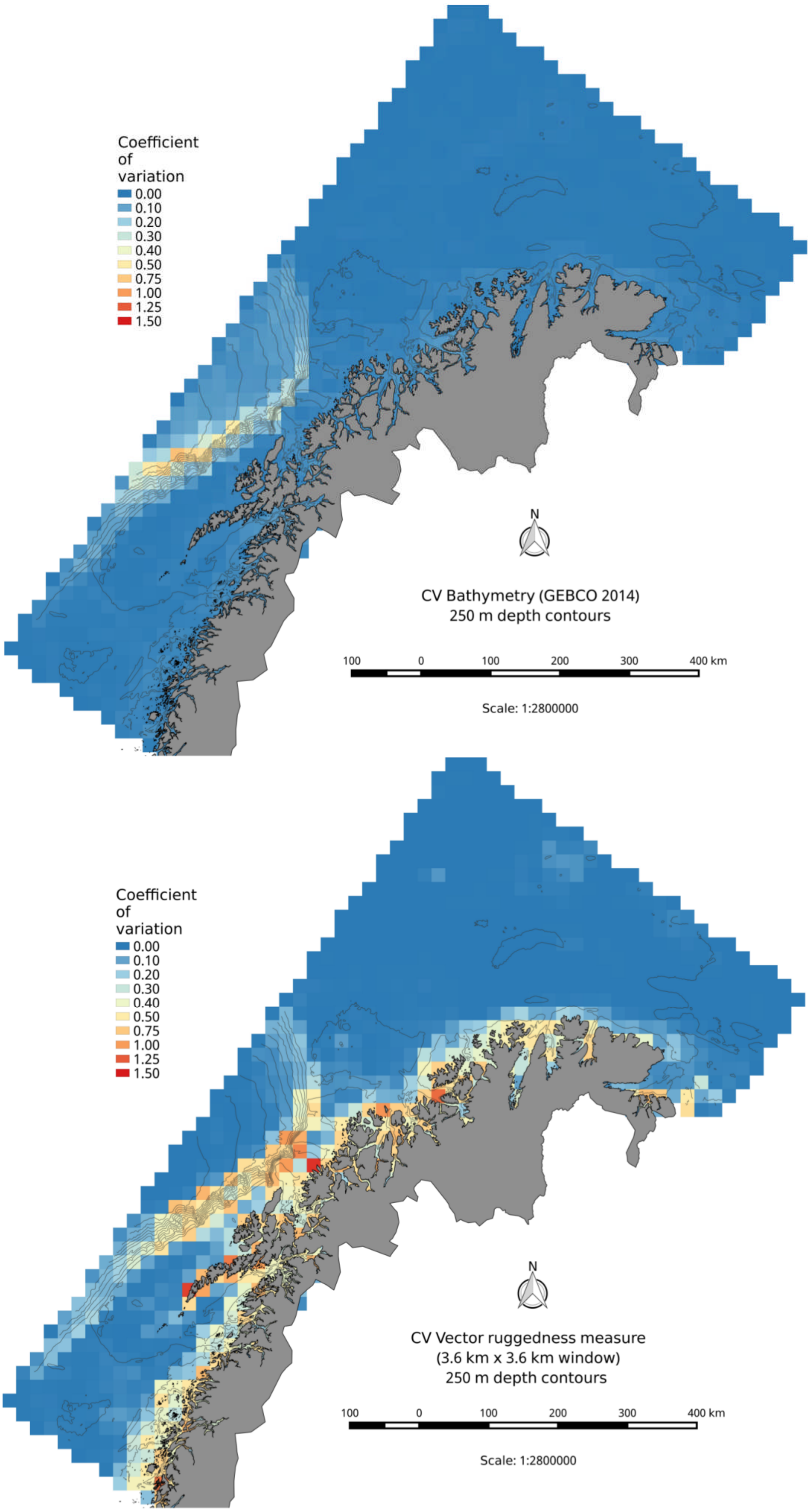
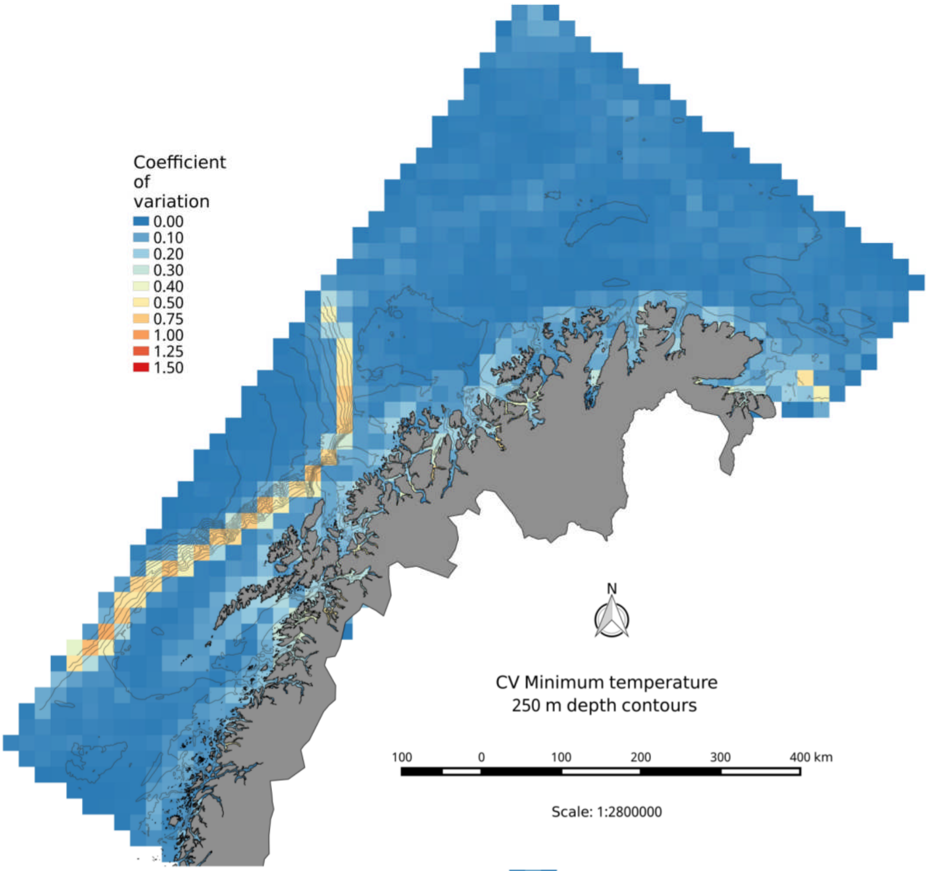


Figure 2. Raster maps showing spatial variability expressed as coefficient of variation for environmental layers included in the calculation of EVI. The coefficient of variation is calculated within 20 km pixels. (a, top) Bathymetry. (b, bottom) VRM.



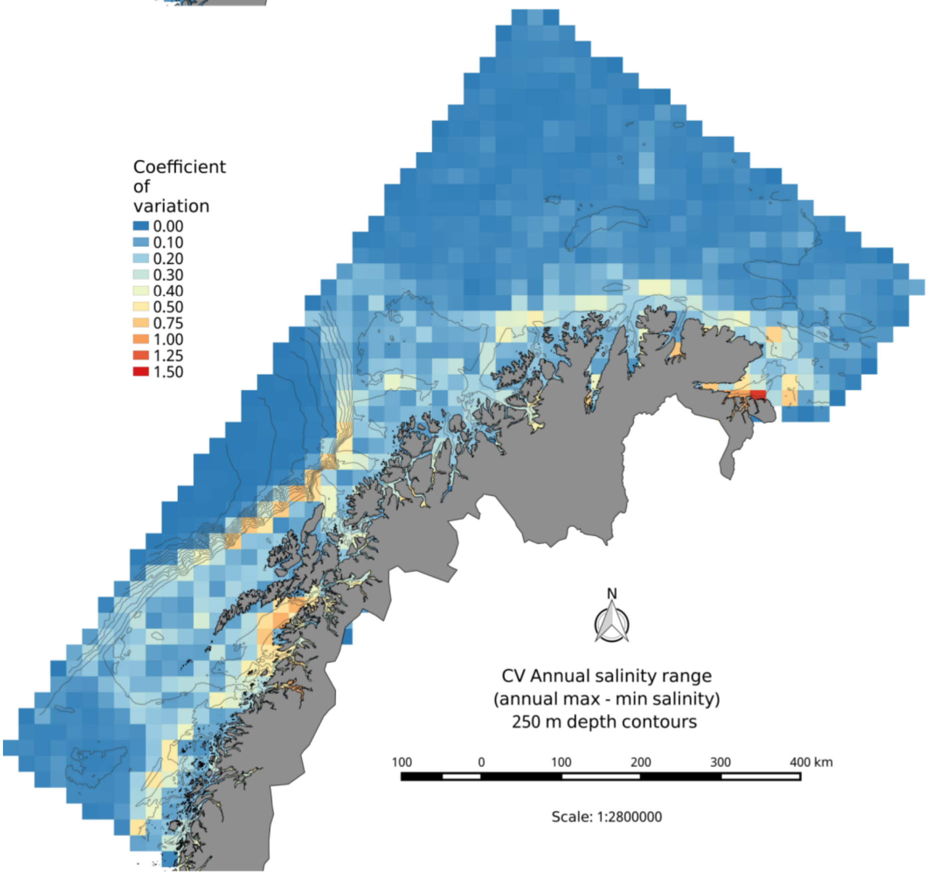
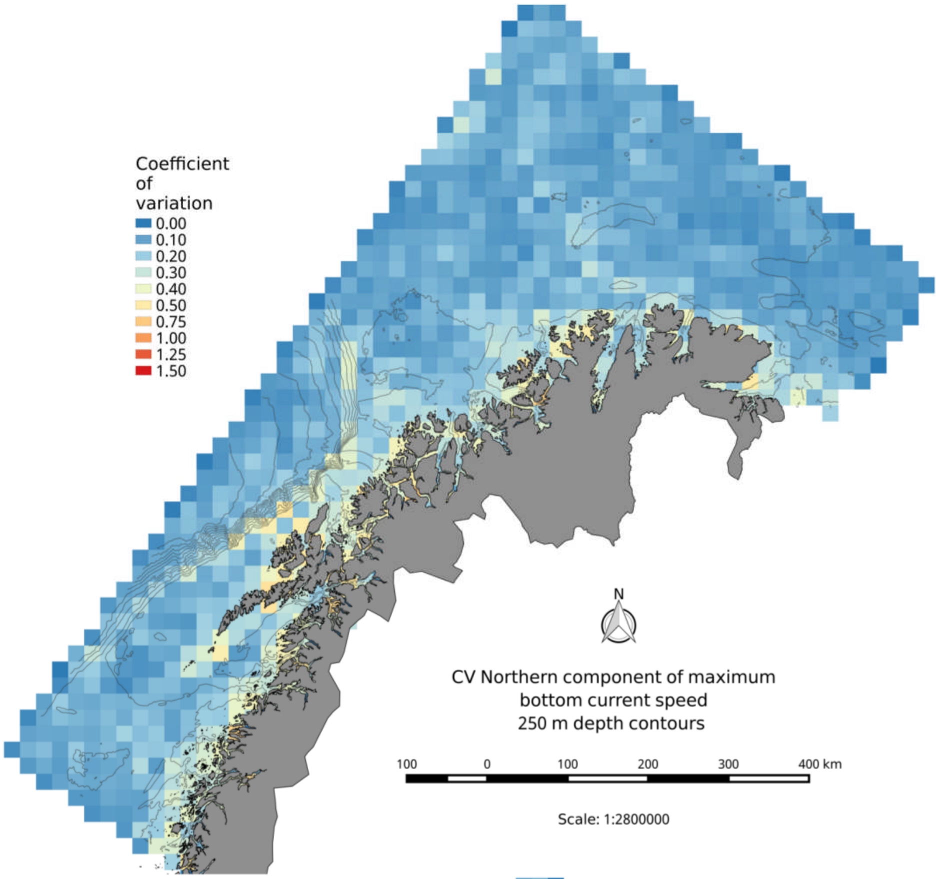


Figure 2 continued. Raster maps showing spatial variability expressed as coefficient of variation for environmental layers included in the calculation of EVI. The coefficient of variation is calculated within 20 km pixels. (c, top) Minimum temperature. (d, bottom) salinity range.



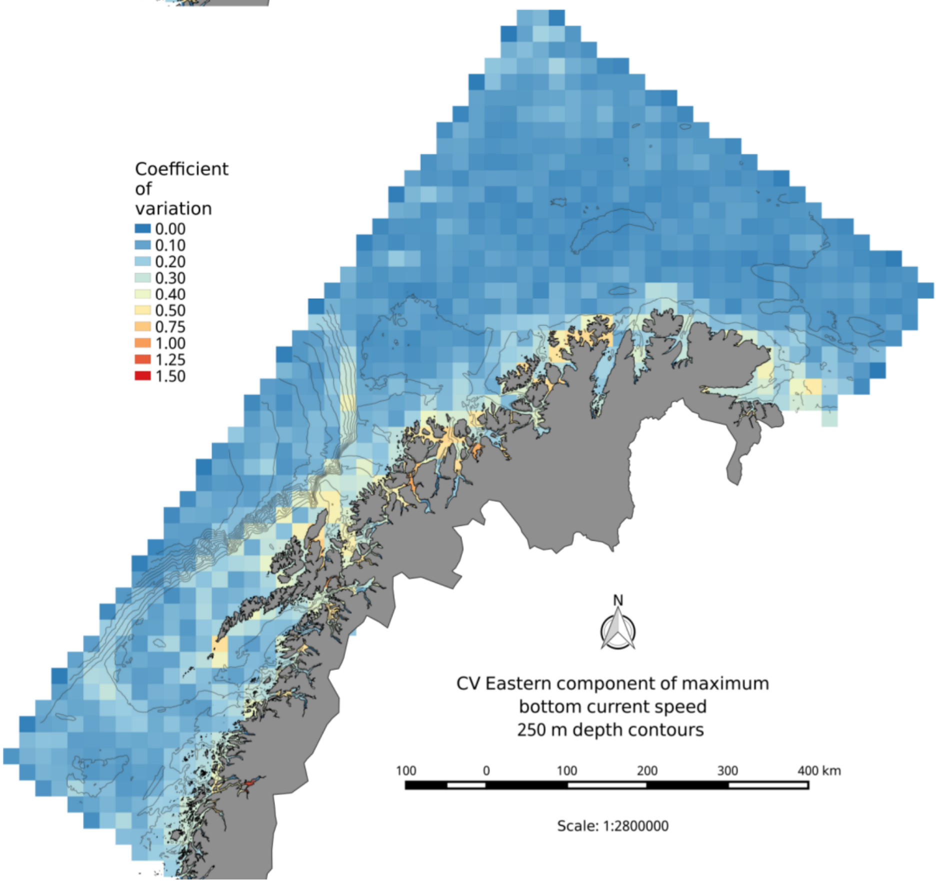


Figure 2 continued. Raster maps showing spatial variability expressed as coefficient of variation for environmental layers included in the calculation of EVI. The coefficient of variation is calculated within 20 km pixels. (e, top) Northern and eastern (f, bottom) component of maximum bottom current speed.

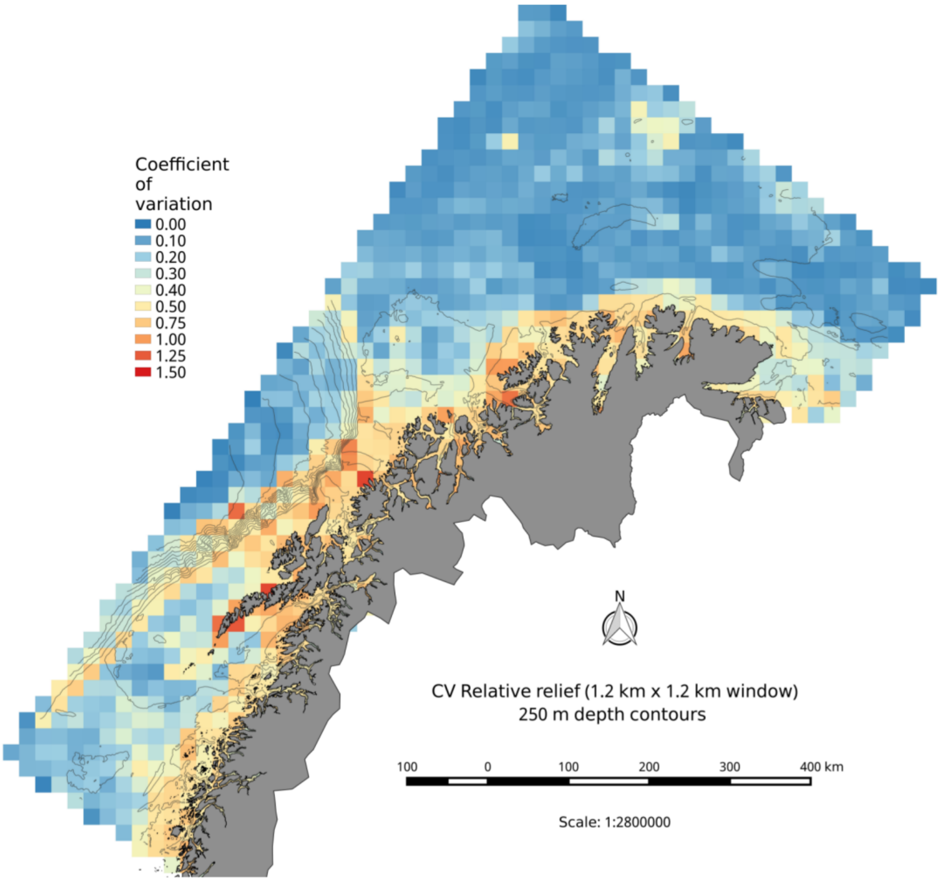


Figure 2 continued. Raster maps showing spatial variability expressed as coefficient of variation for environmental layers included in the calculation of EVI. The coefficient of variation is calculated within 20 km pixels. (g, top) Relative relief.

The EVI and the linearly scaled sampling effort have been calculated for an area covering the continental shelf and slope between the Varangerfjord in the northeast and more or less the island of Røst in the south. The EVI and sampling effort was calculated at aggregation levels of 10 km (Figure 3 a-b), 20 km (Figure 3 c-d), 25 km (Figure 3 e-f), and 30 km (Figure 3 g-h). The results show that if the level of aggregation is too fine, EVI will not be able to pick up the environmental variability properly, only focusing on really fine-scale variability. By contrast, if the level of aggregation is too coarse, none of the fine-scale variability will be detected, resulting in an EVI with too little variation.

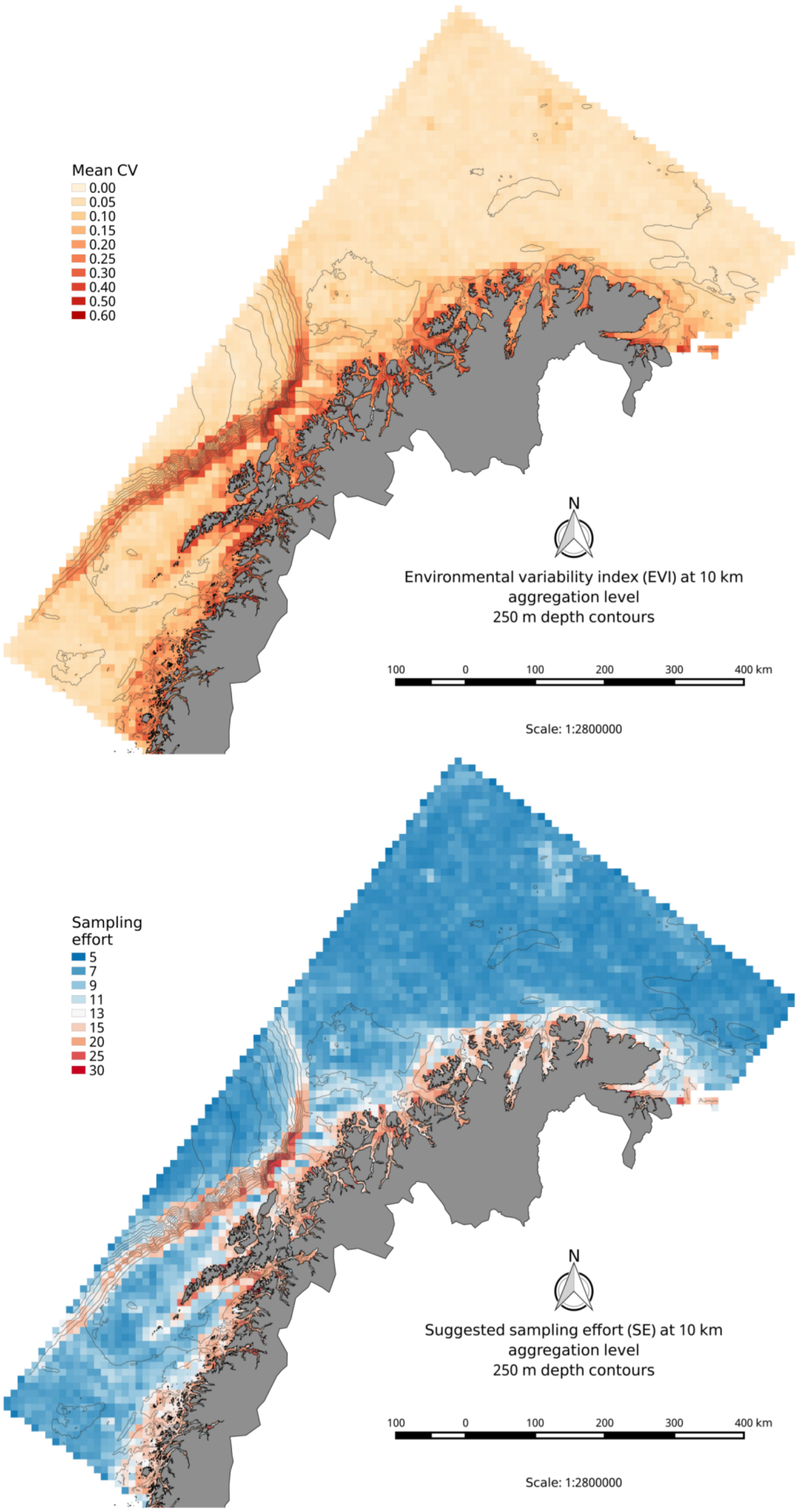
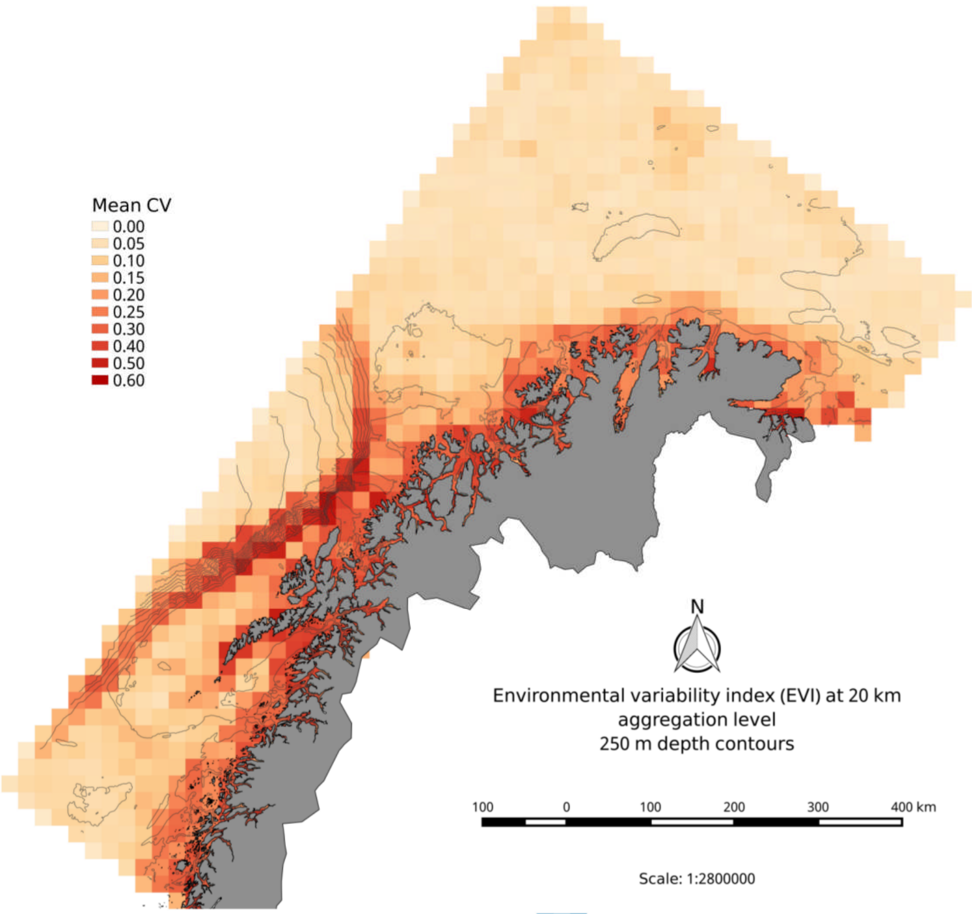


Figure 3. Raster maps showing the estimated EVI and corresponding linearly scaled sampling effort for 10 km aggregation level. (a, top) EVI. (b, bottom) suggested SE.



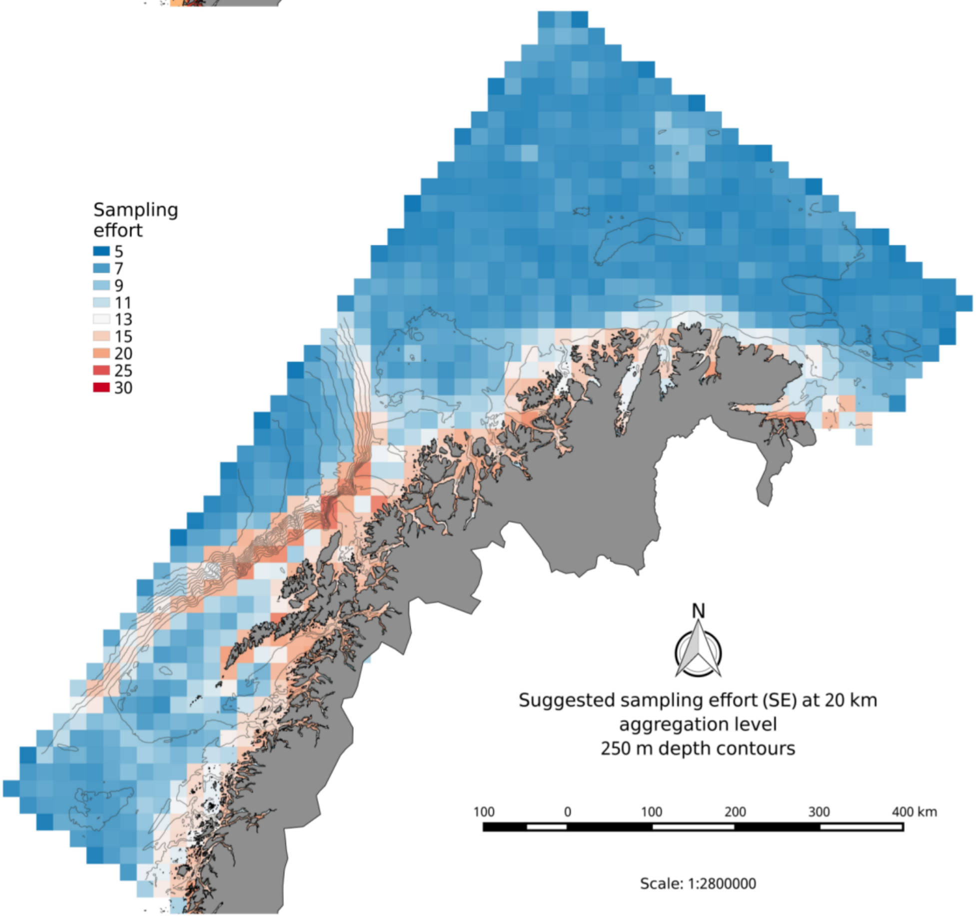
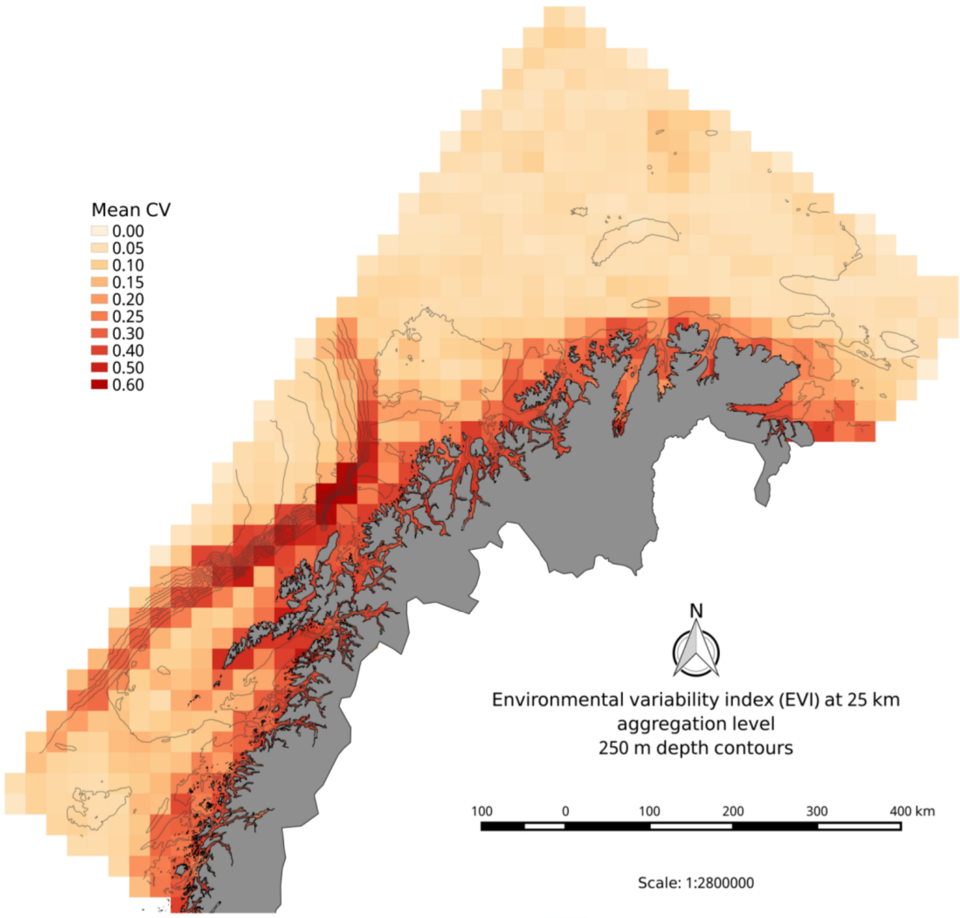


Figure 3 continued. Raster maps showing the estimated EVI and corresponding linearly scaled sampling effort for 20 km aggregation level. (c, top) EVI. (d, bottom) suggested SE.



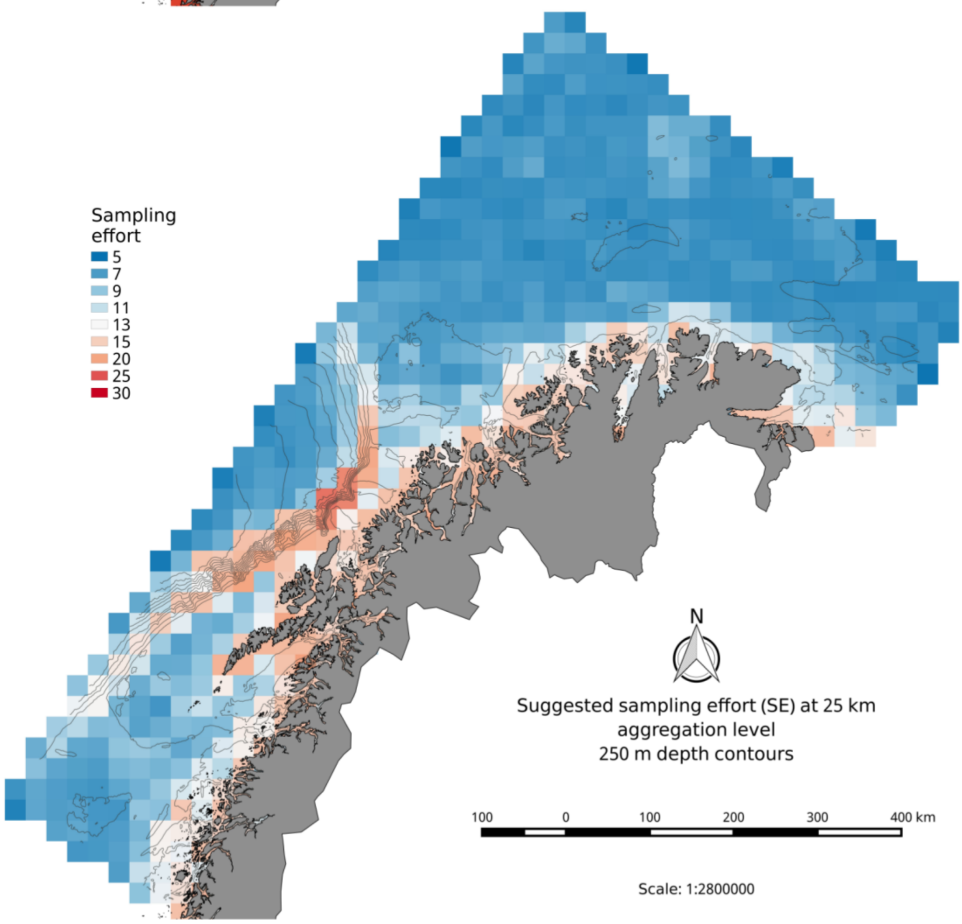
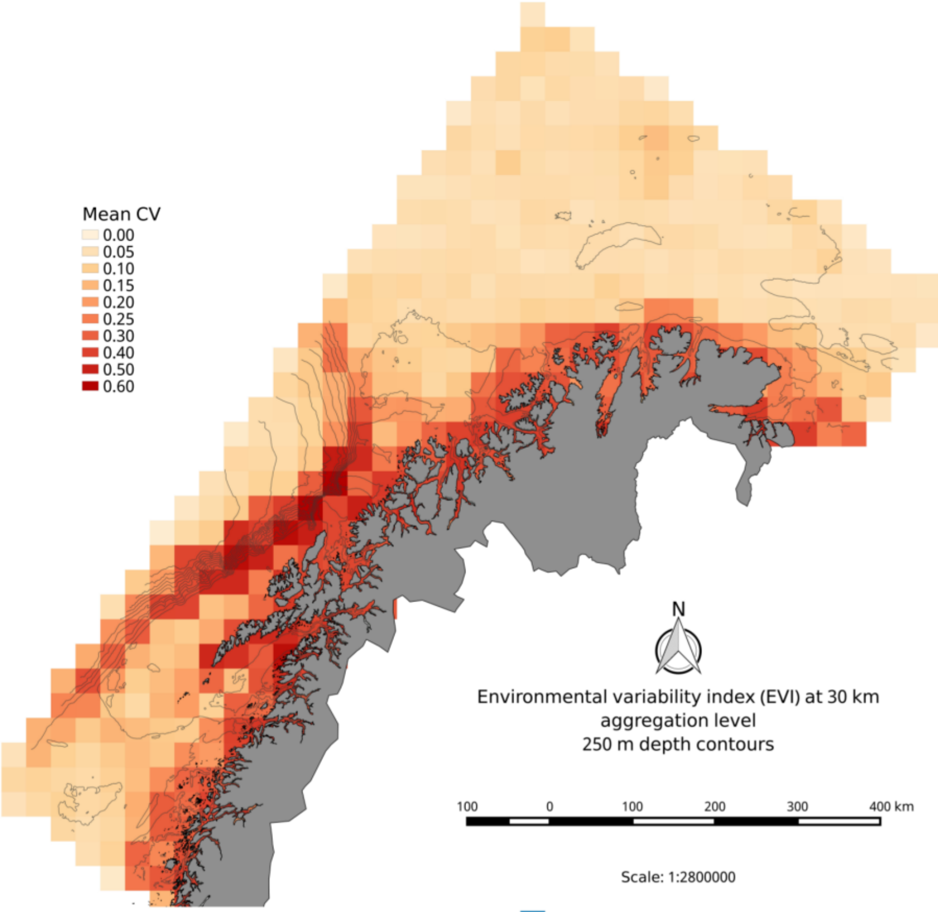


Figure 3 continued. Raster maps showing the estimated EVI and corresponding linearly scaled sampling effort for 25 km aggregation level. (e, top) EVI. (f, bottom) suggested SE.



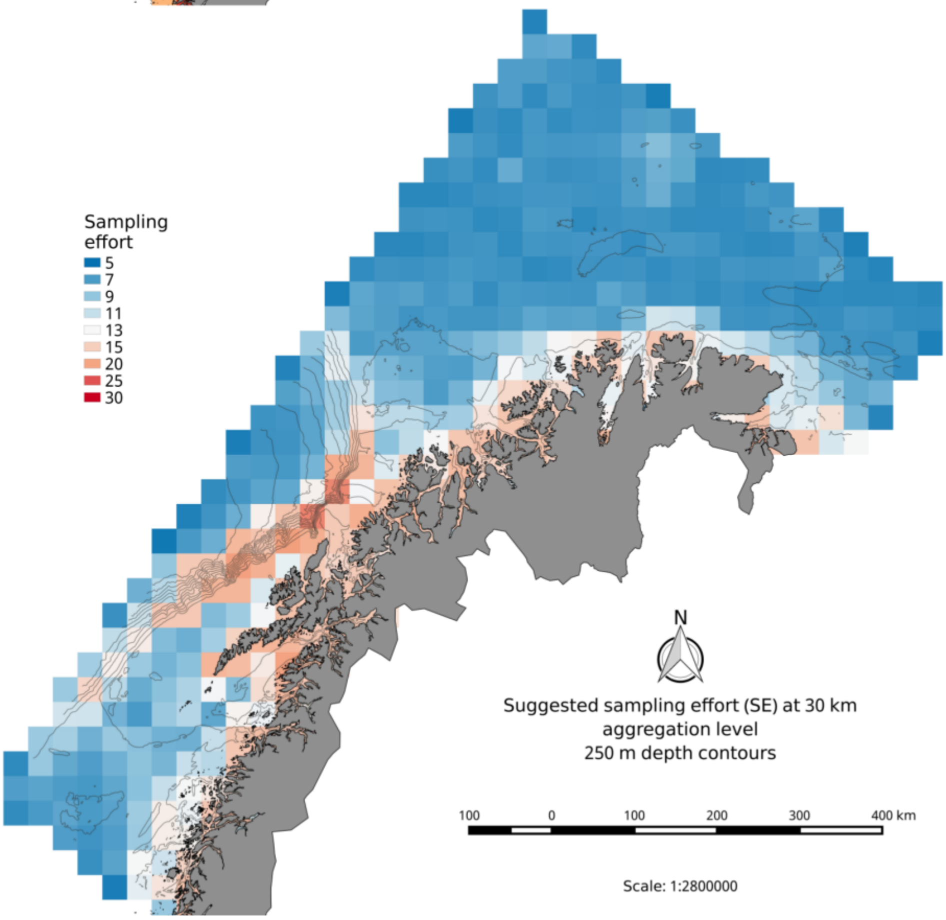


Figure 3 continued. Raster maps showing the estimated EVI and corresponding linearly scaled sampling effort for 30 km aggregation level. (g, top) EVI. (h, bottom) suggested SE.

7. HOW EVI CAN GUIDE THE TOTAL SAMPLING EFFORT FOR A GIVEN STUDY AREA

We have seen in the previous section that EVI provides a useful summary of the environmental variability across MAREANO mapping areas, and that it can be extended to provide recommendations for pixel-wise sampling effort. However, EVI has primarily been developed for its practical application in guiding sampling effort for long-term strategic regional planning and for detailed cruise planning, i.e., within a study area, not on a pixel-by-pixel basis. EVI can provide a mean and total sampling effort for a given study areas by using the pixel-based sampling effort estimations (Figure 4). The mean sampling effort, i.e., the mean number of video lines per 1000 km^2, is simply the mean of the sampling effort pixel values within a given study area (boxes outlined in black in Figure 4). We find the suggested total sampling effort within the study area, i.e., the total number of video lines, by multiplying the mean sampling effort with the extent divided by 1000 km^2.

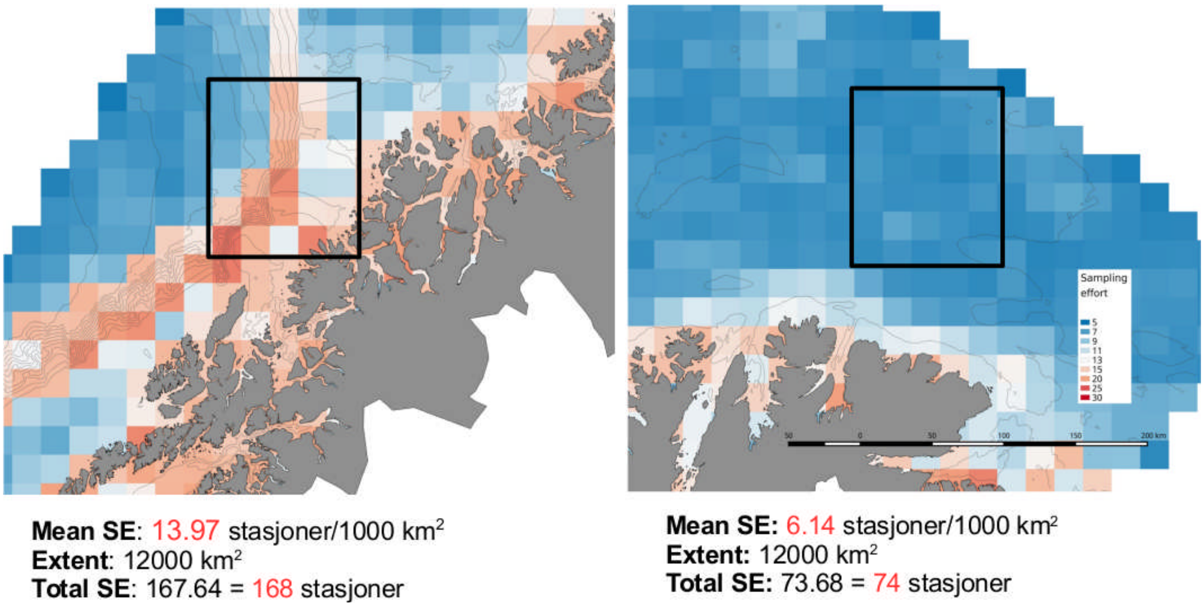


Figure 4. Demonstration of how estimated EVI can guide and suggest sampling effort in two areas (black squares) differing in environmental and spatial variability (left: high variability, left: low variability). Aggregation level 20 km.

The EVI clearly demonstrates its ability to distinguish between areas that we know are different in terms of environmental variability. Here we have demonstrated how EVI estimations can provide an objective suggestion for the mean and total sampling effort within a given area. Detailed cruise planning, i.e. the actual positioning of the stations within the area, depends on finer-scale environmental variability which may be captured by environmental layers with higher resolution e.g. full resolution multibeam data. The environmental layers used for detailed cruise planning are not necessarily the same as those used in the estimation of EVI. For example, the finest resolution of oceanographic data at the moment is at 800m (resampled to 400 m to match the GEBCO data, but that does not change the actual resolution), which may be too coarse to be of use in detailed planning unless there is a great deal of variation in the data within the survey area. Detailed cruise planning is beyond the scope of this report but thoroughly described in a related report (van Son et al. in prep.) in the MAREANO methods review series being published in 2015.

8. RELATION TO BIOLOGICAL AND GEOLOGICAL DIVERSITY

8.1 Biological diversity

The EVI's relation to observed diversity in nature is an important test of its ability to appropriately scale sampling effort. Two measures emerge as candidates for testing EVI's relation to biology, (1) species richness and (2) species turnover (beta diversity).

Species richness is simply a count of observed species and due to its lack of information regarding the identity of species and therefore also of the composition of species, it will often not be appropriate for testing the correlation of EVI to biological diversity. What is needed is a measure that says something about the rate of change in species composition, namely species turnover, i.e. beta diversity. For illustrative purposes, imagine four stations located within an EVI pixel indicating high spatial variability that may each have a relatively low species richness and may lead us to wrongly conclude that EVI is not correlated with biological diversity. However, if we look into how the species composition changes between these four stations, i.e., how many species they share, how many species are unique to each particular station, and how does the relative abundance of species change between stations, we may see that the species turnover is actually very high.

Beta diversity is a measure first proposed by Whittaker (1960, 1972) that takes into account such changes in species composition between stations. It is a measure of variation in species composition between sites, or more specifically a measure of the difference between the assemblages present at each site taking into account the identities of all species. It can be decomposed into two components: species replacement (i.e. turnover), which consists of the substitution of species in one site by different species in the other site; and species loss (or gain), which implies the elimination (or addition) of species in only one of the sites, and leads to the poorer assemblage being a strict subset of the richer one (a pattern called nestedness; Baselga 2000). There exist different dissimilarity indices that account for the two phenomena in different ways.

Here we test the EVI results against biological diversity estimations derived from existing MAREANO data. The comparisons between biological diversity and the EVI are based on the 20 km aggregation level. For all the video stations inside the same 20 km pixel we calculated multiple-site (as opposed to pairwise) overall betadiversity (including both turnover and nestedness components) using the Sørensen dissimilarity index: β_{SOR}. The particularity of this measure of dissimilarity relative to others is that it accounts for turnover and nestedness as being equivalent, as both turnover and nested patterns make alpha-diversity lower than gamma-diversity. We used the package betapart (Baselga and Orme 2012) in R (R Core Team 2014). β_{SOR} varies between 0 and 1, with higher values representing sets of sites with assemblages more dissimilar to one another.

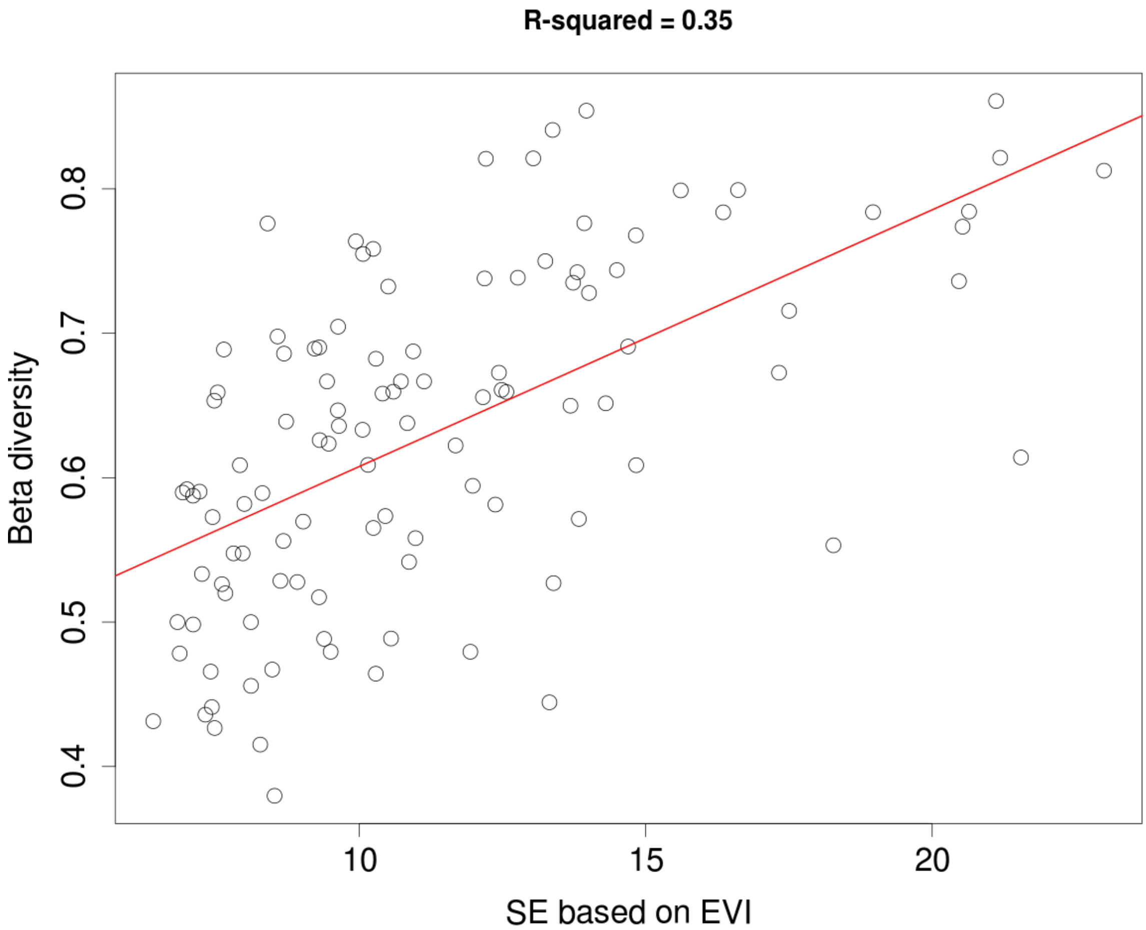


Figure 5. Correlation between beta diversity and sampling effort (SE) based on linearly scaled EVI values aggregated at 20 km (SE scale: 5 – 25). The beta diversity is estimated from changes in species composition between 700 m video line observations within 20 km pixels.

The biological data was derived from analysed video footage with all taxa pooled for each single line, including all the epibenthic fauna visible from the video captured along the line (megafauna), as well as fauna visible from the footage at parked positions. Taxa are identified at a varied range of taxonomic resolutions, as it is not always possible to classify organisms at species level from video data. The data used were reduced to presence/absence.

For EVI to be a good predictor of biological diversity, it needs to correlate well with beta diversity. In order to test this we calculated beta diversity from video lines using the Sørensen dissimilarity index and paired them with corresponding sampling effort values that were linearly rescaled EVI values (as described above). Then we fitted a linear model using sampling effort as the only predictor and beta diversity as response. In cases when facing spatial autocorrelation, the generalised least-squares model provides a promising alternative. The residuals of the linear model were randomly distributed, so there appear to be no issues of spatial autocorrelation.

The fitted linear model exhibited an $R^2 = 0.35$, indicating that the EVI (here aggregated at 20 km) is a good predictor for beta diversity (Figure 5). Anderson et al. (2006) also found a positive correlation between a measure of multivariate dispersion of normalised environmental variables and the Sørensen dissimilarity calculated from biological data from the Norwegian shelf. These two results clearly show the potential of spatial variability to guide sampling effort.

8.2 Geological diversity

We have also tested the correlation between geological diversity and EVI in a similar way as for biological diversity. Just as for biological diversity we do not expect EVI to be best correlated with richness of geological attributes, but rather to the turnover (and nestedness) of them. This hypothesis was tested using maps of the distribution of sediment grain size from areas already mapped by MAREANO (Figure 6). Using the 20 km aggregation level of EVI as a basis, we divided each 20 km pixel into four 10 km pixel and calculated the percent cover of each sediment grain size class within each of these 10 km pixels. In order to calculate a multi-pixel beta geodiversity the percent cover data had to be reduced to presence/absence. Using the same procedure as for beta diversity of biological features, we derived the beta geodiversity measure by using data from the four 10 km pixels within each 20 km pixel.

We then fitted a linear model using sampling effort as the only predictor and the beta geodiversity as reponse (Figure 7). The fitted linear model yielded an $R^2 = 0.20$, indicating again that there is a relationship between EVI and geological diversity.

Even though both linear models have very low p-values, neither of the two R^2 values are particularly impressive. However, we have to consider the level of noise in a data set such as this. There are issues related to the recording of the video lines that introduce noise in the data such as a none-standardised distance of the campod to the seabed that affects the detectability of species. Furthermore, there are also a wide range of noise-related issues identified in the backscatter data, which are one of the main data types used for sediment classification along with bathymetry, sub-surface data, video and physical samples. Likewise, it is important to remember that the sediment grain size map is a generalisation to 1:100 000 map scale and is,

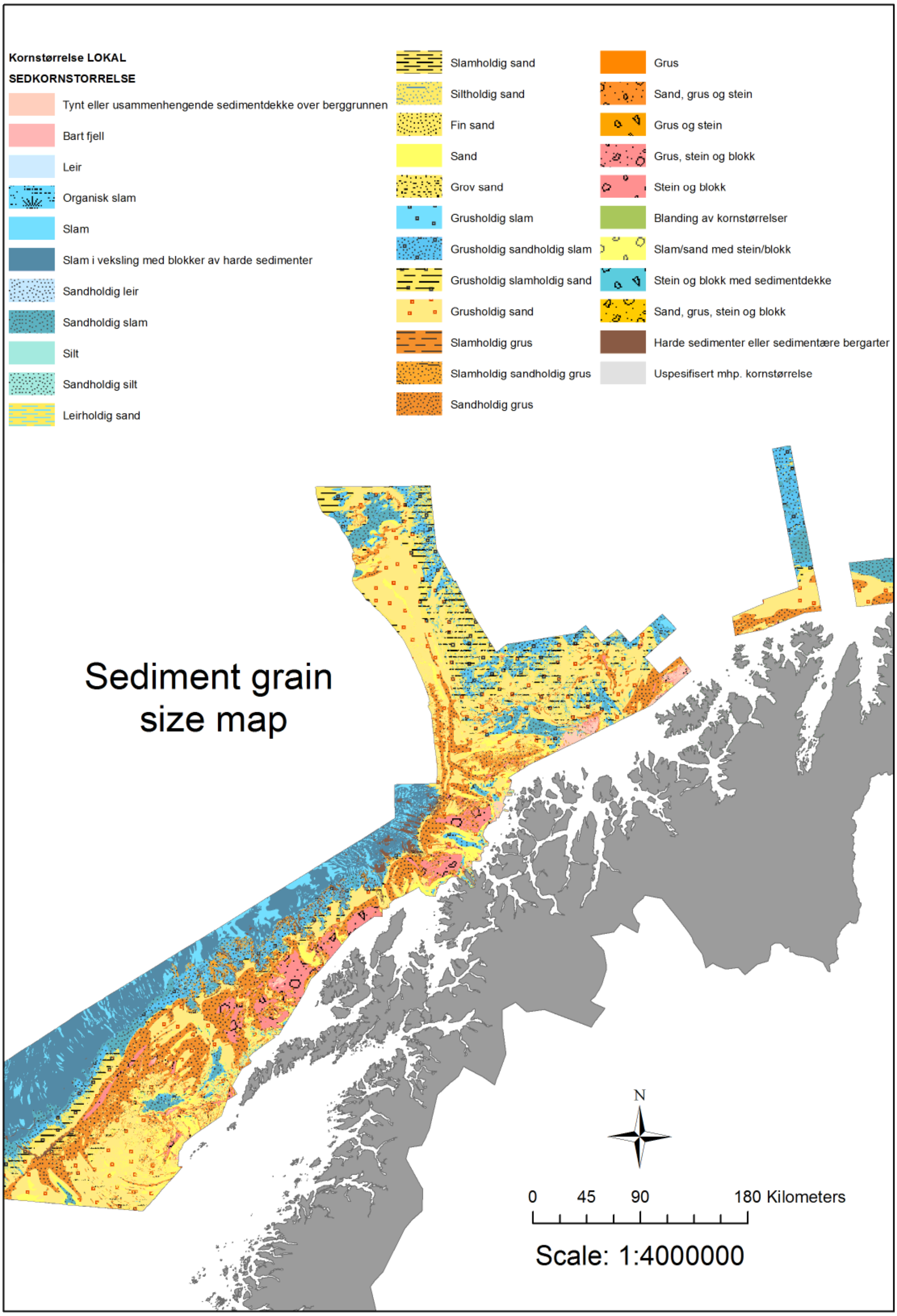


Figure 6. Sediment grain size map used in estimating beta geodiversity.

like any map, an approximation of the truth. Furthermore, by reducing the data to presence/absence [required by the beta.multi() function in the betapart package in R] the abundance and percent cover information inherent in the data is lost. Hence, the fact that our results show a relatively clear signal among this noise is encouraging.

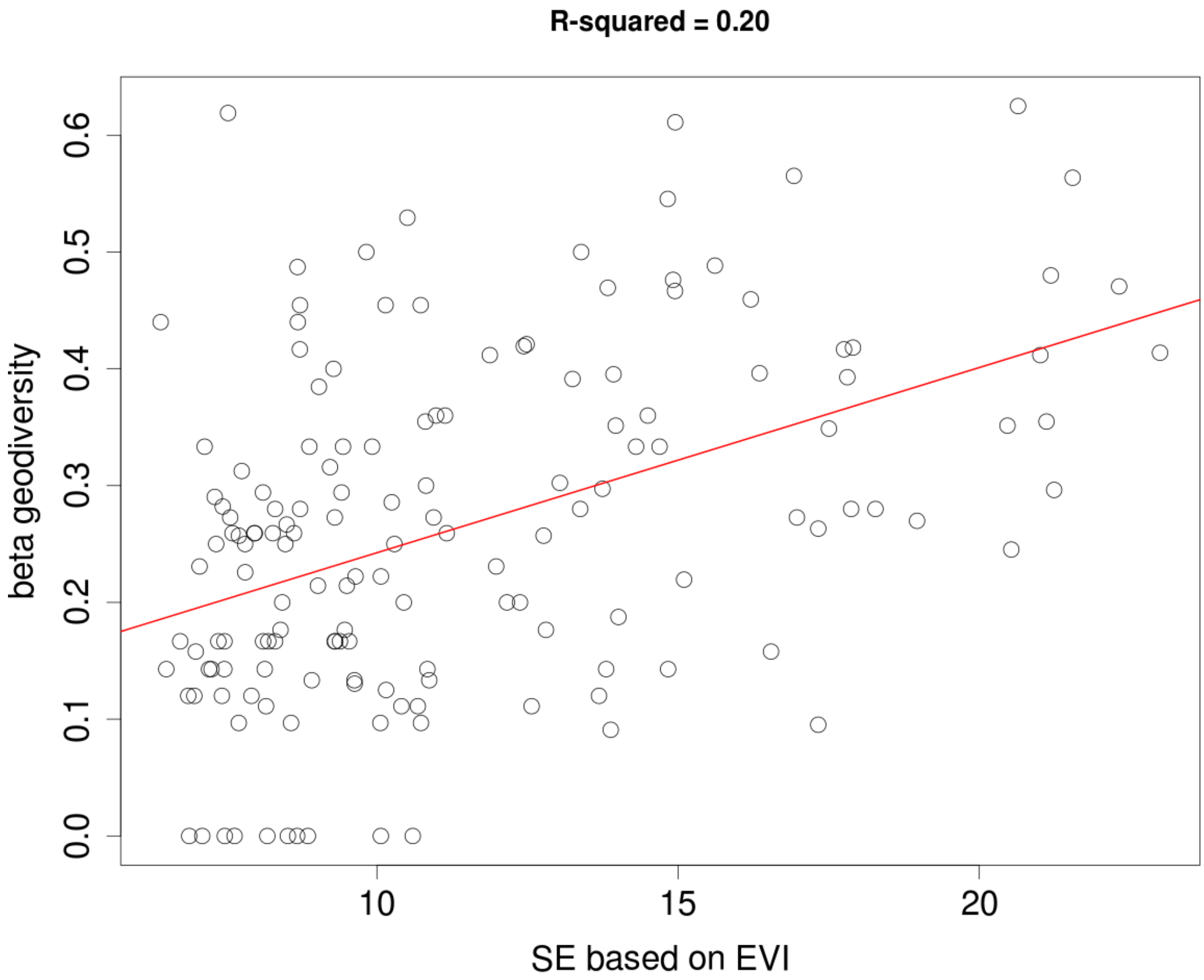


Figure 7. Correlation between beta geodiversity and sampling effort (SE) based on linearly scaled EVI values aggregated at 20 km (SE scale: 5 – 25). The beta geodiversity is estimated from changes in sediment grain size cover in 10 km pixels nested within 20 km pixels.

9. FUTURE DEVELOPMENTS AND DIRECTIONS

The relatively good correlation of EVI with beta- and geobeta diversity shows that the EVI is capturing important environmental and spatial variability and as such is suitable for both regional long-term and detailed cruise planning. However, it should still be regarded as work in progress, and further investigation to fine-tune and optimise EVI estimations will take place. One of the things that require further investigation concerns the optimal level of aggregation.

The use of a measure of multivariate dispersion may provide an alternative way of estimating environmental and spatial variability. For example, Anderson et al. (2006) found a significant and positive correlation between the Euclidean distance to the group centroid of normalised environmental data and the Sørensen distance (dissimilarity) to the group centroid of biological data.

Despite the good correlation found between EVI and biological diversity from our preliminary testing, substantial effort needs to be put into testing and refinement of the modelling stage. Testing of a wider range of measures for both biological and geological diversity is required, as well as testing and evaluation of the effect of EVI aggregation level and its relation to diversity measures. In relation to geological diversity, we intend to spatially combine (overlay) different measures of geological features, which will provide us with composite classes that potentially may have a better correlation structure with the EVI than sediment classes alone. Future work also needs to address the issues introducing noise in the data in order to reduce them to a minimum. This is crucial for delivering the best available data for both management and research purposes in the future.

Access to good bathymetric and oceanographic data has been crucial to the development of a reliable EVI. Ideally, data need to be available several years in advance to be included in strategic planning of sampling. To directly optimise the cost-effectiveness of sampling cruises and facilitate detailed planning all data, including multibeam and oceanographic model data should be available for mapping areas several months in advance of the sampling cruise. Finally, optimal allocation of sampling effort will not only make MAREANO surveys more cost-effective but also provide a better basis for the production of MAREANO deliverables.

10. ACKNOWLEDGMENTS

Thanks to Vidar Lien at IMR for compiling and providing models of oceanographic variables.

11. REFERENCES

Anderson MJ, Ellingsen KE, McArdle BH (2006). Multivariate dispersion as a measure of beta diversity. Ecology Letters 9: 683 – 693.

Baselga A 2010. Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography 19: 134 – 143.

Baselga, A, Orme, CDL (2012). betapart: an R package for the study of beta diversity. Methods in Ecology and Evolution 3: 808–812.

Clements AJ, Strong JA, 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.

Elvenes S, Dolan MFJ, Buhl-Mortensen P, Bellec V (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.

Erikstad L, Bakkestuen V, Bekkby T, Halvorsen R (2013). Impact of Scale and Quality of Digital Terrain Models on Predictability of Seabed Terrain Types. Marine Geodesy 36: 2-21.

Grabowski RC, Droppo IG, Wharton G (2011). Erodibility of cohesive sediment: the importance of sediment properties. Earth-Science Reviews 105: 101-120.

Hale R, Calosi P, McNeill L, Mieszkowska N, Widdicombe S (2011). Predicted levels of future ocean acidification and temperature rise could alter community structure and biodiversity in marine benthic communities. Oikos 120: 661-674.

Harbitz A, Aschan M, Sunnanå K (1998). Optimal effort allocation in stratified, large area trawl surveys, with application to shrimp surveys in the Barents Sea. Fisheries Research 37: 107-113.

Harris PT, Baker EK (2012). Seafloor Geomorphology as Benthic Habitat. GeoHAB Atlas of Seafloor Geomorphic Features and Benthic Habitats. Elsevier, p. 936.

Hiscock K, Southward A, Tittley I, Hawkins S (2004). Effects of changing temperature on benthic marine life in Britain and Ireland. Aquatic Conservation: Marine and Freshwater Ecosystems 14: 333-362.

Hobson RD (1972). Surface roughness in topography: quantitative approach. Pages 221–245 in RJ Chorley, editor. Spatial analysis in geomorphology. Harper and Row, New York, New

York, USA.

Johnson MP, Frost NJ, Mosley MWJ, Roberts MF, Hawkins SJ (2003). The area-independent effects of habitat complexity on biodiversity vary between regions. Ecology Letter 6: 126-132.

Pike RJ, Evans IS, Hengl T (2009). Geomorphometry: A brief guide. Developments in Soil Science 33: 3-30.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rosenberg R, Nilsson HC, Hellman B, Agrenius S (2000). Depth correlated benthic faunal quantity and infaunal burrow structures on the slopes of a marine depression. Estuarine, Coastal and Shelf Science 50: 843-853.

Rosenberg R, Agenius S, Hellman B, Nilsson HC, Norling K (2002). Recovery of marine benthic habitats and fauna in a Swedish fjord following improved oxygen conditions. Marine Ecology Progess Series 234: 43-53.

Rudberg S (1968). Geology and morphology. In: Sømme A. (ed.) A geography for Norden. Universitetsforlaget, Oslo, s. 31–47.

Sappington JM, KM Longshore, DB Thomson (2007). Quantifying Landscape Ruggedness for Animal Habitat Anaysis: A case Study Using Bighorn Sheep in the Mojave Desert. Journal of Wildlife Management 71: 1419 -1426.

van Son TC, Halvorsen R, Norling K, Bakke T, Kaurin M, Melsom F (2014). Identification of fine-scale marine benthic ecoclines by multiple parallel ordination. Journal of Marine Biology Volume 2014: Article ID 462529: 23 pages.

van Son TC, Gonzales-Mirelis G, Dolan M, Thorsnes T, Bjarnadóttir LR, Buhl-Mortensen P (in prep). Future sampling strategy in MAREANO: towards a more objective methodology. NGU report 2015.001.

Stephens D, Diesing M (2014). A comparision of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. PLoS ONE 9: e93950.

Todd BJ, Greene HG (2007). Mapping the Seafloor for Habitat Characterization. Geological Association of Canada Special Paper 47. Geological Association of Canada, St. Johns, Newfoundland.

Whittaker RH (1960). Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 30: 279-338.

Whittaker RH (1972). Evolution and measurement of species diversity. Taxon 21: 213-251.

12. APPENDICES

Appendix 1 (R Markdown document): Development of an Environmental Variability Index (EVI)

Appendix 2 (R Markdown document): Correlation between EVI and biological and geological diversity

Appendix 1: Development of an Environmental Variability Index (EVI)

Thijs van Son & Margaret Dolan

21/04/2015

This document documents the R code related to the development of an Environmental Variability Index (EVI) for MAREANO. The main purpose of the EVI is to, based on measures of environmental and spatial variability, guide both long-term strategic cruise planning and total sampling effort going into detailed cruise planning. For developing this EVI we used an area covering the Norwegian coast from the Varangerfjord in the Northeast to the island of Røst (south of Lofoten) in the South.

Loading packages and functions

```
library(magrittr)
library(ncdf)
library(raster)
library(rgdal)
library(gdalUtils)

source('../R/functions.R')
```

Importing bathymetry data from GEBCO

A new version of GEBCO was released December 2014 (http://www.gebco.net/about_us/news_and_events/gebco_2014_grid.html). The downloaded extent of the data was made sure to cover the extent of our oceanographic variables. At the time of downloading the data, GEBCO provided data in netCDF format (now also available as ESRI ASCII grids and geoTiffs) and has a latlong coordinate reference system. For our purpose we reprojected it to UTM 33. We downloaded the 30 arcsec version of the data, which has a resolution of about 425 m. While reprojecting the data to UTM 33 we also resampled it to 400 m. This was done to match the oceanographic data of 800 m (subsequently disaggregated to 400 m). The GEBCO netCDF file is imported using the raster() function before its exported to geotiff (appropriate format for the gdalwarp() function).

```
## Import netcdf file directly using raster()
gebco2014tmp <- raster('../input/GEBCO_2014_2D_5.0_63.0_40.0_75.0.nc')

## Exporting to geotiff
gebco2014tmp %>%
  writeRaster(filename = '../input/GEBCO2014_latlong',
              format = 'GTiff',
              overwrite = TRUE)

## Reprojecting to utm33
## Using gdalwarp() from gdalUtils
# Resampling to 800m res in the same call
# Works, but get extra padding with 0's... Should be able to remove with crop and mask
  gdalwarp(srcfile = '../input/GEBCO2014_latlong.tif',
```

```
           dstfile = '../input/GEBCO2014_utm33_400m.tif',
           t_srs = 'EPSG:32633',
           tr = c(400, 400),
           r = 'bilinear',
           dstnodata = NA,
           output_Raster = FALSE,
           overwrite = TRUE,
           verbose = TRUE)
```

Loading rasters

```
## input directory for oceanographic data
inputOcean <- '/home/thijs/Dropbox/Geodata/mareano/oceanography/utm33_800m/nanRemoved/'

## GEBCO 2014 bathymetry
gebco2014_400m_full <- '../input/GEBCO2014_utm33_400m.tif' %>%
  raster

## Oceanography
salMin <- 'S_min_800m_utm33.tif' %>%
  paste(inputOcean, ., sep = '') %>%
  raster
res(salMin) <- 800

salMax <- 'S_max_800m_utm33.tif' %>%
  paste(inputOcean, ., sep = '') %>%
  raster
res(salMax) <- 800

tempMin <- 'T_min_800m_utm33.tif' %>%
  paste(inputOcean, ., sep = '') %>%
  raster
res(tempMin) <- 800

tempMax <- 'T_max_800m_utm33.tif' %>%
  paste(inputOcean, ., sep = '') %>%
  raster
res(tempMax) <- 800

currVMax <- 'CurrV_max_800m_utm33.tif' %>%
  paste(inputOcean, ., sep = '') %>%
  raster
res(currVMax) <- 800

currUMax <- 'CurrU_max_800m_utm33.tif' %>%
  paste(inputOcean, ., sep = '') %>%
  raster
res(currUMax) <- 800
```

Calculating derived variables

Temperature and salinity ranges

Temperature and salinity ranges were calculated using their respective max and min rasters.

```
salRange <- salMax - salMin
tempRange <- tempMax - tempMin
```

Combining current speed variables

When two variables are highly collinear, a solution can be to scale them (mean = 0 and sd = 1) and then take the average of them. Here, the current variables are on the same scale, so we only need to average them. In the end we chose to use both variables alone, and not the combined one. The question is if, in the context of EVI, collinearity is a problem or not? We believe its not a big problem since the EVI is standardised by the number of variables used as input. Recall that we are simply interested in distinguishing the environmental and spatial variability between areas. It should not pose a problem if two variables show somewhat similar patterns in terms of environmental and spatial variability.

```
## Using overlay() to take the mean of the two current layers
currUVMax <- overlay(x = currUMax,
                     y = currVMax,
                     fun = mean)
```

Disaggregating/resampling oceanographic data to 400 m resolution

The oceanographic variables need to be disaggregated to 400 m resolution to match the GEBCO data. The disaggregate() function in the raster package is relatively slow, maybe the gdalwarp() function in the gdalutils package is an option.

```
salRange_400m <- salRange %>%
  disaggregate(fact = 2,
               method = 'bilinear')

tempRange_400m <- tempRange %>%
  disaggregate(fact = 2,
               method = 'bilinear')

tempMin_400m <- tempMin %>%
  disaggregate(fact = 2,
               method = 'bilinear')

currUMax_400m <- currUMax %>%
  disaggregate(fact = 2,
               method = 'bilinear')

currVMax_400m <- currVMax %>%
  disaggregate(fact = 2,
               method = 'bilinear')

currUVMax_400m <- currUVMax %>%
  disaggregate(fact = 2,
               method = 'bilinear')
```

Cropping and masking the GEBCO data

The GEBCO data need to be of the exact same extent (and resolution) as the oceanographic data.

```
## First cropping
gebco2014_400m_crop <- gebco2014_400m_full %>%
  crop(extent(salRange_400m))

## Then aligning the extent
extent(gebco2014_400m_crop) <- extent(salRange_400m)

## Then masking values outside the new extent
gebco2014_400m_mask <- mask(gebco2014_400m_crop, salRange_400m)
```

Vector ruggedness measure (VRM) and relative relief

Both VRM and relative relief are derived from the bathymetry data and needed to be calculated after the GEBCO data had been cropped and masked. We used a neighbourhood of 9 for the VRM and a neighbour hood of 3 for the relative relief. Code for the custum-made relRelief() function can be found in the Function repository at the end of the this document.

```
## Loading VRM
vrm9_400m <- '../input/vrm_gebco2014_400m_9.tif' %>%
  raster

## Calculating relative relief
relRelf3_400m <- relRelief(rast = gebco2014_400m_mask,
                           size = 3)
```

Storing environmental layers with 400 m res in a raster stack

Now that all variables have the same resolution and extent, they can be stored in a raster stack. Storing in a raster stack is like storing in a list and is convenient when repetitive analyses will be applied to each layer.

```
vars400m_Orig_stack <- stack(list(gebco2014_400m_mask,
                                  relRelf3_400m,
                                  vrm9_400m,
                                  salRange_400m,
                                  tempMin_400m,
                                  currVMax_400m,
                                  currUMax_400m,
                                  currUVMax_400m))

names(vars400m_Orig_stack) <- c('gebco2014_400m_mask',
                                'relRelf3_400m',
                                'vrm9_400m',
                                'salRange_400m',
                                'tempMin_400m',
                                'currVMax_400m',
                                'currUMax_400m',
                                'currUVMax_400m')
```

Linearly rescaling rasters between 5 and 260

Rasters are on different numeric scales and need to be linearly rescaled to the same scale in order to be able to compare their variability. Often such linear rescaling is done using an 8-bit scale, i.e., between 0 and 255. However, we do not want values close to zero, because they can cause troubles when estimating the coefficient of variation. Therefore, we shift the scale upwards by five units yielding a scale between 5 and 260. A custom-made function rang() is used to achieve this (see Function repository).

```
## Copying the metadata of vars400_Orig_stack
vars400m_260_stack <- vars400m_Orig_stack

## Looping over every environmental layer in stack
for(i in seq(nlayers(vars400m_260_stack))) {
  var.i <- vars400m_Orig_stack[[i]]
  scaled.i <- calc(var.i,
                  fun = function(x) x %>% rang)
  vars400m_260_stack[[i]] <- scaled.i
}

names(vars400m_260_stack) <- c('gebco2014_400m_mask',
                               'relRelf3_400m',
                               'vrm9_400m',
                               'salRange_400m',
                               'tempMin_400m',
                               'currVMax_400m',
                               'currUMax_400m',
                               'currUVMax_400m')
```

Calculation of coefficient of variation

The coefficient of variation (CV; defined as SD/mean) is calculated for each environmental variable in the raster stack. The calculations are based on the linearly rescaled rasters and pixel-wise CV values are aggregated at 20 km. Other aggregations are also possible and have been tested (not shown here). Aggregation is necessary for having observations to estimate variability from. A custom made function aggrFunc() is used to achieve this. The function's code is appended to the function repository at the end of this document.

```
## Coefficient of variation for each env variable
cvs_20km_400m_260 <- aggrFunc(rs = vars400m_260_stack, fact = 50)
```

EVI and recommended sampling effort (SE) calculations

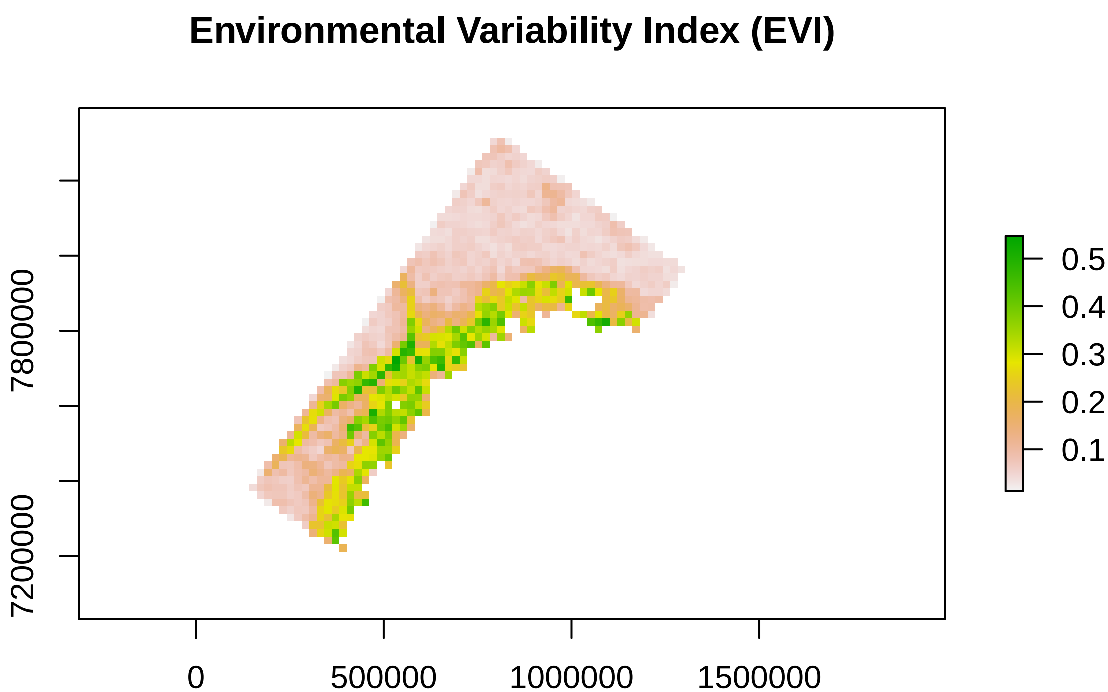
The EVI calculates the pixel-wise mean environmental dispersion based on the input environmental variables. It is possible to give weight to certain environmental variables if deemed appropriate, then EVI is standardised by the sum of weights. By default, each variable is given weight 1. Each pixel returns the proportion of dispersion (variability) in relation to the mean of the variables (recall that CV = sd/mean). The EVI values is then translated into suggested/recommended SE using a linear model that scales the SE between the minimum and maximum desired SE. A custom-made function eviScaledSE() takes a raster stack as input and min and max desired SE as arguments. The function returns EVI, suggested SE, and the mean SE (to control for total sampling effort within a larger area). Function code available in the function repository.

```r
## EVI and SE
# All variables (except the combined current variable)
SE20km_400m_All <- eviScaledSE(rsCV = cvs_20km_400m_260[[-8]],
                               seMin = 5,
                               seMax = 23)
SE20km_400m_All
```

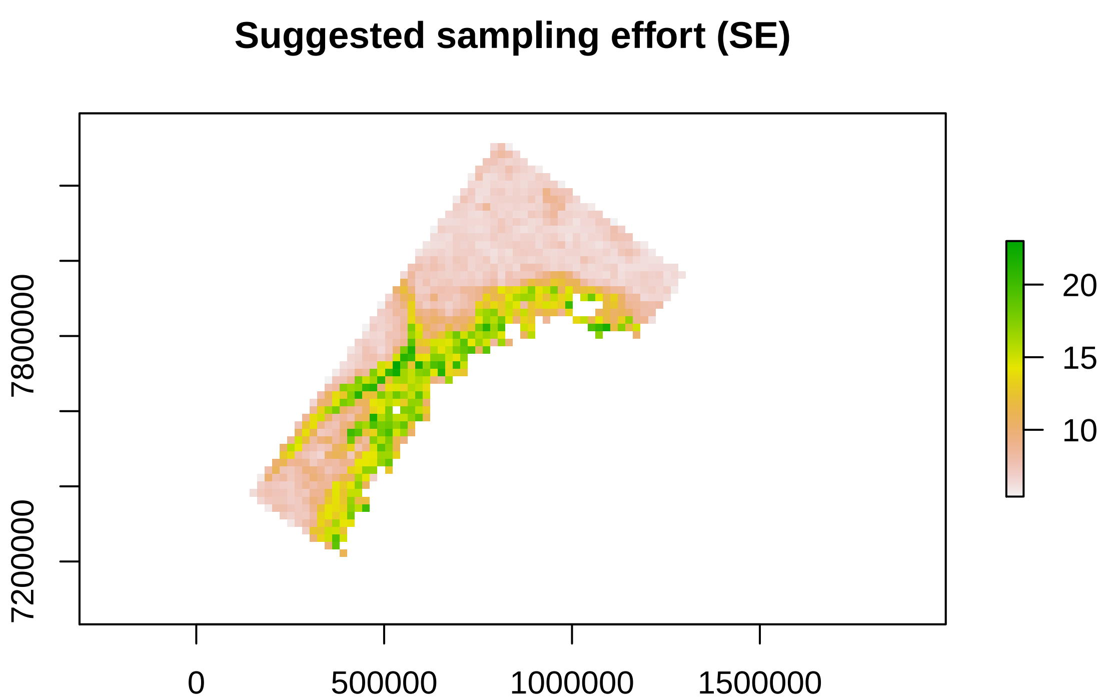
```
## $EVI
## class       : RasterLayer
## dimensions  : 68, 72, 4896  (nrow, ncol, ncell)
## resolution  : 20000, 20000  (x, y)
## extent      : 122031.7, 1562032, 7032714, 8392714  (xmin, xmax, ymin, ymax)
## coord. ref. : +proj=utm +zone=33 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0
## data source : in memory
## names       : layer
## values      : 0.01217788, 0.5475778  (min, max)
##
##
## $SE
## class       : RasterLayer
## dimensions  : 68, 72, 4896  (nrow, ncol, ncell)
## resolution  : 20000, 20000  (x, y)
## extent      : 122031.7, 1562032, 7032714, 8392714  (xmin, xmax, ymin, ymax)
## coord. ref. : +proj=utm +zone=33 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0
## data source : in memory
## names       : layer
## values      : 5.400312, 23  (min, max)
##
##
## $wts
##                 envVar weight
## 1 gebco2014_400m_mask      1
## 2       relRelf3_400m      1
## 3           vrm9_400m      1
## 4       salRange_400m      1
## 5        tempMin_400m      1
## 6       currVMax_400m      1
## 7       currUMax_400m      1
##
## $meanSE
## [1] 9.978159
```

Plotting EVI and SE

```r
## EVI
plot(SE20km_400m_All$EVI,
     main = 'Environmental Variability Index (EVI)')
```




```
## SE
plot(SE20km_400m_All$SE,
     main = 'Suggested sampling effort (SE)')
```



Function respository

```r
## Relative relief function

# relRelief() uses a wrapper function to focal(), focalWrap().
# focal() is a moving window (neighbourhood) function in the raster package.

focalWrap <- function(rast, size, func, ...) {
  ## rast: the raster to be analysed
  ## size: the size of the neighbourhood (moving window)
  ## w: a matrix defining the moving window size and weight given to each
  # pixel in the neighbourhood
  ## fun: function to be applied to the neighbourhood
  focal(x = rast,
        w = matrix(1, nrow = size, ncol = size),
        fun = func # ,
        # pad = TRUE,
        # padValue = mean(rast)
        )
}

relRelief <- function(rast, size, ...) {
  ## rast: raster to be analysed
  ## size: size of neighbourhood (moving window)
  mn <- focalWrap(rast = rast,
                  size = size,
                  func = min)
  mx <- focalWrap(rast = rast,
                  size = size,
                  func = max)
  relRel <- mx - mn
  relRel
}


## Ranging/Scaling function
# Scales a values of a raster between 5 and 260

rang <- function(x) {
  ## x: the raster to be rescaled
  rng <- rep(NA_real_, length(x))
  minx <- min(x, na.rm = T)
  maxx <- max(x, na.rm = T)
  for(i in seq(length(x))) {
    rng[i] <- (((x[i] - minx) / (maxx - minx)) * 255) + 5
  }
  return(rng)
}


## Function for calculating coefficient of variation (CV) of rasters

aggrFunc <- function(rs,
                     fact,
                     # expand = TRUE,
```

```r
                     # na.rm = TRUE,
                     ...) {
  ## rs: raster stack of variables
  ## fact: aggregating factor

  # Basic info of raster stack
  n <- nlayers(rs)
  nms <- names(rs)

  # raster stack template
  template <- rs[[1]] %>%
    aggregate(fact = fact,
              fun = mean,
              expand = TRUE,
              na.rm = TRUE)
  # Filling the output raster stack temporarily with the template
  templateList <- replicate(n, template)
  aggrCv <- stack(templateList)

  # Calculating the CVs
  for(var in seq(n)) {
    rs.var <- rs[[var]]
    mn.var <- rs.var %>%
      aggregate(fact = fact,
                fun = mean,
                expand = TRUE,
                na.rm = TRUE)
    sd.var <- rs.var %>%
      aggregate(fact = fact,
                fun = sd,
                expand = TRUE,
                na.rm = TRUE)
    cv.var <- sd.var / mn.var
    aggrCv[[var]] <- cv.var
    names(aggrCv)[var] <- nms[var]
  }
  aggrCv
}


## Function for calculating EVI, suggested SE, and mean SE

eviScaledSE <- function(rsCV,
                        seMin,
                        seMax,
                        wts = rep(1, nlayers(rsCV))) {
  ## rs: rasterstack of CVs aggregated at a certain spatial level
  ## seMn: the minimum scaled sampling effort desired
  ## seMx: the maximum scaled sampling effort desired
  ## wts: the desired weight given to environmental layers (Default is equal weight)

  # EVI computation
  EVI <- calc(rsCV * wts, fun = sum) / sum(wts)
```

```
  # Extract max EVI
  eviMax <- EVI %>%
    values %>%
    max(na.rm = TRUE)

  # Linearly scaled sampling effort based on EVI
  SE <- seMin + ((seMax - seMin) / eviMax) * EVI

  # Output
  list(EVI = EVI,
       SE = SE,
       wts = data.frame(envVar = names(rsCV),
                        weight = wts),
       meanSE = SE %>% values %>% mean(na.rm = TRUE))

}
```

Appendix 2: Correlation between EVI and biological and geological diversity

Thijs van Son, Genoveva Gonzales-Mirelis & Margaret Dolan

21/04/2015

Herein we document how we tested the relationship between the environmental variability index (EVI) and biological and geological diversity. For reasons explained more thoroughly in the main report, we are not focusing on the richness of biological and geological features, but rather the change in composition of them over space (turnover and nestedness of features). In ecology this is called beta diversity and in geology we have coined the term beta geodiversity to represent the geological equivalent. We first document the relationship between biology and EVI and subsequently between geology and EVI.

Correlation between biological diversity and EVI

The biological diversity is measured from abundances of species observed in video lines of about 700 m length. We use beta diversity, which in our case measures the changes (in terms of turnover and nestedness) in species composition between multiple lines (three or more) that fall within an EVI pixel/grid cell. The beta.multi() function in the betapart package allows for calculating beta diversity from multiple observations (video lines in our case) after the species abundance data has been reduced to presence-absence.

Loading packages

```
library(betapart)
library(magrittr)
library(raster)
library(vegan)
library(rgeos) #gIntersects
```

Loading spatial data

```
## EVI values
evi20_all <- '../input/SE20km_linScal_GEB14_All.tif' %>%
  raster

## Species compositional data
sppdata <- '../input/AllSpeciesData.csv' %>%
  read.csv(header = TRUE)

## stations with positional reference
stations <- '../input/stations.csv' %>%
  read.csv(header = TRUE)

## Converting stations into SpatialPointsDataFrame
coordinates(stations)<-~pos_long+pos_lat
```

```
proj4string(stations) <-
  CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")

## Reprojecting to UTM33
stations <-
  spTransform(stations,
              CRS=CRS("+proj=utm +zone=33 +datum=WGS84 +units=m +no_defs
                      +ellps=WGS84 +towgs84=0,0,0"))
```

Assigning biological stations to groups according to EVI grid cells

Here we are counting the number of stations that are within the boundaries of different EVI grid cells, i.e., within grid cells of 20 km.

```
## Clipping
grid <-rasterToPolygons(evi20_all)

int <- gIntersects(stations, grid, byid = T)

clipped <- apply(int == F, MARGIN = 2, all)

stations.cl <- stations[which(!clipped), ]

stations <- stations.cl

rm(stations.cl)

## Spatial join (including count of stations in each grid cell)

int <- gIntersects(stations, grid, byid = T)  # re-run the intersection query
b.indexes <- which(int, arr.ind = T)

b.indexes <- cbind(b.indexes, stations@data[, 2:3])
colnames(b.indexes) <- c('gridid',
                         'pointid',
                         'refstation_no',
                         'sample_no')

b.count <- aggregate(b.indexes$sample_no,list(b.indexes$gridid), length)
colnames(b.count)<-c('gridid','count_of_stations')
```

Calculating beta diversity

As explained above, a single value for beta diversity is calculated for each EVI grid cell that contains at least three video lines.

```
## Select cells with enough stations (let's say >= 3; max is 10)
sel<-b.count[b.count$count_of_stations > 2, ][ ,1]

## Filtering species data
sppdata.sel<-sppdata[sppdata$Taxon_class > 0,]
```

```r
sppmatrix<-xtabs(sppdata.sel$Value ~
                   sppdata.sel$xSampleNumber +
                   sppdata.sel$Taxon_class)

sppmatrix<-as.data.frame.matrix(sppmatrix)

## Calculating beta diversities
iterations = sel
variables = 2

output <- matrix(ncol = variables,
                 nrow = length(iterations))
for(i in 1:length(iterations)) {
  output[i,2] <- beta.multi(
    decostand(sppmatrix[rownames(sppmatrix) %in%
                          b.indexes[b.indexes$gridid == sel[i], 4], ], "pa"),

    index.family = "sor")[[3]]
}

output[, 1] <- sel

colnames(output)<-c("grid_ID","oa_betad")
```

Extracting EVI in pixels that contain beta diversity values

EVI values for grid cells containing three or more video lines were extracted for later use in linear modelling.

```r
### EVI extraction
x_all <- raster::extract(evi20_all, grid)
x_all <- unlist(x_all)

## Grid ID extraction
gridid <- as.numeric(rownames(grid@data))

## Data frame with model data
model.data <- data.frame(grid_ID = gridid,
                         evi20_all = x_all)

model.data1 <- model.data[sel, ]

model.data1 <- merge(output, model.data)
```

Fitting a linear model to test the correlation between EVI and beta diversity

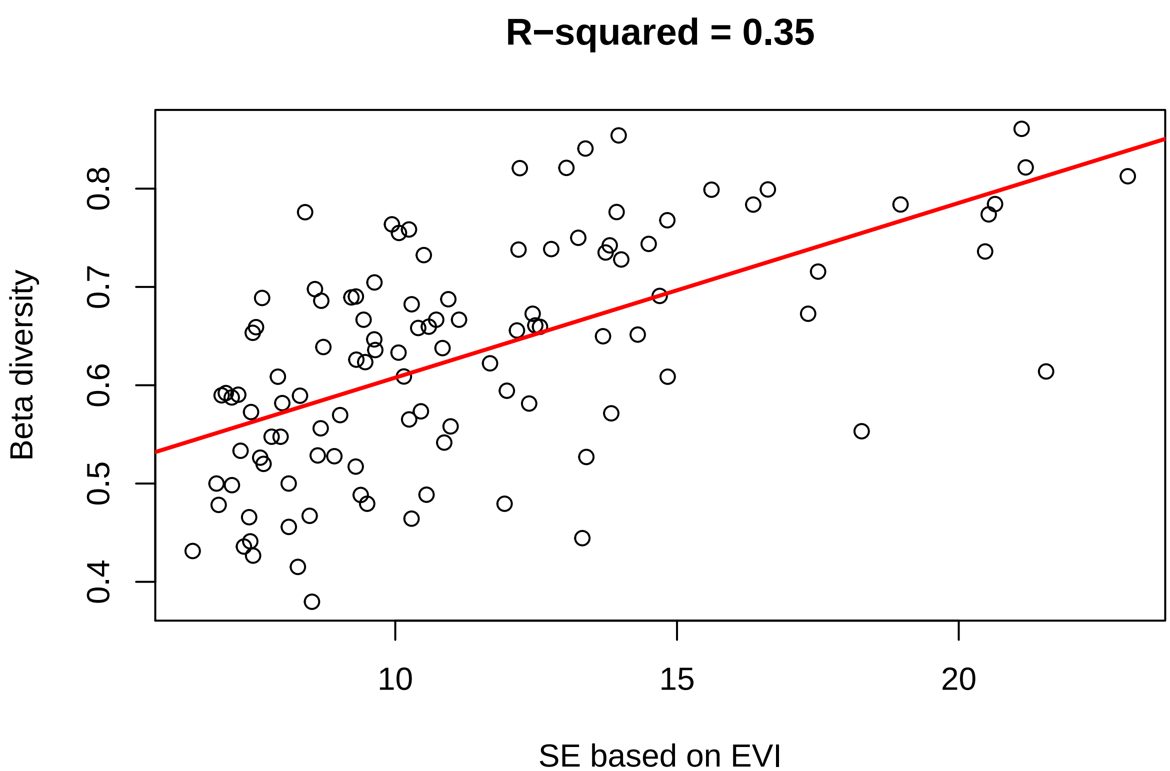
```r
summary(lm_evi20_all <- lm(oa_betad ~ evi20_all, data = model.data1))
```

```
##
## Call:
```

```
## lm(formula = oa_betad ~ evi20_all, data = model.data1)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.222297 -0.060736  0.009731  0.063137  0.196771
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.429984   0.027709  15.518  < 2e-16 ***
## evi20_all   0.017775   0.002321   7.658 8.95e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09287 on 107 degrees of freedom
## Multiple R-squared:  0.354,  Adjusted R-squared:  0.348
## F-statistic: 58.64 on 1 and 107 DF,  p-value: 8.949e-12
```

Plotting the relation between EVI and beta diversity

```r
with(model.data1,
  plot(oa_betad ~ evi20_all,
       xlab = 'SE based on EVI',
       ylab = 'Beta diversity',
       main = 'R-squared = 0.35'))
abline(lm_evi20_all,
       col = 'red',
       lwd = 2)
```



Correlation between geological diversity and EVI

For this exercise we used interpreted sediment grain size maps for estimating beta geodiversity. The basis for this estimation was to calculate the percent cover for each sediment class for pixels of size 10km, which were nested within the 20 km EVI pixels. In similar ways as for calculation of beta diversity, the beta geodiversity is based on the changes in composition (in terms of percent cover) of sediment grain size classes among the 10 km pixels.

Loading packages not already loaded above

```
library(sp)
library(rgdal)
library(ggplot2) #fortify
library(tidyr) #gather
library(dplyr)
library(raster)
```

Loading spatial data

Loading a 20 km fishnet/grid (made in ArcMap) that is perfectly aligned spatialy with the EVI grid cells. In addition, a point shapefile with a resolution of 10 km having the points nested within the 20 km grid is also loaded. The latter will be used for nesting data of the distribution of sediment grain size within the 20 km grid. Finally, a the sedData data set contains the spatial percent cover of the different sediment grain size classes within the 10 km grid.

```
## Shapefiles
fish10k_sedData <- readOGR(dsn = '../input/',
                           layer = '10k_fishnet_label')
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "../input/", layer: "10k_fishnet_label"
## with 19584 features and 3 fields
## Feature type: wkbPoint with 2 dimensions
```

```
fish20k <- readOGR(dsn = '../input/',
                   layer = '20k_fishnet')
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "../input/", layer: "20k_fishnet"
## with 4896 features and 2 fields
## Feature type: wkbPolygon with 2 dimensions
```

```
## Percent cover of sediment grain size classes at 10k resolution
sedData <- '../input/tabIntesect_10kfishnet_sedkornJ15_33_SE.csv' %>%
  read.csv
```

Adding FID's to shapefiles

It is crucial to add the FIDs (i.e. unique feature IDs) of both 10 and 20 km grids. Otherwise it would be impossible to correctly nest the smaller 10 km features within the larger 20 km features.

```
## fish10k with sed data
fish10k_sedData@data$FID_10 <- 0:(nrow(fish10k_sedData) - 1)

## fish20k polygons
# First use fortify
fish20k_fort <- fortify(fish20k)
```

```
## Regions defined for each Polygons
```

```
# Adding FID_20k
fish20k@data$FID_20 <- fish20k_fort$id %>%
  unique %>%
  as.numeric
```

Spatial overlay between 10 and 20k fishnets

Finding which 10 km grids that are nested within the 20 km grids.

```
## Nesting FID_10 within FID_20

## Between fish20k and sedData
overFID20_FID10 <- fish20k %>%
  over(y = fish10k_sedData,
       returnList = T) %>%
  lapply(function(x) x[[4]]) %>%
  unlist %>%
  matrix(nrow = nrow(fish20k@data),
         byrow = T) %>%
  data.frame(FID_20 = 0:(nrow(fish20k) - 1),
             .) %>%
  gather(sample, FID_10, X1:X4)
```

Joining the overlay with the sediment grain size cover data

Now that we know how the smaller grids are nested within the larger ones, we can add the sediment grain size cover data to the same data frame.

```
## Retaining only rows/observations in both sets

## Sediment cover
sedCover <- inner_join(x = overFID20_FID10,
                       y = sedData,
                       by = 'FID_10')
```

Selecting observations with high coverage and converting NAs to 0

We need to convert NAs in the data set that are actually zeros. Subsequently, we make a selection on the smaller grids that excludes every grid in which the total sediment cover is less than 90%. We also need to convert the sediment cover data to presence-absence for the same reason as for the biological data.

```
## Converting NAs to 0
sedCover[is.na(sedCover)] <- 0

## Selecting 10k pixels with cover >= 90%
sedCover_sel10 <- filter(sedCover, totalCover >= 90)

## 20k pixels with at least 3 10k pixels with cover >=90%
## Sediment
FID20_sel_sed <- group_by(sedCover_sel10, FID_20) %>%
  summarise(count = n()) %>%
  filter(count >= 3)


## Data frame based on FID20_sel
sedCover_sel20 <- filter(sedCover_sel10, FID_20 %in% FID20_sel_sed$FID_20)

## Convert to presence absence
sedCover_sel20_pa <- decostand(x = sedCover_sel20[, 4:16], method = 'pa')

# Adding FID_20
sedCover_sel20_pa$FID_20 <- sedCover_sel20$FID_20
```

Calculating beta geodiversity

Now we can calculate the extent of spatial turnover and nestedness (beta geodiversity) of sediment grain size.

```
betaSedDiv_20k <-
  data.frame(FID_20 = rep(NA_integer_,
                          length(FID20_sel_sed$FID_20)),
             betaseddiv = rep(NA_real_,
                              length(FID20_sel_sed$FID_20)))

for(i in seq(length(FID20_sel_sed$FID_20))) {
  fid.i <- FID20_sel_sed$FID_20[i]
  test.i <- filter(sedCover_sel20_pa, FID_20 %in% fid.i) %>%
    select(-FID_20) %>%
    beta.multi(index.family = 'sorensen')
  betaSedDiv_20k[i, 1] <- fid.i
  betaSedDiv_20k[i, 2] <- test.i$beta.SOR
}
```

Coupling beta geodiversity values with correponding EVI values

```
## Selecting grids with betageodiv values
# Also removing unnecessary variables
```

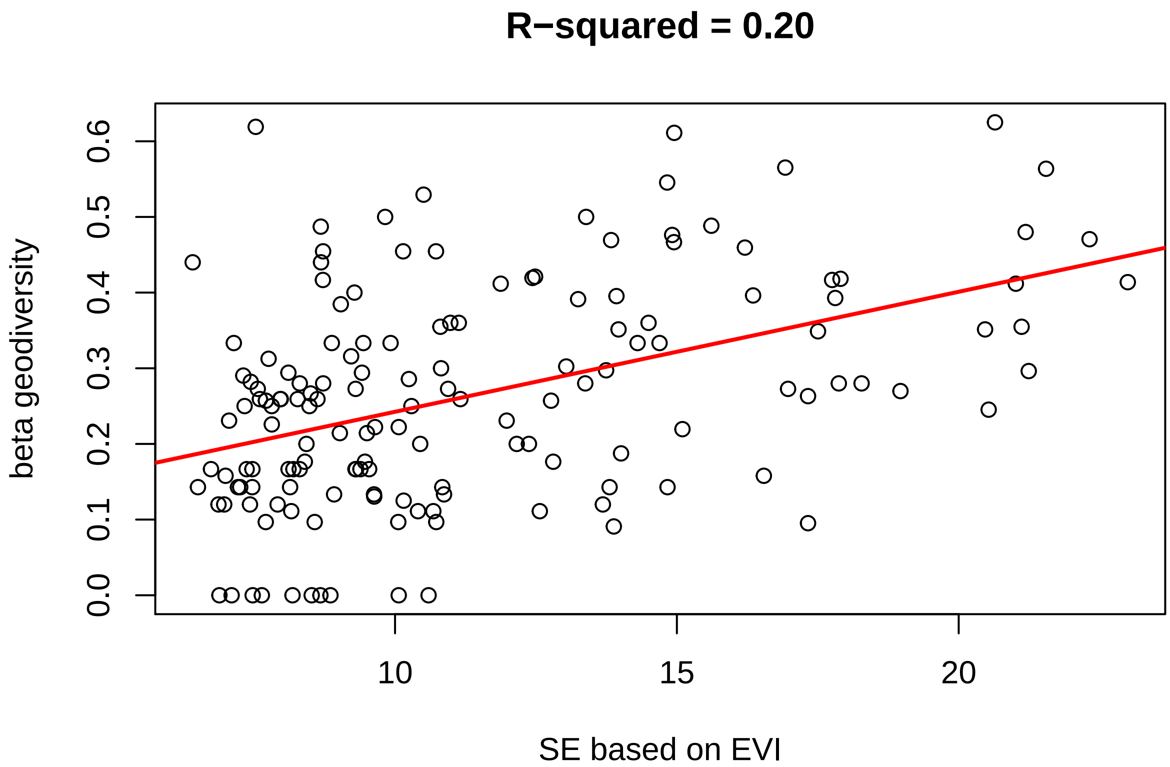
```r
# Adding one to the index, because FIDs start at 0
fish20k_sel_sed <- fish20k[(FID20_sel_sed$FID_20 + 1), -c(1:2)]

## Adding betageodiv values
fish20k_sel_sed@data$betaseddiv <- betaSedDiv_20k$betaseddiv

## Extracting EVI values for selected betageodiv pixels
## Sediment
fish20k_sel_sed@data$evi20k_all <- evi20_all %>%
  raster::extract(fish20k_sel_sed) %>%
  unlist %>%
  as.vector

## Removing observations having NAs
## Sediment
fish20k_selSedFinal <-
  fish20k_sel_sed@data[complete.cases(fish20k_sel_sed@data), ]
```

Linear modelling of the relationship between beta geodiversity and EVI

```r
## Sediment
summary(lm_betageoEVI <-
          lm(betaseddiv ~ evi20k_all, data = fish20k_selSedFinal))
```

```
##
## Call:
## lm(formula = betaseddiv ~ evi20k_all, data = fish20k_selSedFinal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26344 -0.09008 -0.01266  0.08264  0.41576
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.083890   0.031417   2.670  0.00842 **
## evi20k_all  0.015859   0.002598   6.105 8.39e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1305 on 150 degrees of freedom
## Multiple R-squared:  0.199,  Adjusted R-squared:  0.1937
## F-statistic: 37.28 on 1 and 150 DF,  p-value: 8.386e-09
```

```r
with(fish20k_selSedFinal,
     plot(betaseddiv ~ evi20k_all,
          xlab = 'SE based on EVI',
          ylab = 'beta geodiversity',
          main = 'R-squared = 0.20'))
abline(lm_betageoEVI, col = 'red', lwd = 2)
```



R−squared = 0.20
beta geodiversity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
10
15
20
SE based on EVI

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/