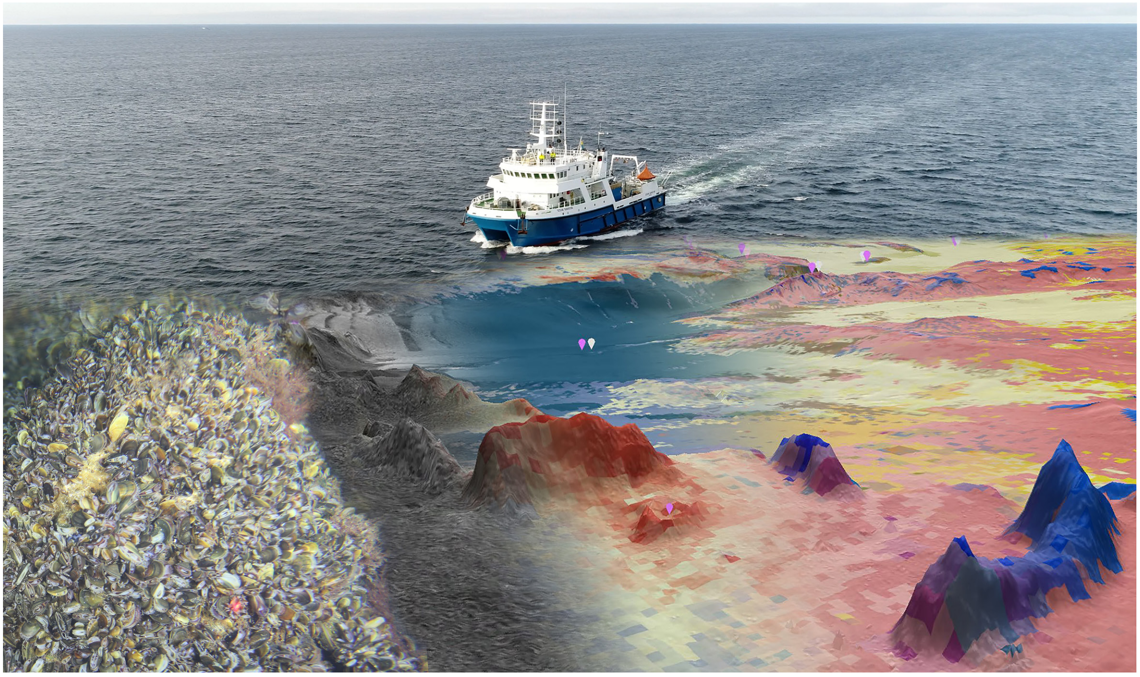
High-resolution benthic habitat mapping of Hoburgs bank, Baltic Sea

Gustav Kågesten, Finn Baumgartner & Francis Freire

November 2020

SGU-rapport 2020:34



SGU Geological Survey of Sweden

Cover image: Collage containing a drone image of SGU's research vessel R/V *Ocean Surveyor* (upper), an underwater image of a boulder covered by blue mussels (bottom left), and a digital seafloor model (bottom right).

Photographer: SGU

Author: Gustav Kågesten, Finn Baumgartner and Francis Freire
Reviewed by: Lovisa Zillén Snowball and Alexander Bäckström
Responsible Head of Department: Lovisa Zillén Snowball
Co-financed by the Swedish Agency for Marine and Water Management, SWAM, (id: 2234-16, 3901-16)

Editor: Johan Sporrong

Geological Survey of Sweden
Box 670, 751 28 Uppsala
tel: 018-17 90 00
email: sgu@sgu.se

www.sgu.se

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The large region of fairly shallow seafloor extending south from the island of Gotland to the central and south-western parts of the Baltic Proper, is relatively unaffected by coastal pollution, and contains important spawning, feeding and resting areas for many marine species, including seabirds and the critically endangered Baltic population of Harbor porpoises. Though connected, the area is often described by three large shoals or banks, the Northern and Southern Midsea banks (Norra- och Södra Midsjöbanken), and Hoburgs bank. Like most of the Baltic Sea, the ecosystems that depend on this region are under pressure from climate change, pollutants and eutrophication, as well as direct human activities such as fishing, energy production, sand extraction, infrastructure and shipping, with busy shipping lanes surrounding and crossing over the shoals.

This region, apart from the shallower areas of the Southern Midsea bank, was recently designated a Natura 2000 marine protected area, making it the largest of its kind in Sweden. Although the ecological importance of this region is well known, a more detailed understanding is missing, which means there is also a lack of critical knowledge on how to best manage and regulate it. In an effort to reduce this knowledge gap, the Geological Survey of Sweden (SGU) surveyed 1 344 km^2 of seafloor on Hoburgs bank from 2016–2017 using hydroacoustic techniques, and collected sediment samples and high-resolution underwater images. This information was then used to produce full coverage benthic habitat maps that illuminated a complex seascape that had never been described in this detail before.

Maps were created by integrating both geological and biological data to produce flexible continuous coverage (%) models of substrate and biological components. These continuous models were combined to produce benthic habitat maps according to different classification schemes (e.g. HELCOM HUB and Natura 2000) but can also be used individually according to end-user needs. In addition, the maps and models produced by this method are high-resolution (5 m), allowing them to be upscaled to relevant management scales, while still retaining important information captured by the high-resolution data.

The overall accuracy of HELCOM HUB maps were 80.5% for level 3 (substrate), 81.9% for level 4 (community structure), 62.3% for level 4–5 (characteristic community), and 53.2% for level 4–6 (dominating taxa). The overall accuracy for Natura 2000 reef/sandbanks were 87.7%. The mapped area on Hoburgs bank ranged from 10–63 m in depth, with a majority of the area lying between 15 and 35 m. Sand, gravel, and pebbles and stones dominated the area covering approximately 80% of the seafloor surface, while hard bottom (i.e. large stones-large boulders, and hard clay) covered approximately 19.3%. The remaining 0.7% consisted of softer sediments (i.e. soft clay-silt). 39% of the bank was classified as a Natura 2000 reef, while the remaining area (61%) was classified as a Natura 2000 sandbank. Wave and currents create a dynamic seafloor environment in the area, and evidence of sand transport was found across the whole bank. 90% of the area had either sand with ripples formed (49%), or coarse-hard bottom with sand present indicating that sand transport was occurring. 9% of the seafloor had significant post-glacial sand deposits (generally deeper areas below 40 m), with no ripples indicating transport. Approximately 1% of the seafloor habitats were completely devoid of sand. These areas included the upper parts of large individual boulders, as well as a series of distinct moraine ridges (0.7% of the bank according to HELCOM HUB level 3 rock and boulder), and hard clay features (0.1% of the bank according to HELCOM HUB level 3). The structurally complex reefs were observed, from drop camera imagery, to be associated with a higher abundance and diversity of fish compared with other seafloor types on the bank. The hard bottom areas were generally colonised with organisms. Colonised seafloor covered 26%

of the bank with epibenthic bivalves (i.e. *Mytilus* spp. 10.7% cover) and filamentous red algae (14.8% cover) dominating the benthos. The shallowest areas (10–15 m) had a more diverse algal assemblage than the deeper areas and most annual algae species were found here (0.6% cover). The deepest record of algae was a filamentous red alga at 38 m depth. Observations in the deeper (~40–60 m) fine sand areas showed increased frequency of crawl marks and notable aggregations of *Macoma balthica* (now referred to as *Limecola balthica*) mussels, detritus and small crustaceans compared with shallower more dynamic sand areas, suggesting that it is important to consider the function of both deeper and shallower sand habitats when assessing the ecosystems in the area.

The results from the mapping suggested that modelling continuous variables of geology and biology to describe habitats at multiple thematic and spatial scales provided more accurate information to the end-user than only providing predefined thematic maps at a fixed scale. Moreover, the accuracies computed for thematic products derived from the continuous modelling approach were equivalent to or higher than products derived by a direct thematic modelling approach. The value of combining high-resolution depth with backscatter data from multibeam sonar was found to be instrumental to the success of the models, highlighting that access to backscatter data needs to be prioritised if the maps of Hoburgs bank are to be expanded, for example, to cover the larger Natura 2000 area with a similar level of thematic and spatial detail.

This report reproduces some of the content presented in a scientific publication by Kågesten et al. 2019 but expands on information about the environments found on Hoburgs bank. It discusses the importance of legacy data in the modelling process and the use of modelled sediment-type predictor variables as inputs in producing the final habitat maps as a means to improve map accuracy. Finally, the report highlights the limitations of existing regional habitat classification schemes and underlines the need for more standardised national and regional mapping initiatives that will allow for better integrated coastal management. All the maps as well as survey data from the project are freely available through SGU's customer service. Methods to publish the amount of material collected during the Hoburgs bank survey through online services are not yet available at SGU but hopefully will be in the near future.

SAMMANFATTNING

En omfattande region av relativt grund havsbotten sträcker sig söderut från Gotland till de centrala och sydvästra delarna av Egentliga Östersjön, och inkluderar tre stora utsjöbanker; Norra- och Södra Midsjöbanken, och Hoburgs bank. Området är relativt opåverkad av kustföroreningar, och innehåller viktiga livsmiljöer för bland annat sjöfågel och den akut hotade Östersjöpopulationen av tumlare. Med bakgrund av detta blev Hoburgs bank och Midsjöbankarna (förutom den grunda delen av Södra Midsjöbanken) i december 2016 utpekat som Natura 2000-område. Området ligger delvis inom Kalmar och Gotlands län och är med sina 10 511 km^2 det i särklass största marina reservatet i Sverige idag. Som de flesta områdena i Östersjön är denna regions ekosystem påverkade av klimatförändringar, föroreningar, och övergödning. De är också direkt påverkade av mänskliga aktiviteter som fiske, infrastruktur och sjöfart, med väl trafikerade farleder som både går runt och korsar bankerna och även gasledningen Nord Stream som går igenom området. I framtiden kan även aktiviteter som energiproduktion och sandutvinning vara aktuellt. Idag är skyddet bara ”på pappret” då arbetet med utformningen och implementeringen fortfarande pågår. Även om den ekologiska betydelsen av denna region var välkänd redan när området pekades ut 2016 så saknades en mer detaljerad förståelse inklusive tillförlitlig rumslig information om arter och habitat som behövs för skötselplaner och regleringar. I ett försök att minska denna kunskapslucka genomförde Sveriges geologiska undersökning (SGU) på uppdrag av Havs- och vattenmyndigheten en högupplöst kartläggning av 1 344 km^2 havsbotten på Hoburgs bank 2016–2017, som denna rapport behandlar, samt vidare en kartering av Norra Midsjöbanken 2018 av liknande omfattning (under bearbetning). Heltäckande kartläggning utfördes med hydroakustiska metoder då projektområdet ej var kartlagt med moderna sjömätningsmetoder tidigare, samt sedimentprover och högupplösta bilder av havsbotten. Därtill gjorde SLU Aqua undersökningar av fisk i samarbete med SGUs kartering 2016. Informationen användes sedan till att producera heltäckande bentiska habitatkartor som beskriver ett komplext och varierande undervattenslandskap som aldrig tidigare har visats i sådan detalj, samt även avsevärt förändrar vår övergripande bild över hur stora ytor som täcks av olika habitattyper samt utbredningen av mer sällsynta och mindre habitat som hårda lerformationer och moränformationer av sten och block. Själva undersökningen samt analysarbetet med Hoburgs bank är därtill ett stort kliv framåt för hur SGU samlar in data och tar fram kartor av både geologisk och biologisk karaktär. Projektet har visat hur man kan integrera många olika undersökningstyper och syften i en och samma fältinsats och med hjälp av bland annat maskininlärning sedan kunna skapa en rik mängd med bearbetad information för många olika syften. Projekten från 2016 och fram till idag har även visat på vikten av kontinuitet i riktad habitatkartering för utövare såsom SGU då erfarenhet av dessa relativt komplexa mätningar och analyser är av stor betydelse. Denna rapport beskriver således både detta teknikspräng samt den nya informationen om miljön på Hoburgs bank.

Kartor genererades utifrån sjömätning samt geologiska och biologiska data som kombinerades till kontinuerliga täckningsmodeller (%) av substrat och biologiska komponenter. Den insamlade informationen tillsammans med experttolkningar och historiska data gjorde de geologiska modellerna till en stabil grund som de biologiska modellerna (med mindre träningsdata tillgängligt) sedan kunde bygga vidare på. Grundbulten i experttolkningarna av substrat var högupplöst sjömätningsdata (sedimentekolod samt djup och sonarmosaiker från multistråligt ekolod med 0,5 m upplösning). Värdet av att kombinera högupplöst djupdata med sonarmosaiker (dvs. backscatterdata) från ett multistråligt ekolod visade sig även vara en högt bidragande faktor till modellernas framgång där ca 50 % förklaringsgrad kom från backscatter. De kontinuerliga modellerna kombinerades sedan för att producera bentiska

habitatkartor som möter olika klassifikationsteman (t.ex. HELCOM HUB och Natura 2000) och kan också användas direkt beroende på behov. Vi rekommenderar således att alla analyser som relaterar till täckningsgrader görs direkt på dessa underlag istället för de generaliserade tematiska kartorna. Kartorna är högupplösta (5 m pixlar, modellerat från undervattensbilder med motsvarande utbredning på havsbotten), vilket tillåter dem att bli omarbetade till relevanta förvaltningsskalor, samtidigt som de i möjlig mån behåller information insamlad av högupplösta data. Dock visade test av olika skalor att mindre vanliga habitat (såsom små områden med hårda lerformationer, eller mindre rev bestående av moränryggar) minskade i omfattning redan vid omräkning av de tematiska kartorna till 10 m upplösning. Resultaten från kartframställningar visar att modellering av kontinuerliga variabler av geologi och biologi ger slutanvändaren mer korrekt information för vidare analyser än att endast utgå från de tematiska kartor (exempelvis HELCOM HUB och Natura 2000-klasser) som efterfrågas för exempelvis rapportering och uppföljning av olika direktiv. Vidare indikerar analys av kartornas träffsäkerhet att den kontinuerliga modelleringsmetoden är lika bra eller bättre än tematiska modelleringsmetoder för slutprodukterna, även de av tematisk karaktär. Den negativa sidan av att jobba direkt med kontinuerliga data (dvs. % täckningsgrader i detta fall) är att den data som behövs är mer mödosam att tolka fram utifrån observationerna jämfört med förenklade klasser av geologi och biologi. Här finns också stora möjligheter till vidare utveckling av analysverktygen av bilder och prover för att inkludera tekniker som exempelvis datorseende.

Den karterade delen av Hoburgs bank är 10–63 m djupt, med majoriteten av området på ett djup mellan 15 och 35 m. Sand och grus täcker ca 80 % av havsbotten i området, medan hårdbotten (stenar–stora block, och hårdlera) täcker ca 19,3 %. De resterande 0,7 % består av mjuka sediment (silt-lera). 39 % av banken klassas som Natura 2000-rev när man beräknar det för 5 m upplösning, medan de resterande områdena (61 %) klassades som Natura 2000-sandbank. Vid aggregeringar till grövre skalor ökar andelen som klassas som rev succesivt då den genomsnittliga förekomsten av hårdbotten och musslor för banken som helhet resulterar i klassningen rev. Vågor och undervattensströmmar skapar en dynamisk havsbottenmiljö, och tecken på sandtransport observerades över hela banken. 90 % av området hade antingen sand med sandvågor (49 %), eller grov hårdbotten med sand närvarande vilket indikerar sandtransport. För 9 % av havsbotten observerades signifikanta post-glaciala sandavlagringar (generellt i områden djupare än 40 m), utan sandvågor vilket indikerar depositionsområden för sand. Endast på ca 1 % av havsbotten var sand helt frånvarande. Dessa områden inkluderar stora individuella block, samt en serie av distinkta moränformationer (0,7 % av banken enligt HELCOM HUB nivå 3 sten och block), och hårda lerstrukturer (0,1 % av banken enligt HELCOM HUB nivå 3 hård lera). Undervattensobservationerna i närheten och på de mer strukturellt komplexa reven indikerade ökad biomassa och mångfald av fisk jämfört med andra havsbottentyper på banken. Bland annat observerades torsk gömma sig under större stenblock vid ett flertal tillfällen. Hårdbottenytorna var generellt koloniserade och täckte 26 % av banken (beräknat utifrån de kontinuerliga kartorna av täckningsgrader), där de dominerande organismerna var blåmusslor (dvs. *Mytilus* spp. 10,7 % täckning) och fintrådiga rödalger (14,8 % täckning). De grundaste områdena (10–15 m) hade högre mångfald av alger än de djupare områdena och flest ettåriga arter av alger kunde hittas här (0,6 % täckning). Den djupaste observationen av alger var en fintrådig rödalg på 38 m djup. Observationer i de djupare (~40–60 m) sandområdena visade krypspår och större aggregeringar av Östersjömussla (*Limecola balthica*), dött organiskt material (detritus) och pungräkor, jämfört med grundare och mer dynamiska sandområden. Detta visar att för en mer komplett förståelse av vilka habitat som är skyddsvärda behöver man titta på både grunda och djupa områden. Den strikta definitionen av Natura 2000-sandbankar har en nedre djupgräns (~30 m), för utsjöbankarna kan det vara värt att utvidga den definitionen. På grund av svårigheten att på ett ekologiskt relevant sätt använda sig av de nuvarande definitionerna av Natura 2000-sandbankar är djup inte med som en avgränsade faktor i de kartor som tagits fram, dock finns djupdata att tillgå för vidare analyser och indelningar av användarna.

Den genomsnittliga noggrannheten i HELCOM HUB-kartorna var 80,5 % för nivå 3 (substrat), 81,9 % för nivå 4 (generell täckningsgrad av biota), 62,3 % för nivåer 4–5 (karakteristiskt organismsamhälle), och 53,2 % för nivåer 4–6 (dominerande taxa). För Natura 2000-rev/sandbankar var noggrannheten 87,7 %. Som förväntat sjunker noggrannheten i kartorna när man har många klasser och där många närbesläktade klasser gör att de lätt förväxlas med varandra.

I en majoritet av svenska vatten är vår kunskap om havsbottens livsmiljöer begränsade av otillräckliga data samt grova och generaliserade kartor. Kartläggningen av Hoburgs bank har öppnat ett fönster (~stort som Öland) till hur komplexa och vackra våra havsbottnar kan vara. Ett naturligt nästa steg är att kartera hela Natura 2000 området som underlag för framtida förvaltning. Det finns idag heltäckande moderna mätningar hos Sjöfartsverket för merparten av den återstående delen av Natura 2000-området Hoburgs bank och Midsjöbankarna, vilket behöver kompletteras med provtagning och kompletterande sjömätningar i mindre omfattning, för att möjliggöra modellering av högupplösta habitatkartor för hela området. Dock är en svårighet att rådatan från Sjöfartsverkets högupplösta sjömätningar, inklusive backscatterdata, inte är tillgängliggjord för de analyser som behövs för att matcha metodiken som presenteras i denna rapport. Som exempel noterades att endast backscatter i kombination med högupplöst (~0,5–1 m) djupdata möjliggjorde grundläggande identifikation av hård vs mjukbottnar i platta områden som var mycket snarlika i sonardata men inte i verkligheten. Backscatter är idag endast en biprodukt till sjökartsframställning (då endast djup behövs för sjökort) med ingen eller dålig uppföljning av kvalitet och tillgängliggörande som resultat i de undersökningar som endast riktar in sig på djupdata. Det finns därför ett stort behov att prioritera tillgången till nationell backscatterdata om livsmiljökartorna över Hoburgs bank och pågående Norra Midsjöbanken ska kunna utökas till att täcka återstående ~75 % Natura 2000-området med samma tematiska och rumsliga detaljnivå som karteringen av Hoburgs bank utan resurskrävande ommätningar.

Denna rapport återproducerar en del av det innehåll som presenterades i den vetenskapliga publikationen av Kågesten m.fl. 2019 men expanderar med information och beskrivningar om miljöerna funna på Hoburgs bank. Rapporten undersöker metodiken att utgå från kontinuerliga modeller även för tematiska kartor, samt diskuterar vidare möjligheten att först modellera geologi med hjälp av olika kvaliteter av data (historisk data, expert bedömningar och högupplöst data) som ett sätt att förbättra noggrannheten även i de biologiska kartorna där tillgången på data är lägre. Publiceringen innefattar även huvuddelen av den källkod (R) som använts i detta projekt. Slutligen belyser publiceringen begränsningar i de habitatklassificeringssystem som används i projektet. I dagsläget saknas möjlighet att publicera det omfattande materialet i sin helhet direkt på webben, men de producerade kartorna samt karteringsdata från projektet är tillgängliga genom SGUs kundtjänst.

INTRODUCTION

The geology of the Baltic seabed varies from soft clays to bedrock and forms the basis for unique communities of benthic and pelagic organisms. Integration of information on seabed geology, geomorphology, biotopes, and species is a prerequisite for long-term sustainable development and management of the Baltic Sea. This includes assessing the impact of human activities on the ecosystem and aiding decisions on whether to develop or protect a particular area. Marine benthic habitat maps are essential tools for marine spatial planning (Ward et al. 1999), and a growing number of requests are being made for full coverage seabed information to assist and develop marine spatial planning processes. Even so, few full coverage, high-resolution maps describing both geological and biological properties of the Swedish seabed are available.

In 2016, the Geological Survey of Sweden (SGU) and the Swedish Agency for Marine and Water Management (SWaM) agreed to co-fund a pilot mapping project under the National Marine Mapping Project (NMK), to produce high-resolution, full coverage benthic habitat maps that could be used to aid environmental protection, marine spatial planning, improve conservation of Natura 2000 areas, and also guide decisions on green marine and coastal infrastructure. In addition, the Department of Aquatic Resources at the Swedish University of Agricultural Sciences (SLU Aqua) collaborated to collect data and model fish and infauna distributions. Producing full coverage information was seen as a means of meeting the requirements of the EU directives on marine spatial planning (European Parliament & Council of the European Union 2014) and habitats (Council of the European Communities 1992), as well as two of the Swedish environmental objectives ("Ocean in balance and a living coast and archipelago" and "A rich flora and fauna").

The focus area was the large offshore shoal Hoburgs bank (~1,200 km^2), a Natura 2000 area with previously identified high conservation value (Larsson 2016, Naturvårdsverket 2010), whose northernmost part is located ~5 nm (~10 km) south of the island of Gotland in the Baltic Sea Proper (Fig. 1). There is substantial interest in the region for fishing, energy production, sand extraction, and infrastructure and shipping, with busy shipping lanes surrounding and crossing over areas of the bank (Forsman 2017, Larsson 2016). Shipping can impact the Natura 2000 area negatively through underwater noise, oil pollution, and the introduction of invasive species, which can disturb seabirds, porpoises, and benthic organisms (Heinänen et al. 2018, Larsson 2016, Larsson & Karlsson 2018). For this reason, the impacts and possibility of rerouting shipping lanes around Hoburgs bank are being assessed in the national marine spatial planning process (Forsman 2017, Heinänen et al. 2018, Larsson 2016, Larsson & Karlsson 2018).

Hoburgs bank is a fairly shallow shoal, with depths ranging from 10 to 35 m (Länstyrelsen i Gotlands län 2005). Earlier studies conducted in 2004 by SGU, which produced geological maps at a scale of 1:500 000 (Nyberg 2016), showed that Hoburgs bank consists of sedimentary bedrock covered by boulder clay superimposed with glacial clay. Glacial clay tends to be the predominant sediment type, with sporadic occurrences of boulder clay. Sand and gravel deposits occur in depressions and declines into deeper water, particularly in the southern part of the bank. Some areas consist of alternating layers of boulder clay and glacial clay, indicating the effect of an oscillating ice margin or a stranding iceberg during the most recent glaciation phase. Shallower areas generally consist of residual material in the form of sand, gravel and cobbles with locally abundant boulder fields also occurring. Medium and coarse sand generally predominate on slopes and in valleys, whereas more mobile medium and fine sand occur as deposits or thin veneers in deeper areas.

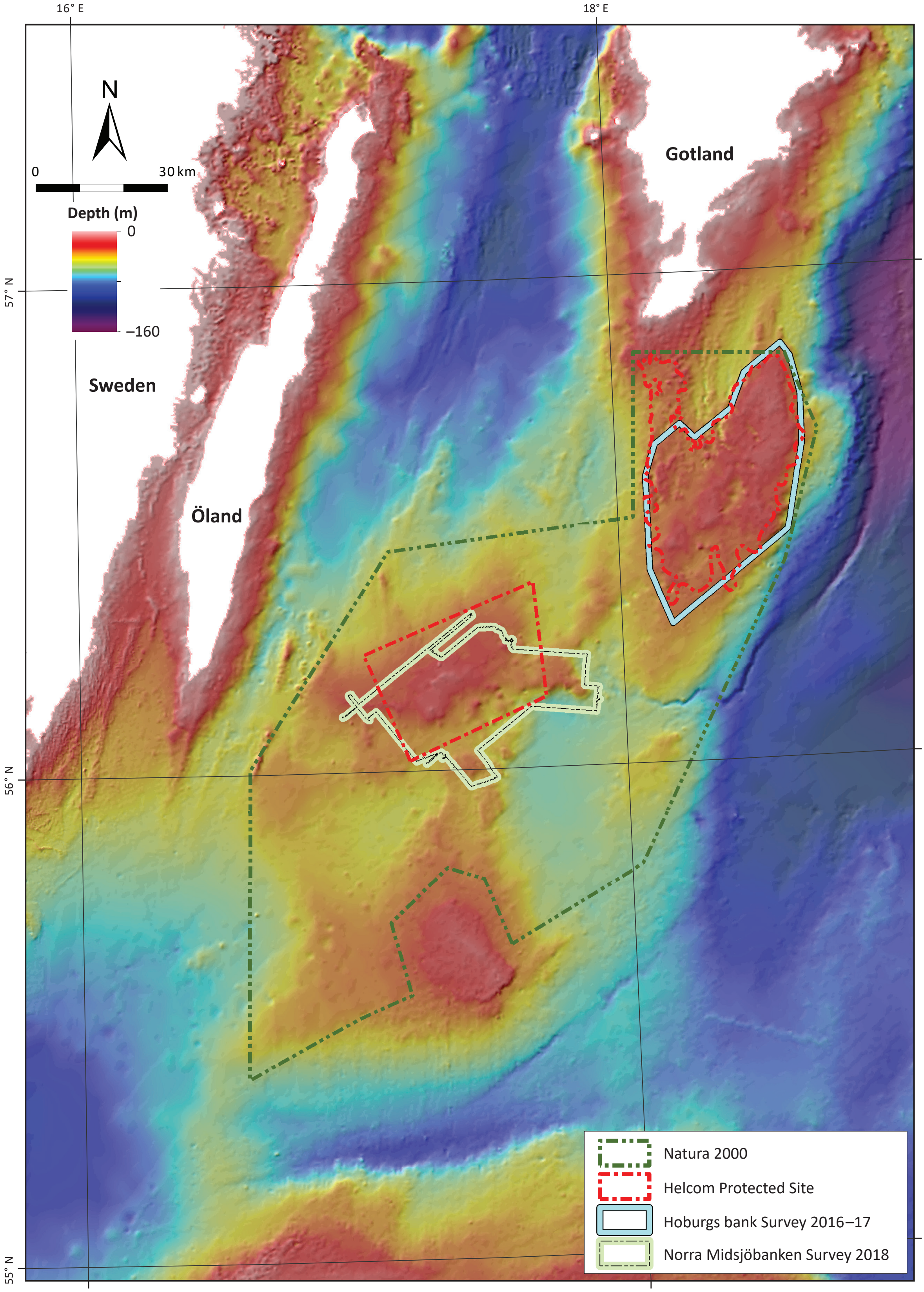


Figure 1. Areas within the red dashed lines represent the original Natura 2000 areas for Hoburgs bank and the Northern Midsea bank (Norra Midsjöbanken). Area within the green dashed line represents the enlarged Natura 2000 area including both Hoburgs bank and the Northern Midsea bank. Area within the blue ang light green line represents the area surveyed by SGU in 2016 and 2017 (Hoburgs bank) and 2018 (Norra Midsjöbanken).

Mean currents over the bank are fairly weak. However, short-term wind events tend to produce intense short-lived currents (Axe & Lindow 2005). These intense currents have probably led to wave erosion and surging. These processes have removed or relocated the finer fractions of the glacial till deposits on the bank, leaving behind residues of larger fractions, such as gravel, cobbles and boulders, surrounded by fairly large areas of sand (Naturvårdsverket 2008, 2010). Sand areas in relatively deep areas have clear ripple marks, indicating the effect of wave energy on the area (Kautsky 2000). This unstable wave-sorted substrate type supports a diverse range of habitats (Kautsky et al. 2017, Länstyrelsen i Gotlands län 2005). The area is of great importance for several seabird species, including the red-listed long-tailed duck (*Clangula hyemalis*), which forages for food on the easily accessible bank (Larsson 2016, Skov et al. 2011). Approximately 25% of the entire long-tailed duck population of the northern hemisphere overwinters here (Skov et al. 2011). Large seabird populations appear to have had a major impact on blue mussel populations, with mussel coverage tending to be lower than on similar banks (e.g. the Midsea banks) (Kautsky 2000, Naturvårdsverket 2008). This has led to a generally high biomass of algae dominated by *Battersia* sp., with filamentous and fleshy red algae species being fairly widespread (Kautsky 2000, Länstyrelsen i Gotlands län 2005). The bank is also home to a high diversity and density of red-listed fish species (Naturvårdsverket 2010). Finally, the area around Hoburgs bank and the nearby Midsea banks are important habitats for the critically endangered population of harbour porpoise (*Phocoena phocoena*) (Carlström & Calen 2016, Carlén et al. 2018).

In 2005 Hoburgs bank was designated a Natura 2000 area under the EU Birds Directive (European Parliament & Council of the European Union 2009) and Habitats Directive (Council of the European Communities 1992) down to the 35 m depth curve on sea charts (Forsman 2017, Länstyrelsen i Gotlands län 2005). However, this area was enlarged in 2016 combining the Hoburgs bank and the Northern Midsea bank (Norra Midsjöbanken designated a Natura 2000 area in 2009) areas in light of their importance as breeding grounds for the critically endangered harbour porpoise (Carlén et al. 2018, Carlström & Calen 2016, Forsman 2017, HELCOM 2013) (Fig. 1). Although this is the largest Natura 2000 area classified as such by Sweden (Regeringskansliet 2016), exceptions to some regulations for certain opposing interests such as shipping lanes have had to be made (Forsman 2017).

The overall goal of this project was to map the heterogeneous and patchy benthic habitats of Hoburgs bank through simultaneous collection of hydroacoustic data (multibeam bathymetry and backscatter as well as sub-bottom profiles) as well as ground truthing data on geology, and benthic fauna and flora. Data were then used to produce high-resolution full coverage maps of depth and backscatter, and, using modelling techniques, continuous coverage (%) maps of surface sediment types, and benthic fauna and flora that could be combined to produce classified maps according to the HELCOM Underwater Biotope Classification System, i.e. HELCOM HUB (HELCOM 2013), and Natura 2000 (European Commission 2008). Legacy data from previous survey expeditions were also assessed for their value in modelling. In addition to these main objectives, a wealth of information on benthic and pelagic fish as well as oceanographic data were collected in line with the adage "collect once, use many times".

METHODS

This section provides an overview of the different methods used to collect, process and model data. A more detailed description of the collection and processing of field data is provided in Appendix 1, "Bio-geophysical survey methods for habitat mapping of Hoburgs bank". Field surveys were conducted by SGU's research vessel R/V *Ocean Surveyor* in 2016 (August–October)

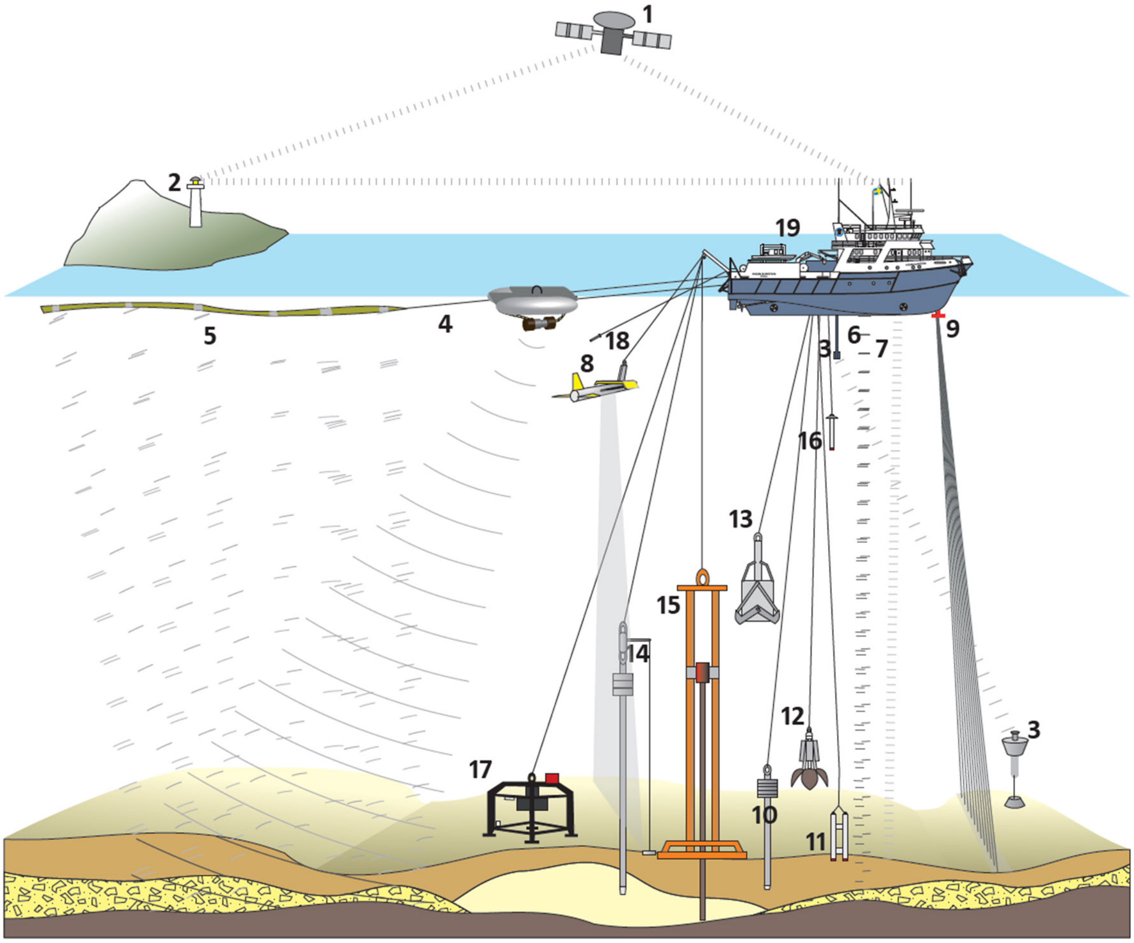


Figure 2. Diagram of R/V *Ocean Surveyor's* instruments. Instruments used in the project were: **1)** Satellite positioning system (GPS); **2)** Reference station for differential GPS; **3)** Hydroacoustic positioning system; **6)** Sub-bottom profiler; **9)** Multibeam echosounder; **12)** Orange peel bucket and Van Veen sediment grabs; **16)** CTD-probe; **17)** Drop camera system; **18)** Moving vessel profiler (MVP).

and 2017 (May–August), collecting both remotely sensed data via hydroacoustic methods, covering an area of approximately 1,344 km^2, and ground truthing data (Fig. 2). In general, a 24h routine was followed where hydroacoustic data were collected during the evening and night (4pm–8am) and ground-truthing data collected during the day (8am–4pm) based on the recently collected hydroacoustic data.

Hydroacoustic

Seafloor bathymetry

Full coverage seafloor bathymetry data were collected using an EM2040D dual-head multibeam echosounder (MBES) (Kongsberg Maritime AS, Kongsberg). It was operated at a frequency of 300 kHz in continuous wave (CW) pulse mode with a 10° overlap. Equidistant beam spacing mode was used, and the beam angles were continuously adjusted (ranging from 76° to 83° for both receiver heads), depending on the depth and distance between survey lines. Raw bathymetry (.all) and water column (.wcd) data files were recorded from the MBES with the acquisition software Seafloor Information System (SIS) (v.4.x Kongsberg Maritime AS, Kongsberg). Depth processing was conducted in CARIS HIPS and SIPS (v10.3.1, Teledyne

CARIS) to produce final bathymetric (referred to below as “depth”) layers at 0.5 m, 1 m and 2.5 m resolutions. These layers, as well as seafloor statistics derived from them, were used as environmental predictor variables (referred to below as “predictors”) in the modelling process.

Sound velocity profiles

Sound velocity profiles (SVPs) were collected to determine the speed of sound in the water column, which is essential to the collection of accurate depth and backscatter data. In 2016, SVPs were collected using a MIDAS SVX2 Combined CTD/SVP (Valeport Ltd, Devon), attached to a manual winch with a depth reader. On average, new profiles were taken every 30 minutes, at intervals varying from 15 minutes to 4 hours, depending on the hydrographic conditions of the survey area. In 2017, however, a new moving vessel profiler (MVP) equipped with a mini SVS Sound Velocity Sensor (Valeport Ltd, Devon) was installed, allowing more frequent SVPs to be collected (up to every 5 minutes) whilst the vessel was in motion. Sound Speed Manager (Masetti et al. 2018) was used to open, edit and upload the SVPs to SIS.

Backscatter

The MBES simultaneously collected acoustic backscatter information together with depth data. Backscatter is the intensity of the sonar return from the seafloor, which provides important information on different seafloor characteristics such as hardness, surficial sediment character, and roughness (Lurton 2010). It is an important source of information in many marine applications, including marine habitat mapping (Brown & Blondel 2009), and was an important predictor used in modelling. Backscatter data were processed using the Fledermaus geocoder toolbox (FMGT) (v7.7.6, QPS, Zeist). Backscatter spatial products produced from the processing included backscatter mosaics, backscatter statistical derivative products, and seafloor characterisation products (i.e. Angular Range Analysis – ARA), which were exported as floating-point geotiff grids at 0.5 m, 1 m and 2.5 m resolutions.

Sub-bottom profiler

Sub-bottom profiler (SBP) data were collected using an Echoes 3500 T3 SBP (chirp 1.7–5.3 kHz, iXblue, Saint-German-en-Laye). SBPs transmit sound pulses that penetrate the sub-seafloor sediment layers and capture information on sediment characteristics (i.e. type and thickness). Data were processed in Meridata Processing Software (MDPS) (v5.2, Meridata, Lohja). Every 2nd to 4th line was processed, and the resulting data interpolated to provide a model of estimated postglacial sand depth, a predictor used for modelling.

Ground truthing (GT)

Ground truthing (GT) data were collected to capture detailed geological and biological information directly from the seabed. Two general methods were used: 1) Sediment sampling; and 2) Underwater observations (UW-obs) from images and videos according to Havs- och vattenmyndigheten 2015 (unpublished report).

Site selection

GT sites were selected using the Sampling Design Tool for ArcGIS 10 (Buja & Menza 2007). The tool selects sites from a simplified habitat raster surface using stratified random sampling. The simplified habitat raster surface was created by combining the derived surfaces from depth and backscatter data (2 m resolution) using a classification analysis built in ArcGIS model builder (v10.5.1, Esri, Redlands). The derived surface included raster data sets with various depth intervals, substrate types (using backscatter intervals) and seafloor shape and complexity (combination of depth standard deviation and bathymetric position index (BPI)). Site selection was conducted on a daily basis as soon as each survey block (see Appendix 1, Fig. 1) was completed. A small number of expert sites where added to capture specific features of interest not identified in the random sampling design

A dynamic positioning system was used to keep the vessel in position when sampling. Vessel position was collected with a real time kinetics (RTK) GPS (Seapath 330 GNSS-RTK, Kongsberg Maritime AS, Kongsberg) with corrections from SWEPOS base stations received via satellite internet, then adjusted to the cable breakpoint of the moonpool and A-frame winches, located amidships and aft respectively, approximately 15 m apart (see Fig. 2 instruments 12 and 17). Horizontal position uncertainty of sediment samples and UW-obs at the seafloor varied with depth and currents but were generally noted to be within ± 2 m when distinct features such as clay reef spurs were visited. In total, 434 sediment samples and 559 UW-obs were collected.

Sediment sampling

Sediment sampling was used to directly examine seafloor surface sediments and was conducted from the moonpool using two types of samplers: 1) Van Veen grab used for sampling sandy sediments to a depth of ~15 cm and 2) Orange-peel bucket for sampling both fine and coarse sediments to a depth of ~40 cm. All 434 sediment samples were analysed onboard by a geologist for composition, particle size, occurrence of flora and fauna as well as probable sedimentation environment. Sediment samples from 117 sites were retained and sent for laboratory grain size analysis (sieve analysis) to the Department of Soil and Environment at the Swedish University of Agricultural Sciences (SLU).

Underwater observations

UW-obs, both images and video, were collected using a drop camera system built in-house by SGU that was deployed from an A-frame at the rear of the vessel. Video footage and images were recorded by two digital cameras: 1) a Canon EOS 6D DSLR and 2) a GoPro Hero 4. Two parallel red lasers placed 30 cm apart, one on either side of the DSLR camera, provided a scale reference in images. A predefined script-driven pattern that covered 360° in the horizontal direction at a number of vertical angles allowed images and video footage to be recorded over an area of ~15 m^2. A total of 23 images were recorded by the DSLR camera. These were manually analysed to extract information on the absence, presence and coverage estimates of substrate and benthic organisms (see image analysis section below).

The drop camera system was also equipped with a mini CTD (Valeport Ltd, Devon), which recorded water temperature and salinity, a ZPulse® Doppler Current Sensor (Aanderaa Data Instruments AS, Bergen), which recorded water current, along with an Oxygen Optode 4835 (Aanderaa Data Instruments AS, Bergen), which recorded water oxygen concentration. These data were interpolated to produce oceanographic predictors for water temperature, salinity, current speed and direction, and O_2 concentration near the seafloor.

Image analysis

Image analysis to extract seafloor information was conducted by both Aquabiota Water Research AB and SGU following the recommendations of Havs- och vattenmyndigheten 2015 (unpublished report) with some modifications (Gullström et al. 2017). Absence and presence of substrate type and benthic organisms were recorded using 21 of 23 images. For coverage estimates, however, a total of 10 randomly selected images from the 17 images that covered an area of ~15 m^2 were selected from each site. A point intercept method was used to estimate coverage, whereby 10 points were placed onto each of the 10 images using Photoquad (v1.4, Trygonis and Sini 2012), and the substrate type and benthic species were recorded for every point (n = 100 points per site). Images recorded at higher vertical angles were cropped i.e. the top 25% of each image was removed to provide coverage estimates encompassing an area of ~5 m^2, as outlined in Havs- och vattenmyndigheten 2015 (unpublished report). Each point represented 1% coverage. However, in the case of benthic organisms a single point could intersect several organisms (e.g. a bivalve covered by algae). It was therefore possible for the coverage of benthic organisms to total to more than 100% for a site. Substrate coverage, however, always added up to 100%. Images from each site that were used to estimate cover (n = 17) were stitched into 360° photo mosaics using Autopano Giga (v4.4.1, Kolor/GoPro). This made it easier to assess the size of larger substrate fractions (i.e. large stones, boulders, and large boulders) that were often larger than a single image (Table 1, Fig. 3A, B). Point intercept annotations could not differentiate between finer grain-sizes than the difference between gravel and sand. The sand class from UW-obs therefor also included finer sediment fractions i.e. silt and soft clay when present (Table 1). Information on finer sediment fractions was instead obtained from the sieve analysis of sediment samples.

In addition to the main dataset, opportunistic presence of sand ripples and their size, *Macoma balthica* shells, crawl tracks, and fish were also recorded. Sediment samples, as well as

Table 1. Substrate fraction grain size ranges (mm) and their respective class names according to ISO standard 14688-1:2017, substrate type, substrate fraction for Hoburgs bank, and ground-truthing analysis method.

Grain size (mm)	ISO standard	Substrate type	Substrate fraction	Lab Analysis	Image Analysis
-	-	Hard	Bedrock [1]	-	Point intercept
> 600	Large boulder		Large boulders	-	Point intercept
> 200–600	Boulder		Boulders	-	Point intercept
> 60–200	Cobble		Large stones	-	Point intercept
> 20–60	Coarse gravel	Soft	Pebbles & stones	-	Point intercept
> 2–20	Fine & medium gravel		Gravel	-	Point intercept
> 0.6–2	Coarse sand		Coarse sand	Sieve	Point intercept [2]
> 0.2–0.6	Medium sand		Medium sand	Sieve	Point intercept [2]
> 0.06–0.2	Fine sand		Fine sand	Sieve	Point intercept [2]
> 0.002–0.06	Silt		Silt	Sieve	Point intercept [2]
≤ 0.002	Clay		Soft clay	Sieve	Point intercept [2]
≤ 0.002	Clay	Hard	Hard clay	-	Point intercept

[1] Not found in the study area.
[2] Fine substrate fractions (i.e. soft clay, silt, and fine, medium and coarse sand) in images were classified according to the predominant grain size, which was always sand (i.e. > 0.06–2 mm) in the study area.

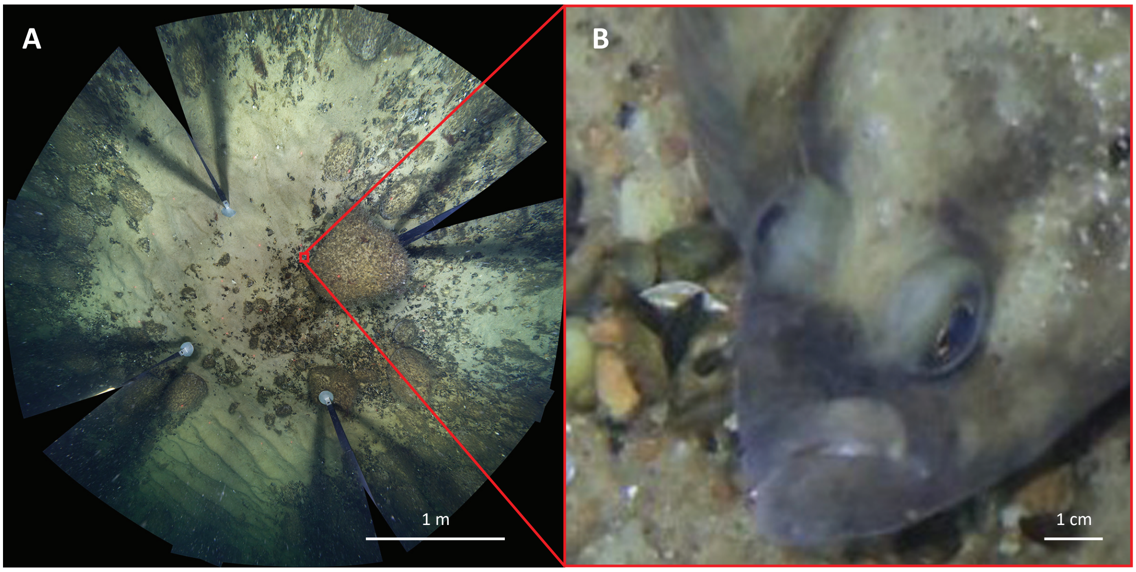


Figure 3. Photo mosaic. Example of a typical mixed hard and soft seafloor habitat on Hoburgs bank. **A.** Full photo-mosaic covering an area of 15 m^2. **B.** Detail zoom showing a flatfish represented by the red square in **A.**

high-resolution depth and backscatter data, were used to help identify the substrate component, particularly in images where the substrate was covered by benthic organisms. All interpreted sites were checked by a second expert to minimise error between interpreters. However, no statistical efforts were made to calculate uncertainty estimates from this process.

Legacy data and expert annotations

To further improve the training dataset for substrate models, legacy UW- obs and sediment samples, collected by SGU in 2004, and video transects collected by Stockholm University and SGU in 2005, were processed and re-annotated with substrate information (n = 449). Video transects were digitised from DV band format to digital mp4 format. Due to the old format, the position and other sensor data were stored in the analogue track, with position and time stored in a separate text file. The data were converted and exported to ArcGIS (ArcGIS Desktop v10.5.1, ArcGIS Pro v2.1.2, Esri, Redlands), using the ArcGIS Full motion video module (FMV) (v1.3.1, Esri, Redlands), which allowed samples/images and video transects to be interpreted together with backscatter and depth data to verify and, if needed, adjust the location of some sites with positional errors.

Additional expert annotations of substrate components (n = 550) were conducted based on the high-resolution depth and backscatter data and available GT for areas with a small number of samples, under-represented habitats and obvious artifacts in early model outputs. Table 2 provides a description and the number of points that were extracted from each GT method and the models/maps in which they were used.

Table 2. List of sources of training and testing data, method of analysis, and data use in models.

Data source	N	Analysis method	Model use
UW-obs (2016–2017)	559	Point intercept[1]	Coverage (%) of benthic organisms
UW-obs (2016–2017)	559	Point intercept[1]	Coverage (%) of substrate fractions (i.e. hard clay, sand, gravel, pebbles & stones, large stones, boulders, and large boulders)
Legacy UW-obs, video transects, sediment samples (2004–2005)	449	Estimated coverage	Coverage (%) of substrate fractions
Expert annotation (depth, backscatter)	550	Estimated coverage	Coverage (%) of substrate fractions
Sediment samples (2016–2017)	117	Sieve	Coverage (%) of fine substrate fractions (i.e. soft clay, silt, and fine, medium and coarse sand)
Sediment samples (2016–2017)	434	Expert	Thematic substrate classes (e.g. silty gravelly sand)

[1] Only coverage (%) from point intercept analyses were included in the testing dataset.

Habitat models and maps

This section describes the process of turning the processed survey data into habitat maps that accord with multiple classification schemes and contain information on geomorphology, geology and biology, using remote sensing and modelling techniques. The primary approach was to model the principal components (i.e. seabed coverage (%) of substrate types and benthic organisms) of the classified thematic maps as continuous variables, an approach that has recently been applied in the Baltic Sea (Herkül et al. 2017). Conditional statements were then used to combine the continuous outputs into thematic habitat maps according to HELCOM HUB (HELCOM 2013), Natura 2000 (European Commission DG environment 2013) and SGU substrate (Hallberg et al. 2010). The purpose of this approach was to harmonise the maps with existing classification schemes, while also allowing the user to go into as much detail as the survey data supported if a certain component or scale was of more interest than the habitat maps themselves. An overview of the whole analysis workflow is provided in Figure 4.

Environmental predictor variables for modelling

In the marine environment it is challenging to collect accurate, high-resolution spatial data for modelling, especially in complex environments such as Hoburgs bank. In general, two approaches have been used in the past for high-resolution habitat mapping: 1) Pixel-based image analysis (PBIA); and 2) Object-based image analysis (OBIA). PBIA is based directly on raster images and uses various statistics calculated around the neighborhood of each pixel as predictor variables. It is a simple and flexible approach, widely used in remote sensing. However, PBIA has some disadvantages in the marine environment, since data quality is often inferior to land-based data. Moreover, it can be challenging to place each pixel into context with the surrounding environment. OBIA, on the other hand, vectorises the high-resolution remote sensing data into features, using techniques such as edge detection algorithms that allow the user to derive statistics on the shape, form and variability of data within these features (e.g. a moraine ridge). OBIA has been successfully used in several habitat mapping projects (Costa & Battista 2013, Kågesten et al. 2015), and retains certain aspects that mimic more traditional manual mapping approaches. However, one drawback of OBIA is that it limits the models to

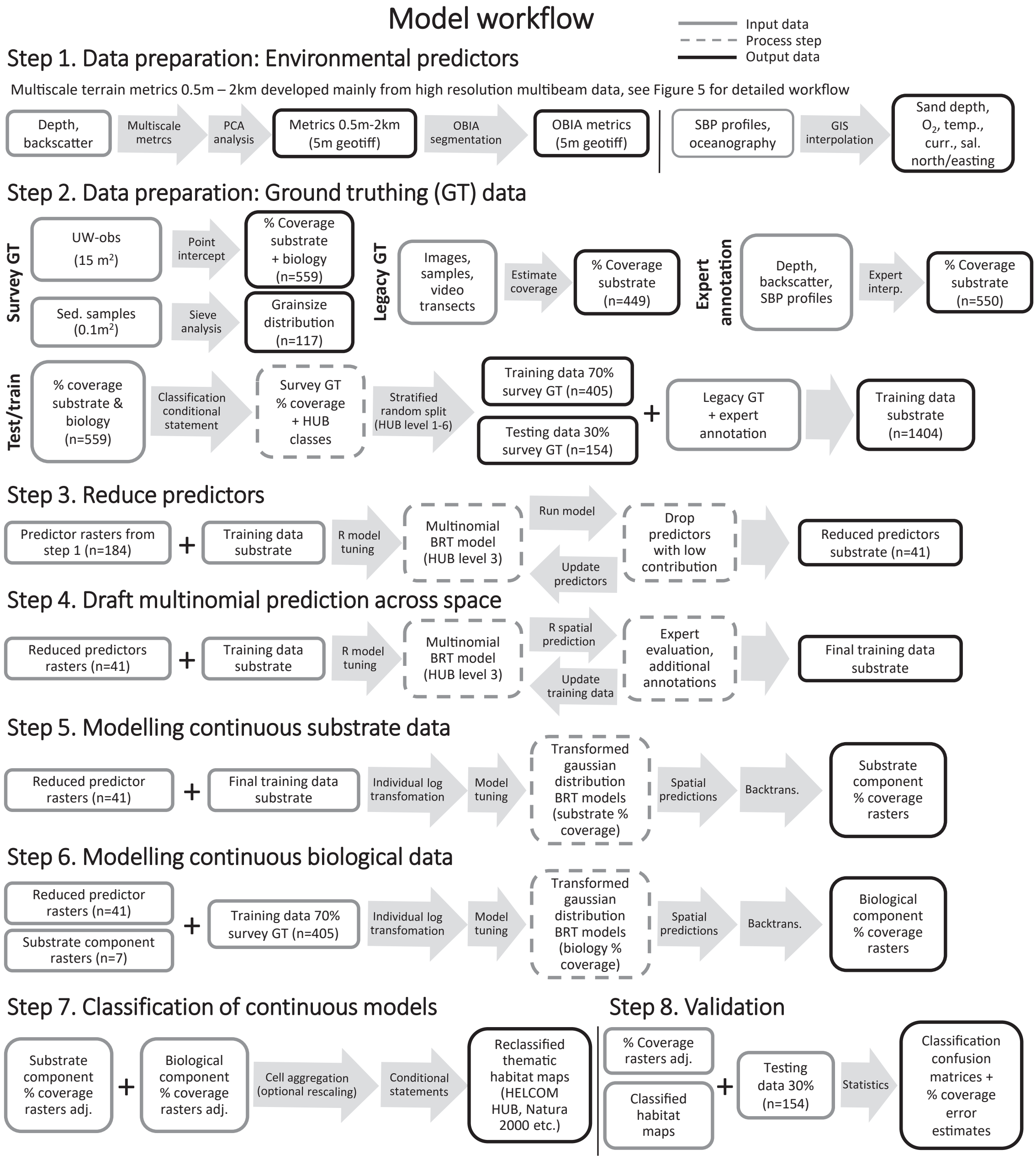


Figure 4. Overview of process for producing habitat models and maps. R scripts used for modelling are published and freely available online in Kågesten et al. 2019.

predefined segments; another is that there tends to be a disconnect between the scale of the training data and the minimum mapping unit of the segments produced.

This project combined the strengths of PBIA and OBIA by producing predictors that used both approaches, and then reclassifies the predictors into rasters. In addition, the modelling approach focused on producing maps of the highest possible resolution. The resolution of the final maps was primarily limited by the resolution and quality of the hydroacoustic data (i.e. depth and backscatter), UW-obs, and the computing power needed to run the spatial implementation of the models. Given these limitations, the highest resolution for the modelling was set at 5 m pixels (i.e. 25 m^2), which roughly corresponds to the scale of the UW-obs (~15 m^2).

Predictors used for modelling substrate and biological components were predominantly derived from high-resolution (0.5 m) depth and backscatter data. Several other predictors were included such as northing and easting, postglacial sand depth (from interpreted and interpolated SBP data) and near seafloor oceanographic variables (O_2, temperature, salinity, current speed and direction interpolated from GT data). Depth and backscatter data were recalculated to include multiple resolution scales (0.5 m, 1 m, 2.5 m, 5 m, 20 m, 50 m, 100 m, 200 m, 500 m, 1 km and 2 km) relevant for the model building, referred to below as multi-scale metrics. To represent scales smaller than 5 m, terrain metrics (*morphometrics*) were first developed (Pittman et al. 2009) in ArcGIS (standard deviation, profile, planform, standard curvature, slope, slope-of-slope, terrain-surface roughness, and surface area to planar area) from the high-resolution grids (i.e. 0.5 m, 1 m, 2.5 m). The information was then aggregated to 5 m using minimum, maximum, range and mean values. The same metrics were also applied using 5 m depth and backscatter (with the addition of median value and BPI). A Gaussian low pass filter was used in MATLAB (v9.0, MathWorks, Natick) for scales > 5 m to smooth the depth grid to represent each scale neighbourhood (i.e. 20 m to 2 km), while maintaining the spatial resolution at 5 m, followed by calculations of selected terrain metrics (i.e. slope, slope of slope, BPI, and slope direction). In order to adjust for potential angle artifacts, uncertainty data retrieved from multibeam depth, and the Euclidean distance to the survey lines were also included as predictors (Fig. 5, Fig. 4 step 1).

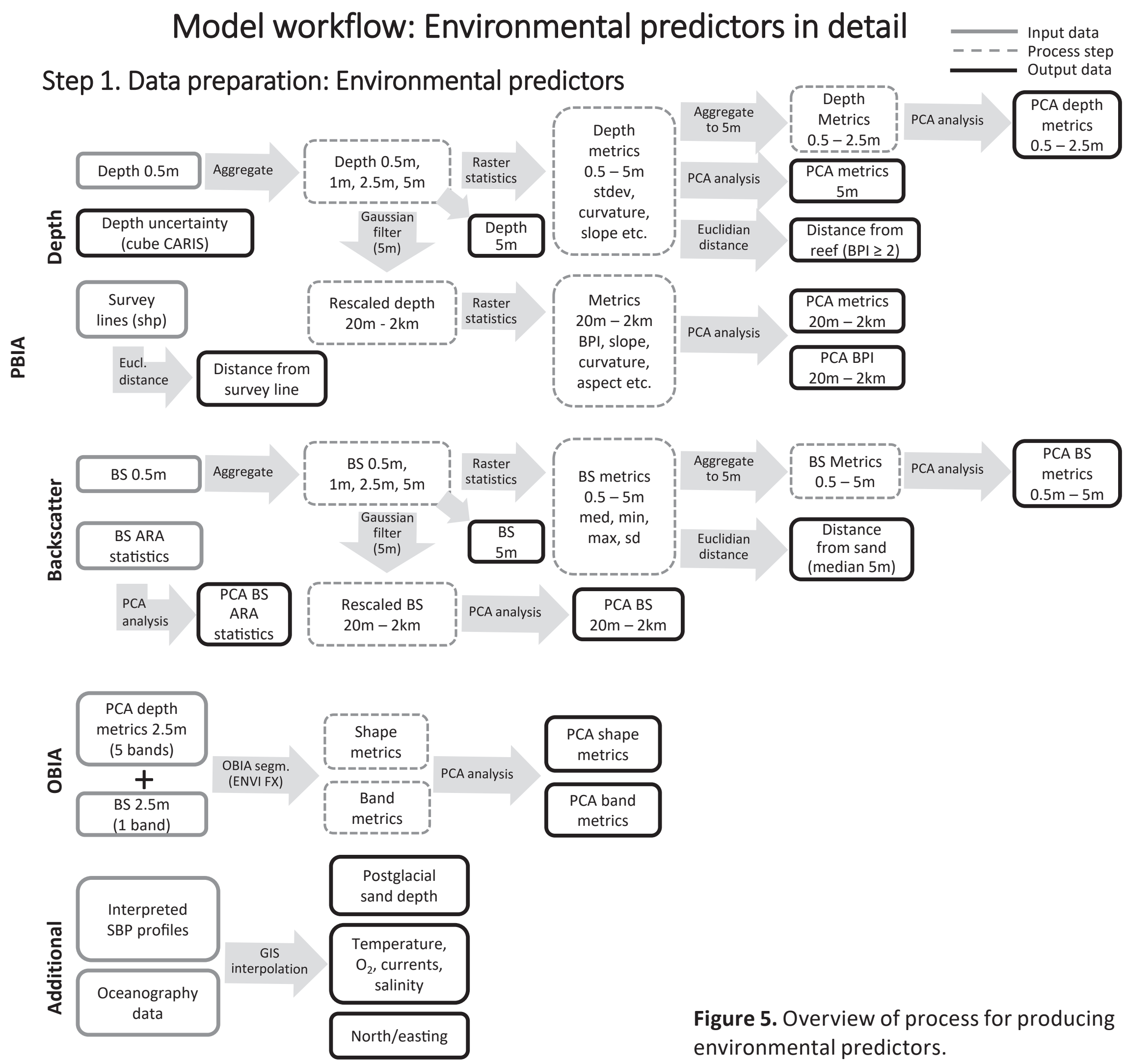


Figure 5. Overview of process for producing environmental predictors.

In addition to predictor metrics derived from PBIA (above), OBIA was conducted using the ENVI Feature Extraction module (v5.4, Harris Geospatial Solutions, Broomfield) to generate segments from high-resolution (0.5 m) depth and backscatter data and derive statistics from those segments, a process similar to previous semi-automated coral reef habitat mapping (Costa & Battista 2013, Kågesten et al. 2015). Segments were developed from a composite image (principal component compressed depth metrics in five bands and backscatter in one) at 2.5 m resolution. ENVI segmentation thresholds were set at segmentation 11 and merge 91, using an edge detection algorithm. The resulting segment statistics (spatial, textural and spectral attributes (ENVI 2008)) were exported as raster images and then aggregated to the modelling resolution of 5 m using a mean value. An alternative method for treating habitat features as objects was also used by calculating the Euclidean distance to features (often consisting of moraine ridges) seen in fine scale BPI (inner radius 5 m, outer radius 25 m, BPI ≥ 2), as well as distance to fine–medium sand patches (using a threshold reclassified backscatter mosaic).

To reduce collinearity and the number of predictors, so allowing more efficient computation of models, a principal component analysis (PCA) was used for each group of predictors (e.g. depth, backscatter etc.). The principal components whose combined contribution to the total variance in each group was less than 5% were excluded (Hengl 2009). Selected predictor variables, such as depth, were excluded from the PCA process in order to better understand how they contributed to the models (Fig. 5).

Additionally, expert evaluation served to further reduce the number of predictors from each group to 41 predictors. A full list of the predictor variables and how they were grouped is included as supplemental information in the in Kågesten et al. 2019 paper which can be downloaded from this site, https://www.mdpi.com/2076-3263/9/5/237#supplementary.

Modelling

Habitat modelling based on classification schemes is difficult since the hierarchical structure and the potential for many combinations of these hierarchies are not well suited for modelling directly. This section describes how these challenges were worked around to produce a flexible result that could fit uses and classification schemes of many kinds. Model algorithms were run with R (v3.4.1 (R Core Team, 2017)). The main scripts used are available as supplemental information in Kågesten et al 2019 paper (https://www.mdpi.com/2076-3263/9/5/237#supplementary). Boosted Regression Trees (BRTs) were used for both multinomial (classified) and continuous models, implemented by the gbm (Elith et al. 2008) and caret packages (Kuhn 2008). BRT is a flexible nonparametric statistical machine learning algorithm suitable for regression and classification problems that cope with non-linear relationships (De'ath 2007, Friedman 2001, Kuhn & Johnson 2013). It has been successfully used in very different contexts, such as coral reef mapping (Pittman et al. 2009, 2012; Pittman & Brown 2011) and marine protected area design (Leathwick et al. 2008, Stamoulis et al. 2018). High performance of this technique is reached by ensembling a large number of trees (Fernández-Delgado et al. 2014). The ensemble is built starting from a single tree produced by a random subset of the data. Further trees are then added sequentially to the previous one to find a better fit (De'ath 2007, Friedman 2001, Kuhn & Johnson 2013), represented by loss in predictive performance (e.g. deviance) (Elith et al. 2008). The final prediction is calculated as a weighted average across all trees (Elith et al. 2008). Models were tuned using the number of trees, interaction depth and shrinkage as parameters. The best combination of these parameters was selected using root mean squared error (RMSE) (continuous models) and overall accuracy (OA) (multinomial models). Following an initial test of model performance and computing time, the maximum model complexity in the tuning grid was limited to 1,500 trees with 15 splits interaction depth and 0.005 shrinkage. The modelling workflow was divided into four main sections: 1) Data preparation; 2) Modelling; 3) Classification; and 4) Validation.

Data preparation

Prior to modelling, UW-obs data from each site were manually classified according to HELCOM HUB levels 1–6. Data were then split into 70% training data (n = 405) and 30% testing data (n = 154) via stratified random sampling according to their HELCOM HUB class (Fig. 4 Step 2 Test/train), to ensure adequate evaluation of map classes. Training data were used to train the models whereas testing data were used as an independent data set for model validation. The same testing data were used for all models except those using sediment samples (i.e. fine sediment fractions), which used all data for training due to the small number of sieve analysed samples (n = 117). In addition, presence absence data were transformed into coverage (%) data by assigning 0.1% coverage to presence with coverage less 1%, and 0.001% coverage to absence. Continuous coverage data from image analyses of benthic organisms were combined according to HELCOM HUB classes in levels 5 and 6, and then modelled as biological components (Table 3, see the section Habitat map descriptions for more detail on how components were combined to HELCOM HUB biotopes). This was partly because data were not sufficient to accurately model some species separately, but also because one of the main aims of the project was to produce HELCOM HUB maps. For multinomial models, testing and training data for each site were classified into thematic classes for HELCOM HUB and Natura 2000, before modelling each level (i.e. HELCOM HUB levels 3, 4, 5, 4–6, and Natura 2000).

Substrate components were the foundation of the modelling approach. GT training data of coverage (%) for substrate components were derived from analyses of UW-obs images, historical UW-obs as well as expert annotation using remotely sensed layers (depth, backscatter, and SBP profiles) leading to a larger number of GT training data (n = 1404) for substrate components compared to biological components (n = 405). Substrate components were assumed to be important predictors for benthic organisms, in addition to depth, since most benthic organisms are dependent on substrate, and were therefore modelled first (Fig. 4 Steps 3–5).

Table 3. Benthic species included in each biological component derived from HELCOM HUB classes in level 5 and 6.

Biological component	Benthic species included in component
Perennial algae	*Battersia arctica, Coccotylus/Phyllophora* (complex), *Delesseria sanguinea,* Filamentous Rhodophyceae, *Furcellaria lumbricalis, Polysiphonia/Rhodomela* (complex), Unidentified Rhodophyceae.
Perennial filamentous algae	Battersia arctica, Filamentous Rhodophyceae, *Polysiphonia/Rhodomela* (complex).
Epibenthic bivalves	*Mytilus* spp.
Epibenthic cnidarians	Hydrozoa (*Cordylophora caspia*)
Epibenthic moss animals	*Electra* sp.
Annual algae	*Ceramium tenuicorne, Ceramium* sp., C*horda filum, Dictyosiphon/Stictyosiphon* (complex), *Ectocarpus/Pylaiella* (complex), Filamentous Phaeophyceae, *Halosiphon tomentosus*

Modelling: Multinomial data

Multinomial models were first explored to produce thematic HELCOM HUB level 3 substrate maps directly from classified GT training data (i.e. instead of using continuous coverage (%) data to model each substrate component). This provided a first visual overview of the output, which was then used to further reduce the number of predictors (Fig. 4 Step 3) and to guide expert annotation of areas where the models were weak (Fig. 4 Step 4). This was particularly obvious for sandy areas, where survey artifacts resulted in a false hard substrate classification but also for hard clay areas classified as rock and boulders due to the low number of training sites (n = 3) for clay reefs. Draft classification model outputs (Fig. 4 Step 4) were used to guide a few rounds of expert annotation, and then models were re-run. BRTs were then applied to predict thematic maps directly from classified training data (HELCOM HUB levels 3, 4, 5, 4–6, and Natura 2000). The purpose of this was to provide a comparison between thematic maps modelled from classified data and thematic maps classified from continuous models. Finally, all habitat data that only existed as classified data were also modelled using this method (i.e. *Macoma balthica* shells, sand ripples, and crawl tracks).

Modelling: Continuous data

Two BRT modelling approaches were explored for modelling continuous substrate and fine substrate components: 1) Individual models of each substrate component; and 2) A compositional workflow that modelled all substrate components simultaneously. The latter approach explicitly considers that substrate components must add up to 100% in each pixel (Stephens & Diesing 2015). For individual models, coverage (%) GT training data were logit transformed to improve normality and each individual substrate component modelled separately using BRTs (Fig. 4 Step 5). For compositional models, the same logit transformation was conducted. However, a balance function was then performed, and all substrate components modelled simultaneously. Each modelling approach had different strengths and weaknesses depending on the substrate component, so a combination of the two approaches was used.

Resulting substrate component models were back-transformed to coverage (%), and these layers were then used as predictors to model the biological components (Fig. 4 Step 6). Training data for biological components were also transformed, modelled and back-transformed in the same manner as substrate components according to the individual model workflow (Fig. 4 Step 6).

Classification of continuous models

To combine the continuous models into reclassified thematic maps, all substrate and biological components had to be adjusted to ensure that the model outputs accorded with the rules of the point intercept image analysis. When individual components are modelled separately, each model optimises the accuracy of that specific component without considering whether the sum of all components will add up to 100% coverage. Since coverage (%) of surface substrate components (i.e. hard clay, sand, gravel, pebbles and stones, large stones, boulders, and large boulders) always added up to 100% in the GT training data, this also had to be the case for the model outputs. Adjustments were made so that the sum of all substrate components in each pixel equalled 100% by dividing each individual component by the sum of all substrate components. Similarly, the fine substrate components modelled from sediment samples (i.e. soft clay, silt, and fine, medium and coarse sand) were adjusted to fit the sand component model, by multiplying the proportion of each fine substrate component by the sand component. Biological components had to be adjusted in a slightly different way, so that uncolonised substrate and biological cover had to add up to 100% or more, and no single biological component when added together with uncolonised substrate could add up to more than 100%.

Resulting adjusted models of continuous substrate and biological components were combined using conditional statements in R to produce classified HELCOM HUB (HELCOM 2013), Natura 2000 (European Commission 2013, Havs- och vattenmyndigheten 2017, unpublished report), and SGU substrate (Hallberg et al. 2010) maps (Fig. 4 Step 7). For classes with unclear definitions of coverage (%) or terrain metrics (i.e. Natura 2000 SGU subtypes, SGU substrate), text descriptions were relied on to define a conditional statement. Since scale was not clearly defined for any of the classification schemes, we used not only the full resolution of the models (5 m), but also created classified maps in other resolutions up to 250 m by aggregating the percentage coverage models before the conditional statements were run.

Validation

All maps produced at 5 m resolution were validated using GT testing data from UW-obs (n = 154; same subset of sites used for all models). Substrate models were also evaluated with training data using bootstrap prior to spatial predictions. Thematic maps and reclassified thematic maps were assessed with confusion matrices of observed versus predicted values for HELCOM HUB levels 3, 4, 5, 4–6 and Natura 2000, calculated in R (Fig. 4 Step 8). Not all classes could be assessed, because in some cases insufficient GT data were available (e.g. the class characterised by annual algae only had two GT sites in total). Confusion matrices capture map uncertainty to allow the user to better understand how to use these results. However, the uncertainty values also include the spatial and thematic uncertainty of the actual sample data, which were not analysed separately in this project. Overall accuracy (OA), Producer accuracy (PA), and User accuracy (UA) were calculated directly from confusion matrices (Story & Congalton 1988 in Kågesten et al. 2015). Confusion matrices consist of a square array of numbers arranged in rows and columns. OA was calculated as the sum of values in the major diagonal (i.e. correct classifications, values where the same map class intersects) divided by the total number of accuracy assessment samples (n_{tot}).

PA and UA were calculated to provide a measure of the classification accuracy of individual map classes. PA measures how well the model classifies a map class (e.g. proportion of times that substrate ground-truthed as sand was correctly mapped as sand). PA was calculated by dividing the value in the major diagonal by its column total (n_{col}). UA measures how often the pixels of a map class were classified correctly (e.g. proportion of times that a pixel classified as sand was ground-truthed as sand). UA was calculated by dividing the value in the major diagonal by its row total (n_{row}).

Continuous models were assessed using the coefficient of determination (r^2), root mean square error (RMSE), mean absolute error (MAE) and mean error (ME) (see equations (1), (2), and (3)).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(obs_i - pred_i)^2} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(obs_i - pred_i)| \quad (2)$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}(obs_i - pred_i) \quad (3)$$

Where *obs* represents the observed value and *pred* the corresponding predicted value, *i* is the replicate number and *n* is the total number of replicates.

Continuous models were also evaluated following the recommendations of Piñeiro et al. 2008, so that regression analysis of the observed values (GT testing data) against the corresponding predicted values (model data) for each component was performed, and the significance of a slope = 1 and y-intercept = 0 tested (i.e. comparison of regression model against a 1:1 line where GT testing data and model data agree perfectly). In addition, continuous models were classified into four coverage classes (< 10%, ≥ 10 < 50%, ≥ 50 < 90%, and ≥ 90 ≤ 100%), as well as absence and presence classes and evaluated using the same confusion matrices as multinomial models (Fig. 4 Step 8). ME was calculated for each of the four coverage classes to provide an indication of possible bias in the models along different parts of the scale. Continuous models of fine substrate components were only evaluated using bootstrap prior to the spatial prediction due to the relatively low number of sieve analysed sediment samples.

HABITAT MAP DESCRIPTIONS

Habitat mapping/classification schemes provide a standardised method of describing habitats at a national and regional level. At a European level, common habitat mapping schemes are the European Union Nature Information System (EUNIS) (Davies & Moss 2004), the HELCOM Underwater Biotope and Habitat Classification System (HELCOM HUB) (HELCOM 2013) and Natura 2000 (European Commission 2008). These classification schemes are aimed at meeting the objectives of the European Marine Strategy Framework Directive (MSFD) (European Parliament & Council of the European Union 2008) and the Habitats Directive (Council of the European Communities 1992). The intended use of these schemes is to classify marine areas through several steps; first, according to the environmental setting (e.g. vertical zone), next, by geological features (e.g. soft/hard substrate), and finally by biological components (e.g. communities, dominant species).

Classification schemes, although an important tool in management, are a simplification of nature and do not capture the entire complex nature of marine habitats (Kågesten et al. 2019). It is important that the user understands what the HELCOM HUB and Natura 2000 maps are capable of capturing, and the uncertainty involved.

This section provides a short overview of the HELCOM HUB (HELCOM 2013) and Natura 2000 (European Commission 2013, Havs- och vattenmyndigheten 2017, unpublished report) classification schemes, which were used to map the habitats on Hoburgs bank, together with the corresponding SGU definitions and descriptions used to capture other aspects of the benthic environment. Habitat class definitions are provided for both the habitat scheme and the conditional statements used by SGU to classify from continuous models.

Visual examples of the different habitats found on Hoburgs bank are provided where available, based on UW-obs collected by SGU in 2016 and 2017. The thematic resolution of the continuous maps (i.e. grain size ranges, and combination of species into groups) were limited by the method used to collect and interpret samples and underwater images. This is described in more detail in the methods chapter above.

HELCOM HUB

HELCOM HUB, was developed to provide a common framework for defining the marine environment within the Baltic Sea region (HELCOM 2013). It was designed to be compatible with the European-wide, EUNIS framework. The classification scheme is hierarchical in nature, and structured into six levels of habitats and biotopes by applying split rules like a

Table 4. Summary of HELCOM HUB split rules.

Level	Description
HELCOM HUB habitats	
1 Baltic (Letter)	- Baltic
2 Vertical Zone (Letter)	- Photic - Aphotic
3 Substrate (Letter)	- Coverage of a specified substrate type ≥ 90% - Coverage < 90%, Mixed
HELCOM HUB biotopes	
4 Community Structure (Number)	- Coverage of Macroscopic vegetation or sessile macroscopic epifauna ≥ 10% - Coverage > 0% < 10%, Sparse - Coverage = 0%, No vegetation or macro fauna present
5 Characteristic Community (Letter)	- Coverage of a specified taxonomic group ≥ 10% - Coverage ≥ 10% but not of a specified taxonomic group, Mixed community - Coverage = 0%, No macroscopic community
6 Dominating Taxa (Number)	- Biomass/biovolume of some specified taxa ≥ 50%

dichotomous key (Table 4). Habitats are the abiotic environment while biotopes are the abiotic environment coupled with the characteristic organism community. Levels 1–3 describe habitats, whilst levels 4–6 describe biotopes. Each level is represented by a character which serves as HELCOM HUB codes.

Note: Habitat maps were classified following the HELCOM HUB scheme, however, just prior to publishing this report, information was received that the interpretation used in this project to classify to HELCOM HUB level 6 diverged from what was intended by the scheme (see Results section HELCOM HUB levels 4–6 for a brief description). Therefore, it is emphasised that the following sections describe how maps have been classified according to the SGU definition of HELCOM HUB level 6 and not to that which was intended by the scheme.

HELCOM HUB habitats

HELCOM HUB level 1: Baltic

HELCOM HUB definition: The whole Baltic Sea.
SGU definition: The whole Baltic Sea.

HELCOM HUB level 2: Vertical Zone

Photic benthos (code A)

HELCOM HUB definition: The vertical zone in which the amount of light is sufficient for photosynthesis.
SGU definition: Areas where the depth is less than or equal to 43.39 m, which is based on the maximum depth that photosynthetic organisms were observed at GT sites, and where the biological components perennial algae, annual algae and soft crustose algae together cover at least 1% of the seabed.

Aphotic benthos (code B)

HELCOM HUB definition: The vertical zone below which the amount of light is not sufficient for photosynthesis.

SGU definition: Areas where the depth is greater than 43.39 m, which is based on the maximum depth that photosynthetic organisms were observed at GT sites, and where the biological components perennial algae, annual algae and soft crustose algae together cover less than 1% of the seabed (Fig. 6).

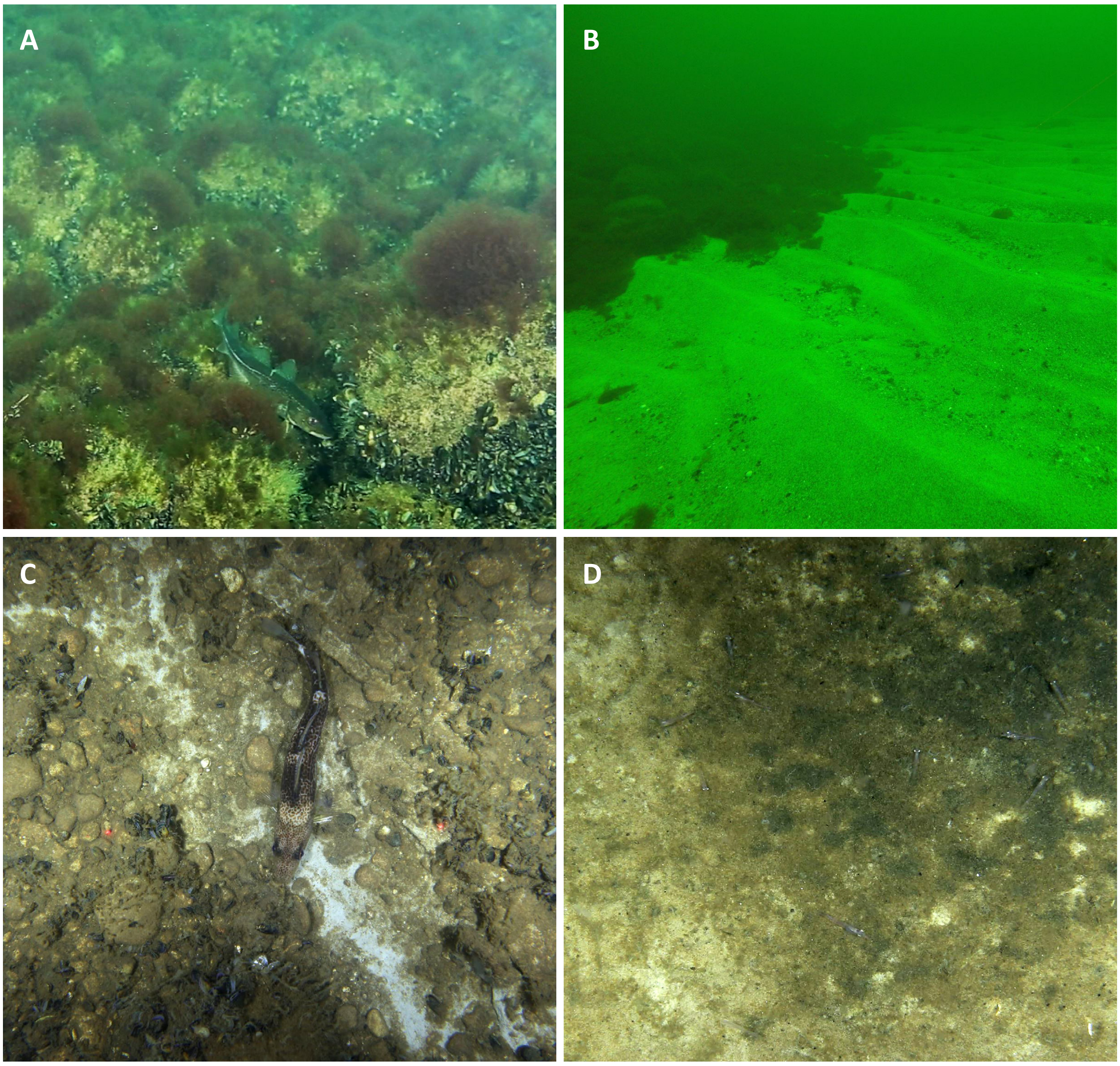


Figure 6. A. Photic hardbottom covered with algae and mussels and a cod feeding on animals hiding between the rocks. **B.** Unclolonised photic softbottom shown with nearby hardbottom covered in algae. **C.** Aphotic hardbottom colonized by cnidarians and blue mussels with a cod blending in with its surrounding. **D.** Aphotic softbottom with shrimp feeding on the accumulated detritus.

HELCOM HUB level 3: Substrate

Rock and boulders (code A)

HELCOM HUB definition: At least 90% coverage of rock, boulders or stones of more than 63 mm in diameter.
SGU definition: Areas where the substrate components large stones (> 60–200 mm), boulders (> 200–600 mm) and large boulders (> 600 mm) together cover at least 90% of the seabed (Fig. 7).

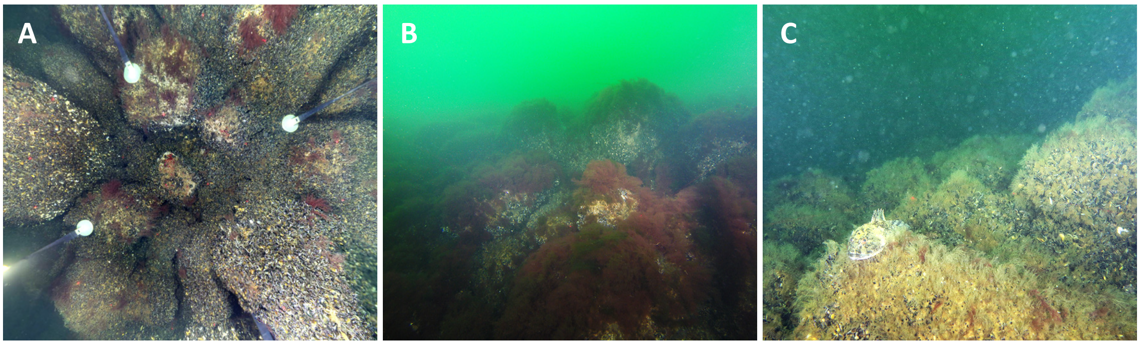


Figure 7. A. Photo example of mussel covered site classified as rock and boulders. **B & C.** Oblique view of rock and boulder reefs.

Hard clay (code B)

HELCOM HUB definition: At least 90% coverage of hard clay.
SGU definition: Areas where the substrate component hard clay covers at least 90% of the seabed (Fig. 8).

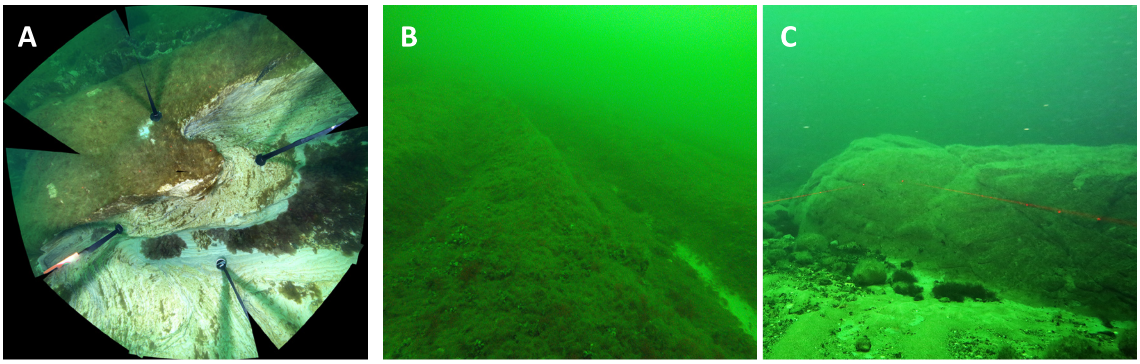


Figure 8. Photo mosaic example of a site classified as hard clay. **A.** Top view. **B & C.** Oblique view. The hard clay reefs often showed a spur and grove like pattern, and were quite easily identified by combining depth and backscatter data.

Coarse sediment (code I)

HELCOM HUB definition: At least 90% coverage of coarse sediment. Coarse sediment has less than 20% of mud/silt/clay fraction (< 63 µm), and the proportion of gravel and pebbles (grain size 2–63 mm) exceeds 30% of the combined gravel and sand fraction.
SGU definition: Areas where the substrate components soft clay (≤ 0.002 mm), silt (> 0.002–0.06 mm), sand (> 0.06–2 mm), gravel (> 2–20 mm), and pebbles and stones (> 20–60 mm) together – i.e. soft sediment – cover at least 90% of the seabed, and soft clay (≤ 0.002 mm) and silt (> 0.002–0.06 mm) together make up less than 20% of the soft sediment, and the

proportion of gravel (> 2–20 mm) and pebbles and stones (> 20–60 mm) combined is at least 30% of sand (> 0.6–2 mm), gravel (> 2–20 mm) and pebbles and stones (> 20–60 mm) combined (Fig. 9).

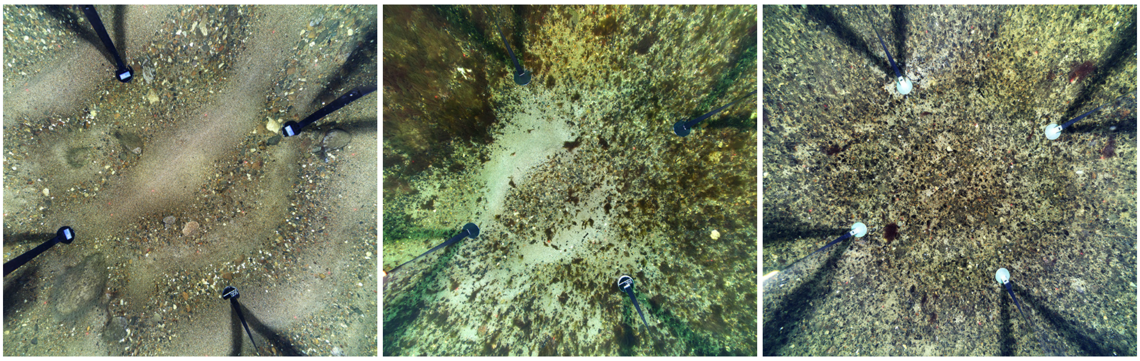


Figure 9. Photo mosaic examples of sites classified as coarse sediment.

Sand (code J)

HELCOM HUB definition: At least 90% coverage of sand. Sand has less than 20% of mud/silt/clay fraction (< 63 µm), and the proportion of sand (grain size 0.063–2 mm) exceeds 70% of the combined gravel and sand fraction.

SGU definition: Areas where the substrate components soft clay (≤ 0.002 mm), silt (> 0.002–0.06 mm), sand (> 0.06–2 mm), gravel (> 2–20 mm), and pebbles and stones (> 20–60 mm) together – i.e. soft sediment – cover at least 90% of the seabed, and soft clay (≤ 0.002 mm) and silt (> 0.002–0.06 mm) together make up less than 20% of the soft sediment, and the proportion of gravel (> 2–20 mm) and pebbles and stones (> 20–60 mm) combined is less than 30% of sand (> 0.6–2 mm), gravel (> 2–20 mm) and pebbles and stones (> 20–60 mm) combined (Fig. 10).

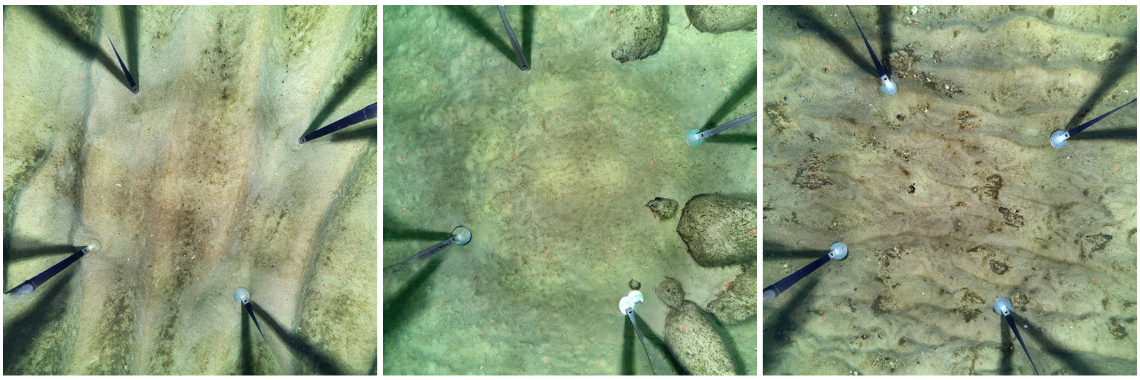


Figure 10. Photo mosaic examples of sites classified as sand.

Mixed substrate (code M)

HELCOM HUB definition: Less than 90% coverage of a certain substrate type. Mixed substrates comprise any proportion of mix of any substrate type of soft/mobile and/or hard/non-mobile substrates.

SGU definition: Areas where the substrate components hard clay or large stones (> 60–200 mm), boulders (> 200–600 mm) and large boulders (> 600 mm) together – i.e. rock and boulders – or soft clay (≤ 0.002 mm), silt (> 0.002–0.06 mm), sand (> 0.06–2 mm), gravel (> 2–20 mm), and pebbles and stones (> 20–60 mm) together – i.e. soft sediment – cover less than 90% of the seabed respectively (Fig. 11).

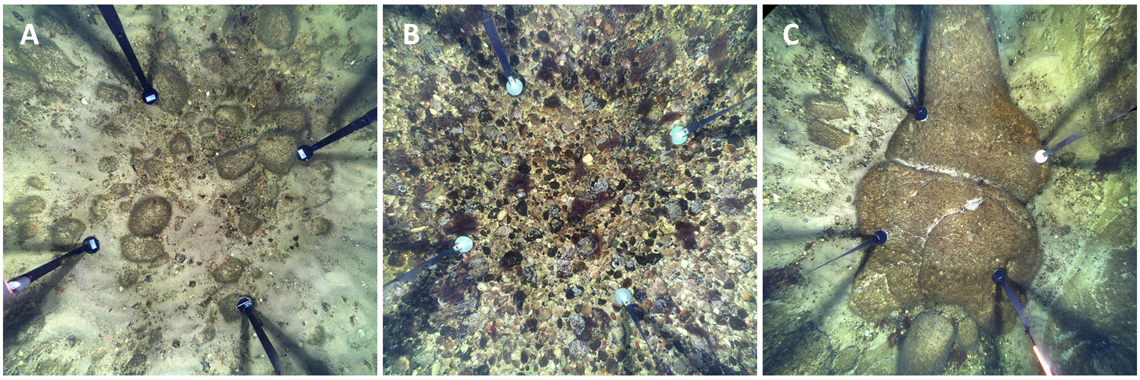


Figure 11. The range of sites classified as mixed sediment in HELCOM HUB is large and include areas dominated by: **A.** Sand. **B.** Pebbles and stones. **C.** A mix of hard clay, rock and boulders and sand.

HELCOM HUB biotopes

HELCOM HUB level 4: Community structure

Characterised by macroscopic epibenthic biotic structures (Code 1)

HELCOM HUB definition: At least 10% coverage of macroscopic vegetation or sessile macroscopic epifauna.

SGU definition: Areas where the biological component colonised substrate covers at least 10% of the seabed.

Characterised by sparse macroscopic epibenthic biotic structures (Code 2)

HELCOM HUB definition: Coverage of macroscopic vegetation or sessile macroscopic epifauna is greater than 0% and less than 10%.

SGU definition: Areas where the biological component colonised substrate is present (defined by a nominal threshold of at least 0.1% seabed cover in the model) but covers less than 10% of the seabed.

Characterised by no macroscopic epibenthic biotic structures (Code 4)

HELCOM HUB definition: No macroscopic vegetation or sessile macroscopic epifauna.

SGU definition: Areas where the biological component colonised substrate is absent (defined by a nominal threshold of less than 0.1% seabed cover in the model).

HELCOM HUB levels 5 and 6: Characteristic community and Dominating taxa

Characterised by perennial algae (code C)

HELCOM HUB definition: Perennial attached algae cover at least 10% of the seabed and more than other perennial attached erect groups.

SGU definition: Areas where the biological component perennial algae cover at least 10% of the seabed and more than epibenthic bivalves, epibenthic cnidarians, and epibenthic moss animals respectively (Fig. 12).

Species included in group: *Battersia arctica, Coccotylus/Phyllophora* (complex), *Delesseria sanguinea,* Filamentous Rhodophyceae, *Furcellaria lumbricalis, Polysiphonia/Rhodomela* (complex), Unidentified Rhodophyceae.

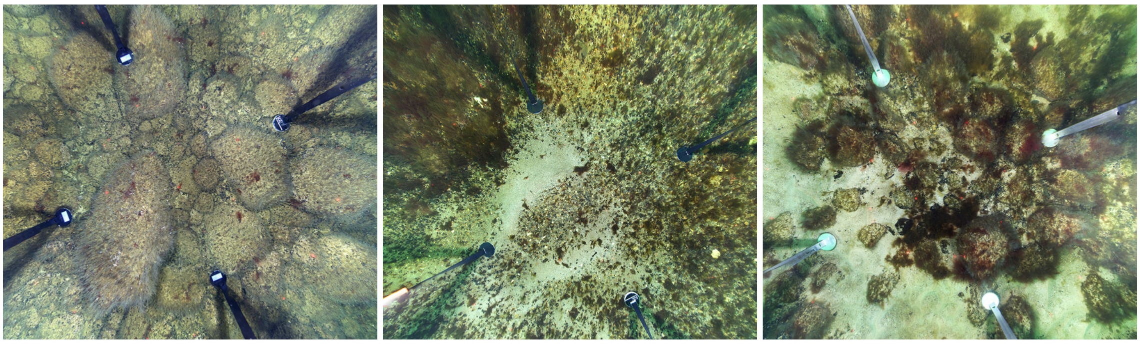


Figure 12. Photo mosaic examples of sites classified as being characterised by perennial algae.

Dominated by perennial filamentous algae (code C5)

HELCOM HUB definition: Perennial attached algae cover at least 10% of the seabed, and more than other perennial attached erect groups. Out of the perennial attached algae perennial filamentous algae constitute at least 50% of the biovolume.

SGU definition: Areas where the biological component perennial algae cover at least 10% of the seabed and more than epibenthic bivalves, epibenthic cnidarians, and epibenthic moss animals respectively, and perennial filamentous algae cover at least 50% of the seabed (Fig. 13)

Species included in group: *Battersia arctica,* Filamentous Rhodophyceae, *Polysiphonia/Rhodomela* (complex).

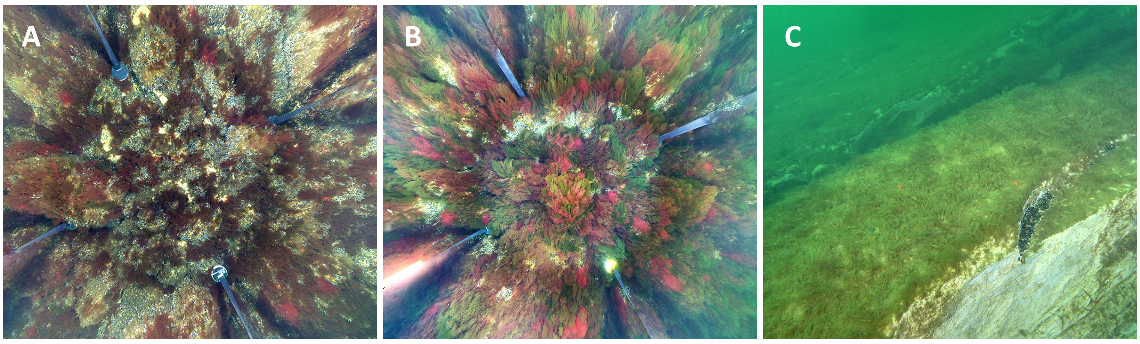


Figure 13. Photo mosaic examples of sites classified as being dominated by filamentous perennial algae.
A & B. Perennial algae attached to large rocks and boulders. **C.** Perennial algae attached to hard clay substrate.

Characterised by epibenthic bivalves (code E)

HELCOM HUB definition: Sessile/semi-sessile epibenthic bivalves cover at least 10% of the seabed and more than other perennial attached erect groups.

SGU definition: Areas where the biological component epibenthic bivalves cover at least 10% of the seabed and more than perennial algae, epibenthic cnidarians, and epibenthic moss animals respectively (Fig. 14).

Species included in group: *Mytilus* spp.

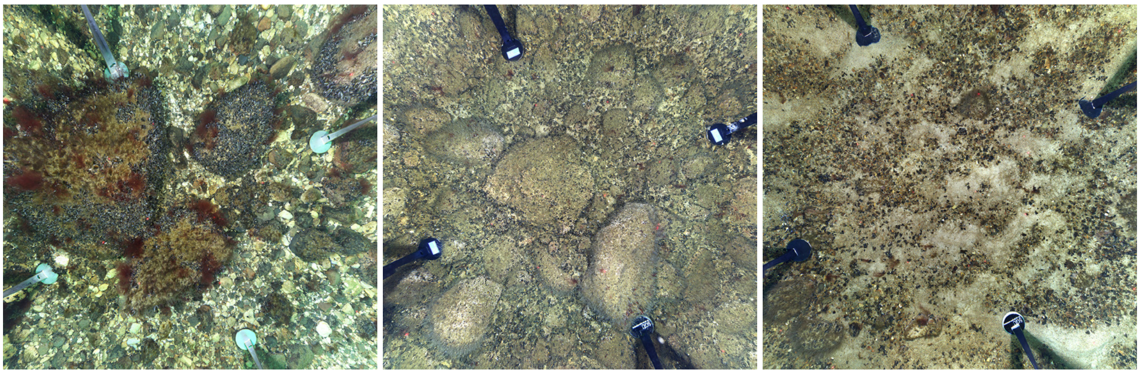


Figure 14. Photo mosaic examples of sites classified as being characterised by epibenthic bivalves.

Dominated by Mytilidae (Code E1)

HELCOM HUB definition: Epibenthic bivalves cover at least 10% of the seabed and more than other perennial attached erect groups. Out of the epibenthic bivalves Mytilidae constitute at least 50% of the biomass.

SGU definition: Areas where the biological component epibenthic bivalves cover at least 10% of the seabed and more than perennial algae, epibenthic cnidarians, and epibenthic moss animals respectively, and epibenthic bivalves cover at least 50% of the seabed (Fig. 15).

Species included in group: *Mytilus* spp.

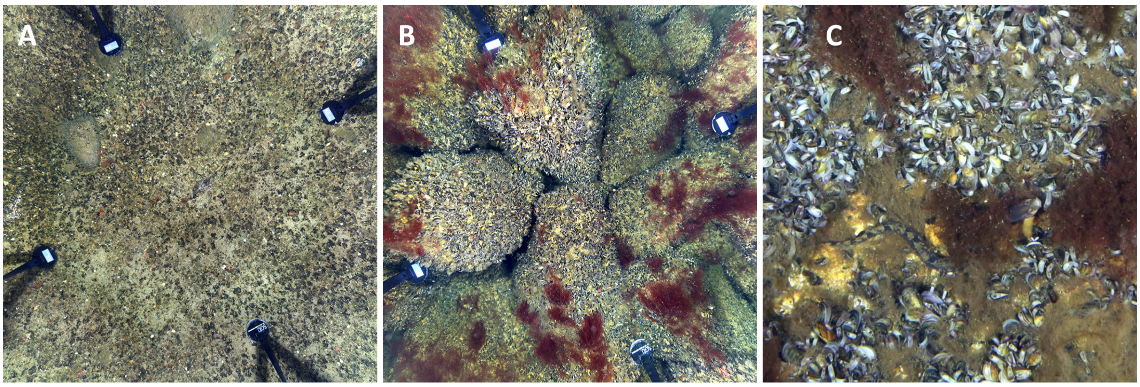


Figure 15. A & B. Photo mosaic examples of sites classified as being dominated by Mytilidae. **C.** Area completely covered by bivalves with fish and algae.

Characterised by epibenthic cnidarians (Code G)

HELCOM HUB definition: Sessile/semi-sessile epibenthic cnidarians cover at least 10% of the seabed and more than other perennial attached erect groups.
SGU definition: Areas where the biological component epibenthic cnidarians cover at least 10% of the seabed and more than perennial algae, epibenthic bivalves, and epibenthic moss animals respectively (Fig. 16).

Species included in group: Hydrozoa *(Cordylophora caspia)*.

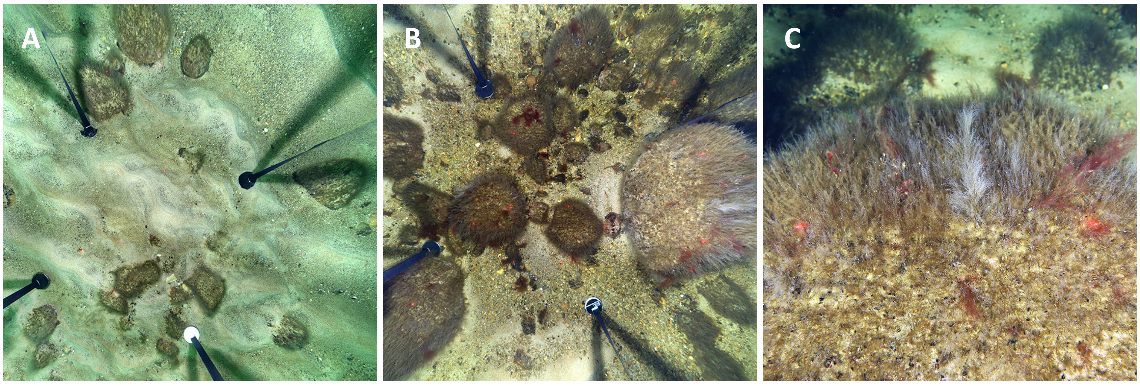


Figure 16. Photo mosaic examples of sites classified as characterised by cnidarians. **A & B.** Generally found on hardbottom substrates where algae growth is limited by light availability. **C.** Oblique view of rock with cnidarians and filamentous algae. Note, distance between two red, pointer lights in photo C is ~30 cm.

Dominated by hydroids (Code G1)

HELCOM HUB definition: Epibenthic cnidarians cover at least 10%, and more than other perennial attached erect groups. Out of the attached epibenthic cnidarians hydroids represent at least 50% of the biomass.
SGU definition: Areas where the biological component epibenthic cnidarians cover at least 10% of the seabed and more than perennial algae, epibenthic bivalves, and epibenthic moss animals respectively, and epibenthic cnidarians cover at least 50% of the seabed (Fig. 17).

Species included in group: Hydrozoa *(Cordylophora caspia)*.

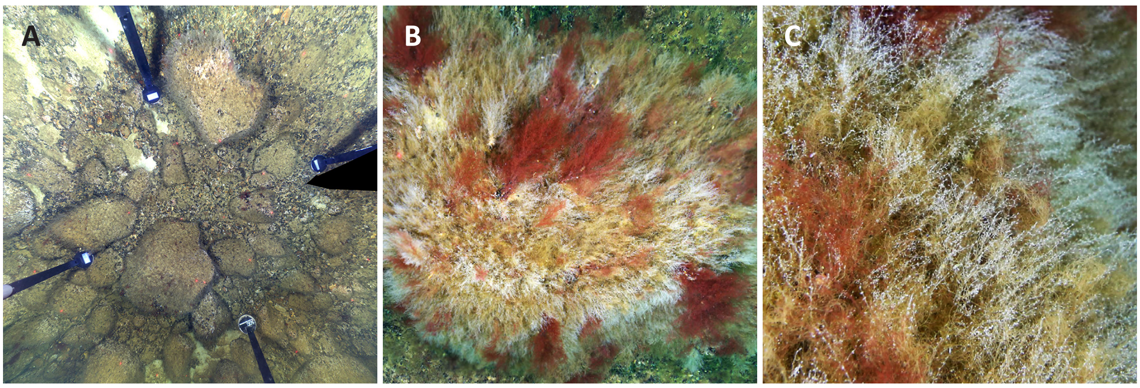


Figure 17. Photo example of site classified as being dominated by hydroids. **A.** Closeup view of a rock dominated by hydroids that is interspersed with filamentous algae (**B & C**).

Characterised by epibenthic moss animals (Code H)

HELCOM HUB definition: Sessile/semi-sessile epibenthic moss animals cover at least 10% of the seabed and more than other perennial attached erect groups.
SGU definition: Areas where the biological component epibenthic moss animals covers at least 10% of the seabed and more than perennial algae, epibenthic bivalves, and epibenthic cnidarians respectively (Fig. 18).

Species included in group: *Electra* sp.

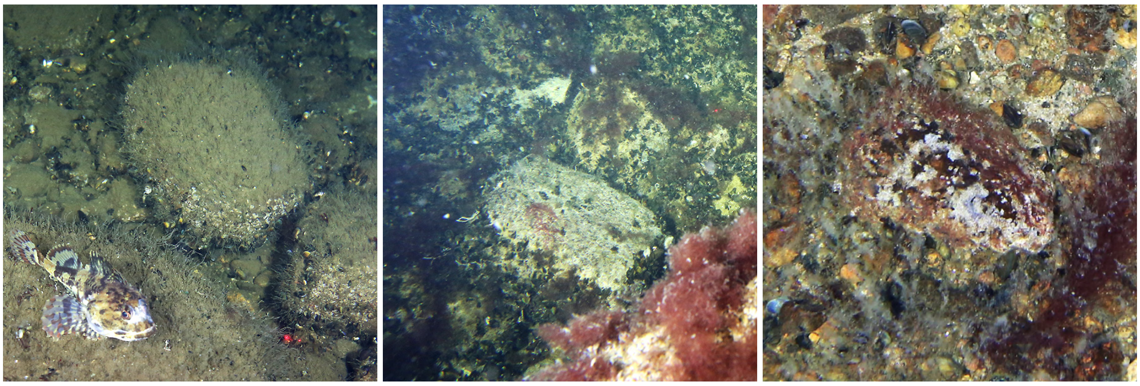


Figure 18. Photo examples of epibenthic moss animals (*Electra* sp.). No sites were characterised as epibenthic moss animals, however the cover of moss animals were greater than 10% in some locations. .

Characterised by annual algae (Code S)

HELCOM HUB definition: Annual algae cover at least 10% of the seabed, while all other epibenthic biotic structures cover less than 10%.
SGU definition: Areas where the biological components perennial algae, epibenthic bivalves, epibenthic cnidarians, and epibenthic moss animals cover less than 10% of the seabed respectively and annual algae covers at least 10%.

Species included in group: *Ceramium tenuicorne, Ceramium* sp., *Chorda filum, Dictyosiphon/Stictyosiphon* (complex), *Ectocarpus/Pylaiella* (complex), Filamentous Phaeophyceae, *Halosiphon tomentosus.*

Dominated by filamentous annual algae (Code S1)

HELCOM HUB definition: Annual algae cover at least 10% of the seabed, while all other vegetation covers less than 10%. Out of the annual algae, filamentous annual algae constitute at least 50% of the biovolume.
SGU definition: Areas where the biological components perennial algae, epibenthic bivalves, epibenthic cnidarians, and epibenthic moss animals cover less than 10% of the seabed respectively and annual algae cover at least 50% (Fig. 19).

Species included in group: *Ceramium tenuicorne, Ceramium* sp., *Chorda filum, Dictyosiphon/Stictyosiphon* (complex), *Ectocarpus/Pylaiella* (complex), Filamentous Phaeophyceae, *Halosiphon tomentosus.*

Note: *H. tomentosus* and *C. filum* are not considered part of the filamentous annual algae group. However, they only occurred at a small number of sites (n = 10 and 2 respectively) and had very low coverage (< 4%), so the annual algae biological component was considered a proxy for filamentous annual algae, despite its inclusion of *H. tomentosus* and *C. filum.*

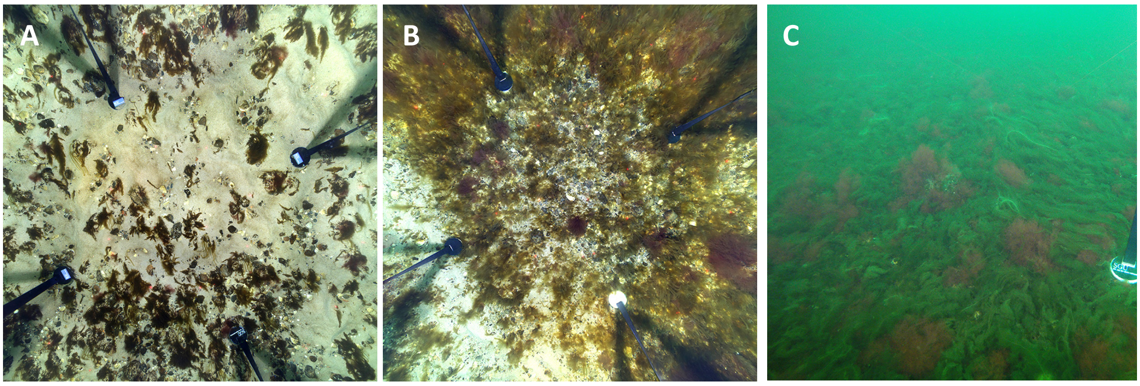


Figure 19. A. Photo mosaic example of site classified as being characterised by annual algae. **B.** Photo mosaic examples of site classified as being dominated by filamentous annual algae. **C.** Photo from site dominated by filamentous annual algae (mainly filamentous Phaeophyceae).

Characterised by mixed epibenthic macrocommunity (Code V)

HELCOM HUB definition: Macroscopic vegetation or sessile macroscopic epifauna is present but none of them cover more than 10% of the seabed.
SGU definition: Areas where the biological components perennial algae, epibenthic bivalves, epibenthic cnidarians, epibenthic moss animals and annual algae cover less than 10% of the seabed respectively, but together cover at least 10% (Fig. 20).

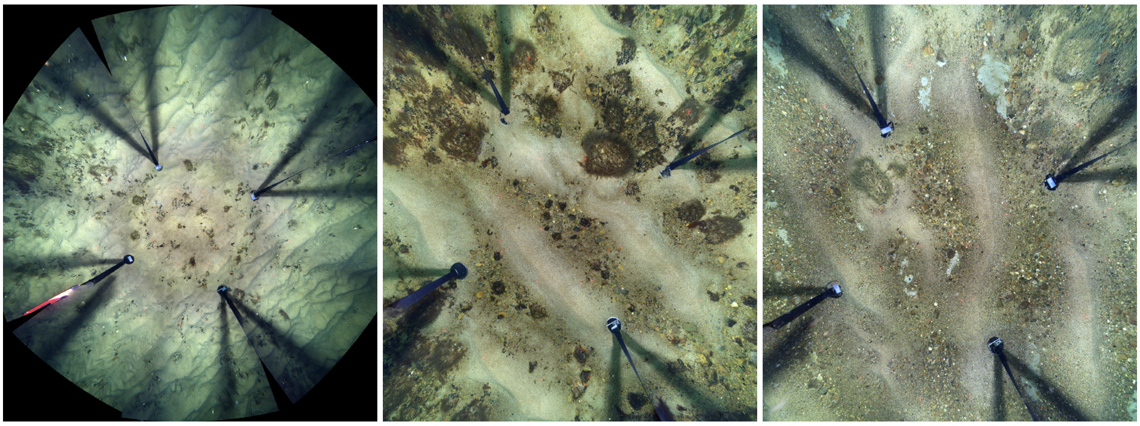


Figure 20. Photo mosaic examples of sites classified as being characterised by mixed epibenthic macrocommunity.

Characterised by sparse epibenthic macrocommunity (Code T combined with Code 2 from level 4)

HELCOM HUB definition: Coverage of macroscopic vegetation or sessile macroscopic epifauna is greater than 0% and less than 10%.
SGU definition: Areas where the biological component colonised substrate is present (defined by a nominal threshold of at least 0.1% seabed cover in the model) but covers less than 10% of the seabed (Fig. 21).

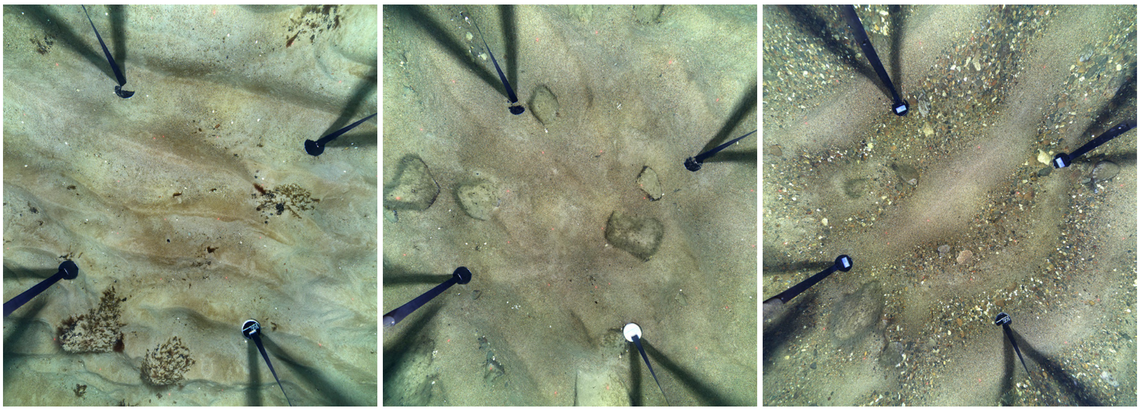


Figure 21. Photo mosaic examples of sites classified as being characterised by sparse epibenthic macrocommunity.

Characterised by no macrocommunity (Code U combined with Code 4 from level 4)
HELCOM HUB definition: No macroscopic vegetation or sessile macroscopic epifauna is present.
SGU definition: Areas where the biological component colonised substrate is absent (defined by a nominal threshold of less than 0.1% seabed cover in the model) (Fig. 22).

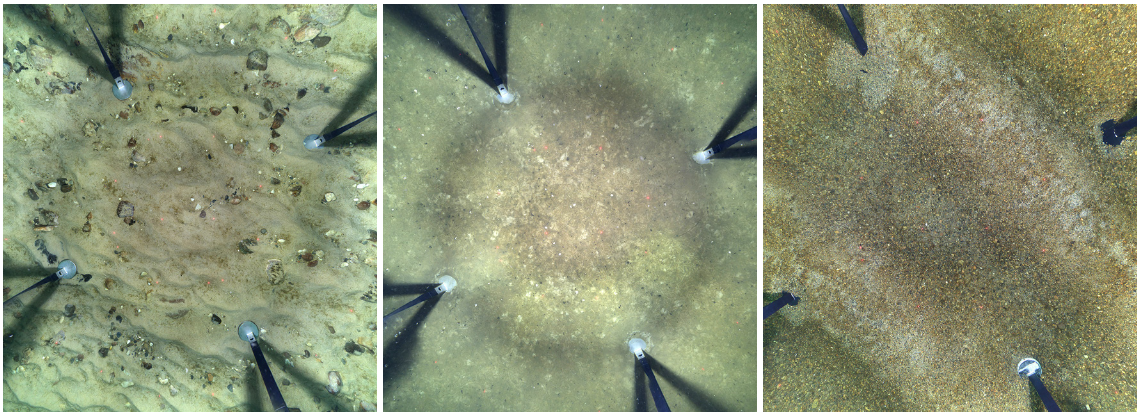


Figure 22. Photo mosaic examples of sites classified as being characterised by no macrocommunity.

HELCOM HUB levels 1–6 (combined)

The final HELCOM HUB map, consisting of different combinations of the previously described classes from HELCOM HUB levels 1–5 or 1–6.

Natura 2000

Natura 2000 is a network of environmentally sensitive areas across all 28 EU countries, both on land and at sea, that are protected by EU-wide legislation because they serve as core breeding and resting sites for rare and threatened species (approximately 2000), and areas of rare natural habitat types (approximately 230). The aim of the network is to ensure the long-term survival of Europe's most valuable and threatened species and habitats, listed under both the Birds Directive and Habitats Directive.

Each EU member state identifies sites for conservation from the listed habitat types and species occurring on their territory based on ecological grounds and commonly agreed scientific criteria (European Commission 2008). Habitat types consist of a 4 digit code and are redefined in terms of the specific EU member state along with subtypes and typical and characteristic species that occur within the nature types, and the member nation that contain a specific code for the nature type, and lastly, a definition of the subtype if required. 16 different habitat types occur in Swedish coastal and marine areas, Hoburgs bank consisted of two of these habitat types (European Commission 2008).

Reef (Code 1170)

Natura 2000 general definition: Reefs can be either biogenic concretions or of geogenic origin. They are hard compact substrata on solid and soft bottoms, which arise from the sea floor in the sub-littoral and littoral zone. Reefs may support a zonation of benthic communities of algae and animal species as well as concretions and corallogenic concretions.

Swedish Natura 2000 interpretation: Reefs are delimited in relation to surrounding areas if the seabed has a soft bottom cover greater than 50% and/or if biogenic formations (mussels and/or oysters) cover less than 10%.

SGU definition: Areas where the substrate components hard clay, large stones (> 60–200 mm), boulders (> 200–600 mm) and large boulders (> 600 mm) combined – i.e. hard bottom – cover at least 50% of the seabed or where hard bottom covers less than 50% of seabed but where the biological component epibenthic bivalves cover at least 10%.

SGU Reef subtypes

SGU developed reef subtypes based on features found in the project area and general Natura 2000 descriptions. Due to a lack of explicit detail in the general definitions, classification thresholds were based on SGU expert interpretation and knowledge of the area.

Flat

SGU definition: Reef areas with low fine-scale topographic complexity i.e. the remaining reef areas that did not fit the descriptions for rock boulder, ridge, or clay (described below).

Rock boulder

SGU definition: Reef areas with a fine-scale BPI of less than 2 (BPI with 5 m and 25 m inner and outer radius respectively) and the substrate components boulders (> 200–600 mm) and large boulders (> 600 mm) combined cover at least 1% of the seabed, or large boulders cover at least 0.1% of the seabed.

Ridge

SGU definition: Reef areas with a fine-scale BPI of at least 2 (BPI with 5 m and 25 m inner and outer radius respectively) where the substrate component hard clay covers less than 50% of the seabed.

Clay

SGU definition: Reef areas where the substrate component hard clay covers at least 50% of the seabed.

Mussels > 10%

SGU definition: Reef areas where the biological component epibenthic bivalves covers at least 10% of the seabed.

Mussels < 10%

SGU definition: Reef areas where the biological component epibenthic bivalves covers less than 10% of the seabed.

Geogenic
Natura 2000 definition: Reefs of geogenic origin, i.e. formed by non-biogenic substrata
Biogenic
Natura 2000 definition: Biogenic concretions, encrustations, corallogenic concretions and bivalve mussel beds originating from dead or living animals, i.e. biogenic hard bottoms that provide habitats for epibiotic species.

Sandbank (Code 1110)

Natura 2000 general definition: Sandbanks are elevated, elongated, rounded or irregular topographic features, permanently submerged and predominantly surrounded by deeper water. They consist mainly of sandy sediments, but larger grain sizes, including boulders and cobbles, or smaller grain sizes including mud may also be present on a sandbank. Banks where sandy sediments occur in a layer over hard substrata are classified as sandbanks if the associated biota are dependent on the sand rather than on the underlying hard substrata.

Swedish Natura 2000 interpretation: Sandbanks are delimited in relation to surrounding areas if the seabed has a sandy sediment cover of less than 50% and/or if biogenic formations cover more than 10% of the seabed.

SGU definition: Areas where hard bottom is less than 50% and the substrate components sand (> 0.06–2 mm), gravel (> 2–20 mm) and pebbles and stones (> 20–60 mm) together – i.e. sand and gravel – are greater than the combined coverage of silt (> 0.002–0.06 mm) and soft clay (≤ 0.002 mm) – i.e. soft bottom – and the biological component epibenthic bivalves cover less than 10% of the seafloor.

Note: According to both the general definition and Swedish interpretation, sandbanks seldom occur at depths greater than 20 m but may extend deeper than this. This threshold was not considered relevant for Hoburgs bank and was therefore not included. However, if required, the threshold can be implemented through combination with the bathymetric map.

SGU Sandbank subtypes

SGU developed additional sand bank modifiers based on features found in the project area. These included blue mussels and sand ripples presence/absence. The latter was observed to be a driver of biological assemblages in the area. Sand ripples (or sand waves) are considered to be an important factor in describing seafloor habitats in other regions, with some species specifically targeting dynamic sand environments (Greene et al. 2017).

Ripples
SGU definition: Sandbank areas where small to large ripples occur in the additional model sand ripples.
No ripples
SGU definition: Sandbank areas where no ripples occur in the additional model sand ripples.
Mussels < 10%
SGU definition: Sandbank areas where the biological component epibenthic bivalves are present (defined by a nominal threshold of 0.1% cover in the model) but cover less than 10% of the seabed (areas with at least 10% epibenthic bivalve cover are classified as reef).

Additional map products

SGU developed additional map products that captured other observed features that were not covered by any of the classification schemes used. These four map products capture postglacial sand deposits, sand transport, *Macoma balthica* shells and crawl tracks. These products were not intended for setting a new classification standard but rather to encourage exploration of potentially important habitat features.

Postglacial sand depth

Gives the depth of postglacial sand in metres. Postglacial sand depth was interpreted from SBP profiles that were then interpolated.

Sand ripples

Provides an estimate of the magnitude of sand transport by classifying the occurrence of sand ripples into different size classes. The larger the distance between sand ripples the greater the presumed sand transport. Sand ripple size is defined as the distance between the peak of one sand ripple and the peak of an adjacent sand ripple. This was determined visually from GT images (Fig. 23).

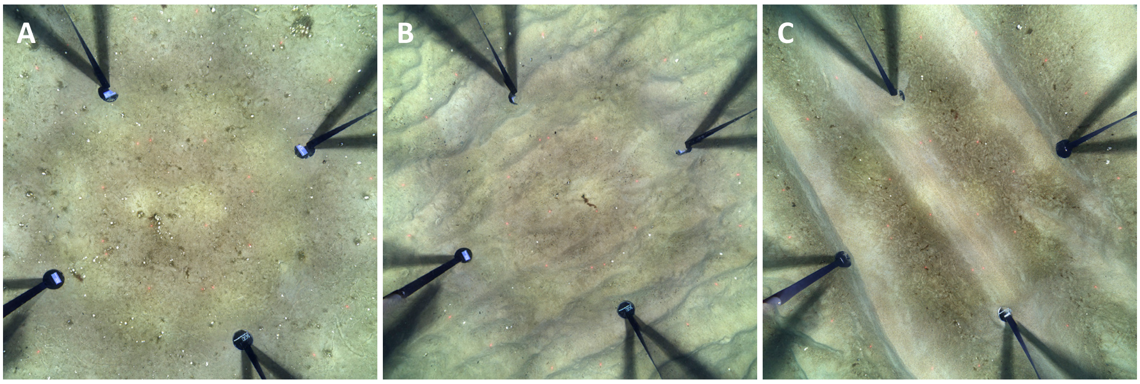


Figure 23. Photo mosaic examples of sand ripple sites classified as: **A.** No ripples. **B.** Small ripples (≥ 1 < 50 cm). **C.** Medium ripples (≥ 50 < 100 cm).

Macoma balthica shells

Provides a rough estimate of the prevalence of *Macoma balthica* shells (includes both live and dead individuals), determined visually from GT images (Fig. 24).

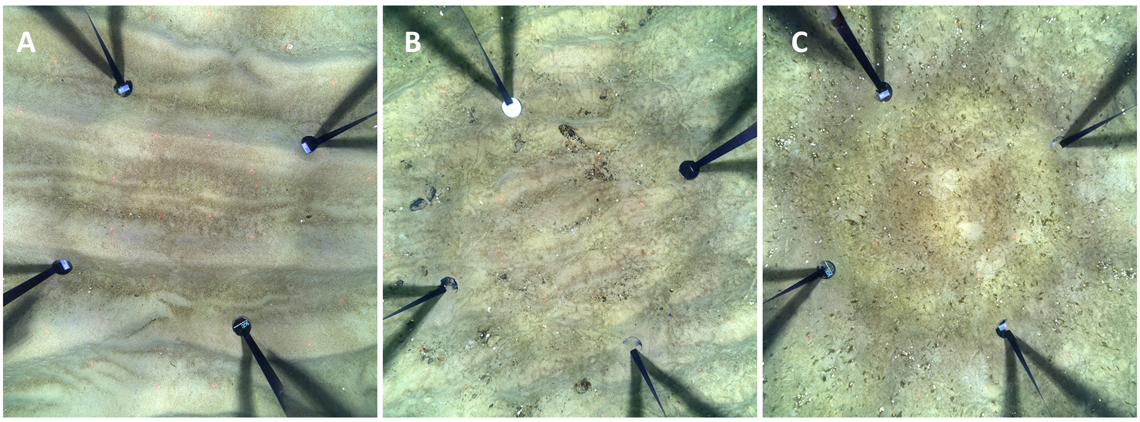


Figure 24. Photo mosaic examples of sites classified as **A.** sparse **B.** common and **C.** very common for the prevalence of Macoma balthica shells.

Crawl tracks

Indicates areas of bioturbation based on the prevalence of crawl tracks, determined visually from GT images (Fig. 25).

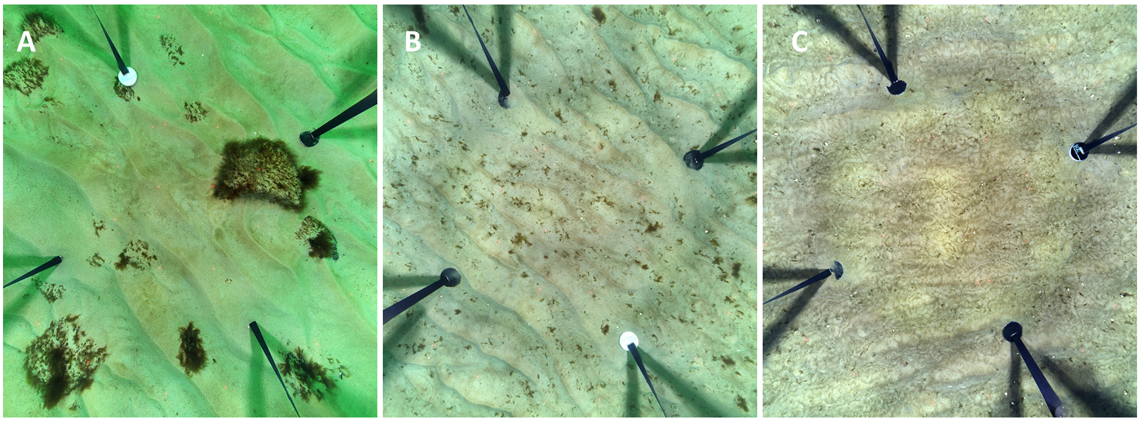


Figure 25. Photo mosaic examples of sites classified as **A.** sparse **B.** common and **C.** very common for the prevalence of crawl tracks.

RESULTS

Below we describe the results from the project structured along three main themes: 1) Data overview, maps and statistics describing the Hoburgs bank area; 2) End-user applications, including the effect of scale and predefined thresholds; and 3) Technical comparison of the methods used and the importance of different predictor variables.

Maps and statistics

This project has resulted in the first full coverage high-resolution benthic habitat maps of the Hoburgs bank area, mapping ~1,344 km^2 of complex seabed habitats. The spatial resolution is 5 m and the thematic resolution covers geomorphological, substrate and biological components, captured both as continuous variables of coverage (%) where possible, and as classified HELCOM HUB, Natura 2000, and SGU substrate maps. The overall accuracy of the HELCOM HUB maps were 80.5% for level 3 (substrate), 81.9% for level 4 (community structure), 62.3% for levels 4–5 (characteristic community), and 53.2% for levels 4–6 (dominating taxa). The overall accuracy for Natura 2000 reef/sandbanks was 87.7%. Detailed statistics of all models, both continuous percent coverage and classified thematic maps, are presented in the sections below.

Survey data overview

The survey of Hoburgs bank resulted in a full coverage bathymetry model as well as a backscatter mosaic, both at a resolution of 0.5 m, 589 drop camera sites with video and high-resolution images, of which 508 sites had oceanographic measurements recorded (i.e. O_2, current speed and direction, salinity, and temperature), 443 sediment samples, 385 salinity and temperature profiles from CTD casts and approximately 2,200 SVPs from CTD and MVP casts, 30 Secchi disk measurements, as well as some sampling of infauna (see Karlsson et al. 2017) and fish (data available in SLU Aqua's kustfiskedatabas-KUL).

Tables 5 and 6 provide an overview of the data and formats that have been collected, the products that have been derived from the data, and where data and products can be accessed. Due to the development of numerous new products, SGU is still (at the time of the writing) working on how to best distribute this data. For the latest update, go to the project webpage or contact SGU directly through customer service. Some data will also be published through the Svenskt Havsarkiv (SHARK) database hosted by the Swedish Meteorological and Hydrological Institute (SMHI). Data collected on Hoburgs bank is published under CC0, and will be available free of charge, aside from possible administrative fees associated with distribution.

Table 5. Overview of data collected during the project.

	Instrument type	Equipment	Data type	Format	Date	License	Data owner
Survey	Multibeam echosounder	Kongsberg EM2040D	Bathymetry raw	.all	2016–2017	CC0	SGU*
	Multibeam echosounder	Kongsberg EM2040D	Backscatter raw	.all	2016–2017	CC0	SGU*
	Multibeam echosounder	Kongsberg EM2040D	Water column raw	.wcd	2016–2017	CC0	SGU*
	Sub-bottom profiler	iXblue Echoes 3500 T3 SBP	Raw data	.segy	2016–2017	CC0	SGU*
	Split-beam sonar	Kongsberg EK60	Water column raw	.raw	2017	CC0	SGU*
	SV sensor	Valeport mini SVS	Surface sound velocity	.svp	2016–2017	CC0	SGU*
	CTD/SVP probe	Valeport MIDAS SVX2	Sound velocity, pressure, temperature, conductivity	.txt	2016–2017	CC0	SGU*
	MVP probe	Valeport mini SVS	Sound velocity, pressure, temperature	.txt	2017	CC0	SGU*
	Secchi disk	20 cm plate	Secchi depth	.csv	2017	CC0	SGU*, SHARK**
Sampling	UW-drop camera CTD	Valeport mini CTD	Pressure, temperature, conductivity	.txt	2016–2017	CC0	SGU*
	UW-drop camera oxygen optode	Anderaa Oxygen Optode 4835	Dissolved oxygen	.txt	2016–2017	CC0	SGU*
	UW-drop camera doppler current sensor	Anderaa ZPulse® Doppler Current Sensor 4420	Water current speed and direction	.txt	2016–2017	CC0	SGU*
	UW-drop camera Compass	OceanServer Technology Inc. OS5000-S Fluxgate Compass	Heading	.txt	2016–2017	CC0	SGU*
	UW-drop camera camera 1	Canon EOS 6D DSLR	Image, video, heading	.jpg, .mov	2016–2017	CC0	SGU*
	UW-drop camera camera 2	GoPro hero 4 black	Image, video	.jpg, .mp4	2016	CC0	SGU*
	UW-drop camera camera 3	GoPro hero 5 black	Image, video	.jpg, .mp4	2017	CC0	SGU*
	Sediment sampling	Orange-peel bucket	Image, grain size interpretation	.jpg, .csv	2016–2017	CC0	SGU*
	Sediment sampling	Van Veen grab	Image, grain size interpretation	.jpg, .csv	2016–2017	CC0	SGU*

Table 5 continues

	Instrument type	Equipment	Data type	Format	Date	License	Data owner
	Sediment sampling	Van Veen grab	Grain size sieve analysis	.csv	2016–2017	CC0	SGU*
	Sediment sampling	Van Veen grab	Infauna analysis	.csv	2016–2017	CC0	SLU Aqua*, SHARK
	Fish sampling	Gill net	Fish analysis	.csv	2016	CC0	SLU Aqua*, KUL
Legacy data	Video transects	Video sled	Video (incl time and position)	.mpg, mplex	2005	CC0	SGU*
	Samples	Orange-peel bucket, Vibro corer	Image, interpreta-tion	.jpg, pdf	2004	CC0	SGU*
	UW-drop camera	Canon PowerShot G10 DSLR	Image	.jpg	2004	CC0	SGU*

* available on request

** not yet available, expected update 2019/2020

Habitat overview

The mapped area on Hoburgs bank ranged from 10 to 63 m in depth with a majority lying between 15 and 35 m. Sand and gravel were the predominant substrate types, covering approximately 80% of the seafloor surface, while hard bottom (hard clay and substrate types with a grain size ≥60 mm) covered approximately 19.3% (Fig. 26). The remaining 0.7% consisted of softer sediments (silt–clay fractions) that often occurred within sediments dominated by sand and gravel. 39% of the mapped area was classified as Natura 2000 Reef (1170), when using full resolution models, while the remaining area (61%) was classified as Sandbank (1110; all sand areas, no depth limitation). Due to the dynamic nature of the wave and current affected seabed, sand is transported across the bank and deposited in deeper areas. Areas where sand and gravel formed ripples covered 49% of the bank, while sand without ripples covered 9%. Sand movement also affected hard bottom habitats where a thin veneer of sand often covered parts of the seabed. It is likely that some of these areas shift between a soft bottom or hard bottom habitat as sand moves across the seabed. The main hard bottom habitats where sand was absent were large boulders scattered over the bank (0.3% of the area), as well as a series of distinct moraine ridges (0.7% of the bank according to HELCOM HUB level 3 rock and boulders, and 1.7% according to the SGU implementation of the ridge reef definition in Natura 2000 subtypes), and hard clay features (0.1% of the bank according to HELCOM HUB level 3). These structurally complex reefs (especially the moraine ridges) were noted to have a higher abundance and diversity of fish compared with other seafloor types based on data from drop camera imagery, which included sporadic observations of cod hiding or feeding near or on the reefs. Though the total cover for these reef features was only 1–2% they were scattered over a large area, where almost half of total area surveyed had a reef feature within a few hundred metres.

Epibenthic bivalves (*Mytilus* spp.) and filamentous red algae dominated the benthic flora and fauna in hard bottom areas, while the sandy areas were mostly uncolonised by sessile organisms. Colonised seafloor covered 26% of the mapped area. The shallowest areas (10–15 m) had a more diverse algal assemblage than the deeper areas and most annual algae species where found there. Observations in the deeper (~45–60 m) fine sand areas showed notable aggregations of *Macoma balthica*, crawl marks, microphyte detritus, and small crustaceans, while the observations from shallower more dynamic areas indicated lower abundances of these organisms. Epibenthic bivalves had an average cover of 10.7% over the mapped area and occurred almost exclusively on hard substrates (> 60 mm) and pebbles and stones (> 20–60 mm), with the densest aggregations found on moraine ridges. 35.6% of the mapped area had an epibenthic bivalve coverage between 10 and 50%, and 1.5% a coverage exceeding 50%. Epibenthic bivalves were found at all depths on the bank where hard bottom was present, though often at lower densities below 35–40 m. Algae covered approximately 15.5% of the bank, dominated by perennial red filamentous algae (14.8% cover). Annual algae covered only about 0.6% and was found mainly in the shallowest (10–15 m) areas. The deepest record of algae was a filamentous red alga at 38 m depth.

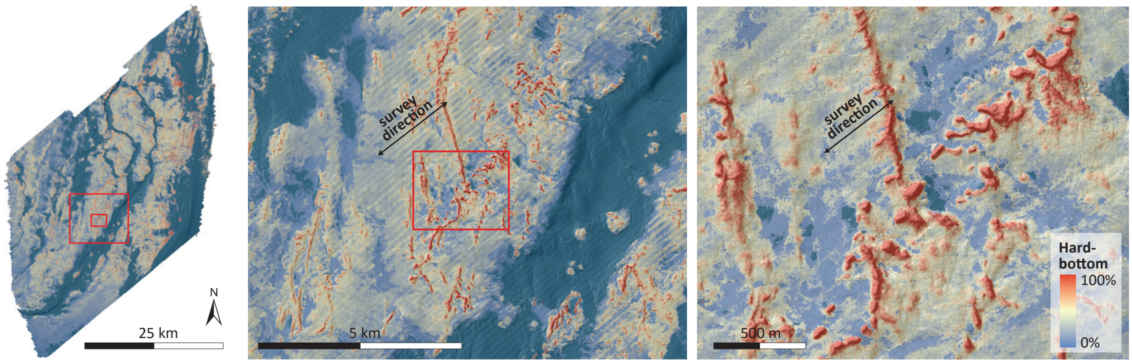


Figure 26. Model of percent coverage of hard bottom substrate on Hoburgs bank, showing detail at different zoom levels. The resolution of all substrate and biota models are 5 m, covering 1 344 km² of seafloor. Some visual artefacts, seen as striped pattern in the direction of the survey lines (SW/NE direction), remain in some places. The artefacts are most pronounced in areas with flat seafloor with hard–coarse sediment (challenging to predict), and where oceanographic conditions affected the quality of the sonar data.

Table 6. Overview of products produced from data collected during the project.

Product theme	Product	Unit	Format	Resolution	License	Data owner
Bathymetry	Bathymetry model	m, rh2000	geotiff	0.5 m, 1 m, 5 m, 10 m	CC0	SGU*, SwAM
Backscatter	Backscatter mosaic	dB	geotiff	0.5m, 1m, 5m, 10m	CC0	SGU*, SwAM
UW-observations	Substrate components	Coverage (%)	.csv		CC0	SGU*, SHARK**
UW-observations	Benthic organisms	Coverage (%)	.csv		CC0	SGU*, SHARK**
UW-observations	Fish observations	Presence	.csv		CC0	SGU*, SLUAqua
Geological model	Substrate hard clay (≤ 0.002 mm)	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Geological model	Substrate soft clay (≤ 0.002 mm)	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Geological model	Substrate silt (> 0.002–0.06 mm)	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM

Table 6 continues

Product theme	Product	Unit	Format	Resolution	License	Data owner
Geological model	Substrate sand (> 0.06–2 mm)	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 25 0m	CC0	SGU*, SwAM
Geological model	Substrate gravel (> 2–20 mm)	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Geological model	Substrate pebbles & stones (> 20–60 mm)	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Geological model	Substrate large stones (> 60–200 mm)	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Geological model	Substrate boulders (> 200–600 mm)	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Geological model	Substrate large boulders (> 600 mm)	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Geological model	Fine sand (> 0.06–0.2 mm)	Coverage (%)	geotiff	5 m	CC0	SGU*, SwAM
Geological model	Medium sand (> 0.2–0.6 mm)	Coverage (%)	geotiff	5 m	CC0	SGU*, SwAM
Geological model	Coarse sand (> 0.6–2 mm)	Coverage (%)	geotiff	5 m	CC0	SGU*, SwAM
Geological model	Sediment sample classes	Classes	geotiff	5 m	CC0	SGU*
Geological model	Postglacial sand depth	Metre	geotiff	5 m interpolated	CC0	SGU*, SwAM
Geological model	SGU surface substrate	Classes	geotiff	5 m	CC0	SGU*
Geological model	Sand ripples	Classes	geotiff	5 m	CC0	SGU*, SwAM
Geological model	Uncertainty substrate	SD coverage (%)	geotiff	5 m	CC0	SGU*, SwAM
Biological model	Uncolonised substrate	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Detritus	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Epibenthic sponges	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Epibenthic moss animals	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Epibenthic cnidarians	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Annual algae	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Perennial algae	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Perennial filamentous algae	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Perennial foliose red algae	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Perennial non-filamentous corticated red algae	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Soft crustose algae	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Biological model	Epibenthic bivalves	Coverage (%)	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM

Table 6 continues

Product theme	Product	Unit	Format	Resolution	License	Data owner
Biological model	Crawl tracks	Classes	geotiff	5 m	CC0	SGU*, SwAM
Biological model	Macoma balt-hica shells	Classes	geotiff	5 m	CC0	SGU*, SwAM
Biological model	Uncertainty biology	SD coverage (%)	geotiff	5 m	CC0	SGU*, SwAM
Habitat map	HELCOM HUB level 2	Classes	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Habitat map	HELCOM HUB level 3	Classes	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Habitat map	HELCOM HUB level 4	Classes	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Habitat map	HELCOM HUB level 4–5	Classes	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Habitat map	HELCOM HUB level 4–6	Classes	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Habitat map	HELCOM HUB level 1–6	Classes	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Habitat map	Natura 2000	Classes	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM
Habitat map	Natura 2000 subtypes	Classes	geotiff	5 m, 10 m, 25 m, 50 m, 250 m	CC0	SGU*, SwAM

* available on request
** not yet available, expected update 2019/2020

Map products

The following section describes all maps produced from data collected during the project, and the associated uncertainty.

Bathymetry

Depth in the mapped area ranged from 10.4 to 62.9 m with an average depth of 26.6 m (Fig. 27A). Overall, the survey fulfilled International Hydrographic Organization (IHO) S-44 requirements for Special Order (see Appendix 1). Line artifacts were present due to high beam width and reduced overlap.

Backscatter

Backscatter intensity in the mapped area ranged from -0.1 to -45.0 dB (Fig. 27B). Beam angle artifacts were observed, particularly at the nadir. In general, however, the map gave a good indication of sediment characteristics, with higher backscatter intensity, i.e. darker shading, relating to coarser or harder substrates (e.g. gravel, cobbles and boulders) and lower backscatter intensity i.e. lighter areas relating to softer sediments (e.g. silt, sand, and hard clay) (Fig. 27B). Some confusion occurred in areas of higher intensity, where additional depth metrics were required to further distinguish between different grain size fractions. Backscatter images allowed features such as sand ripples (depending on the survey direction relative to sand ripple direction) and individual large boulders to be observed directly from the 0.5 m resolution mosaics.

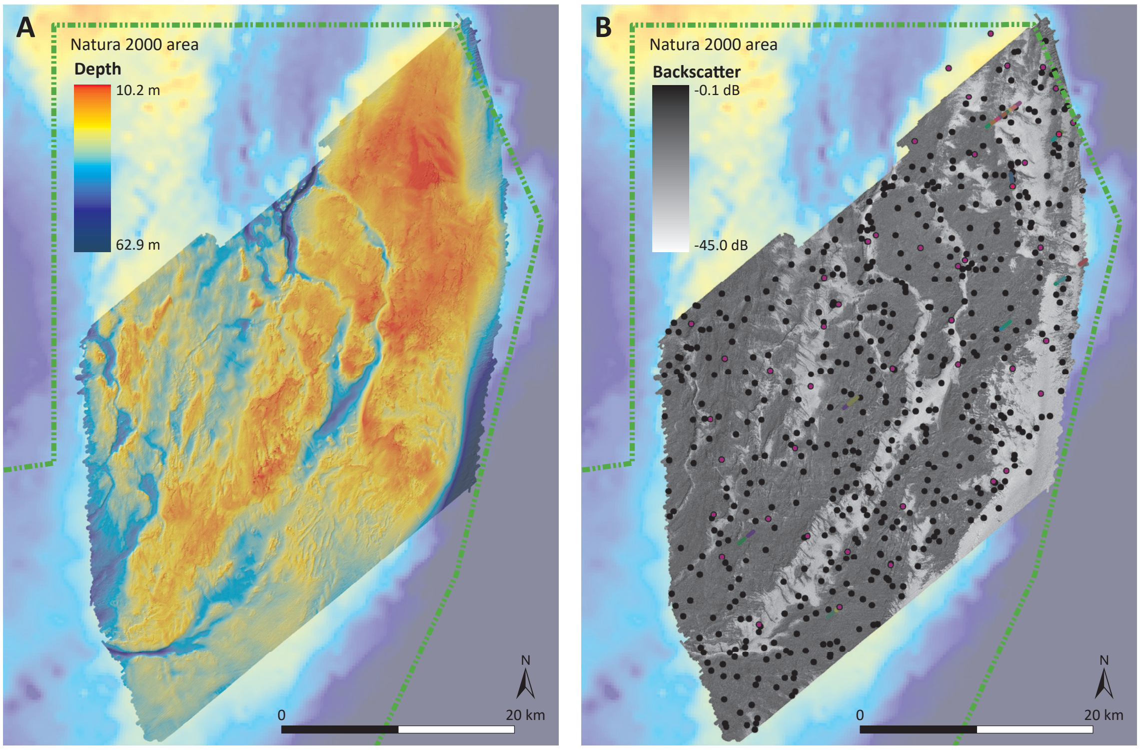


Figure 27. High-resolution (0.5 m) maps of Hoburgs bank from MBES data shown with Emodnet 3 bathymetry as background. **A.** Depth (bathymetric) grid. **B.** Backscatter mosaic and ground truthing sites (black dots denote 2016–2017 survey, coloured dots denote legacy data from 2005 survey).

Substrate components

In general, Hoburgs bank consisted of fairly shallow areas, with a patchy mixture of gravel, pebbles and stones, large stones and boulders in varying proportions, with large boulders present in smaller areas. Small areas with large stones, boulders and large boulders were locally abundant (Fig. 28E–G). In deeper areas below approximately 30 m, sand was the dominant substrate fraction (Fig. 28B and Fig. 27A).

Mean absolute error (MAE) ranged from 2.0% (hard clay) to 10.5% (pebbles and stones) across all substrate component models. However, the low MAE for hard clay was probably a result of a low occurrence of hard clay in the area. On average, there did not appear to be a substantial amount of bias in the models, with mean error (ME) ranging from -3.5% (gravel) to 1.6% (pebbles and stones). Regression analysis of observed and predicted substrate cover showed varying degrees of fit, depending on the substrate component, with r^2 ranging from 0.28 (pebbles and stones) to 0.81 (sand) (Table 7). To varying degrees, all models tended to underestimate towards the lower end of the cover scale and overestimate towards the upper end. This was exemplified by significant differences between the y-intercept and slope ($p < 0.05$, Table 7) compared to the 1:1 line for all models, apart from large boulders for slope, coupled with the visual assessment that regression lines tended to be above the 1:1 line at the lower end of the scale and below the 1:1 line at the upper end (Fig. 29). Moreover, positive ME in the < 10% class and a predominance of increasingly negative ME values in classes ≥ 10% also indicated this effect (Table 8). When predicted values were reclassified into interval classes < 10%, ≥ 10 < 50%, ≥ 50 < 90% and ≥ 90 ≤ 100% (OA classes) and presence-absence ≥ 0.01% (OA abs–pres), overall accuracy (OA) were generally high ranging from 75.3% (sand, pebbles and stones, and boulders) to 92.2% (hard clay), and 79.9% (large boulders) to 87.7% (large stones) respectively (Table 9).

Table 7. y-intercept, slope, and coefficient of determination (r^2) of substrate component regression models fitted to observed GT values against corresponding predicted values (n = 154). All regression models were significant (P < 0.001). Values of the y-intercept and slope in bold indicate significant differences from 0 and 1, respectively (P < 0.05).

Substrate component	y-intercept	Slope	r^2
Hard clay (≤ 0.002 mm)	**1.43**	**0.52**	0.36
Sand (> 0.06–2 mm)	**6.88**	**0.87**	0.81
Gravel (> 2–20 mm)	**2.97**	**0.73**	0.49
Pebbles & stones (> 20–60 mm)	**5.79**	**0.53**	0.28
Large stones (> 60–200 mm)	**3.83**	**0.46**	0.33
Boulders (> 200–600 mm)	**3.48**	**0.78**	0.56
Large Boulders (> 600 mm)	**1.30**	1.10	0.64

Table 8. Mean error (ME) ± standard error (SE) for substrate components where predicted coverage (%) values were divided into interval classes < 10%, ≥ 10 < 50%, ≥ 50 < 90%, and ≥ 90 ≤ 100%. The number of replicates in each class is denoted by n.

Interval class →	< 10%			≥ 10 < 50%			≥ 50 < 90%			≥ 90 ≤ 100%		
↓ Substrate component	n	ME	SE	n	ME	SE	n	ME	SE	n	ME	SE
Hard clay (≤ 0.002 mm)	152	1.3	0.4	0	-	-	0	-	-	2	-45.2	45.7
Sand (> 0.06–2 mm)	62	6.2	2.0	27	3.1	5.3	13	-6.8	8.0	52	-4.6	1.7
Gravel (> 2–20 mm)	107	2.3	0.8	37	-1.8	3.0	10	-13.3	5.3	0	-	-
Pebbles & stones (> 20–60 mm)	90	2.9	1.0	56	-0.9	3.0	7	-34.9	8.7	1	-57.1	-
Large stones (> 60–200 mm)	96	2.0	0.5	49	-8.7	2.9	9	-34.5	7.0	0	-	-
Boulders (> 200–600 mm)	106	3.0	0.9	31	1.1	4.2	15	-14.2	7.4	2	-24.0	9.1
Large Boulders (> 600 mm)	141	1.4	0.5	12	2.1	5.3	1	26.6	-	0	-	-

Table 9. Substrate component accuracy statistics, root mean square error (RMSE), mean absolute error (MAE) and mean error (ME), calculated from observed values from GT testing data against corresponding predicted values from models. Overall accuracy (OA) results from confusion matrices where predicted values were grouped according to interval classes (OA classes) (< 10%, ≥ 10 < 50%, ≥ 50 < 90%, ≥ 90 ≤ 100%) and absence (< 0.01%) presence (≥ 0.01%) classes (OA abs–pres).

Substrate component	RMSE	MAE (%)	ME	OA classes (%)	OA abs–pres (%)
Hard clay (≤ 0.002 mm)	8.8	2.0	0.7	92.2	80.5
Sand (> 0.06–2 mm)	19.3	10.0	0.9	75.3	85.1
Gravel (> 2–20 mm)	12.7	7.1	0.3	76.0	87.0
Pebbles & stones (> 20–60 mm)	18.2	10.5	-0.6	75.3	85.7
Large stones (> 60–200 mm)	16.2	9.4	-3.5	76.6	87.7
Boulders (> 200–600 mm)	16.7	8.5	0.6	75.3	81.8
Large Boulders (> 600 mm)	7.6	3.1	1.6	89.0	79.9

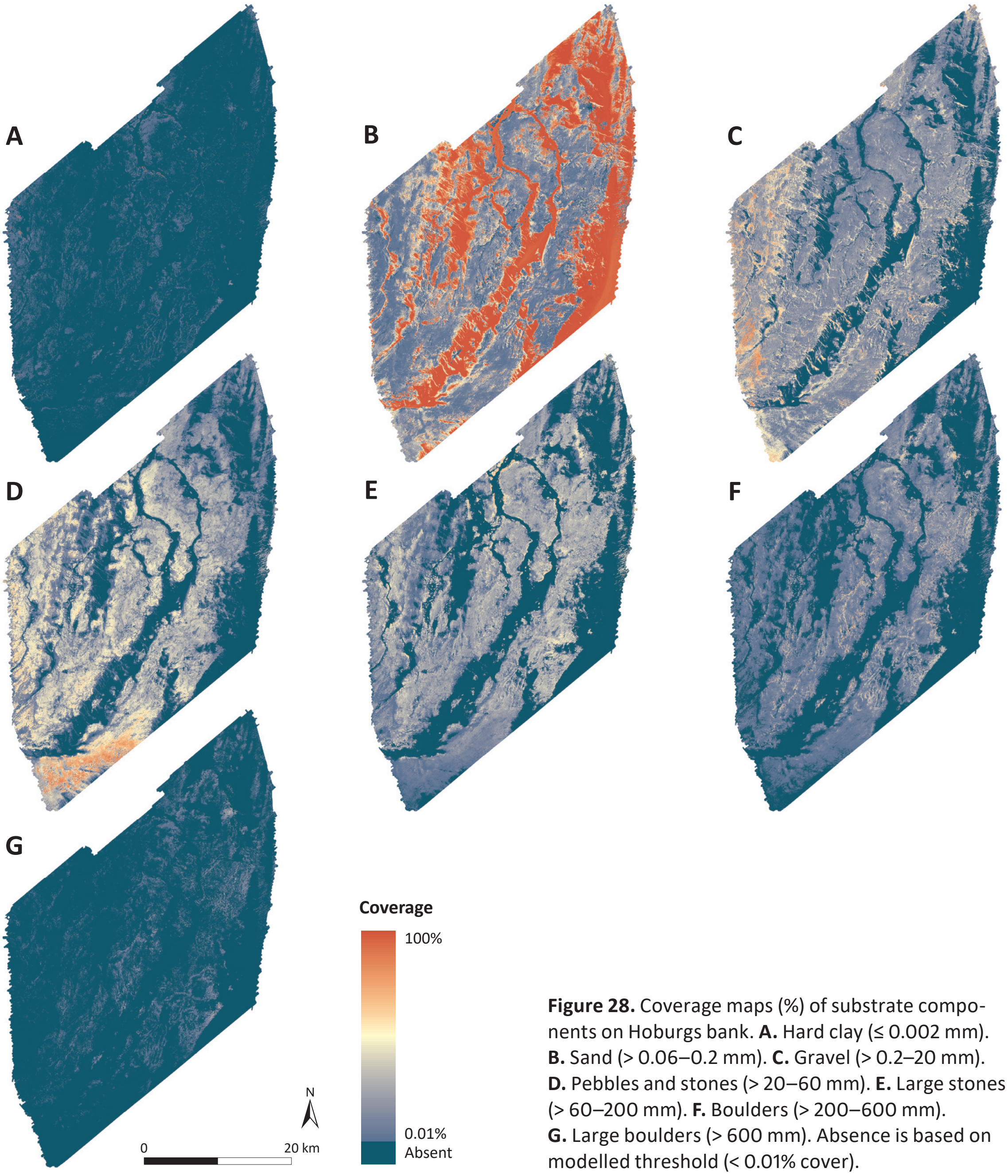
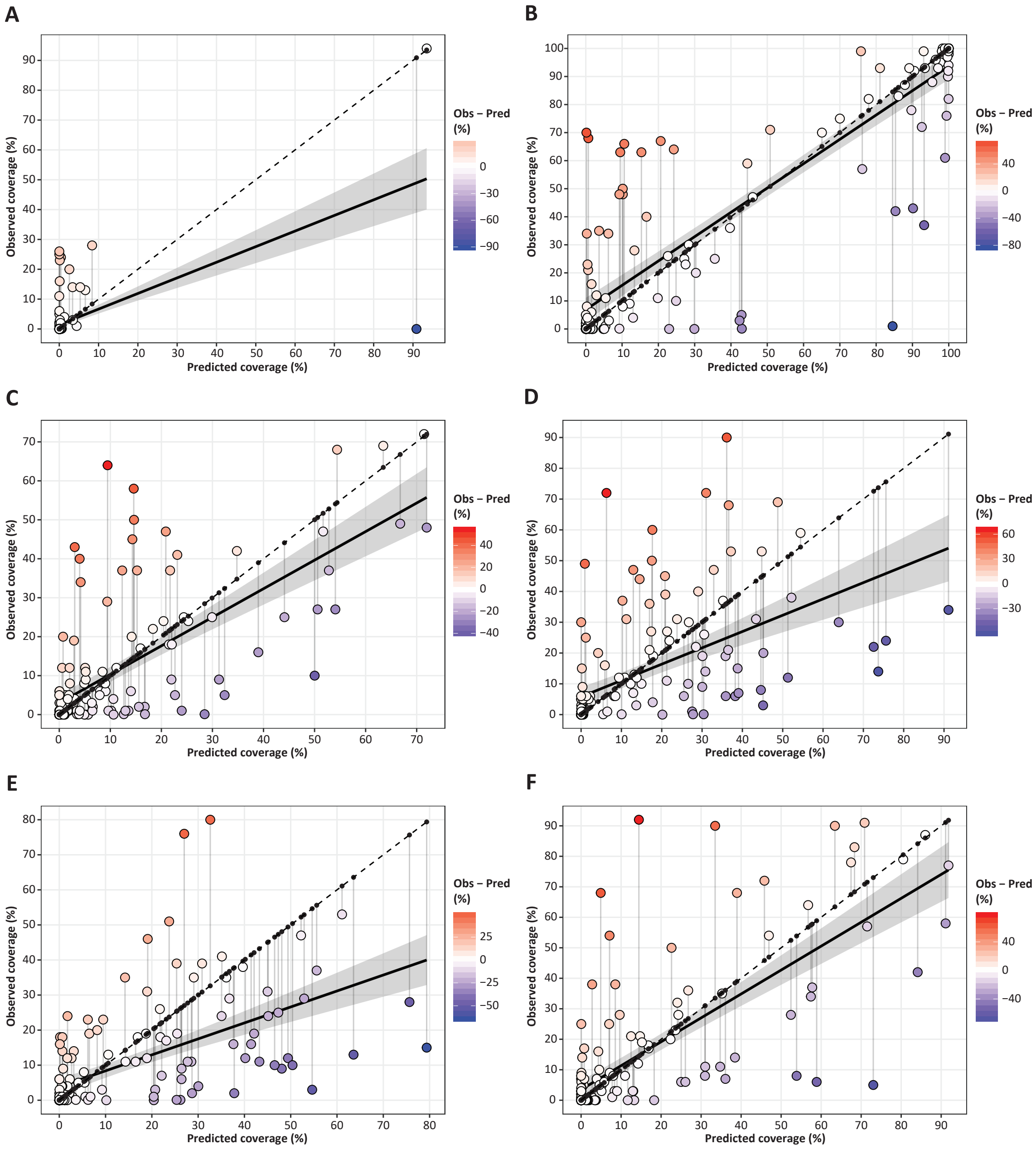


Figure 28. Coverage maps (%) of substrate components on Hoburgs bank. **A.** Hard clay (≤ 0.002 mm). **B.** Sand (> 0.06–0.2 mm). **C.** Gravel (> 0.2–20 mm). **D.** Pebbles and stones (> 20–60 mm). **E.** Large stones (> 60–200 mm). **F.** Boulders (> 200–600 mm). **G.** Large boulders (> 600 mm). Absence is based on modelled threshold (< 0.01% cover).



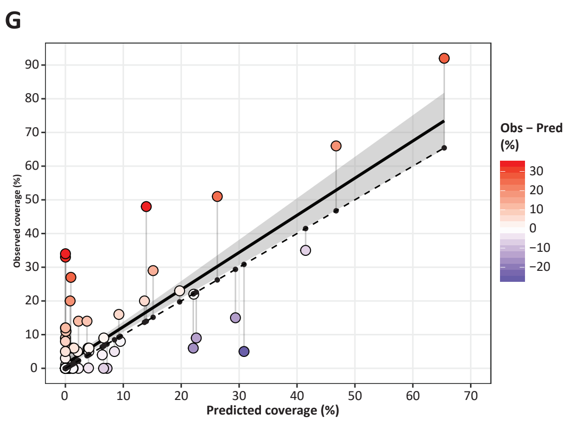


Figure 29. Relationship between observed coverage values (%) from GT testing data (n = 154) and corresponding predicted coverage values (%) from substrate component models. **A.** Hard clay (≤ 0.002 mm). **B.** Sand (> 0.06–0.2 mm). **C.** Gravel (> 0.2–20 mm). **D.** Pebbles and stones (> 20–60 mm). **E.** Large stones (> 60–200 mm). **F.** Boulders (> 200–600 mm). **G.** Large boulders (> 600 mm). Solid black lines represent linear regressions and 95% confidence bands (grey shade) between observed and predicted coverage. Broken lines represent a theoretical 1:1 line where observed and predicted values are equal (i.e. y = x, y-intercept = 0, slope = 1); black points along this line represent the specific predicted coverage values. Point colours represent the magnitude of difference between an observed value and its corresponding predicted value on the 1:1 line (black point).

Fine substrate components

Maps of fine substrate components (sieve analysis of fine substrate fractions) adjusted to the sand model are shown in Figure 30, with accuracy assessments using bootstrap shown in Table 10. In a visual assessment of the outputs, it was noticed that silt and fine sand accumulated in areas where postglacial sand deposits were observed in the SBP profiles. This observation often also corresponded with sediments with no sand ripples (included in Figure 41), and the occurrence of *Macoma balthica* shell aggregations.

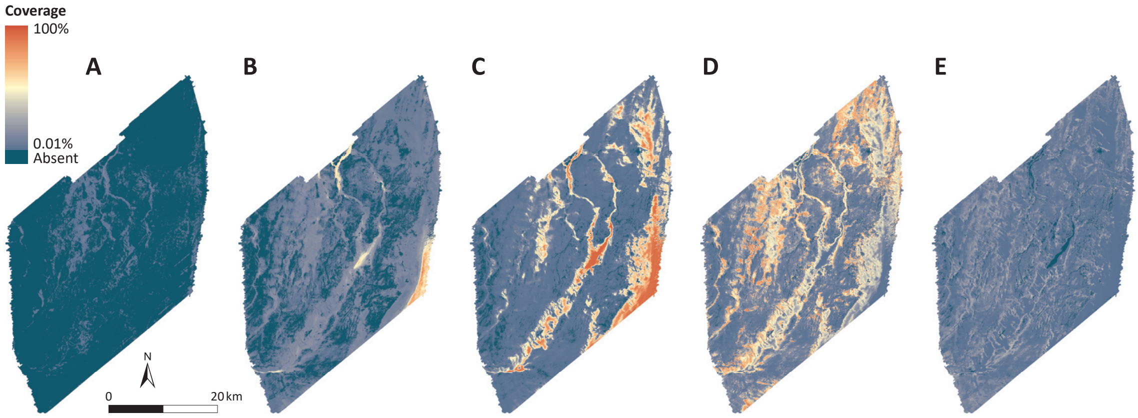


Figure 30. Coverage maps (%) of fine substrate components on Hoburgs bank modelled from sieved grab samples (n = 117). **A.** Soft clay (≤ 0.002 mm). **B.** Silt (> 0.002–0.06 mm). **C.** Fine sand (> 0.06–0.2 mm). **D.** Medium sand (> 0.2–0.6 mm). **E.** Coarse sand (> 0.6–2 mm). For soft clay and silt the colours extend from 0–10% instead of 0–100%. Absence is based on modelled threshold (< 0.01% cover).

Table 10. Accuracy statistics for fine sediment components, coefficient of determination (r^2), root mean square error (RMSE), mean absolute error (MAE) calculated using bootstrap (n = 25) and training data (n = 117).

Fine substrate component	r^2	RMSE	MAE (%)
Soft clay (≤ 0.002 mm)	0.03	2.02	0.92
Soft clay–silt (≤ 0.002–0.06 mm)	0.27	0.71	0.54
Silt (> 0.002–0.06 mm)	0.32	0.66	0.53
Fine sand (> 0.06–0.2 mm)	0.76	1.07	0.85
Medium sand (> 0.2–0.6 mm)	0.79	0.76	0.59
Coarse sand (> 0.6–2 mm)	0.58	1.41	1.02

Biological components

All biological components were associated with areas where cover of hard bottom (i.e. substrate components hard clay, large stones, boulders and large boulders) was higher (Fig. 28A, E–G). Perennial algae and epibenthic bivalves (i.e. *Mytilus* spp.) were the dominant flora and fauna on the bank (Fig. 31A, B), with epibenthic cnidarians having a similar distribution but with lower cover (Fig. 31C). Annual algae were locally abundant (Fig. 31E) in certain areas although the sporadic distribution could have been a result of strong wind and currents encountered during the survey leading to dislodgement of the algae from the substrate.

Like the substrate components, MAE for HELCOM HUB level 5 biological components were low, ranging from 0.2% (epibenthic moss animals) to 13% (perennial algae) across all

HUB level 5 biological component models. However, the low MAE for epibenthic moss animals was probably a result of extremely low cover of epibenthic moss animals in the area. On average, bias in the models was relatively low, except for colonised substrate, where ME was -6.7% (Table 13). Regression analysis of observed and predicted cover showed varying degrees of fit, depending on the biological component, with r^2 ranging from 0.01 (epibenthic moss animals) to 0.84 (colonised substrate) (Table 11). Like substrate components, models of perennial algae and epibenthic cnidarians tended, to varying degrees, to underestimate towards the lower end of the cover scale, and overestimate towards the upper end (Fig. 32A, C). The epibenthic bivalves model tended to underestimate at the lower end of the scale (Fig. 32B), although the slope of epibenthic bivalves was not significantly different from the 1:1 line (Table 11). The annual algae model did not behave in the same manner, tending to show an underestimation of annual algae cover (Fig. 32E). But this was probably because 149 of 154 of the predicted cover values were less than 2% making cover difficult to assess. In addition, the epibenthic moss animals model was difficult to assess, since all predicted values were less than 0.9% and observed cover was generally low (< 6%, Table 12) and Fig. 32D). The colonised substrate model, on the other hand, indicated a tendency to overestimate the cover of colonised substrate (Fig. 32 F, Table 12), exemplified by negative ME values across all classes (Table 13) as well as a regression line below the 1:1 line (Fig. 32 F). However, the regression analysis suggested no significant difference between the y-intercept and slope (Table 11). When predicted values were reclassified into the classes < 10%, ≥ 10 < 50%, ≥ 50 < 90% and ≥ 90 ≤ 100%, overall accuracy (OA) were generally high, ranging from 68.2% (perennial algae) to 100% (epibenthic moss animals) (Table 12). However, high values for epibenthic moss animals and annual algae were not surprising, since few to no observed values were > 10% (Fig. 32D, E). OAs when predicted values were reclassified to absence presence were high, ranging from 78.6% (Cnidarians) to 96.8% (Table 12).

Table 11. y-intercept, slope, and coefficient of determination (r^2) of HELCOM HUB level 5 biological component regression models fitted to observed GT values against corresponding predicted values (n = 154). All regression models were significant (P < 0.001) except for epibenthic moss animals (P = 0.31). Values of the y-intercept and slope in bold indicate significant differences from 0 and 1, respectively (P < 0.05).

Biological component	y-intercept	Slope	r^2
Perennial algae	**6.11**	0.80	0.52
Epibenthic bivalves	**3.42**	0.94	0.62
Epibenthic cnidarians	**3.73**	0.55	0.12
Epibenthic moss animals	**0.17**	0.74	0.01
Annual algae	0.86	**3.26**	0.37
Uncolonised substrate	-4.71	0.97	0.84

Table 12. HELCOM HUB level 5 biological component accuracy statistics, root mean square error (RMSE), mean absolute error (MAE) and mean error (ME), calculated from observed values from GT testing data against corresponding predicted values from models. Overall accuracy (OA) results from confusion matrices where predicted values were grouped into interval classes (OA classes) (< 10%, ≥ 10 < 50%, ≥ 50 < 90%, ≥ 90 ≤ 100%) and absence (< 0.01%) presence (≥ 0.01%) classes (OA abs–pres).

Biological component	RMSE	MAE (%)	ME (%)	OA classes (%)	OA abs–pres (%)
Perennial algae	**20.4**	13.0	1.8	68.2	90.3
Epibenthic bivalves	**15.5**	9.1	2.3	75.3	87.7
Epibenthic cnidarians	**9.7**	4.7	2.4	77.3	78.6
Epibenthic moss animals	**0.8**	0.2	0.2	100.0	87.7
Annual algae	7.1	**1.5**	1.5	94.8	90.9
Uncolonised substrate	17.9	8.0	-6.7	80.5	96.8

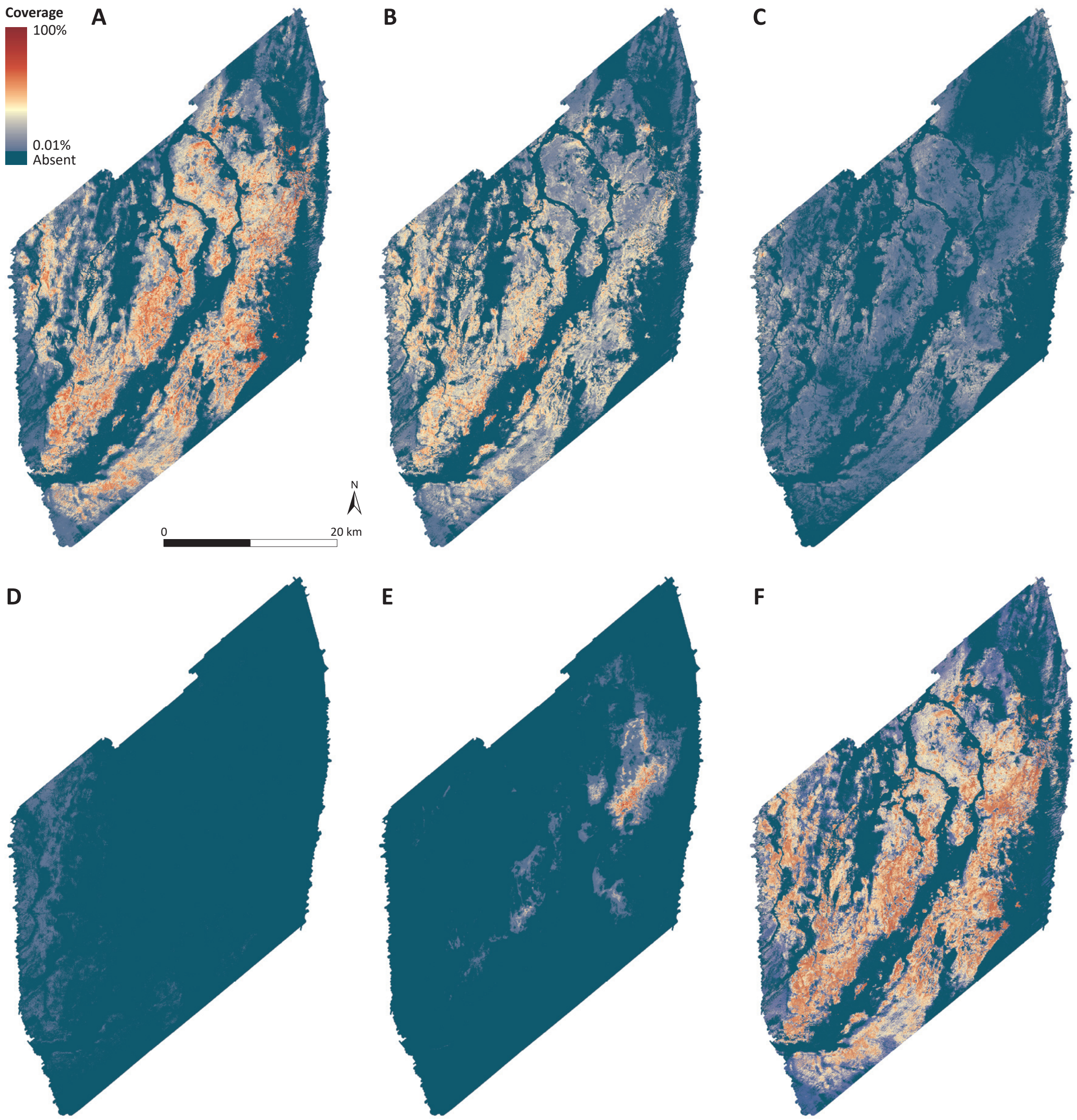


Figure 31. Coverage maps (%) of HELCOM HUB level 5 biological components. **A.** Perennial algae. **B.** Epibenthic bivalves. **C.** Epibenthic cnidarians. **D.** Epibenthic moss animals. **E.** Annual algae. **F.** Colonised substrate. Absence is based on modelled threshold (< 0.01% cover).

Table 13. Mean error (ME) ± standard error (SE) for biological components, where predicted coverage (%) values were divided into interval classes < 10%, ≥ 10 < 50%, ≥ 50 < 90%, and ≥ 90 ≤ 100%. The number of replicates in each class is denoted by n.

Interval class →	**< 10%**			**≥ 10 < 50%**			**≥ 50 < 90%**			**≥ 90 ≤ 100%**		
↓ Biological component	n	ME	SE	n	ME	SE	n	ME	SE	n	ME	SE
Perennial algae	84	4.3	1.2	39	5.0	4.7	31	-8.8	4.1	0	-	-
Epibenthic bivalves	80	2.6	0.9	56	2.5	2.8	18	0.6	4.8	0	-	-
Epibenthic cnidarians	139	3.4	0.7	15	-7.1	3.2	0	-	-	0	-	-
Epibenthic moss animals	154	0.2	0.1	0	-	-	0	-	-	0	-	-
Annual algae	153	1.4	0.6	1	15.7	-	0	-	-	0	-	-
Colonised substrate	27	-0.8	0.4	22	-9.8	3.0	23	-13.7	5.8	82	-5.9	1.7

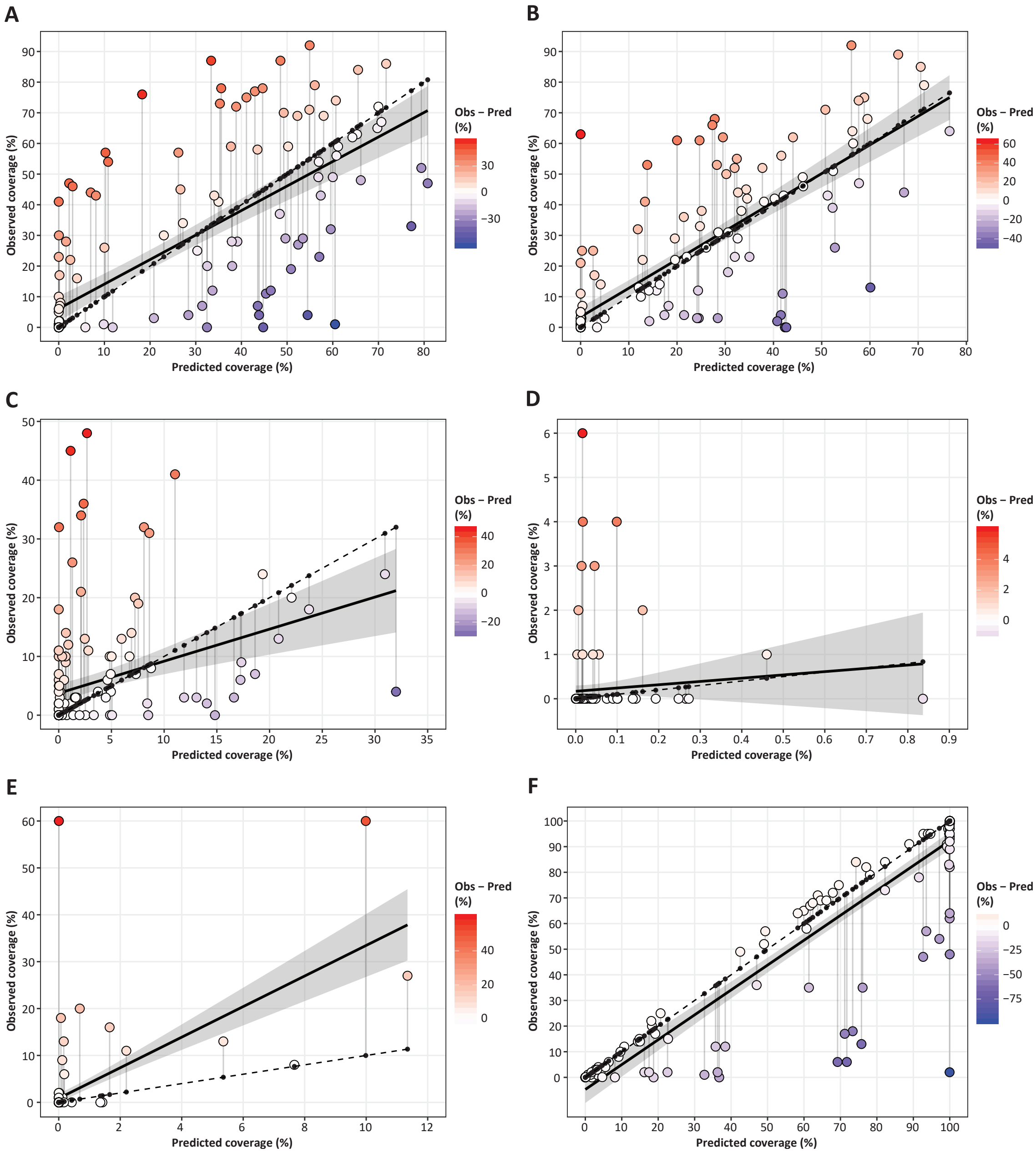


Figure 32. Relationship between observed coverage values (%) from GT testing data (n = 154) and corresponding predicted coverage values (%) from HELCOM HUB level 5 biological component models. **A.** Perennial algae. **B.** Epibenthic bivalves. **C.** Epibenthic cnidarians. **D.** Epibenthic moss animals. **E.** Annual algae. **F.** Colonised substrate. Solid black lines represent linear regressions and 95% confidence bands (grey shade) between observed and predicted coverage. Broken lines represent theoretical 1:1 lines where observed and predicted values are equal (i.e. y = x, y-intercept = 0, slope = 1); black points along dashed lines represent the specific predicted coverage values. Point colours represent the magnitude of difference between an observed value and its corresponding predicted value on the 1:1 line (black point).

Uncertainty of continuous models

Both substrate and biological components generally showed higher uncertainty in the mixed and often patchy areas (Fig. 33 versus Fig. 28 and Fig. 31) whereas more uniform sandy areas (Fig. 29B) where fewer organisms tended to occur (Fig. 31F) were associated with lower uncertainty (dark green areas, Fig. 33A and B).

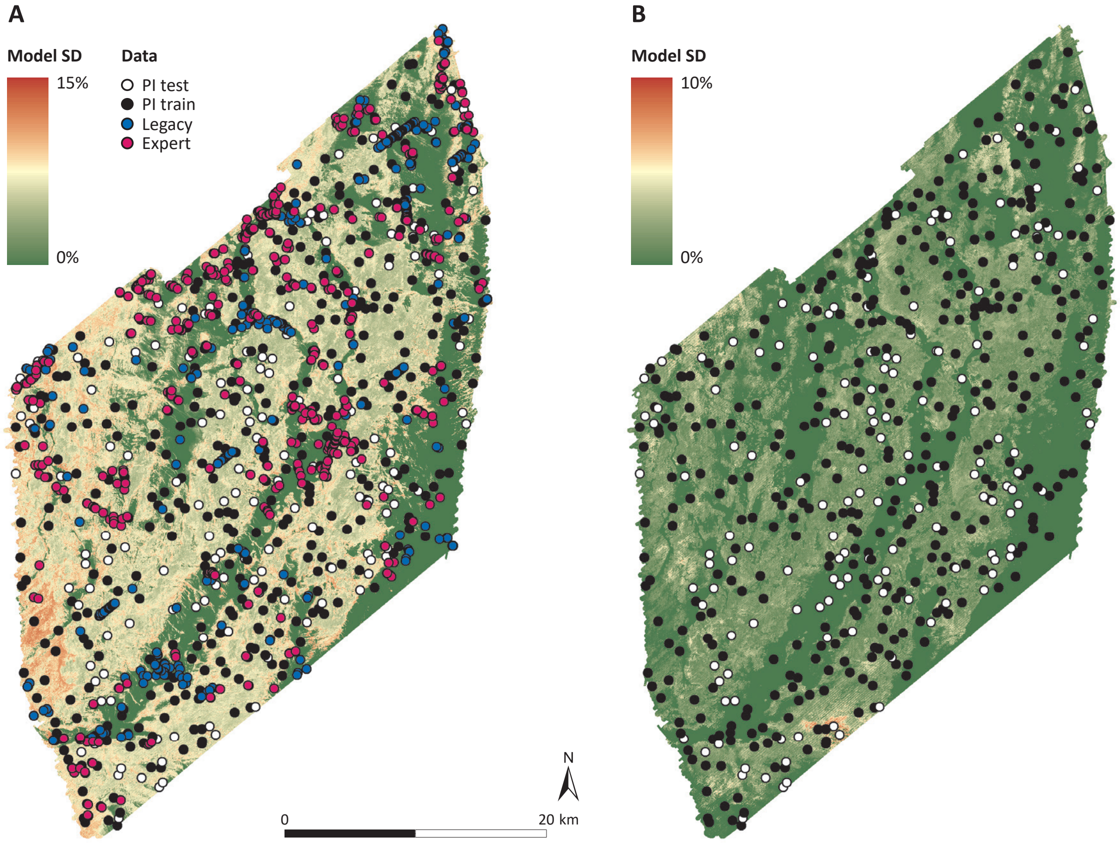


Figure 33. Locations of testing (PI stands for point intercept data from the 2016–2017 survey) and training data used for the modelling, as well as mean standard deviation for all continuous models. **A.** Substrate components. **B.** Biological components. The maps are only intended to capture spatial patterns and cannot be used to compare values between A and B.

HELCOM HUB level 3: Substrate

Five substrate classes were defined in the mapped area which consisted predominantly of mixed substrate (626 km^2) and sand (490 km^2), with coarse sediment (218 km^2), rock and boulders (10 km^2) and hard clay (0.8 km^2) making up the remainder (Fig. 34 and Fig. 35).

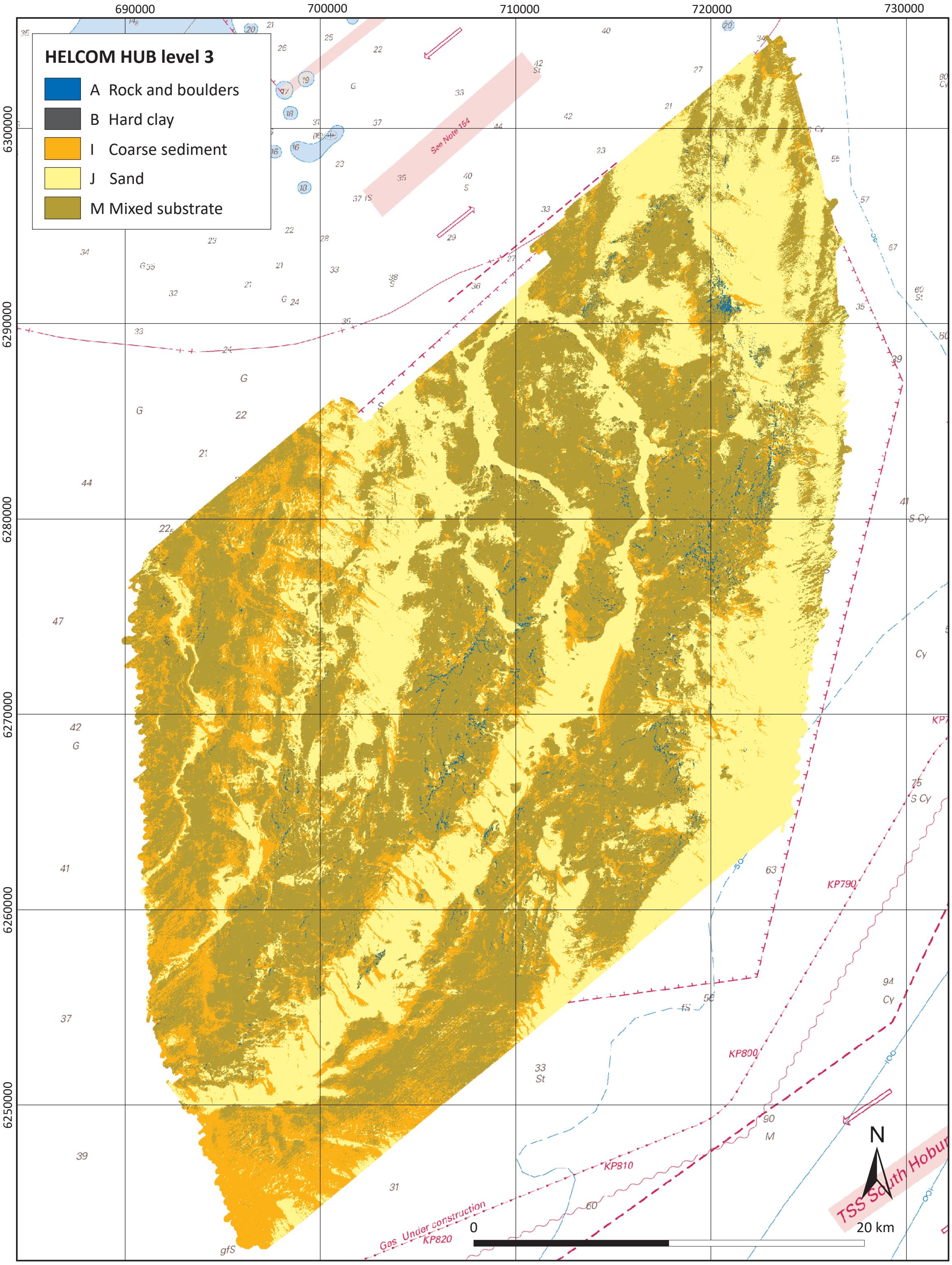


Figure 34. Map of Hoburgs bank classified according to HELCOM HUB level 3: Substrate.

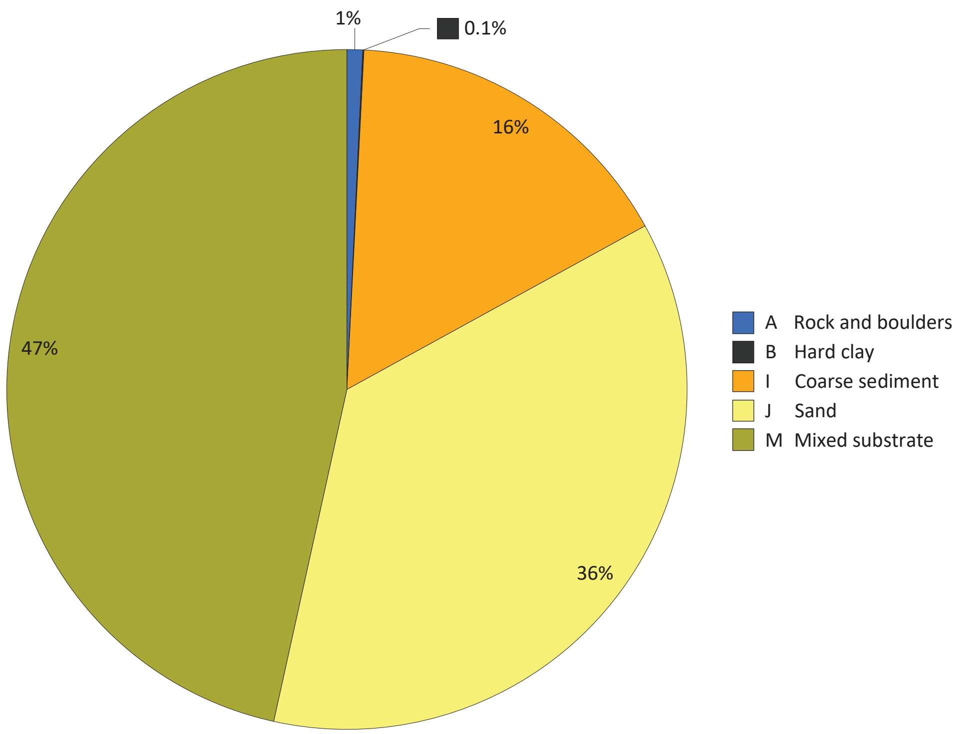


Figure 35. Class proportions (%) for HELCOM HUB level 3: Substrate on Hoburgs bank.

Table 14. Confusion matrix for HELCOM HUB level 3: Substrate. OA denotes overall accuracy, UA user accuracy, and PA producer accuracy. HELCOM HUB codes are in parentheses.

	Rock & boulders (A)	Hard clay (B)	Coarse sediment (I)	Sand (J)	Mixed substrate (M)	n_{row}	UA (%)
Rock & boulders (A)	**15**	0	0	0	1	16	**93.8**
Hard clay (B)	0	**1**	0	0	1	2	**50**
Coarse sediment (I)	0	0	**9**	0	7	16	**56.3**
Sand (J)	0	0	3	**53**	7	63	**84.1**
Mixed substrate (M)	4	0	6	1	**46**	57	**80.7**
n_{col}	19	1	18	54	62	**n_{tot} = 154, OA = 80.5%**	
PA (%)	**78.9**	**100**	**50**	**98.1**	**74.2**		

Confusion matrices for substrate are presented in Table 14. OA at the substrate level was 80.5%. Producer accuracy (PA) and user accuracy (UA) ranged from 50–100% across all substrate classes. In general, classification of rock and boulders, sand and mixed substrate was relatively high (≥ 74.2%) whereas classification of coarse sediment was lower. Hard clay consisted of only one GT data point, so PA and UA values should be interpreted with care.

HELCOM HUB level 4: Community Structure

The dominant community structure in the mapped area was Macroscopic epibenthic biotic structure (EBS) which covered 628 km^2 (47%) of the mapped area, followed by no macroscopic EBS covering 525 km^2 (39%) and sparse macroscopic EBS covering 192 km^2 (14%).

Confusion matrices for community structure are presented in Table 15. OA at the community structure level was 81.9%. PAs and UAs ranged from 31.8–100% across all community structure classes. In general, classification accuracies of macroscopic epibenthic biotic structures (i.e. coverage ≥ 10%) and no macroscopic epibenthic biotic structures were relatively high (> 70.6%), whereas sparse macroscopic epibenthic structures (i.e. coverage > 0.1% and < 10%) had lower accuracy values (PA and UA of 31.8% and 50% respectively).

Table 15. Confusion matrix for HELCOM HUB level 4: Community Structure. OA denotes overall accuracy, UA user accuracy, and PA producer accuracy. HELCOM HUB codes are in parentheses. EBS denotes epibenthic biotic structures.

	Macroscopic EBS (1)	Sparse Macroscopic EBS (2)	No Macroscopic EBS (4)	n_{row}	UA (%)
Macroscopic EBS (1)	**71**	1	0	72	**98.6**
Sparse Macroscopic EBS (2)	7	**7**	0	14	**50.0**
No Macroscopic EBS (4)	6	14	**48**	68	**70.6**
n_{col}	84	22	48	**n_{tot} = 154, OA = 81.9%**	
PA (%)	**84.5**	**31.8**	**100**		

HELCOM HUB levels 4 and 5: Community Structure and Characteristic Community

Cover statistic for HELCOM HUB level 4–5 classes are described in next section (HELCOM HUB level 4-6), and in Figures 36, 37.

Confusion matrices for community structure and characteristic community are presented in Table 16. OA at the characteristic community level was 62.3%. PAs and UAs ranged from 0–100% across all classes. In general, the classes characterised by no macrocommunity and by perennial algae had the highest PAs and UAs, ranging from 60.8–100%. The classes characterised by epibenthic bivalves and by sparse epibenthic macrocommunity were lower, with PAs and UAs ranging from 31.8–50% whereas those characterised by epibenthic cnidarians and by mixed epibenthic macrocommunity had PA and UA values of 0%, meaning classification was missed. Classes characterised by epibenthic bivalves were commonly misclassified as characterised by perennial algae, and vice versa. Moreover, the class characterised by sparse epibenthic macrocommunity was commonly misclassified as characterised by no macrocommunity.

Table 16. Confusion matrix for HELCOM HUB levels 4 and 5: Community Structure and Characteristic Community. OA denotes overall accuracy, UA user accuracy, and PA producer accuracy. Due to space constraints "characterised by" has been omitted from each class, e.g. characterised by perennial algae. HELCOM HUB codes are in parentheses. EM and M denote epibenthic macrocommunity and macrocommunity respectively.

	Perennial algae (1C)	Epibenthic bivalves (1E)	Epibenthic cnidarians (1G)	Mixed EM (1V)	Sparse EM (2T)	No M (4U)	n_{row}	UA (%)
Perennial algae (1C)	**31**	15	4	1	0	0	51	**60.8**
Epibenthic bivalves (1E)	9	**10**	1	0	0	0	20	**50.0**
Epibenthic cnidarians (1G)	0	0	**0**	0	1	0	1	**0.0**
Mixed EM (1V)	0	0	0	**0**	0	0	0	**0.0**
Sparse EM (2T)	4	0	1	2	**7**	0	14	**50.0**
No M (4U)	3	2	1	0	14	**48**	68	**70.6**
n_{col}	47	27	7	3	22	48	**n_{tot} = 154, OA = 62.3%**	
PA (%)	**66.0**	**37.0**	**0.0**	**0.0**	**31.8**	**100**		

HELCOM HUB levels 4–6: Community Structure, Characteristic Community and Dominating Taxa

In this project HELCOM HUB level 6 was classified differently to that which was intended by the scheme. An area was classified to level 6 if a biological component covered at least 50% of the seabed and more than other components. However, the intended classification at level 6 is based on a biomass or biovolume comparison within a level 5 group. For example, two areas where the epibenthic bivalves component covered 10% and 50% respectively and greater than all other biological components were classified as characterised by epibenthic bivalves (i.e. to level 5) and dominated by Mytilidae (i.e. to level 6) respectively in this project. On Hoburgs bank, however, the epibenthic bivalves component consists entirely of *Mytilus* spp., so Mytilidae consists of at least 50% of the biomass of epibenthic bivalves so, according to the scheme, both areas should have been classified to level 6 as dominated by Mytilidae. See Table 17 for how those areas classified to level 5 in the project should have been classified to level 6 according to the intended application of the HELCOM HUB scheme. It should be mentioned that HELCOM HUB level 4–6 and level 1–6 products in this report are based on this project's interpretation of HELCOM HUB and not that which was intended by the scheme.

A total of 12 HELCOM HUB level 4–6 classes were defined within the mapped area in varying proportions (Fig. 36). The proportion of the map area in which the coverage of benthic organisms was less than 10% was 53%, with the class characterised by no macrocommunity (525 km^2) predominating, and the class characterised by sparse epibenthic macrocommunity (192 km^2) accounting for the remainder (Fig. 37). The remaining 47% of the mapped area in which the coverage of benthic organisms was at least 10% was dominated by the class characterised by perennial algae (302 km^2), followed by that characterised by epibenthic bivalves (159 km^2) and dominated by perennial filamentous algae (104 km^2). The classes dominated by Mytilidae, characterised by cnidarians, and mixed epibenthic community were the other notable classes, which covered 12, 19 and 32 km^2 of the map area respectively (Fig. 36 and Fig. 37).

Confusion matrices for community structure, characteristic community and dominating taxa are presented in Table 18. OA was 53.2%.

Table 17 HUB 4–6. Areas classified to HELCOM HUB level 5 in maps according to alternative classification used in project vs. classification intended by HELCOM HUB.

Classified to level 5 in map	Classification to level 6 according to HELCOM HUB
Characterised by perennial algae (1C)	Dominated by perennial filamentous algae (1C5)
Characterised by epibenthic bivalves (1E)	Dominated by Mytilidae (1E1)
Characterised by epibenthic cnidarians (1G)	Dominated by Hydroids (1G1)
Characterised by moss animals (1H)	**Dominated by crustose moss animals (1H1)**
Characterised by annual algae (1S)	Dominated by filamentous annual algae (1S1)
Mixed EM (1V)	N/A
Sparse EM (2T)	N/A
No M (4U)	N/A
No M (4U)	N/A

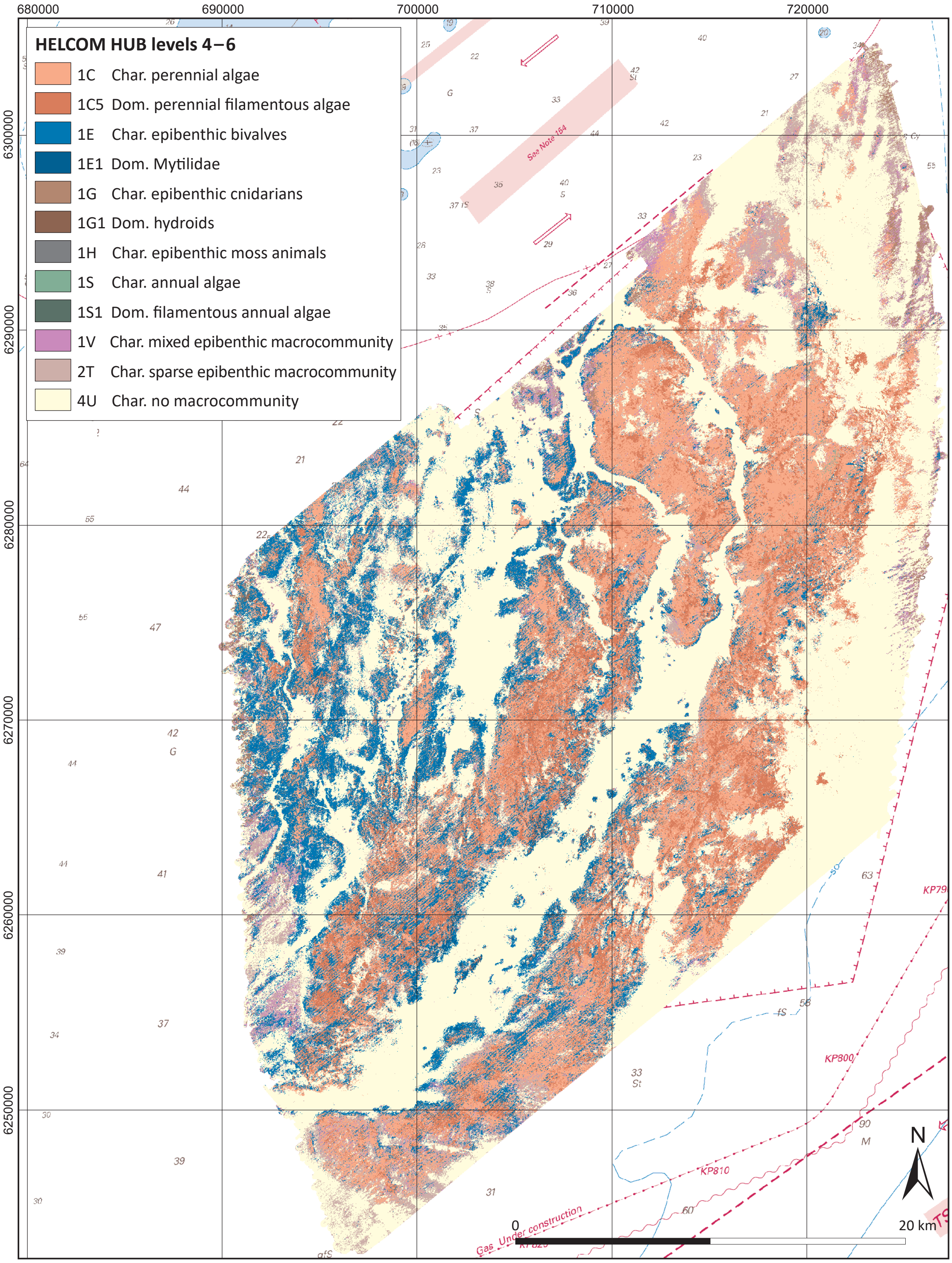


Figure 36. Map of Hoburgs bank classified according to HELCOM HUB levels 4–6: Community Structure, Characteristic Community and Dominating Taxa. Char. and Dom. denote Characterised by and Dominated by respectively.

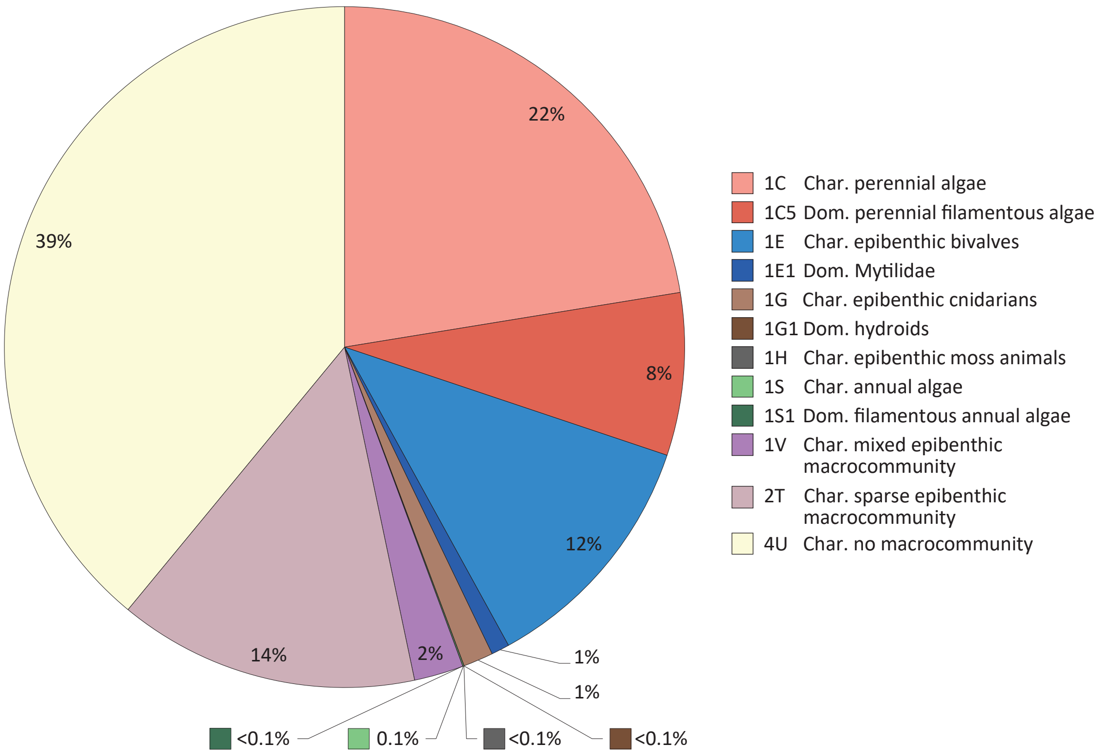


Figure 37. Class proportions (%) of HELCOM HUB levels 4–6: Community Structure, Characteristic Community and Dominating Taxa on Hoburgs bank Char. and Dom. denote Characterised by and Dominated by respectively.

Table 18 HUB level 4–6. Confusion matrix for HELCOM HUB levels 4–6: Community Structure, Characteristic Community and Dominating Taxa. OA denotes overall accuracy, UA user accuracy, and PA producer accuracy. Due to space constraints "dominated by" and "characterised by" have been omitted from each class, e.g. characterised by perennial algae, dominated by Mytilidae. HELCOM HUB codes are in parentheses. EM and M denote epibenthic macrocommunity and macrocommunity respectively.

	Perennial algae (1C)	Perennial filamentous algae (1C5)	Epibenthic bivalves (1E)	Mytilidae (1E1)	Epibenthic cnidarians (1G)	Mixed EM (1V)	Sparse EM (2T)	No M (4U)	n_{row}	UA (%)
Perennial algae (1C)	**5**	8	3	3	3	1	0	0	23	**21.7**
Perennial filamentous algae (1C5)	4	**14**	3	6	1	0	0	0	28	**50**
Epibenthic bivalves (1E)	2	1	**3**	2	1	0	0	0	9	**33.3**
Mytilidae (1E1)	2	4	0	**5**	0	0	0	0	11	**45.5**
Epibenthic cnidarians (1G)	0	0	0	0	**0**	0	1	0	1	**0**
Mixed EM (1V)	0	0	0	0	0	**0**	0	0	0	**0**
Sparse EM (2T)	4	0	0	0	1	2	**7**	0	14	**50**
No M (4U)	3	0	1	1	1	0	14	**48**	68	**70.6**
n_{col}	20	27	10	17	7	3	22	48	**ntot = 154 OA = 53.2%**	
PA (%)	**25**	**51.9**	**30**	**29.4**	**0**	**0**	**31.8**	**100**		

HELCOM HUB levels 1–6

A total of 59 HELCOM HUB classes were defined in the mapped area making for a very complex map (Fig. 38). To provide a simplified overview of this map HELCOM HUB levels 4–6 classes were separated by HELCOM HUB level 3 substrate classes, to provide information on the different biotope assemblages associated with each substrate class (Fig. 39). Only the classes in the photic zone (HELCOM HUB level 2: Code A) were used, since they covered 99% of the mapped area.

Proportions of different biotopes varied, depending on the substrate class (Fig. 39). Photic rock and boulders (10 km^2, 1%) consisted mainly of the class dominated by perennial filamentous algae, followed by those dominated by Mytilidae, with those characterised by epibenthic bivalves and by perennial algae having similar proportions (Fig. 39A). Photic hard clay (0.8 km^2, 0.1%) was largely associated with the classes characterised by perennial algae and by epibenthic bivalves, as well as those dominated by perennial filamentous algae (Fig. 39B). The finer substrate classes photic coarse sediment (218 km^2, 16%; Fig. 39 C) and photic sand (477 km^2, 35%, Fig. 39D) were dominated by classes where the coverage of benthic organisms was less than 10%, i.e. those characterised by sparse epibenthic macrocommunity and by no macrocommunity. Finally, photic mixed substrate (626 km^2, 47%; Fig. 39E) had a similar class composition to photic rock and boulders (Fig. 39A), but with a smaller proportion of the class dominated by Mytilidae and larger proportions of the classes characterised by cnidarians, by mixed epibenthic macrocommunity and by sparse epibenthic macrocommunity (Fig. 39E). The varying amount of coverage of each substrate class would suggest that the photic mixed substrate class was an important habitat. However, despite its small area, rock and boulders appeared to be important for perennial algae and epibenthic bivalves.

Legend Figure 38 (next page)

HELCOM HUB levels 1–6

AA.A1C	Photic rock and boulders char. perennial algae
AA.A1C5	Photic rock and boulders dom. perennial filamentous algae
AA.A1E	Photic rock and boulders char. epibenthic bivalves
AA.A1E1	Photic rock and boulders dom. Mytilidae
AA.A1G	Photic rock and boulders char. epibenthic cnidarians
AA.A1G1	Photic rock and boulders dom. hydroids
AA.A1V	Photic rock and boulders char. mixed epibenthic macrocommunity
AA.A2T	Photic rock and boulders char. sparse epibenthic macrocommunity
AA.B1C	Photic hard clay char. perennial algae
AA.B1C5	Photic hard clay dom. perennial filamentous algae
AA.B1E	Photic hard clay char. epibenthic bivalves
AA.B1E1	Photic hard clay dom. Mytilidae
AA.B1G	Photic hard clay char. epibenthic cnidarians
AA.B1G1	Photic hard clay dom. hydroids
AA.B1V	Photic hard clay char. mixed epibenthic macrocommunity
AA.B2T	Photic hard clay char. sparse epibenthic macrocommunity
AA.I1C	Photic coarse sediment char. perennial algae
AA.I1C5	Photic coarse sediment dom. perennial filamentous algae
AA.I1E	Photic coarse sediment char. epibenthic bivalves
AA.I1E1	Photic coarse sediment dom. Mytilidae
AA.I1G	Photic coarse sediment char. epibenthic cnidarians
AA.I1S	Photic coarse sediment char. annual algae
AA.I1S1	Photic coarse sediment dom. filamentous annual algae
AA.I1V	Photic coarse sediment char. mixed epibenthic macrocommunity
AA.I2T	Photic coarse sediment char. sparse epibenthic macrocommunity
AA.I4U	Photic coarse sediment char. no macrocommunity
AA.J1C	Photic sand char. perennial algae
AA.J1E	Photic sand char. epibenthic bivalves
AA.J1E1	Photic sand dom. Mytilidae
AA.J1G	Photic sand char. epibenthic cnidarians
AA.J1S	Photic sand char. annual algae
AA.J1V	Photic sand char. mixed epibenthic macrocommunity
AA.J2T	Photic sand char. sparse epibenthic macrocommunity
AA.J4U	Photic sand char. no macrocommunity
AA.M1C	Photic mixed substrate char. perennial algae
AA.M1C5	Photic mixed substrate dom. perennial filamentous algae
AA.M1E	Photic mixed substrate char. epibenthic bivalves
AA.M1E1	Photic mixed substrate dom. Mytilidae
AA.M1G	Photic mixed substrate char. epibenthic cnidarians
AA.M1G1	Photic mixed substrate dom. hydroids
AA.M1H	Photic mixed substrate char. epibenthic moss animals
AA.M1S	Photic mixed substrate char. annual algae
AA.M1S1	Photic mixed substrate dom. filamentous annual algae
AA.M1V	Photic mixed substrate char. mixed epibenthic macrocommunity
AA.M2T	Photic mixed substrate char. sparse epibenthic macrocommunity
AB.A1E	Aphotic rock and boulders char. epibenthic bivalves
AB.A1G	Aphotic rock and boulders char. epibenthic cnidarians
AB.A1V	Aphotic rock and boulders char. mixed epibenthic macrocommunity
AB.B1V	Aphotic hard clay char. mixed epibenthic macrocommunity
AB.B2T	Aphotic hard clay char. sparse epibenthic macrocommunity
AB.I2T	Aphotic coarse sediment char. sparse epibenthic macrocommunity
AB.I4U	Aphotic coarse sediment char. no macrocommunity
AB.J2T	Aphotic sand char. sparse epibenthic macrocommunity
AB.J4U	Aphotic sand char. no macrocommunity
AB.M1E	Aphotic mixed substrate char. epibenthic bivalves
AB.M1G	Aphotic mixed substrate char. cnidarians
AB.M1V	Aphotic mixed substrate char. mixed epibenthic macrocommunity
AB.M4U	Aphotic mixed substrate char. no macrocommunity

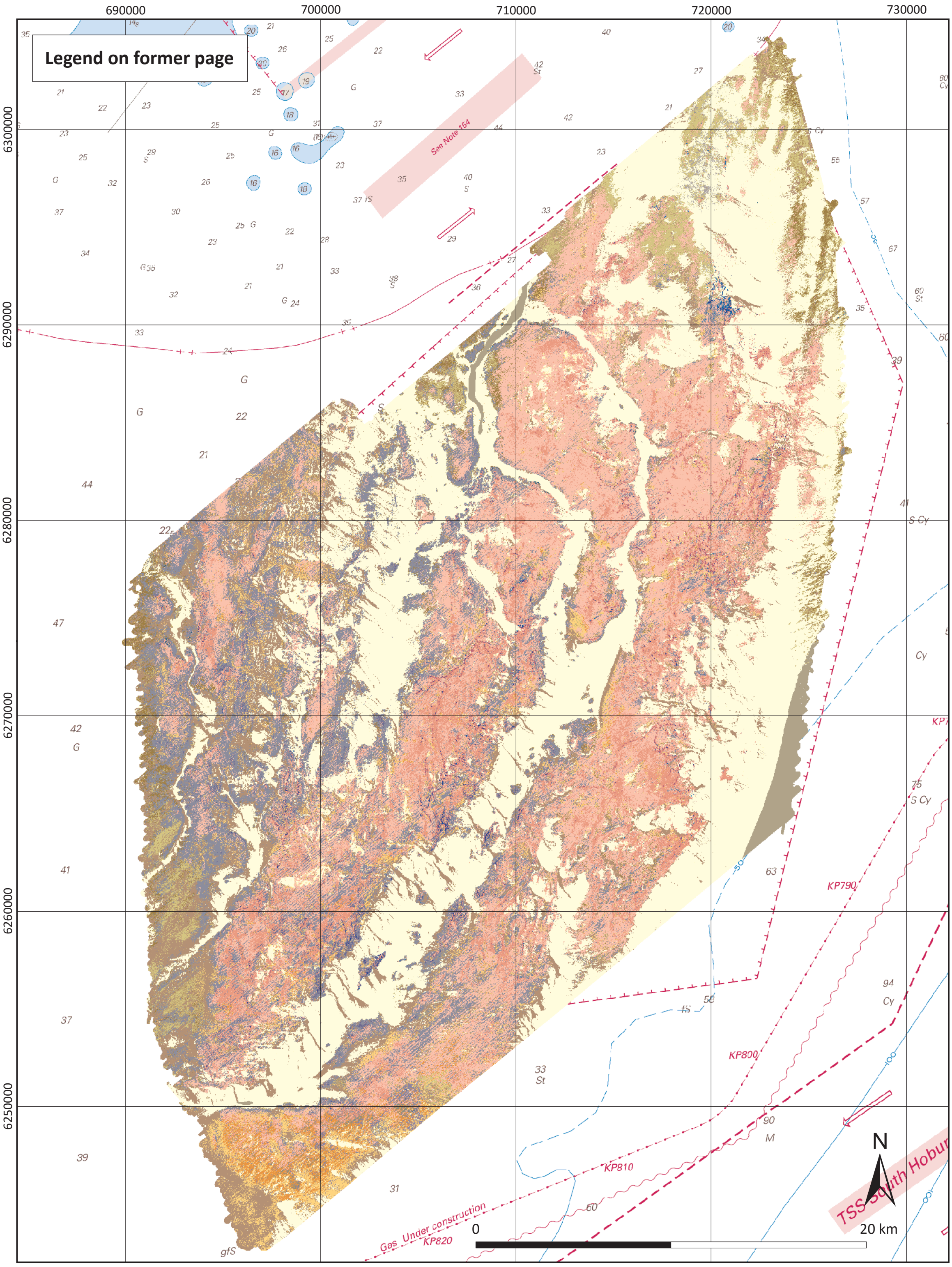


Figure 38. Map of Hoburgs bank classified according to the full HELCOM HUB (levels 1–6). Note: HELCOM HUB level 1: Baltic has been omitted from the legend due to space constraints. Char. and Dom. denote Characterised by and Dominated by respectively.

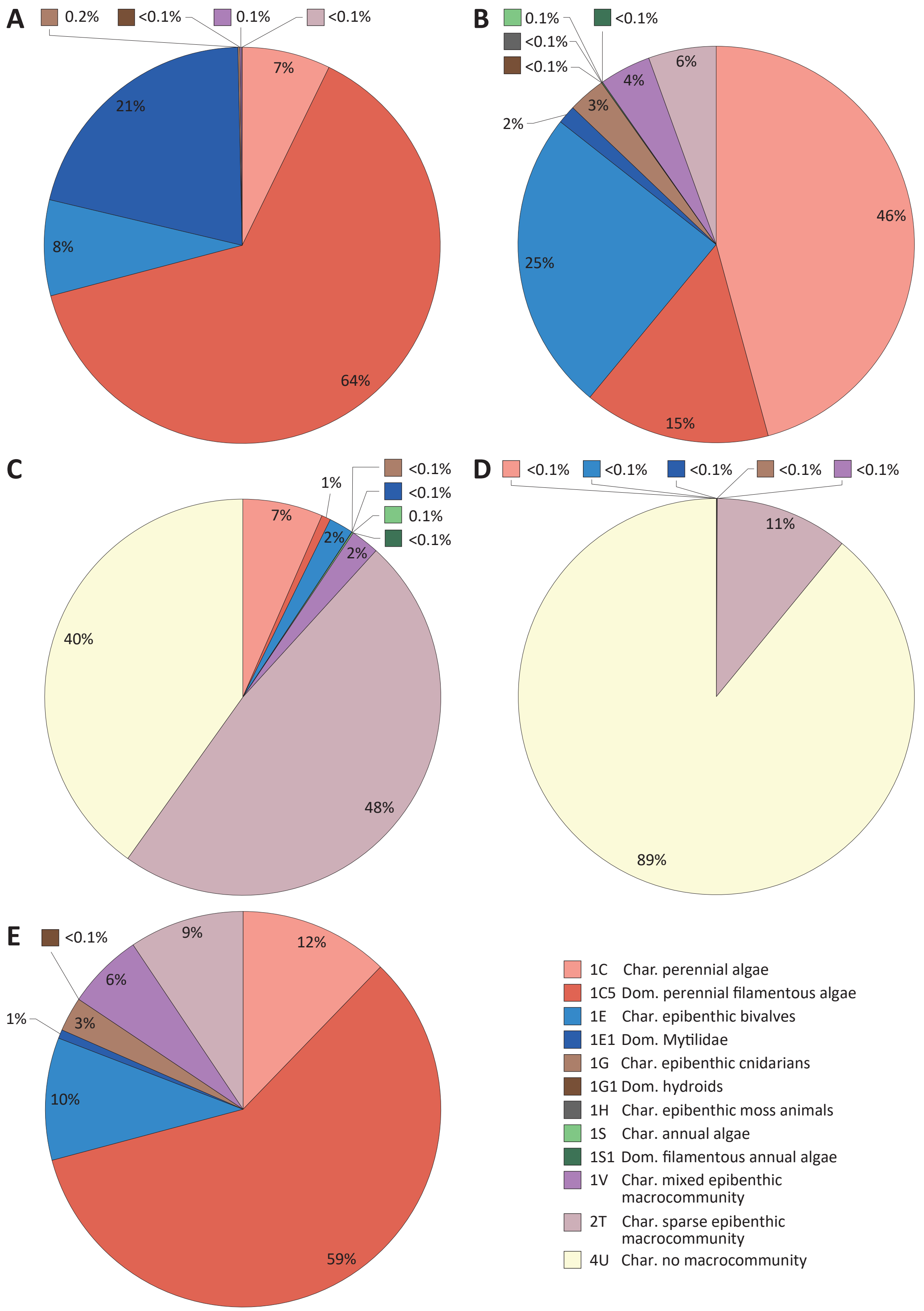


Figure 39. Class proportions (%) of HELCOM HUB levels 4–6: Community Structure, Characteristic Community and Dominating Taxa on different Photic (HELCOM HUB level 2: A) and Substrate (HELCOM HUB level 3) classes. **A.** Photic rock and boulders (A.A), Total Area = 10 km^2 or 1% of mapped area. **B.** Photic hard clay (A.B), Total Area = 0.8 km^2 or 0.1% of mapped area. **C.** Photic coarse sediment (A.I), Total Area = 218 km^2 or 16% of mapped area. **D.** Photic sand (A.J), Total Area = 477 km^2 or 35% of mapped area. and **E.** Photic mixed substrate (A.M), Total Area = 626 km^2 or 47% of mapped area. Char. and Dom. denote Characterised by and Dominated by respectively.

Natura 2000

Natura 2000 map that delineates Reef and sandbank show that Reef (1170) classes covered 38.8% (522 km^2) while sandbank (1110) covered 61.2% (823 km^2) of the mapped area (Fig. 40). Their corresponding confusion matrices are presented in Table 19. OA was 87.7%. All UA and PA values were greater than 85.3%.

Table 19. Confusion matrix for Natura 2000. OA denotes overall accuracy, UA user accuracy, and PA producer accuracy. Natura 2000 codes are in parentheses.

	Reef (1170)	Sandbank (1110)	n_{row}	UA (%)
Reef (1170)	**64**	11	75	**85.3**
Sandbank (1110)	8	**71**	79	**89.9**
n_{col}	72	82	n_{tot} = 154, OA = 87.7%	
PA (%)	**88.9**	**86.6**		

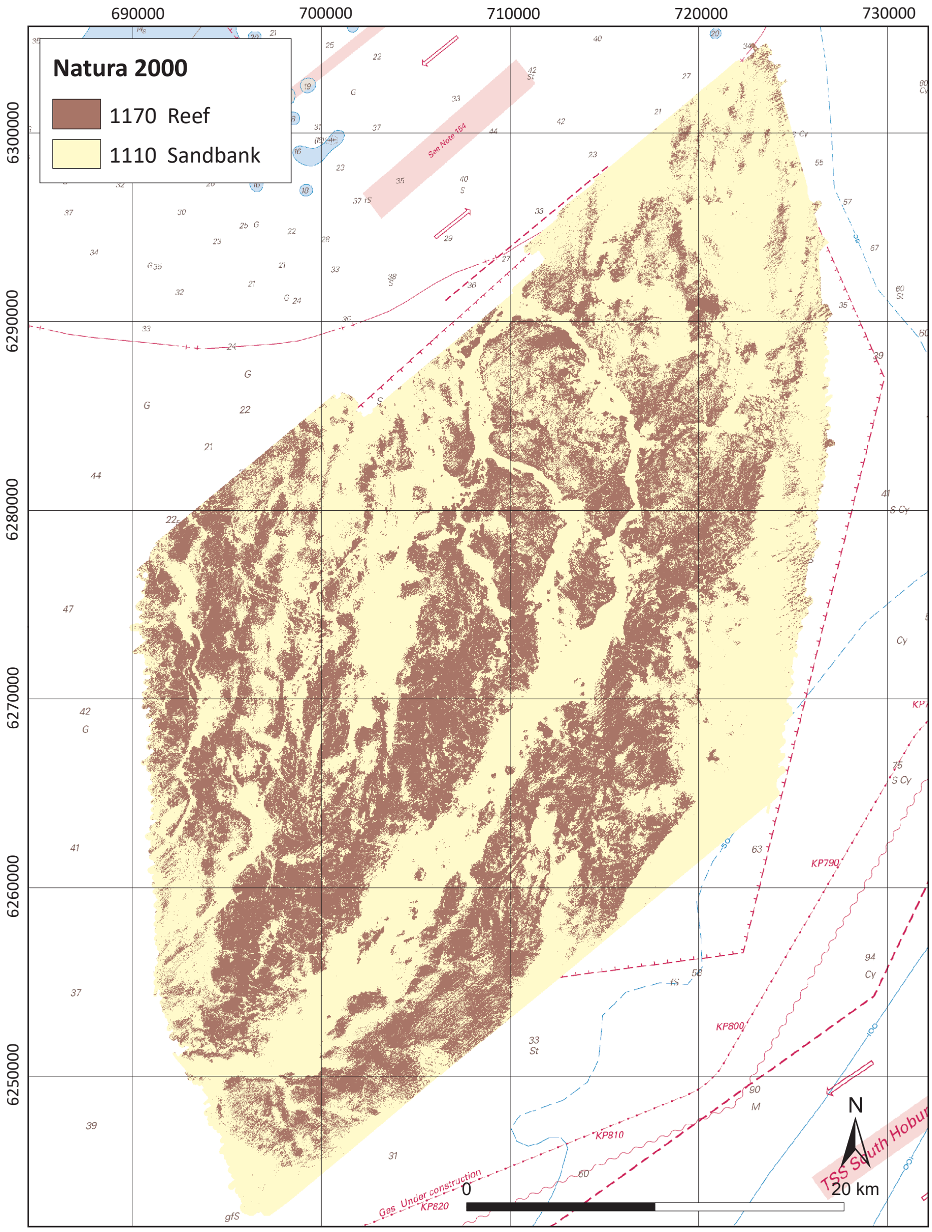


Figure 40. Map of Hoburgs bank classified according to Natura 2000.

Natura 2000 subtypes

The Natura 2000 subtypes map captured 8 reef classes and 4 sandbank classes, revealing features and seascape patterns not seen on the main Natura 2000 map (Fig. 40), including an extensive pattern of ridges (Fig. 41). Rock & boulder reef classes (449 km^2 combined) were

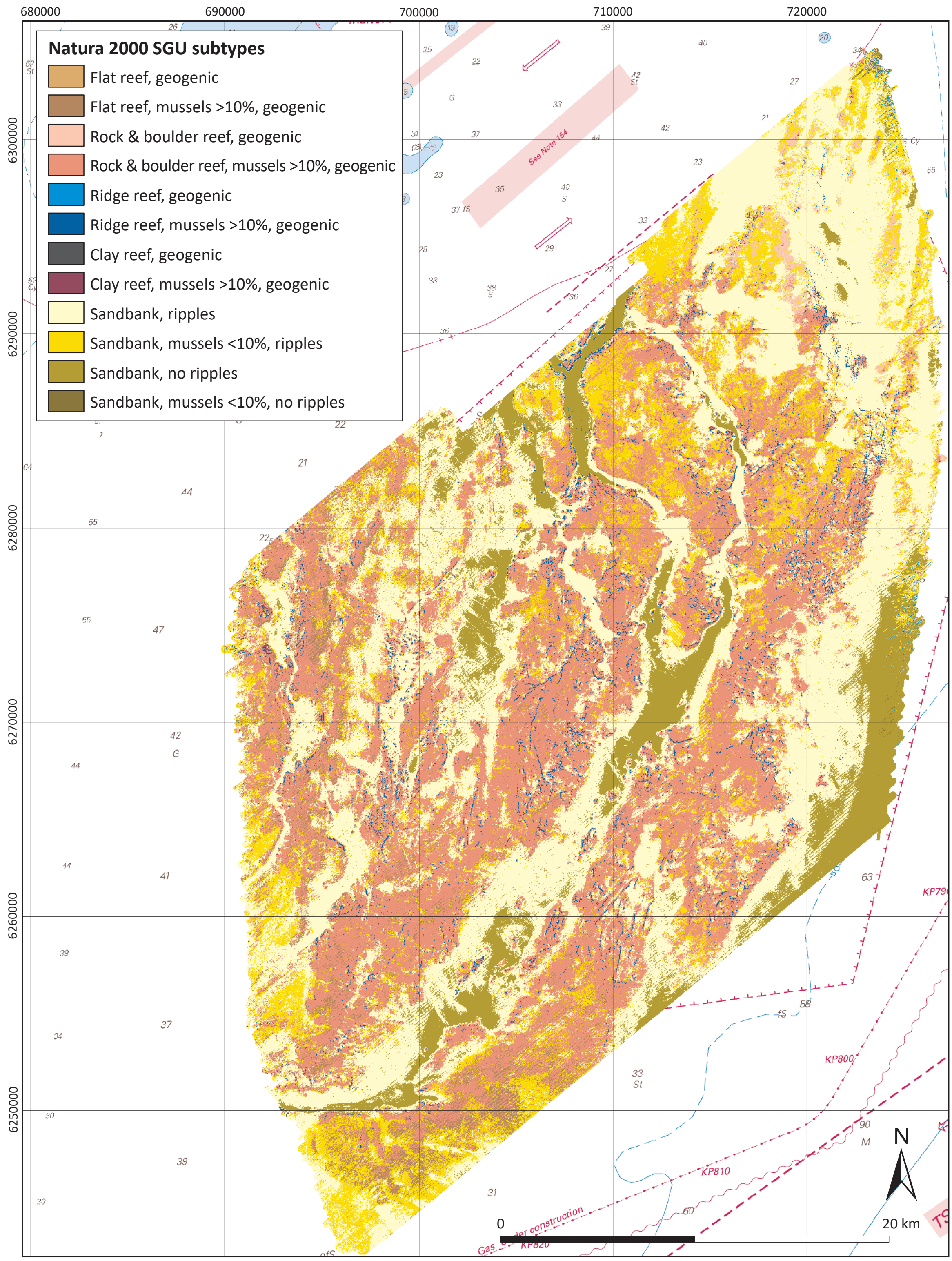


Figure 41. Map of Hoburgs bank classified according to SGU's implementation of Natura 2000.

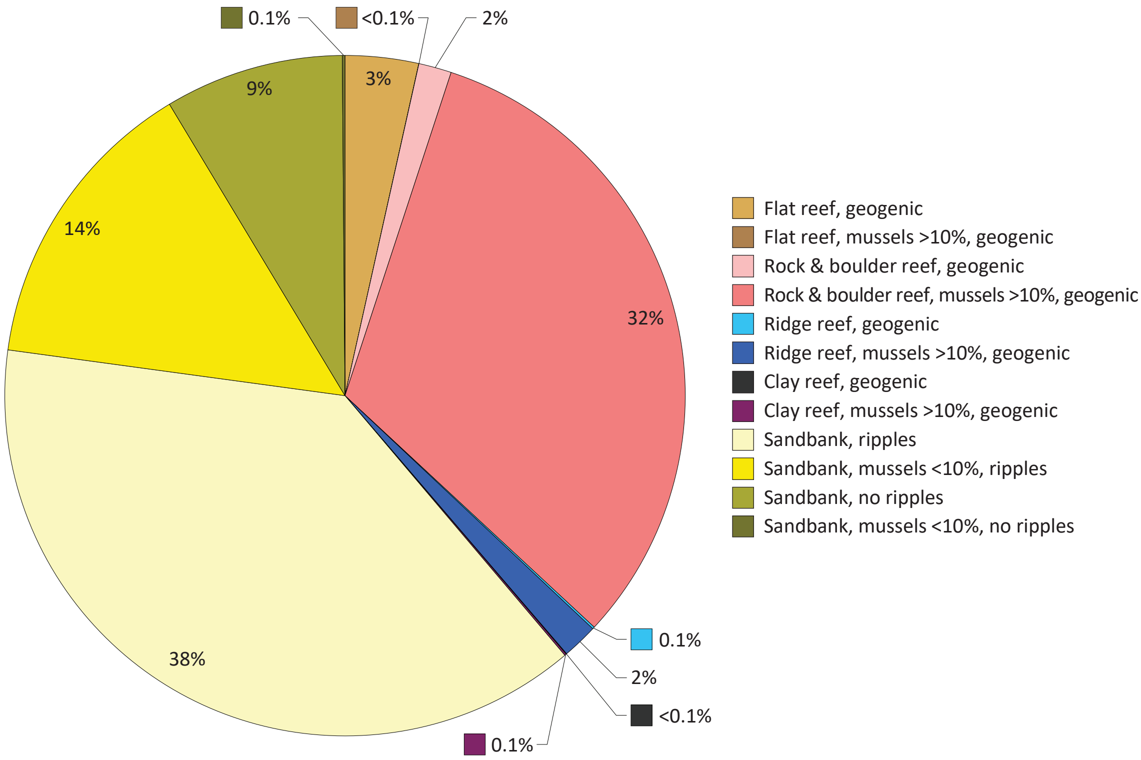


Figure 42. Class proportions (%) of Natura 2000 subtypes map.

the predominate reef type followed by flat reef (47 km^2), ridge reef (24 km^2) and clay reef (1,5 km^2) classes. In most cases each reef classification with > 10% mussel coverage (e.g. rock & boulder reef, mussel > 10%, geogenic) had a greater coverage than those with mussels < 10% except flat reef (e.g. rock & boulder reef, geogenic; Fig. 42). Sandbank, ripples classes (707 km^2 combined) were the dominant sandbank classes with the sandbank, without ripples classes (116 km^2 combined) tending to be restricted to deeper areas of the bank (Fig. 41 and Fig. 42).

Additional map products

This section shows additional map products that were not included in the classification schemes used but were observed during the project. These include postglacial sand depth interpreted from sub-bottom profiles (Fig. 43A), sand ripples (Fig. 43B), crawl tracks (Fig. 43C), and *Macoma balthica* shells (Fig. 43D). Postglacial sand depth was used as a predictor, and sand ripples was used in the classification of SGU's Natura 2000 subtype map (Fig. 41). The Crawl tracks and *Macoma balthica* shells layers were developed to capture biological features observed in different types of sand areas. These four products were not validated via GT.

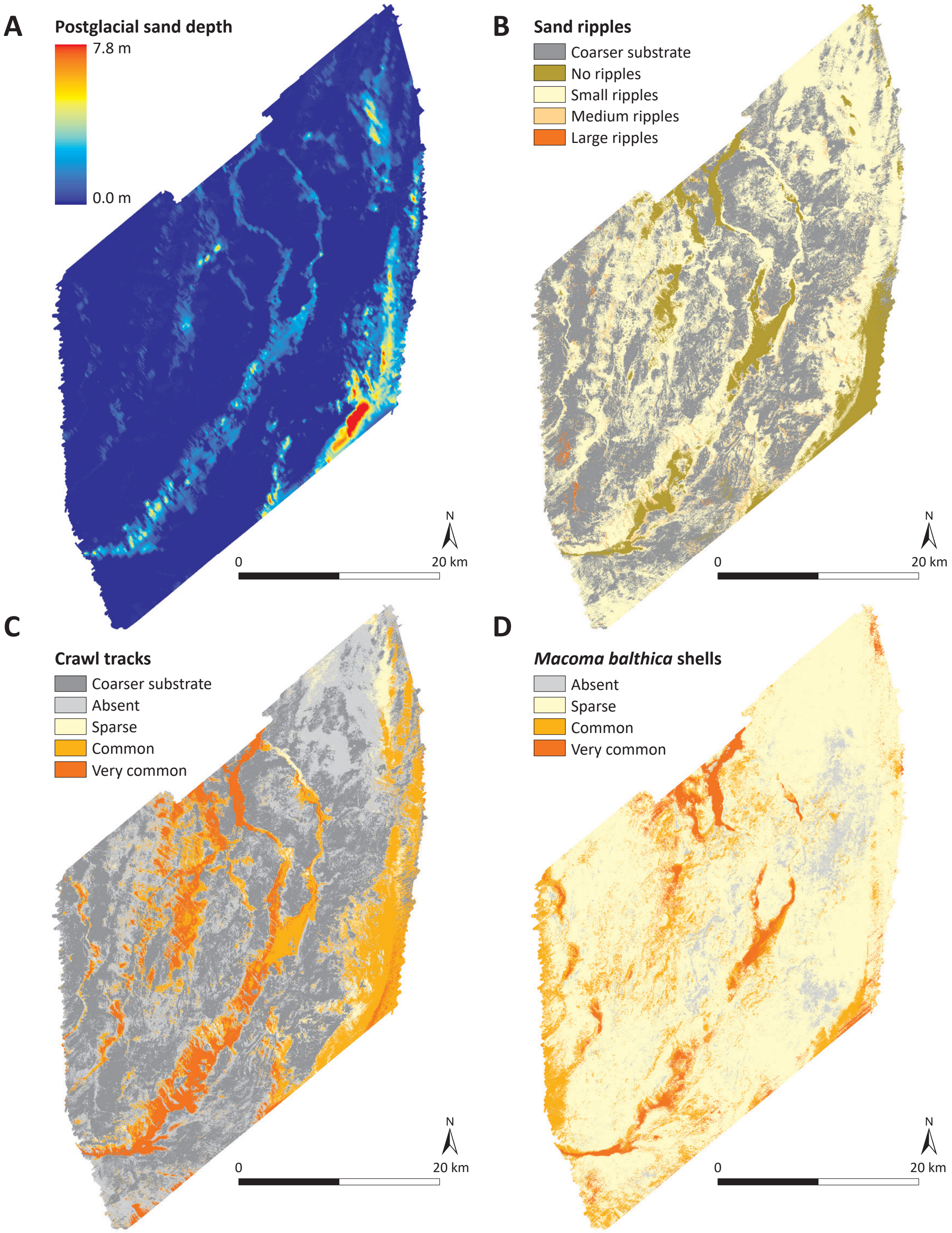


Figure 43. Maps of additional products on Hoburgs bank. **A.** Postglacial sand depth from interpolated interpreted SBP profiles. **B.** Occurrence of sand ripples in different size classes. Coarser substrate = All sediment not classed as sand, No ripples = Sand with no sand ripples, Small ripples = ≥ 1 < 50 cm between sand ripple peaks, Medium ripples = ≥ 50 < 100 cm between sand ripple peaks, Large ripples = > 100 cm between sand ripples peaks. **C.** Prevalence of crawl tracks. **D.** Prevalence of *Macoma balthica* shells.

End-user applications

Accuracy, thematic and spatial resolution are all important factors to consider when using these maps outputs. The section below demonstrates the effect of these factors and is in large part a reproduction of Kågesten et al. 2019.

Impact of scale

The following sections illustrate and highlight the advantages of using high-resolution map products for habitat mapping. The percentage cover of the resulting thematic classes greatly changes depending on the spatial resolution used when reclassifying continuous maps. Figure 44 shows that a HELCOM HUB level 3 classification was applied to a growing-size window, small distinct features either declined in cover or were completely reclassified into more general classes (Fig. 44, Table 20).

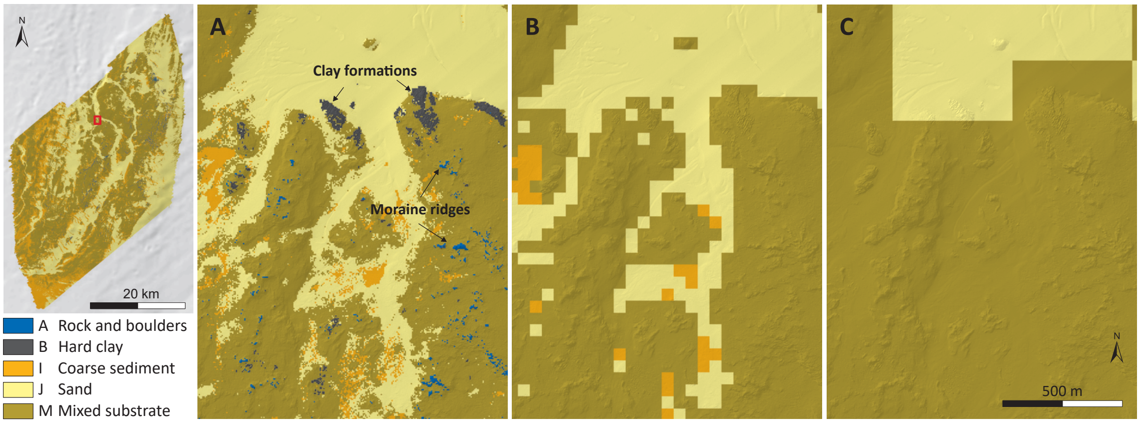


Figure 44. Comparison of HELCOM HUB level 3 substrate maps created using different scales. As the analysis window grows larger features quickly give way to more general classes such as mixed substrate. **A.** 5 m resolution. **B.** 50 m resolution. **C.** 250 m resolution.

Table 20. Changes in cover (%) of HELCOM HUB level 3 substrate classes as the analysis window increases from a resolution of 5 m to 250 m. Significant changes in cover for the rarer classes already occur at a resolution of 10 m.

Substrate class	5 m	10 m	25 m	50m	250 m
Rock & boulders (A)	0.74%, 10 km^2	0.56%, 7.5 km^2	0.32%, 4.3 km^2	0.10%, 1.4 km^2	0.01%, 0.12 km^2
Hard clay (B)	0.06%, 0.81 km^2	0.02%, 0.22 km^2	0.003%, 0.04 km^2	0.00%, 0.0 km^2	0.00%, 0.0 km^2
Coarse sediment (I)	16.2%, 217 km^2	14.3%, 192 km^2	13%, 179 km^2	12,4%, 167 km^2	10,0%, 133 km^2
Sand (J)	36.4%, 469 km^2	35.2%, 472 km^2	35%, 465 km^2	33.9%, 456 km^2	31.3%, 420 km^2
Mixed substrate (M)	46.4%, 623 km^2	50,0%, 672 km^2	51.8%, 696 km^2	53.5%, 719 km^2	58.7%, 789 km^2

In a second example, this time regarding the Natura 2000 map (i.e. Fig. 41), the area occupied by the reef class increased in cover as the map resolution increased (Fig. 45), moving towards 100% cover of reef when using mean cover at the scale of the whole bank. Predicted mean cover of hard bottom (19.3%) was below the threshold for reef (≥ 50%) but epibenthic bivalves in the study area were above the threshold (≥ 10% cover), which then classified it as reef. Conversely, if the mean cover for epibenthic bivalves had been < 10%, the mean cover of the study area would have been classified as sandbank, reversing the pattern of the growing analysis window example in Figure 45.

High-resolution habitat maps provide the possibility of aggregating information to the appropriate management-scale resolution. Applying a management-scale resolution of 250 m to the HELCOM HUB level 3 classification eliminated the rock and boulders and hard clay classes (Table 20). However, if all 250 m pixels containing these classes at the 5 m scale were mapped instead, rock and boulders would cover 32% and hard clay 17.8% (Fig. 46B and Fig. 46C).

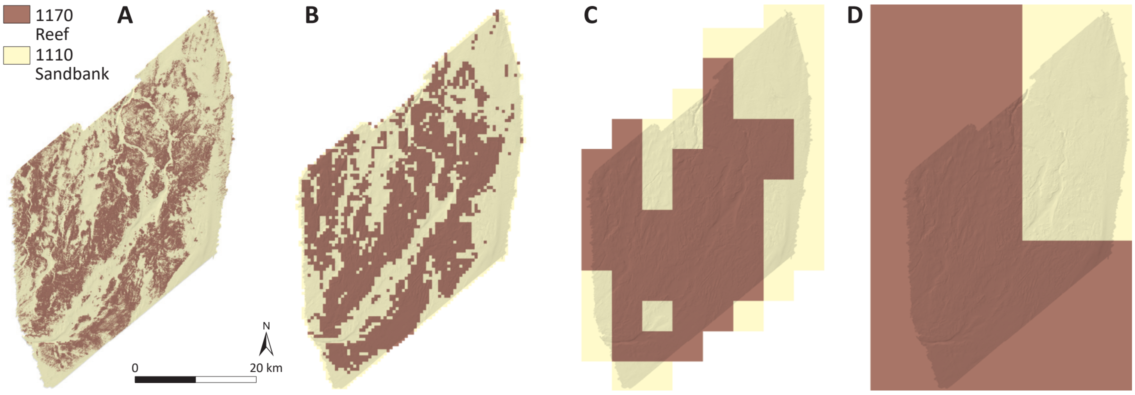


Figure 45. Comparison of Natura 2000 maps created using different scales from 5 m to 25 km. The definition for reef is ≥ 50% hard bottom (the mean hard bottom cover on the bank was 19.3%) or ≥ 10% epibenthic bivalves (the mean bivalve cover on the bank was 10.7%). Scale choice changes the cover of each class. **A.** 5 m resolution (reef 44% coverage). **B.** 500 m resolution (reef 51% coverage). **C.** 5 km resolution (reef 55% coverage). **D.** 25 km resolution (reef 67% coverage). At the scale of the whole bank the reef cover would be 100%.

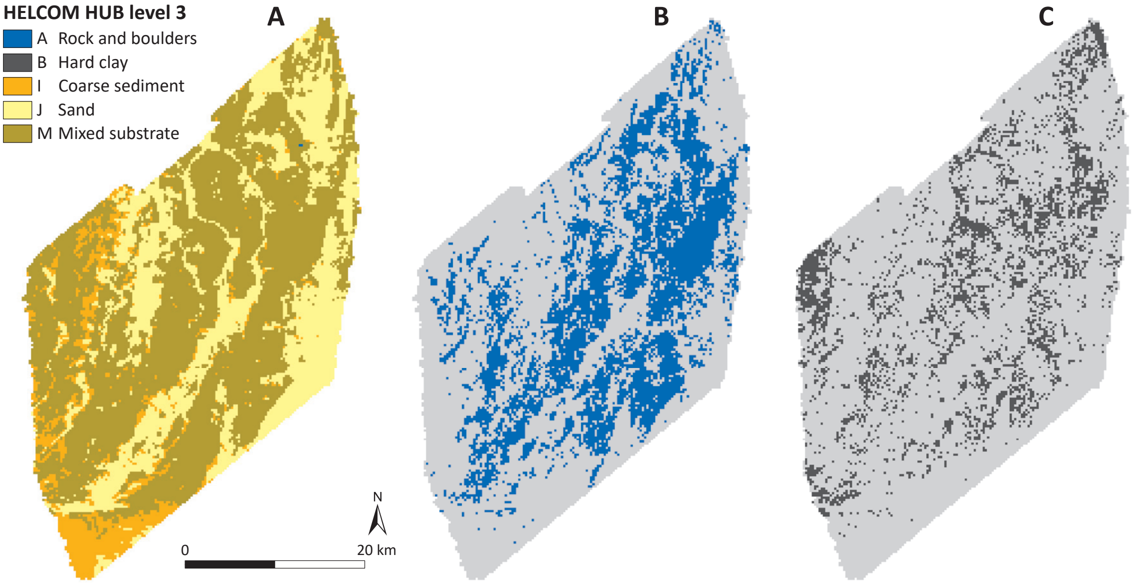


Figure 46. A. HELCOM HUB level 3 substrate map at 250 m scale. **B.** 250 m pixels containing the HELCOM HUB level 3 rock and boulders class (blue) at the 5 m scale. **C.** 250 m pixels containing the HELCOM HUB level 3 hard clay class (black) at the 5 m scale.

Effect of predefined thresholds in HELCOM HUB

Areas classified as being characterised by epibenthic bivalves or dominated by Mytilidae (i.e. ≥ 10% < 50% and ≥ 50% epibenthic bivalve cover respectively) in the reclassified HELCOM HUB map were 159 km^2 (11.8%) and 11.6 km^2 (0.9%) respectively (Fig. 47A). Areas using the same thresholds (i.e. ≥ 10% < 50% and ≥ 50%) but calculated directly from the epibenthic bivalves biological component map were 479 km^2 (35.6%) and 19.8 km^2 (1.5%) respectively (Fig. 47B). Due to overlap with other benthic organisms with higher cover (primarily perennial algae), the HELCOM HUB map was not as useful if the ability to assess the percentage cover of individual organisms is required. If the maps were to be analysed using cover as an input we recommend that the user work directly with the individual models of each substrate or biological component of interest.

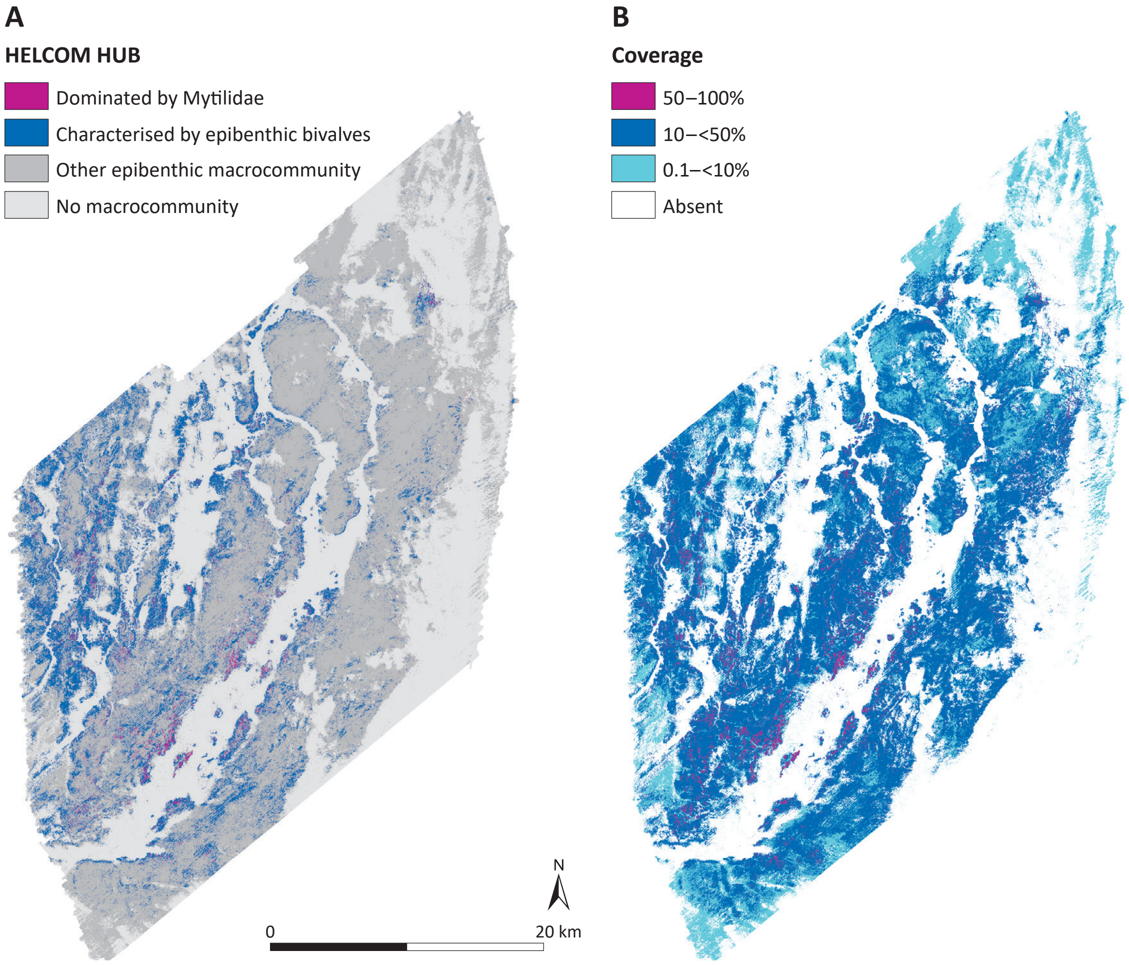


Figure 47. Comparison of epibenthic bivalves maps from HELCOM HUB classes versus the continuous prediction. **A.** Epibenthic bivalves classes from HELCOM HUB level 4–6 map (characterised by epibenthic bivalves is ≥ 10% < 50% cover, dominated by Mytilidae is ≥ 50% cover). **B.** Epibenthic bivalves classes extracted from continuous prediction (absent (< 0.1%), ≥ 0.1% < 10%, ≥ 10% < 50% and ≥ 50%).

SGU and SYMPHONY

Comparing the reclassified map produced in this project against existing SGU thematic surface substrate map showed maps from this project had greater detail and more thematic classes (6 classes) than the legacy surface substrate map which was produced at a scale of 1:500 000 (Fig. 48A versus B). The discrepancy was also observed when it was compared with SYMPHONY (Hammer et al. 2018) products, which are SGU thematic maps created for marine spatial planning (Fig. 48C versus D).

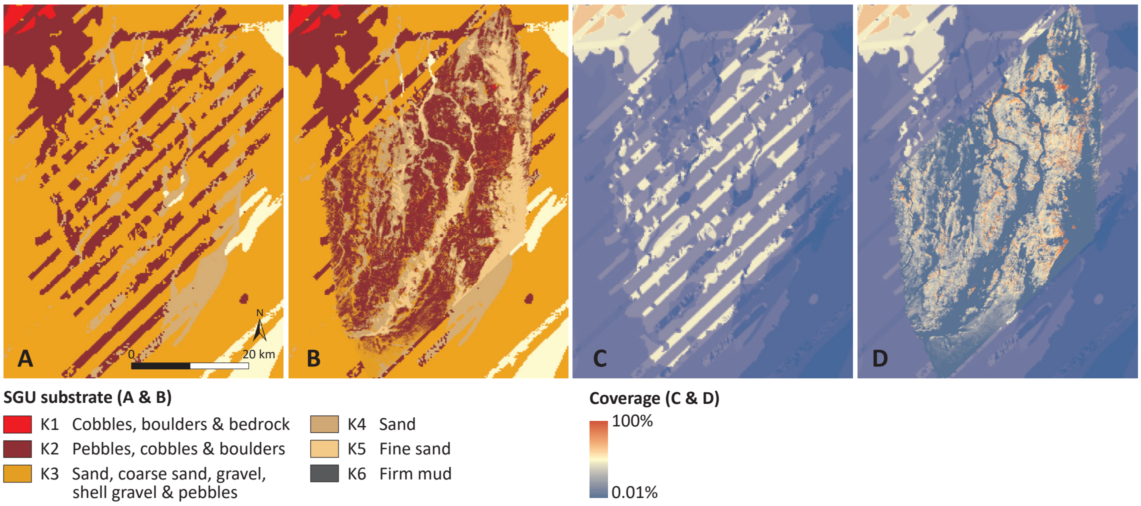


Figure 48. Before and after comparison between existing thematic map products for surface substrate and the new reclassified thematic maps (from continuous substrate models) overlaid. **A.** SGU legacy surface substrate maps (combination of 1:500 000 and 1:100 000 map). **B.** same as **A** with the reclassified thematic map overlaid. **C.** Coverage of hard bottom map (%) developed by SGU from the legacy thematic maps (same as **A**) for the marine spatial planning cumulative impact assessment tool SYMPHONY. **D.** same as **C** with the new percent hard bottom map overlaid.

Technical comparisons

Mapping and modelling both the geology and habitats of Hoburgs bank represents a new developmental direction for SGU. The following two sections investigate how the continuous modelling workflow compares with a more traditional thematic workflow, and how different predictor variables and training data affect the end result.

Thematic versus reclassified thematic maps

Comparisons of HELCOM HUB and Natura 2000 thematic maps (i.e. from multinomial models) with reclassified thematic maps (i.e. from continuous models) indicated that accuracy was similar across all levels of the classification schemes (Table 21), regardless of the BRT workflow used. This observation was further supported by studying the spatial output of the maps together with high-resolution backscatter and depth data, taking into account visual aspects such as survey artifacts.

Overall accuracy (OA) was highest for maps with fewer classes (e.g. Natura 2000), with declining OA as the number of classes increased (Table 21). The number of classes in the reclassified thematic maps was generally higher than the original classes identified in the training data, whereas the thematic maps were restricted to the number of classes in the training data set. For the complete HELCOM HUB map (levels 1–6), the number of classes doubled in the reclassified map from 29 to 59 classes. However, 8 of those classes had fewer than 10 predicted pixels (Table 22).

Table 21. Comparison of overall accuracy (OA) calculated from confusion matrices for thematic maps from multinomial models and reclassified thematic maps from continuous models for HELCOM HUB (HUB) level 3, 4, 4–5 and 4–6, and Natura 2000.

	OA (%)	
Map type →	Thematic	Reclassified
HUB 3	77.9	80.5
HUB 4	79.9	81.9
HUB 4–5	63.0	62.3
HUB 4–6	55.8	53.2
Natura 2000	87.7	87.7

Table 22. Comparison of the number of unique classes found in thematic maps from multinomial models and reclassified thematic maps from continuous models (same number of classes as in the training data) for HELCOM HUB (HUB) level 3, 4, 4–5, 4–6 and 1–6, Natura 2000, and Natura 2000 SGU subtypes. Unique classes with less than 10 predicted pixels are in parentheses.

Map type	HUB 3	HUB 4	HUB 4–5	HUB 4–6	HUB 1–6	Natura 2000	Natura 2000 subtypes
Thematic	5	3	7	9	29	2	-
Reclassified	5	3	8	12	59 (8)	2	12

Predictors, legacy data and expert annotations

Different types of predictors and training data contributed to the accuracy of thematic HELCOM HUB level 3 substrate models in different ways (Table 23). The highest accuracy was observed for the model using all training data (including legacy and expert annotations) and all types of predictor variables (including backscatter, multiscale metrics and OBIA). Models trained using the same predictors but only the point intercept dataset for training showed a large decline in accuracy (OA 77.9% vs. 58.9%). Testing model accuracy by removing OBIA, multiscale metrics and backscatter revealed that models that included a backscatter mosaic had relatively high accuracy even if only 5 m depth metrics were used (73.4%), and the gain when including OBIA and multiscale metrics was relatively low (+ 4.5%). When backscatter was not included, however, both high-resolution metrics (0.5–2.5 m) and multiscale metrics 20 m–2 km clearly contributed to increased accuracy (OBIA was not tested here due to the inclusion of backscatter). The percentage contribution for all predictors used in the thematic models is summarized in Kågesten et al. 2019.

Comparisons between biological models that included substrate models as predictors and those that did not, indicated that the inclusion of substrate increased the accuracy of biological models. For example, accuracy of the epibenthic bivalves model improved with the addition of substrate models, and the mean cover increased from 8% to 10.7% (Table 24). A visual comparison between the outputs showed that the greatest change in bivalve cover occurred in flat areas composed of hard substrates, and areas with coarse soft substrates, where it was difficult to differentiate between the two types of environment in the depth and backscatter data. When the same models were tested but backscatter was also removed, performance decreased further and the mean cover declined to 4.4% (Table 24).

Table 23. Comparison of overall accuracy (OA) from confusion matrices for bootstrap (n = 25) training and validation testing (n =154) of HELCOM HUB level 3 substrate multinomial models using different predictor variables (all groups contained northing/easting) and training data. Predictors: **1)** Depth metrics 0.5 m–2 km, OBIA, and backscatter metrics. **2)** Depth metrics 0.5 m–2 km and backscatter. **3)** Depth metrics 5 m and backscatter mosaic. **4)** Depth metrics 0.5 m–2 km. **5)** Depth metrics 0.5 m–5 m. **6)** Depth metrics 5 m, and **7)** same as **1)** with survey training data only.

Predictors	Training data	OA bootstrap (%)	OA validation (%)
1) All	Survey, legacy, expert	77	77.9
2) Reduced	Survey, legacy, expert	76	77.2
3) Reduced	Survey, legacy, expert	70	73.4
4) Reduced	Survey, legacy, expert	68	64.3
5) Reduced	Survey, legacy, expert	64	61.0
6) Reduced	Survey, legacy, expert	63	55.2
7) All	Survey	73	58.9

Table 24. Epibenthic bivalves model accuracy statistics, coefficient of determination (r^2), root mean square error (RMSE), mean absolute error (MAE) and mean error (ME), calculated from observed values from GT testing data against corresponding predicted values from models using different predictor variables. Overall accuracy (OA) results from confusion matrices where predicted values were grouped according to interval classes (OA classes) (< 10%, ≥ 10 < 50%, ≥ 50 < 90%, ≥ 90 ≤ 100%) and absence (< 0.01%) presence (≥ 0.01%) classes (OA abs–pres). Mean predicted cover (%) of epibenthic bivalves in mapped area. Predictors: **1)** refers to both substrate component models and survey metrics. **2)** same as **1)** excluding substrate component models. **3)** same as **2)** excluding all backscatter derived predictors.

Predictors	r^2	RMSE	MAE (%)	ME (%)	OA classes (%)	OA abs–pres (%)	Mean cover (%)
1) All	0.62	15.53	9.10	2.30	75	89	10.7
2) No substrate	0.60	16.69	9.84	5.61	69	88	8.0
3) No backscatter	0.49	19.60	11.83	7.96	60	75	4.4

DISCUSSION

The publication by Kågesten et al. 2019 has tackled many of the issues encountered in this project and this discussion builds from this work by adding more perspectives and elucidating salient points that could not be accommodated in the paper. This discussion is divided into the following sections; (1) High-resolution multidisciplinary seafloor mapping, which addresses the benefits and challenges associated with multidisciplinary habitat surveys, (2) The value of high-resolution continuous maps, which discusses the value of producing continuous rather than simplified thematic maps, and also highlights limitations of maps created at lower spatial and thematic resolutions, and (3) Issues with classification schemes, which describes some of the issues that were encountered when implementing the Natura 2000 and HELCOM HUB classification schemes.

High-resolution multidisciplinary seafloor mapping

The general setup of the data collection, where both hydroacoustic surveying and groundtruthing were conducted within a 24h period, was an efficient way to map in the offshore environment. It allowed an "on the fly" stratified random sampling that captured both common and rarely occurring seafloor features (that most likely would not have been found without guidance from high-resolution depth and backscatter data), and enabled a cost effective and flexible survey approach making the most of the relatively stable oceanographic conditions at night for hydroacoustic surveys, and used the day time light for sampling and underwater observations. We recognise that it is possible to separate the survey into two parts (survey and sampling) if the hydroacoustic survey has already been conducted and it includes backscatter data.

The high quality, high-resolution depth and backscatter data from multibeam surveys, together with accurately positioned, standardised underwater observations and samples were of paramount importance for producing high-resolution habitat models. The results from the modelling clearly showed the significant role of backscatter data, which greatly improved the quality of our map outputs. This makes it an essential product to collect during the multibeam surveys especially now that new standards and approaches in backscatter collection and processing have been shown to further improve the backscatter map outputs (Malik 2019). With this in mind, it becomes of critical concern that multibeam backscatter data does not have a designated data host in Sweden. SGU together with the Swedish Maritime Administration (SMA) need to further collaborate and raise support to ensure that current data holdings are processed and made available, and that collecting high quality backscatter data is prioritised during multibeam surveys to meet future substrate and habitat mapping needs.

Furthermore, SGU, in collaboration with SLU Aqua, also collected and analysed water column data from the multibeam sonar as well as split beam sonar data designed for fish detection. This analysis showed that it was possible to use the water column data from the multibeam (which is commonly not saved) to map fish (biomass), and that collecting split beam sonar data worked well together with the existing sonar systems on SGU's survey vessel. Notably, it was possible to observe the patterns of schooling fish (daytime) as they spread out for feeding during night time. The current challenge is in storing and processing the relatively large quantities of data, and to have a national data host to serve this data. The raw sonar data collected on Hoburgs bank is now stored at SGU and is freely available for future studies.

During the hydroacoustic survey we were challenged by meeting both project objectives of collecting high-resolution high-quality seafloor information over a large area and conforming to hydrographic standards for charting set by SMA. By closely monitoring oceanographic conditions together with a moving vessel profiler that continuously measured sound velocity changes in the water column we were able to survey with higher coverage and speed within the limits set by the hydrographic standards, roughly doubling the speed of acquisition, especially in the shallow areas between 10 and 25 m. To compensate, we spent significant time collecting cross-lines to be able to statistically show the quality of the depth information provided to SMA. The results showed that in most areas we were still able to produce accurate, high quality, and high-resolution (0.5 m) depth data (i.e. within the limits of IHO Special Order). Still, the data have not yet been fully integrated into the SMA charts, and likewise SGU have had problems accessing and unlocking the full potential of data collected by SMA, including backscatter collected in a standardised way. To sum up, there is a further need to align the objectives between charting and habitat mapping/research surveying to make sure that we can make the most of all multibeam surveys conducted in Swedish waters.

Our underwater observation methods were aligned according to the latest draft of "Visuella undervattensmetoder för uppföljning av marina naturtyper och typiska arter", (Havs- och vattenmyndigheten 2015, unpublished report) which sets the national guidelines on how to collect drop-camera information for biological surveys. Due to the custom-made drop camera system on SGU's survey vessel, we were able to collect detailed and standardised information with high positional accuracy. This, in combination with sediment samples that allowed us to accurately map the fine grain (clay and sand) fractions, was the foundation of mapping patchy habitats in high-resolution. Using a standardised footprint and resolution that aligned with the modelling resolution of our maps helped us to better understand the effect of spatial scale in the final products. Today, most drop-camera surveys in Sweden are done with cheaper and simpler setups designed for fast surveys in shallow waters. Although useful for many applications, this kind of data can have great variation in quality as a result. It is clear that the current standards used for drop-camera surveys need to be developed further to help increase the quality of observations and account for new technologies and challenges associated with deeper/offshore surveys. Although the result produced by the drop-camera and sampling method was satisfactory, the main drawbacks were the slow speed of UW-obs (roughly 10–15 min per site with the rotating camera system), and the difficulty encountered observing the substrate and benthic organisms with only one camera perspective when canopy forming algae limited the view of the seafloor.

The use of legacy data and expert annotations, together with the approach of using substrate model outputs as predictors increased model accuracy and is recommended for use in future studies as reported in Kågesten et al. 2019. Accuracy analysis of the maps produced in this project have utilised various statistical measures, and the results showed a generally high level of accuracy with most of the errors found in patchy, heterogeneous seascapes. Possible errors in point intercept annotations and limitations encountered during data collection and processing that affected the quality of the survey data may have contributed to lower accuracy of the habitat maps. Future developments in sonar technology and computer vision technologies (Beijbom et al. 2012, Berthold et al. 2017, Dumke et al. 2018, González-Rivero et al. 2016) have the potential to further improve the approach and outputs presented in this project. It is, however, pragmatic to accept that there will always be some degree of uncertainty in habitat mapping projects because of natural variability that modelling cannot incorporate (Fiorentino et al. 2018, Rocchini et al. 2011). Even with the innovative use of expert interpretation and historical data for improved substrate models, a notable challenge within Hoburgs bank was the ability to collect and process sufficient, high-quality underwater observations and samples, especially for rare habitats and species. We believe that in the long-term, it would be beneficial to connect high quality observational data between multiple projects and utilise the benefits of large data sets in order to increase the accuracy of model outputs and reduce survey costs. For that to happen we need better standards for both high quality acquisition and interpretation of observational data as well as improved standards for high quality multibeam acquisition of both depth and backscatter. Standardised data from geophysical survey products such as sub-bottom profilers would also be helpful. The ongoing mapping of the nearby Northern Midsea bank will be a first opportunity to show how data collected on Hoburgs bank can help map other areas as well.

Describing the accuracy of our map products was not easy since there were many sources of error and many ways to describe accuracy. In the results section we have provided several accuracy measures to enable the user to better understand the limitations of these products. However, the measured accuracy is also dependent on the accuracy of the positioning of samples/observations, and the accuracy of interpretation of substrate and benthic organisms. The latter is a tricky exercise as we noted quite large differences between interpreters during the project, even between seasoned experts. In the end, one expert had to go through all observations to streamline the interpretations. Future projects would benefit from quantifying the error associated with sampling and underwater observations at the interpretation level. Also, increasing the number of sieve analysed sediment samples would improve the knowledge about finer grain sizes, and in turn enable better understanding of the connection between biota, such as infauna, and the seafloor substrates. Since the study area is extremely patchy in its distribution of substrate and benthic organisms, the maps were created at high-resolution to connect the scale of the observations with the environmental data one to one. There is still likely to be a substantial amount of variation between the modelled maps and reality, but our understanding of the area together with the data showed that the maps do a very good job of describing this complex mosaic of habitats. We believe that future improvements are less in the modelling approach, and all about improving the accuracy and quantity of the survey data. As expected, map accuracy dropped as more habitat levels were combined (i.e. accuracy for HELCOM HUB level 3 (substrate) was 80.5%, 62.3% for levels 4–5 (characteristic community), and 53.2% for levels 4–6 (dominating taxa). This suggests that users are better off working directly with the principal components of the maps (e.g. cover of epibenthic bivalves or the cover of hard bottom) when possible.

Issues with classification schemes

Mapping habitats using HELCOM HUB and Natura 2000 classification schemes was a novel exercise for the SGU habitat mapping team. HELCOM HUB was a new classification scheme and there were no examples of using the scheme for high-resolution mapping when we started the project, hence it was unavoidable that issues were found in its implementation. The Natura 2000 definitions, though quite simple, were challenging to translate into maps as definitions were fairly open to interpretation. This section highlights some specific issues we found in implementing these region-wide classification schemes, in the hope that it can help guide future updates and improvements.

Regarding Natura 2000, there were only two main classes specified in habitat mapping, "reefs" and "sandbanks". The reef class was well defined aside from the guidance on spatial scale. However, we noticed that the definition did not capture the large differences within the class, for example the difference between a flat hard bottom and a rugose complex patch reef. We addressed this by defining and mapping reef subtypes based on general descriptions and our knowledge of the Hoburgs bank area. The "sandbank" class was challenging since it was supposed to apply only to shallow areas. However, we found little evidence that the areas shallower than 20–30 m were more ecologically valuable then the deeper areas, which was technically still shallow relative to the surrounding area. To avoid dealing with this dilemma, the depth part of the definition was not included, leaving it to the user to add this later if relevant (the depth grid is freely available from SGU together with the maps). We also created sandbank subtypes that were defined using the level of energy (i.e. signs of sand transport) at the seabed to separate deeper, low-energy sandy areas and shallower, high-energy areas, since we observed differences in the species composition between these two sand areas.

As previously stated, the project's interpretation of HELCOM HUB level 6 biotopes was different to that which was intended by the scheme (Table 17). However, classification to level 6 biotopes, where the term, "dominated by" is used, provides a false sense of importance as this can amount to a coverage from as low as 5% to as high as 100% for a particular level 6 group, depending on presence of other groups from the same level 5 biotope (e.g. perennial algae). This could mean that an area classified as far as level 5, i.e. "characterised", could be interpreted as being less important than an area classed to level 6, i.e. "dominated by", even though the former may have a more diverse assemblage of organisms but no single level 6 group has a biomass/biovolume greater than 50%. This project's HELCOM HUB level 6 interpretation provided a little more information insofar as an area classified to level 6, meant that a particular level 6 group covered at least 50% of the substrate.

HELCOM HUB recommends that a continuous scale of coverage, instead of discrete classes, is used to decrease the chance of two groups having the exact same cover value, which causes problems in the classification split rules on levels 5 and 6. Despite using a continuous coverage scale, there were several GT sites where two perennial attached groups had the exact same coverage value, which affected classification. While the effect was small and the approach of combining continuous models to HELCOM HUB classes wasn't affected, rules need to be devised to select a specific group over another when an equal coverage occurs between groups. For example, ranking groups within the perennial attached erect group based on some measure of ecological importance.

By classifying from continuous coverage models to HELCOM HUB, new combinations of levels 1–6 were produced that were not found at GT sites nor been previously defined in HELCOM HUB. This can be viewed as a positive effect of the project's method by identifying potentially new biotopes. However, it is stated on page 74 of the HELCOM HUB technical report (HELCOM HUB 2013) in a table that outlines the classes on each level that "the table is not to be used for creating 'new' biotopes by selecting one feature from each level". Several GT sites in this project were identified as hard clay dominated by perennial filamentous algae. If these sites were classified exactly according to HELCOM HUB they would've been hard clay characterised by macroscopic biotic structures (AA.B1) i.e. only down to level 4. However, it is stated on page 34 (HELCOM HUB 2013) that 'no dead-ends should be encountered before level 5'. This project, therefore, took a more pragmatic approach and created 'new' biotopes in order to classify to levels 5 and 6 when the data supported it.

Finally, many of the issues we encountered with regards to implementing the region-wide classification schemes could have been avoided if a verified and complete open source code was made available (for example R or Python) for the end users. This will not only clarify many of the intended outcomes of the scheme but also standardise and prevent implementation differences between end-users.

The value of high-resolution, continuous maps

As discussed in Kågesten et al. 2019, the results from this project suggest that high-resolution continuous and thematic models can help overcome some of the limitations of classification schemes. Accuracy, i.e. overall accuracy of observed vs. predicted values, of classified HELCOM HUB and Natura 2000 maps were similar regardless of whether classified data were modelled directly or reclassified as continuous models of coverage (%) (Table 21). However, producing continuous maps and then applying classifications had a greater value due to the added end-user possibilities, such as simple application of thresholds (conditions) or spatial analysis directly from continuous models (e.g. through raster calculation).

Kågesten et al. 2019 further discussed the advantage of using continuous, high-resolution maps as a basis for producing multi-scale maps of required categories that still accommodate the different definitions in existing classification schemes. This is particularly evident in HELCOM HUB classification since continuous maps offer the possibility to highlight other classes that may be crucial for species management which would otherwise be subsumed if only HELCOM HUB classes are used.

Furthermore, the study pointed out the advantage of using continuous, high-resolution maps to identify distinct features that are essential for ecological function that would have important implications for effective management interventions. Since the new maps of Hoburgs bank show the seascape at an unprecedented level of detail both spatially and thematically, we believe that new scientific questions related to benthic habitats and the larger ecosystem can be further explored and understood. Such questions could include how seabird feeding areas relate to features with high mussel cover, how invasive species such as the mussel feeding fish *Neogobius melanostomus* affect the benthic communities or tracking how fish and mammals move and use the seascape. The detailed maps also mean that our present understanding of nature values only capture simplistic assumptions of management concepts. For example, we often discuss the value of shallow areas where greater light availability allows a higher diversity of plants, algae and associated fauna to exist. However, as more data from surveys becomes available, we find ecologically diverse areas even in places with limited light availability. Notably the deepest sandy areas mapped (40–60 m) had more crawl tracks, shells and an abundance of bottom feeding shrimp, compared with the shallower sandy areas. The extensive system of moraine ridges and hard clay features across the bank was also something that was previously unknown at this level of detail and the underwater observations indicated that these features were linked to both higher concentration of mussels and algae, and diversity and abundance of fish.

We argue that this project has exposed several weaknesses of the most commonly used marine classification systems in Sweden, and perhaps the entire region, and that the current language for marine habitat classification in Sweden has room for improvement. Specifically, HELCOM HUB, though useful for highlighting general habitat types, has been challenging to fully understand and map, particularly at more detailed levels. As outlined in the habitat map descriptions, our implementation of HELCOM HUB (particularly level 6) turned out to be slightly different to that of some other users in other countries around the Baltic Sea. An overview of what HELCOM HUB captures and how it can be practically implemented in mapping activities would be of great use, including the use of biomass vs. percent cover. We recommend that more focus should be placed on mapping the most important geological and biological components of benthic habitats (preferably using continuous variables if feasible), rather than combining them into advanced multilevel thematic classes. We hope the maps developed for Hoburgs bank can help advance future discussion on this theme even as we admit that even within SGU, more work is needed to align our existing geological maps to the new methods described in this report, including the general use of classification schemes for substrate and geology.

CONCLUSION

This project marks a milestone for habitat mapping in Sweden and shows how a combination of multidisciplinary surveys and machine learning technology can be combined to create state of the art information about our seafloor environments. It also marks a new direction for SGU in moving from the traditional, expert-driven, map making process to one where inputs from experts and machine learning algorithms are combined to create maps with many new themes with high detail and accuracy. Simultaneous collection and integration of physical, geological and biological information was key to the success of this project. The project also generated additional benefits such as generating data for use in charting, and observations on how fish utlilise the seafloor habitats. A critical part of the mapping campaign was the use of data from different sources, i.e. multibeam depth and backscatter, drop camera, and sediment sample information, during early stages of the survey. This resulted in more efficient data collection of both common and rare habitat features.

Multibeam backscatter, which was one of the most important variable in the models, is often overlooked in national/regional benthic models due to difficulties to obtain this data from the Swedish Maritime Administration. The access to backscatter data on a national level needs to be addressed to meet this demand, as well as the national standards for multibeam surveys where there is little support for backscatter acquisition today.

Habitat modelling with the "continuous approach" resulted in maps which are not limited to one classification scheme or scale but can be used to capture benthic environments in both thematic maps (e.g. HELCOM HUB classification) and continuous maps (e.g. coverage (%) of epibenthic bivalves). This approach will improve our understanding of the seascape in areas where such maps exists and help researchers and policy makers better understand its benefits and the limitations of existing and future models and maps produced on national and regional levels. High-resolution continuous maps captured many important habitats or features that would otherwise have been lost with lower resolution data. These maps also provided more flexibility in application and could be scale-adjusted and resampled more easily to address a wider range of research questions and management applications.

While the value and convenience of thematic maps is recognised, relying solely on such classification schemes may result in the neglect of habitats or features with great environmental value. It is therefore essential to understand the limitations of any classified maps for marine spatial planning and decision-making, particularly with regard to scale and schema dependence, as well as the associated errors that may be propagated through the mapping process. We suggest that the HELCOM HUB system be revised and simplified, since the current descriptions are not only complicated to fully understand and use, but can also leave some ecologically important features underrepresented. An overview of the national implementation of the Natura 2000 system would also be valuable with the lessons learned in this project in mind. Our survey of Hoburgs bank revealed an area filled with a complex and unique mix of seafloor habitats previously obscured by lower resolution maps. We believe that similar high-resolution maps could be valuable in other areas with complex habitats where national scale models may prove less useful for many applications. Finally, the methodologies and the lessons learned from this project can serve as a valuable starting point for future habitat mapping projects. However, due to the complex nature of both the survey and the analysis involved, this knowledge needs to be combined with a long term commitment to habitat mapping in order to be successful over time.

REFERENCES

Axe, P. & Lindow, H., 2005: *Hydrographic Conditions Around Offshore Banks*. SMHI Oceanografisk Enhet.

Beijbom, O., Edmunds, P. J., Kline, D. I., Mitchell, B. G. & Kriegman, D., 2012: Automated annotation of coral reef survey images. *2012 IEEE Conference on Computer Vision and Pattern Recognition,* 1170–1177. <https://doi.org/10.1109/CVPR.2012.6247798>

Berthold, T., Leichter, A., Rosenhahn, B., Berkhahn, V. & Valerius, J., 2017: Seabed sediment classification of side-scan sonar data using convolutional neural networks. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1–8). <https://doi.org/10.1109/SSCI.2017.8285220>

Brown, C. J. & Blondel, P., 2009: Developments in the application of multibeam sonar backscatter for seafloor habitat mapping. *Applied Acoustics, 70*(10), 1242–1247. <https://doi.org/10.1016/j.apacoust.2008.08.004>

Buja, K. & Menza, C., 2007: Sampling Design Tool for ArcGIS. National Centers for Coastal Ocean Science. <https://coastalscience.noaa.gov/project/sampling-design-tool-arcgis/> accessed 200904.

Carlén, I., Thomas, L., Carlström, J., Amundin, M., Teilmann, J., Tregenza, N., Tougaard, J., Koblitz, J. C., Sveegaard, S., Wennerberg, D., Loisa, O., Dähne, M., Brundiers, K., Kosecka, M., Anker Kyhn, L., Tiberi Ljungqvist, C., Pawliczka, I., Koza, R., Arciszewski, B., Galatius, A., Jabbusch, M., Laaksonlaita, J., Niemi, J., Lyytinen, S., Gallus, A., Benke, H., Blankett, P., Skóra, K. E. & Acevedo-Gutiérrez, A., 2018: Basin-scale distribution of harbour porpoises in the Baltic Sea provides basis for effective conservation actions. *Biological Conservation, 226,* 42–53. <https://doi.org/10.1016/j.biocon.2018.06.031>

Carlström, J. & Calen, I., 2016: Skyddsvärda områden för tumlare i svenska vatten.

Costa, B. M. & Battista, T. A., 2013: The semi-automated classification of acoustic imagery for characterizing coral reef ecosystems. *International Journal of Remote Sensing, 34*(18), 6389–6422. <https://doi.org/10.1080/01431161.2013.800661>

Council of the European Communities, 1992: Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora.

Davies, C. E. & Moss, D., 2004: EUNIS Habitat Classification Marine Habitat Types: Revised Classification and Criteria. *Centre for Ecology & Hydrology, 44*(0), 84.

De'ath, G., 2007: Boosted Trees for Ecological Modeling and Prediction. *Ecology, 88*(1), 243–251. <https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2>

Dumke, I., Purser, A., Marcon, Y., Nornes, S. M., Johnsen, G., Ludvigsen, M. & Søreide, F., 2018: Underwater hyperspectral imaging as an in situ taxonomic tool for deep-sea megafauna. *Scientific Reports, 8*(1), *12860.* <https://doi.org/10.1038/s41598-018-31261-4>

Elith, J., Leathwick, J. R. & Hastie, T., 2008: A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>

ENVI, 2008: ENVI Feature Extraction Module User's Guide, Feature Extraction Module Version 4.6. ITT Visual Information Solutions. <http://www.harrisgeospatial.com/portals/0/pdfs/envi/feature_extraction_module.pdf> accessed 200904.

European Commission, 2008: *NATURA 2000 - protecting Europe's biodiversity.* (S. Wegefelt, Ed.). Oxford: Information Press.< https://doi.org/10.2779/45963>

European Commission, 2013: Interpretation Manual of European Union Habitats. October, (July), 142. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:INTERPRETATION+MANUAL+OF+EUROPEAN+UNION+HABITATS#0> accessed 200904.

Fernández-Delgado, M., Cernadas, E., Barro, S. & Amorim, D., 2014: Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research, 15,* 3133–3181. <https://doi.org/10.1117/1.JRS.11.015020>

Fiorentino, D., Lecours, V. & Brey, T., 2018: On the Art of Classification in Spatial Ecology: Fuzziness as an Alternative for Mapping Uncertainty. *Frontiers in Ecology and Evolution, 6.* <https://doi.org/10.3389/fevo.2018.00231>

Forsman, B., 2017: *Omdirigeringsanalys av sjöfart kring Hoburgs bank och Midsjöbankarna – Underlag inom svensk havsplanering.*

Friedman, J. H., 2001: Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics, 29*(5), 1189–1232. <https://www.jstor.org/stable/2699986> accessed 200904.

González-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Holtrop, T., González-Marrero, Y., Ganase, A., Roelfsema, C., Phinn, S. & Hoegh-Guldberg, O., 2016: Scaling up ecological measurements of coral reefs using semi-automated field image collection and analysis. *Remote Sensing, 8*(1). <https://doi.org/10.3390/rs8010030>

Greene, H. G., Cacchione, D. A. & Hampton, M. A., 2017: Characteristics and Dynamics of a Large Sub-Tidal Sand Wave Field—Habitat for Pacific Sand Lance (Ammodytes personatus), Salish Sea, Washington, USA. *Geosciences.* <https://doi.org/10.3390/geosciences7040107>

Gullström, M., Sundblad, G., Mörk, E., Sjöö, G. L., Naeslund, M., Halling, C. & Lindegarth, M., 2017: Utvärdering av videoteknik som visuell undervattensmetod för uppföljning av marina naturtyper och typiska arter - Metodsäkerhet, precision och kostnader. *Havs- och vattenmyndighetens rapport 2017:8*.

Hallberg, O., Nyberg, J., Elhammer, A. & Erlandsson, C., 2010: Ytsubstratklassning av maringeologisk information. *SGU-rapport 2010:6*. Sveriges geologiska undersökning.

Hammar, L., Schmidtbauer Crona, J., Kågesten, G., Hume, D., Pålsson, J., Aarsrud, M., Mattson, D., Åberg, F., Hallberg, M. & Johansson, T., 2018: Integrerat planeringsstöd för statlig havsplanering utifrån en ekosystemansats. *Havs- och vattenmyndighetens rapport 2018:1*.

Heinänen, S., Chudzinska, M. & Skov, H., 2018: Effekter av omdirigering av sjöfart på alfågel och tumlare vid Hoburgs bank och Midsjöbankarna. *Havs- och vattenmyndigheten dnr: 396-18*.

HELCOM., 2013: Technical Report on the HELCOM Underwater Biotope and habitat classification. Baltic Sea Environment Proceedings.

Hengl, T. (2009). A Practical guide to Geostatistical Mapping. *Government Publications Review* (Vol. 13). University of Amsterdam. <https://doi.org/10.1016/0277-9390(86)90082-8>

Herkül, K., Peterson, A. & Paekivi, S., 2017: Applying multibeam sonar and mathematical modeling for mapping seabed substrate and biota of offshore shallows. *Estuarine, Coastal and Shelf Science, 192,* 57–71. <https://doi.org/10.1016/j.ecss.2017.04.026>

Karlsson, E., Fredriksson, R. & Bergström, U., 2017: Bottenfaunaprovtagning vid Hoburgs bank 2016. *SLU ID: SLU.ua. 2016.5.2-374*. Sveriges lantbruksuniversitet.

Kautsky, H., 2000: Hoburg och Midsjöbankar – naturreservat eller vindmöllepark i utsjön? *Östersjö,* 34–35.

Kautsky, H., Martin, G. & Snoejis-Leijonmalm, P., 2017: The phytobenthic zone. In P. Snoeijs-Leijonmalm, H. Schubert, & T. Radziejewska (Eds.), *Biological Oceanography of the Baltic Sea* (p. 387). Springer Netherlands.

Kuhn, M., 2008: Building Predictive Models in R Using the caret Package. *Journal of Statistical Software, 28*(1), 1–26. <https://doi.org/10.18637/jss.v028.i05>

Kuhn, M. & Johnson, K., 2013: *Applied predictive modeling.* New York; Heidelberg; Dordrecht; London: Springer.

Kågesten, G., Sautter, W., Ann, E. K., Costa, B. M., Kracker, L. M. & Battista, T. A., 2015: Shallow-water benthic habitats of Northeast Puerto Rico and Culebra Island. (N. C. for C. O. S. (U.S.), Ed.). Silver Spring, MD. <https://doi.org/10.7289/V5Z899FH>

Kågesten, G., Fiorentino, D., Baumgartner, F. & Zillén, L., 2019: How do continuous high-resolution models of patchy seabed habitats enhance classification schemes? *Geosciences (Switzerland), 9*(5). <https://doi.org/10.3390/geosciences9050237>

Larsson, K., 2016: Sjöfart och naturvärden vid utsjöbankar i centrala Östersjön - Havsplanering kan reducera konflikter. *Havs- och vattenmyndighetens rapport 2016:24*.

Larsson, K. & Karlsson, P., 2018: Fartygstrafik i och nära skyddade och känsliga havsområden runt Gotland och Öland (Rapporter om natur och miljö). *Rapport nr 2018:11*. Länstyrelsen i Gotlands län.

Leathwick, J., Moilanen, A., Francis, M., Elith, J., Taylor, P., Julian, K., Hastie, T. & Duffy, C., 2008: Novel methods for the design and evaluation of marine protected areas in offshore waters. *Conservation Letters, 1*(2), 91–102. <http://dx.doi.org/10.1111/j.1755-263X.2008.00012.x>

Lurton, X., 2010: *An Introduction to Underwater Acoustics - Principles and Applications* (2nd ed.). Springer-Verlag Berlin Heidelberg.

Länstyrelsen i Gotlands län, 2005: *Bevarandeplan för Natura 2000-område Hoburgs bank SE0340144.*

Malik, M., 2019: Sources and impacts of bottom slope uncertainty on estimation of seafloor backscatter from swath sonars. *Geosciences (Switzerland), 9(4).* <https://doi.org/10.3390/geosciences9040183>

Masetti, B. G., Gallagher, B., Calder, B. R., Zhang, C. & Wilson, M., 2018: Sound Speed Manager: An open-source application to manage sound speed profiles. *The International Hydrographic Review, 0*(17), 31–40.

Naturvårdsverket, 2008: Undersökning av utsjöbankar – Inventering, modellering och naturvärdesbedömning. *Rapport nr 6385.*

Naturvårdsverket, 2010: Inventering av marina naturtyper på utsjöbankar. *Rapport nr 5576.*

Nyberg, J., 2016 Beskrivning till maringeologiska kartan Sydöstra Östersjön. *Sveriges geologiska undersökning K 542.*

Piñeiro, G., Perelman, S., Guerschman, J. P. & Paruelo, J. M., 2008: How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling, 216*(3), 316–322. <https://doi.org/10.1016/j.ecolmodel.2008.05.006>

Pittman, S. J. & Brown, K. A., 2011: Multi-Scale Approach for Predicting Fish Species Distributions across Coral Reef Seascapes. *PLOS ONE, 6*(5), e20583. <https://doi.org/10.1371/journal.pone.0020583>

Pittman, S. J., Costa, B. M., & Battista, T. A. (2009). Using Lidar Bathymetry and Boosted Regression Trees to Predict the Diversity and Abundance of Fish and Corals. *Journal of Coastal Research (JCR), 2009(*10053), 27–38. <https://doi.org/10.2112/SI53-004.1>

Pittman, S. J., Connor, D. W., Radke, L. & Wright, D. J., 2012: Application of Estuarine and Coastal Classifications in Marine Spatial Management. In E. Wolanski & D. McLusky (Eds.), *Treatise on Estuarine and Coastal Science* (Vol. 1, pp. 163–205). Elsevier. <https://doi.org/10.1016/B978-0-12-374711-2.00110-8>

R Core Team., 2017: *R: A language and environment for statistical computing.* R Foundation for Statistical Computing Vienna, Austria. <https://www.r-project.org/> accessed 200904.

Regeringskansliet, 2016: Regeringen skyddar marina områden för tumlare. <https://www.regeringen.se/pressmeddelanden/2016/12/regeringen-skyddar-marina-omraden-for-tumlare> accessed 190426.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A., 2011: Accounting for uncertainty when mapping species distributions: {The} need for maps of ignorance. *Progress in Physical Geography, 35(*2), 211–226. <https://doi.org/10.1177/0309133311399491>

Skov, H., Heinänen, S., Žydelis, R., Bellebaum, J., Bzoma, S., Dagys, M., Durinck, J., Garthe, S., Grishanov, G., Hario, M., Kieckbusch, J. J., Kube, J., Kuresoo, A., Larsson, K., Luigujoe, L., Meissner, W., Nehls, H. W., Nilsson, L., Krag Petersen, I., Mikkola Roos, M., Pihl, S., Sonntag, N., Stock, A., Stipniece, A. & Wahl, J., 2011: Waterbird populations and pressures in the Baltic Sea. *TemaNord 2011:550.* Nordic Council of Ministers.

Stamoulis, K. A., Delevaux, J. M. S., Williams, I. D., Poti, M., Lecky, J., Costa, B., Kendall, M. S., Pittman, S. J., Donovan, M. K., Wedding, L. M. & Friedlander, A. M., 2018: Seascape models reveal places to focus coastal fisheries management. *Ecological Applications, 28*(4), 910–925. <https://doi.org/10.1002/eap.1696>

Stephens, D. & Diesing, M., 2015: Towards Quantitative Spatial Models of Seabed Sediment Composition. *PLOS ONE, 10*(11), *e0142502.* <https://doi.org/10.1371/journal.pone.0142502>

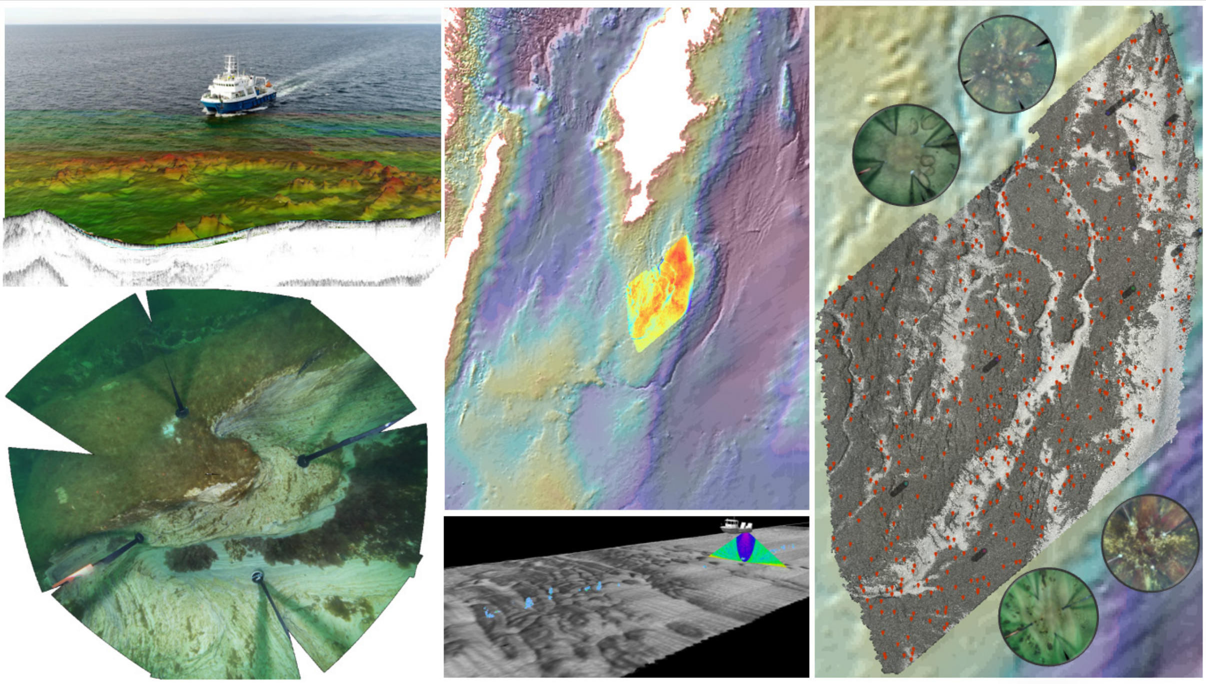
Ward, T. J., Vanderklift, M. A., Nicholls, A. O. & Kenchington, R. A., 1999: Selecting marine reserves using habitats and species assemblages as surrogates for biological diversity. *Ecological Applications, 9*(2), 691–698. <https://doi.org/10.1890/1051-0761(1999)009[0691:SMRUHA]2.0.CO;2>

APPENDIX 1

Bio-geophysical survey methods for habitat mapping of Hoburgs bank

Francis Freire, Gustav Kågesten,
Finn Baumgartner and Annika Dahlgren

June, 2020



Cover image: Collage containing a drone image of SGU's research vessel R/V *Ocean Surveyor*, Location map of Hoburgs bank, backscatter map of Hoburgs bank with samples of underwater picture and their locations, and center bottom is the water column sonar image of the survey. Illustrator: Gustav Kågesten.

Table of contents

ABSTRACT

Hoburgs bank is a shallow water (10-63 m) shoal located south of Gotland that has been identified as an area with high ecological values, and recently became a part of Sweden's largest marine Natura 2000 reserve. Due to its significance and the lack of detailed survey information in the area, the Geological Survey of Sweden (SGU) was tasked by the Swedish Agency for Marine and Water Management (SwAM) to conduct a multiciliary high-resolution biological and geophysical survey 2016–2017. This report presents the field survey methods and data processing used in the bio-geophysical mapping of 1344 km^2 seafloor of Hoburgs bank. In essence, the report builds on information that the Swedish Maritime Authority (SMA) require to evaluate the quality of a multibeam survey, with the addition of similarly detailed technical information of geophysical data such as backscatter data from the multibeam sonar and sediment profiler sonar, oceanographic data and biological data. The information generated from this report was used to create sediment and habitat maps of Hoburgs bank in 5 m resolution. The depth data generated by the survey was also shared to SMA for charting purposes. The survey data is freely available on request to SGU due to its non sensitive location outside the territorial boarders of Sweden, and the shared financing with SwAM. Finally, it is the intention that this report serve as the basis for other, future marine habitat mapping initiatives of SGU and others, and provide inspiration on how much environmental information can be collected during a single field campaign.

SAMMANFATTNING

Hoburgs bank är en grund (10–63 m) utsjöbank strax söder om Gotland med tidigare kända höga naturvärden, och numera en del av Sveriges största marina skyddade Natura 2000-område. Då Hoburgs bank trots sin betydelse saknade detaljerade kartor över geofysiska och biologiska förhållanden har Sveriges geologiska undersökning (SGU) tillsammans med Havs- och vattenmyndigheten (HaV) samarbetat för att kartlägga 1344 km^2 havsbotten i området 2016–2017. SGU har även haft karteringen som en pilot för att utveckla hur topografiska, oceanografiska, geologiska och biologiska förhållanden på havsbotten kan karteras med hög upplösning från ett och samma fartyg i en integrerad multidisciplinär undersökningsmetod. Utöver kartering av bentiska förhållanden samarbetade SGU i delar av projektet även med Institutionen för akvatiska resurser (SLU Aqua) för att samla in information om infauna i sedimenten samt fisk. Den här rapporten är en detaljerad redovisning av de tekniska detaljerna för sjömätningsdelen av projektet. I stort bygger innehållet i rapporten på de krav Sjöfartsverket ställer på djupmätningar för navigation, dock med tillägget att även geofysiska och biologiska undersökningsmetoder är redovisade här. Informationen som samlades in från fältmätningarna har sedan använts för att skapa detaljerade kartor över livsmiljöerna på Hoburgs bank (djup, geologi, biologi). Djupinformationen har även delats med Sjöfartsverket för att möjliggöra att det kan användas för framställning av sjökort. Då Hoburgs bank ligger inom Sveriges ekonomiska zon men utanför territorialgränsen är den insamlade informationen fri från sekretessrestriktioner, och tillhandahålls i alla dess delar via förfrågan till SGUs kundtjänst. Den öppna tillgången (samfinansierat av SGU och HaV), detaljnivån och storleken på denna information gör det till ett unikt marint dataset i Sverige som även visar på hur komplex och varierad havsmiljön ofta kan vara men sällan visas i kartform. Intentionen är att denna rapport bidrar till fortsatt utveckling av kostnadseffektiv och detaljerad kastrering av havsmiljön utifrån de många olika aspekter som behövs för att möta informationsbehoven av en hållbar blå ekonomi.

INTRODUCTION

Hoburgs bank is a large and relatively shallow shoal that is located on the western part of the Baltic Sea approximately 10 nautical miles south of Gotland. Shaped by receding ice during the last deglaciation, it is known as one of Sweden's ecologically complex and diverse marine areas. It was declared a Natura 2000 site that forms part of a network of protected areas across the entire European continent listed under the Habitats Directive (Council of the European Communities, 1992). Its shallow characteristic along the Baltic Sea not only make it an important habitat for many species, but its location within one of the Baltic's busiest sea lanes also makes the shoal a vulnerable to a myriad of anthropogenic disturbances.

In 2016, the Geological Survey of Sweden (SGU) was tasked by the Swedish Agency for Marine and Water Management (HAV) to undertake a comprehensive mapping project of this important ecological region. The main goal of this endeavor was to collect different bio-geophysical spatial information on Hoburgs bank for habitat mapping, and eventually, in marine spatial planning. Habitat mapping can be defined as the collection and synthesis of physical, geological and biological data necessary to differentiate environmental features that can make a particular area suitable for life functions (Kurland & Woodby 2008). Habitat mapping projects require a large number of varying datasets that are used to create better and more effective models for marine spatial planning applications.

The habitat mapping survey of Hoburgs bank studied different aspects of the marine environment. Thus, several survey systems and survey methods were used to collect information about the water column, geology, sediment type, seafloor topography, oceanographic conditions and biology over a two-year span. The data collection work was mainly conducted onboard SGU's ship, the R/V *Ocean Surveyor*, a catamaran-type survey vessel designed and equipped with various sensors for scientific data gathering (Fig. 1). This report focuses predominantly on the description of the technical aspects of data collection that includes the specific methodology employed during the field survey, the technical specifications of the equipment, and methods used in the processing of the data. This report is an accompaniment to another document, "The high-resolution Hoburgs bank benthic habitat mapping report", which details the modeling and production of habitat maps. These output maps will be used in the management of Hoburg's bank and its environment.

HYDROGRAPHIC SURVEY

Bathymetric data or underwater topography maps are a fundamental property of oceans, seas or lakes and are considered to be an integral component in habitat mapping and marine spatial planning projects (Hell et al. 2012).A hydrographic survey using acoustic systems was conducted during autumn 2016 and spring and summer 2017 with the aim of obtaining an accurate model of the underwater topography of Hoburgs bank. Outputs from the hydrographic survey provided baseline knowledge of the geophysical condition of the study area.

This survey was one of the most extensive hydrographic mapping surveys done by Swedish Geological Survey (SGU) in the offshore waters of Sweden. It covered an area of approximately 1344 km^2 and produced high-resolution bathymetric maps of up to 0.5 m^2 in resolution together with other geophysical and geological data, which that took ~500 hours of ship time. The main survey equipment used for the hydrographic survey was an EM2040 multibeam echo sounder (MBES) (Kongsberg Marine AS, Kongsberg) that was installed on the hull of R/V *Ocean Surveyor*.

The aim of the hydrographic survey was to collect high-quality bathymetry data that could not only be used for marine habitat mapping but also to produce data of high enough accuracy that it could be used in the production of navigational charts. Thus, this survey was conducted following the stringent international and Swedish standards for hydrographic surveying. This part of the report details the methods and standards used in the hydrographic survey of Hoburgs bank.

Implementation

Pre-survey planning

The project survey area was divided into 34 smaller areas (maximum cover of up to 3 km x 12 km) referred to as line blocks to help facilitate the survey and data processing (Fig. 1). Three specific line plans, aimed at 100% seabed coverage, were designed for each of the blocks to take into account the different oceanographic conditions. The three plans were designed using 5.5×, 6.3× and 6.8× the mean water depth[1] (including 20% overlap between adjacent lines). Generally, the 5.5× water depth line plan was used in the summer when oceanographic conditions weren't as stable, and 6.3× or 6.8× were used in the spring (2017) and autumn (2016). In some conditions, e.g. rough terrain, the lines were shifted to get an even more conservative line spacing. The survey lines were run at survey speeds between 7 to 9 knots. Swath widths used ranged from 75 to 220 m and were kept constant along each line (using a distance range limit in the acquisition software Seafloor Information System (SIS) (Kongberg Maritime AS, 2013). SIS changed the angle automatically with depth resulting in a max angle range from 65 ° to 85°. This resulted in a mean overlap between adjacent lines of 15% to 35%.

Hydrographic system used

Equipment used for the hydrographic survey of Hoburgs bank are listed in Table 1 and their location in the survey vessel is shown in Figure 2. The main system used was the dual swath Simrad EM2040D multibeam echosournder (MBES) from Kongsberg. It is an ultra wide-swath MBES that has a maximum angle of 200° and can cover up to 6 times the water depth. Auxiliary sensors include a Seapath 330 with MRU5, MK11 and IMU sensors that are used for inertial/GNSS positioning and measure ship heading, attitude and position.

[1] Theoretical mean, max and min depth for each survey line was calculated in ArcGIS using 500 m depth grid data from the Swedish Maritime Administration (Baltic sea bathymetry database, downloaded 2016)

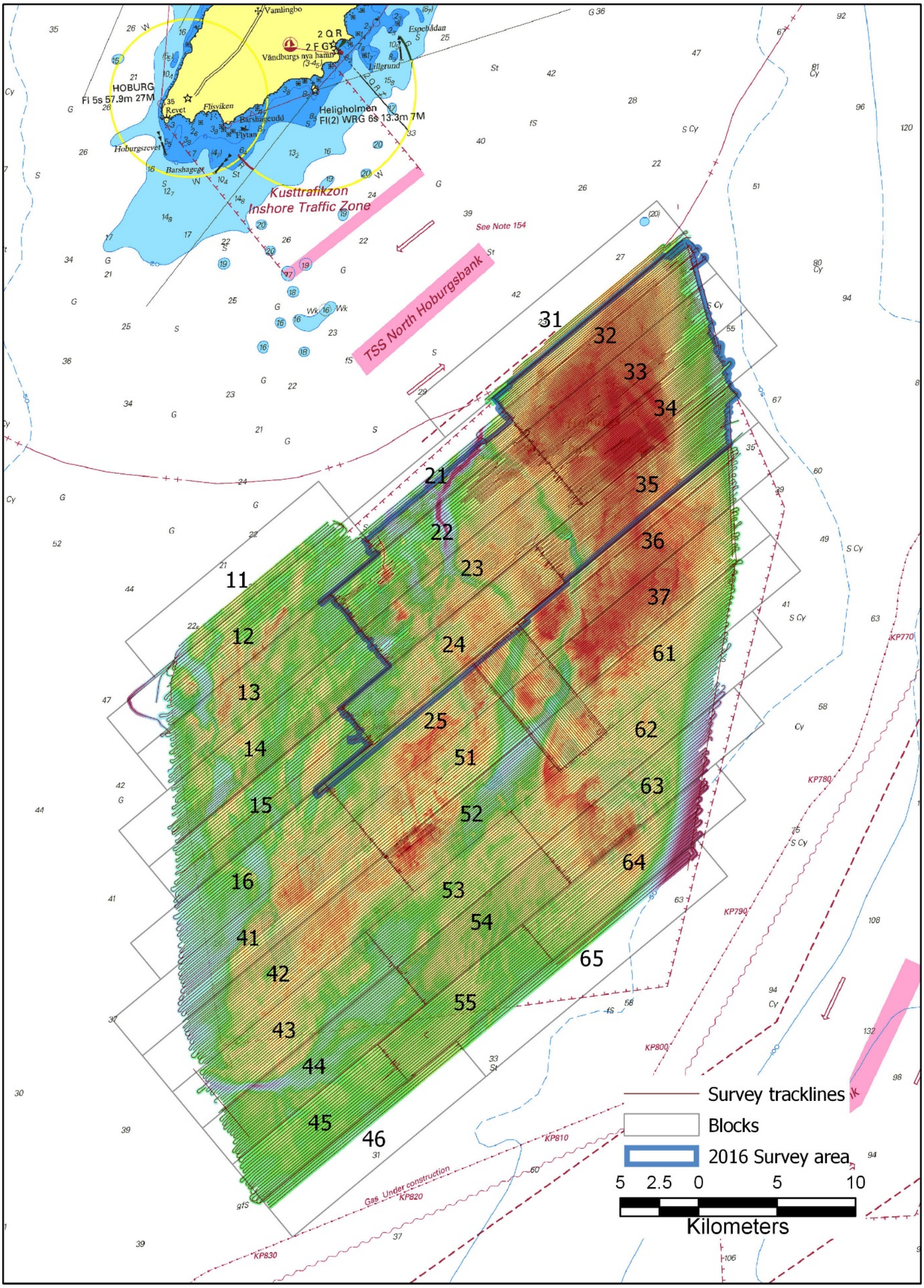


Figure 1. Hoburgs bank project survey area overlain on the nautical chart. The survey area was divided into numbered blocks (thin gray rectangles). Thin red lines represent the actual ship track lines that were followed to produce the bathymetric data (i.e., colored background).

Table 1. R/V *Ocean Surveyor* hydrographic survey equipment's used.

Sensor	Type	Description
MBES	Simrad EM2040D	MBES system used to measure the depth. It has a frequency range of 200-400 kHz and have an output sample rate up to 60 kHz. It has beamwidth of 0.5° for transmit and 1° for receive. It has angular coverage of up to 200° allowing it to cover $5^1/_2$ times the water depth.
MRU	Seapath 330 with MRU5 MK11 IMU sensor	Integrated inertial/GNSS product that provides up to 0.008° RMS roll and pitch accuracy and 2 cm heave accuracy (using PFreeHeave® algorithms)
SVP	(CTD) MIDAS SVX2	Collects vertical sound velocity, temperature and salinity profiles, as well as manual reading of secci-depth
SVP	Valeport mini SVS	Measures the surface sound velocity and water temperature – installed both in the hull to measure surface sound speed and in the MVP

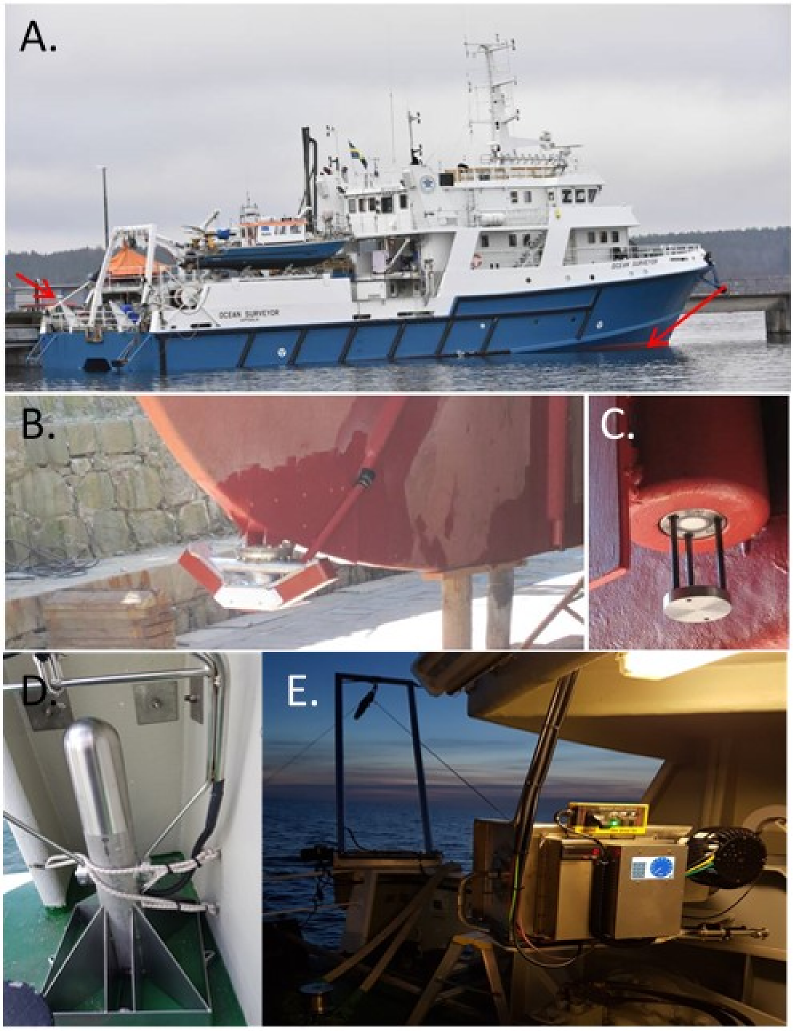


Figure 2. A. Side view of the catamaran R/*V Ocean Surveyor* showing where the EM2040 and MVP profiler are mounted (red arrow). **B.** Image of the installed Kongsberg EM 2040D when R/V *Ocean Surveyor* was drydocked. The Transmitter (Tx) mounted downwards in the longitudinal direction. Starboard receiver (Rx) at angle 40 degrees from horizontal and port Rx oriented 30 degrees from horizontal. **C.** The Valeport mini SVS installed besides the EM2040, to measure the surface sound velocity. **D.** Image of the newly installed MVP with mini SVS that is used to measure the SVP. **E.** The winch system that is mounted on the stern, portside of the ship. Photos: SGU.

Acquisition Settings

A frequency setting of 300 kHz in a CW pulse mode was used by the EM2040 MBES system to ensure the optimal acquisition of both bathymetry and backscatter data. During 2016 the CW pulse was forced to medium pulse length to optimize the backscatter collection, for the 2017 survey the setting was changed to auto mode (resulting in a shorter CW length). After initial testing, results showed slightly higher resolution in the backscatter data that overall were comparable to the results obtained from the 2016 survey. Equidistant beam spacing was used. The swath width settings were set to be limited by distance and not by angle.

Sound Velocity Profiles

Sound velocity profiles (SVP) were collected using the in-house developed, Moving Vessel Profiler (MVP) that was mounted on the stern and port side of the ship. The MVP uses a MiniSVP sound velocity probe (Fig. 2D & 2E) to take vertical profiles of sound speed and temperature. SVPs were taken as often as possible (at least once every fifteen minutes) or when a large difference between the surface sound velocity and the SVP was detected, which was often due to the highly dynamic oceanographic conditions in the area. More than 2,200 SVPs were taken on Hoburgs bank for the 2017 survey season using the MVP as compared with 355 SVPs taken using the manual CTD in 2016 (Fig. 3). The higher number of SVP casts in 2017 allowed a faster and more economical use of survey time that resulted not only in higher data quality but also allowed a larger area to be covered during the survey. The SVP data were uploaded to SIS for use in the gridding of the acquired bathymetric data. A collated/stacked version of all collected SVPs was created using in-house scripts for use in data processing.

SVPs were also collected using CTD cast (MIDAS SVX) once a day in the morning together with other oceanographic parameters namely: secchi disk depth, conductivity and temperature data. This was also used to check the SVP values obtained by the MVP, a total of 30 CTD casts were taken 2017. Salinity values used in the SIS absorption coefficient calculations were taken from the daily CTD casts. A salinity value consisted of the average salinity recorded by the CTD, which was then used as the Salinity absorption coefficient value. In some cases, the CTD cast was not available so the mean salinity value from the previous day was used. However, salinity only varied between 7 and 8 ppt.

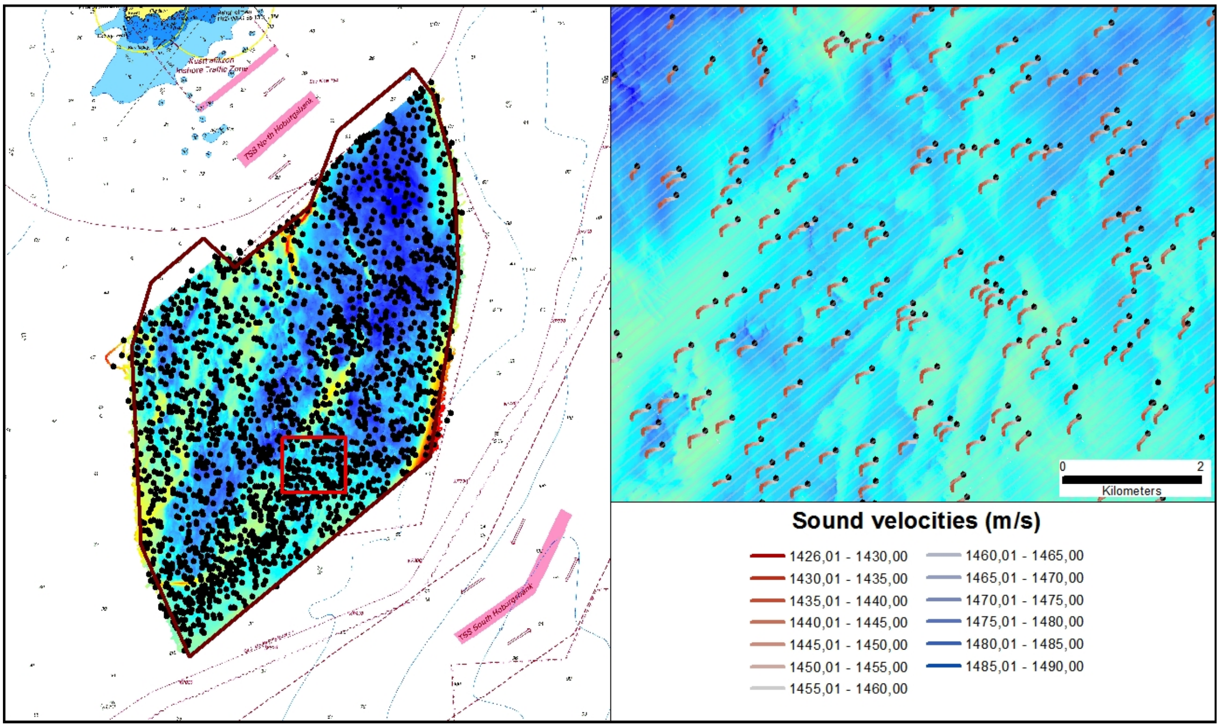


Figure 3. Left image shows a map of plotted MVP casts (black dots) from Hoburgs bank survey area. samples taken throughout the survey area. Image on the right is an enlarged portion showing variable hydrodynamic conditions as shown by the shape of the sound velocity profile.

Positioning systems

Horizontal Positioning system

A Kongsberg Seapath 330 GNSS RTK system was used to measure the heading, attitude and positioning information (Fig. 3). Positioning was done using the Global Navigation Satellite System (GNSS)-RTK mode. RTK corrections transmitted by SWEPOS system were received (through an Iridium satellite internet connection) by the Seapath 330 and used for real-time corrections of the vessel's position. All survey data was referenced to WGS84-UTM 33 datum which had similar parameters as the RH2000 (Swen08 geoidmodel).

Vertical reference system

The tide data used in the survey was derived from the RTK-corrected GPS height data to produce the GPS tide in RH2000 using *Caris Hips and Sips* software (v 10.2.3). GPS tide was manually edited, and then low pass filtered to remove artifacts from RTK dropouts and wave motion. No tide stations were established during the survey. Instead, the survey utilized tide level data from the Swedish Meteorological and Hydrological Institute (SMHI) tide station in Visby and Norra Öland for comparison with GPS tide acquired by the RTK system.

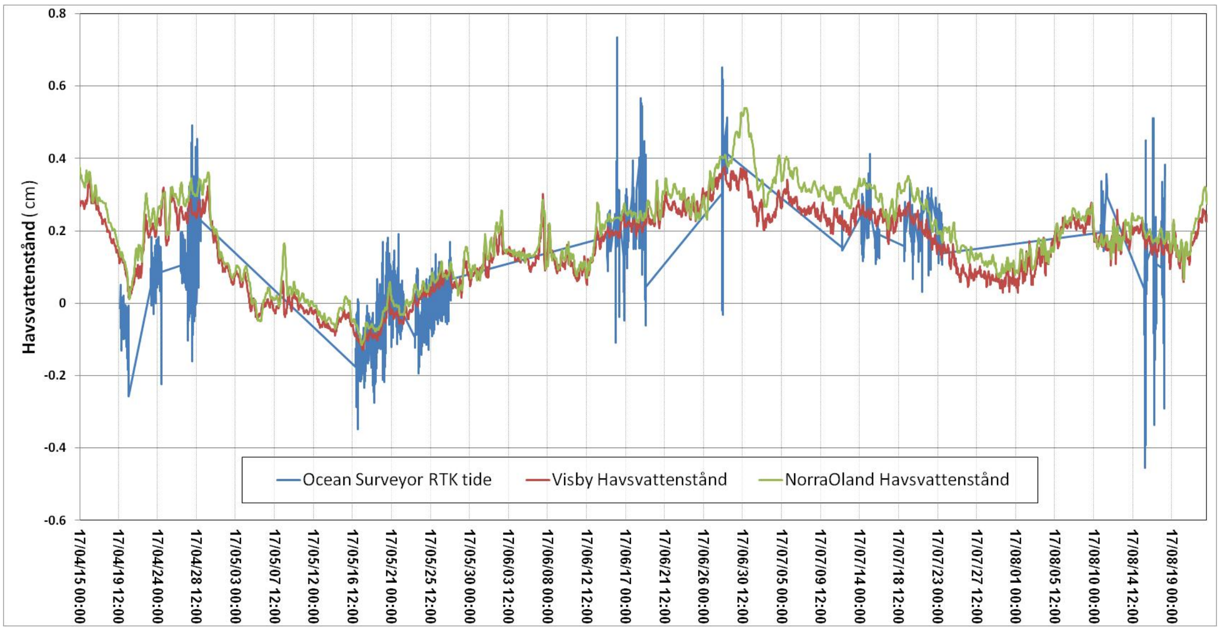


Figure 4. Comparison of the RTK-derived tide (blue line) with the tide level readings from Visby and Norra Öland tide stations (red and green line) using RH2000 datum. RTK-derived tide values were down-sampled and average smoothing applied and included only on dates when actual surveys were conducted.

Results from the comparison show that the GPS tide followed the same general pattern when compared with tide readings from the Visby and Norra Öland tide station (see Fig. 4). However, some noticeable differences in the magnitude of fluctuations were observed from the GPS tide. This difference was likely due to dynamic draft shifts that the low pass filter for wave movement (already accounted for by the MRU) did not, and should not, capture. Reasons for the dynamic draft shifts could be a combination of survey and wind direction, speed and fuel levels. Static and dynamic draft values were not incorporated into the sounding measurements since we used the GPS height directly. The draft data have been collected, so it is possible to calculate an alternative tide solution based on a model from the tide stations. The static and dynamic draft tables are found in Appendix A and B.

Work Chronology

The survey work was undertaken during the autumn of 2016 and spring and summer of 2017 following a two-week deployment and two-week rest schedule. This resulted in a total of 10 weeks of survey time for 2017 that consisted of ~240 hours of actual hydrographic surveying with the remaining hours used for transit, visual, geological and physical sampling and observations, calibration and weather standby.

The amount of survey time allotted for 2017 was almost the same as 2016 but the area covered by the 2017 survey was almost three times as large compared to the survey in 2016. The main reason for the greater coverage in 2017 was due to the installation of a Moving Vessel Profiler (MVP) that allowed continual collection of SVPs (at least once every 10–20 min) without stopping the vessel. This allowed a more liberal line spacing during the survey while maintaining higher quality data, as compared with the 2016 season where SVP casts were fewer (every ~30–40 min). Other factors that contributed to greater coverage in 2017 were better weather conditions and the fact that a large portion of the survey lines for 2017 were conducted in deeper areas of Hoburgs bank.

System Calibration

Patch Test

A patch test was undertaken on the EM2040D in calm weather in an area with a sloping gradient in Mälaren at the beginning of the survey season to calibrate the system errors. A series of parallel and perpendicular survey lines were run to calibrate the roll timing, pitch and heading settings as well as to quantify the system accuracy (Fig. 5). The patch test was conducted following the procedures outlined in the Kongsberg Operations Manual for Dual-head systems (Kongberg Maritime AS 2012).

The order in which the system biases were determined was as follows; roll, timing, pitch and heading. The calibration started by checking the roll offset bias. This was done by conducting three parallel survey lines on a flat seafloor in opposing directions with 100% overlap (Fig. 6A). Then, another three parallel but opposite lines on the sloping part of the seafloor were run on top of each other with the third line at half the normal survey speed to calibrate the timing and pitch (Fig. 6B). Finally, two parallel lines travelling in the same direction across the slope were run to check for the Heading settings (Fig. 6C). The offset values were then determined and applied from SIS using the calibration window. The same offset values were also applied to the Vessel File Settings in the Caris HIPS and SIPS software (v. 10.3) that was later used for data processing.

A second patch test was also undertaken at Oxelösund following the same procedures mentioned above just prior to the start of the survey to verify the result of the previous patch test. This resulted in a slight change to roll and pitch offsets. A summary of all changes to the previous settings are shown in Table 2.

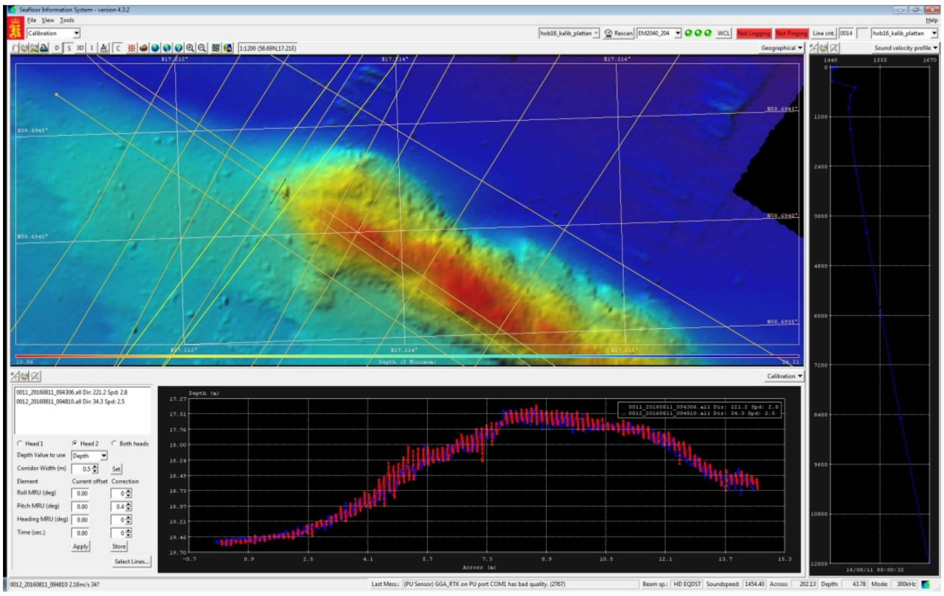


Figure 5. Screen grab of the SIS interface for the patch test.

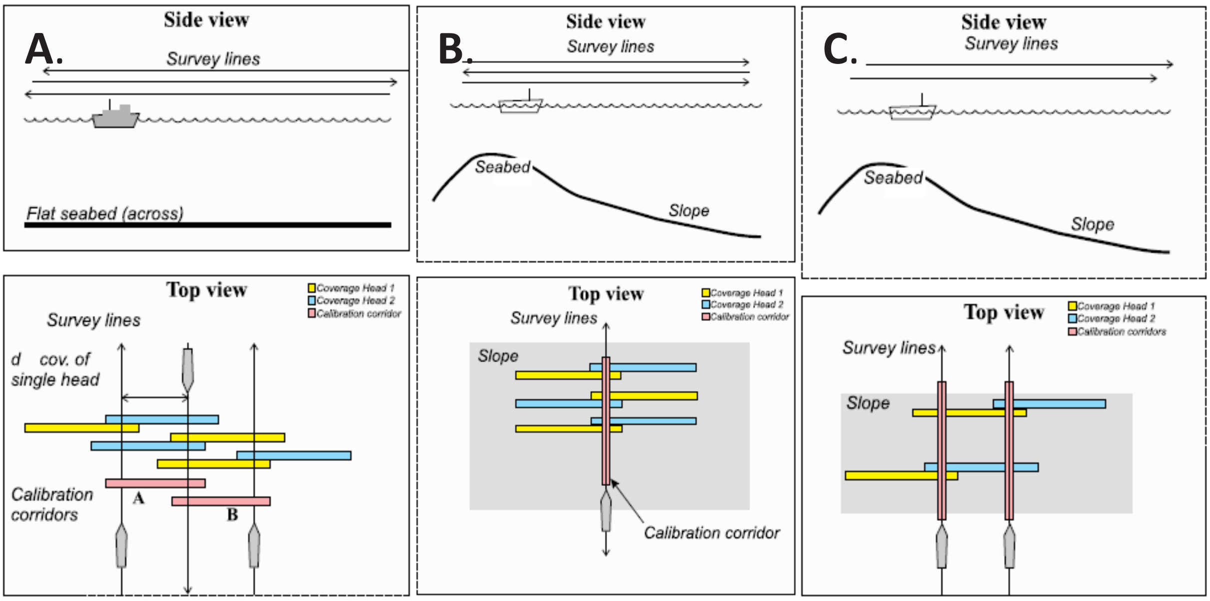


Figure 6. Survey lines run to check for system offsets and calibration corridors in SIS acquisition software. **A** – lines run to check for roll; **B** – lines run to check for pitch and timing; **C** – lines run to check for heading. Images taken from SIS reference manual (Kongberg Maritime AS 2013).

Table 2. Changes applied to the system offset of the R/V *Ocean Surveyor*'s EM2040 MBES system for the 2017 survey.

Time	Area	System	Roll	Pitch	Heading	Timing
2017, March 28-29	Mälaren	RX Transducer port	29.54 (-0.2)	1.46 (-0.3)	358 (0.6)	NC
		RX Transducer starboard	-40.23 (-0.12)	2.21 (-0.3)	0.71 (-0.04)	NC
		TX Transducer	-0.32 (0.04)	1.8 (-0.3)	359.28 (0.2)	NC

*Values in parenthesis were the amount of change from 2016 survey settings

Comparison against a known depth

After the patch tests had been completed, Multiple test lines were again run on Ostplattan in Oxelösund, before and after the survey season to compare our soundings against a known depth (Fig. 7). Ostplattan, established by Sjöfartsverket (SjöV) in 2014, is a known test area for calibration of bathymetric acoustic systems because of its known depth and position. The lines were run parallel and perpendicular to each other at varying distances from Ostplattan to determine the performance of R/V *Ocean Surveyors* EM2040 MBES system. The system settings (i.e., survey speed, frequency, pulse length, etc.) used during the Hoburgs bank survey were also the same settings used in the Ostplattan test survey except for lower swath width settings.

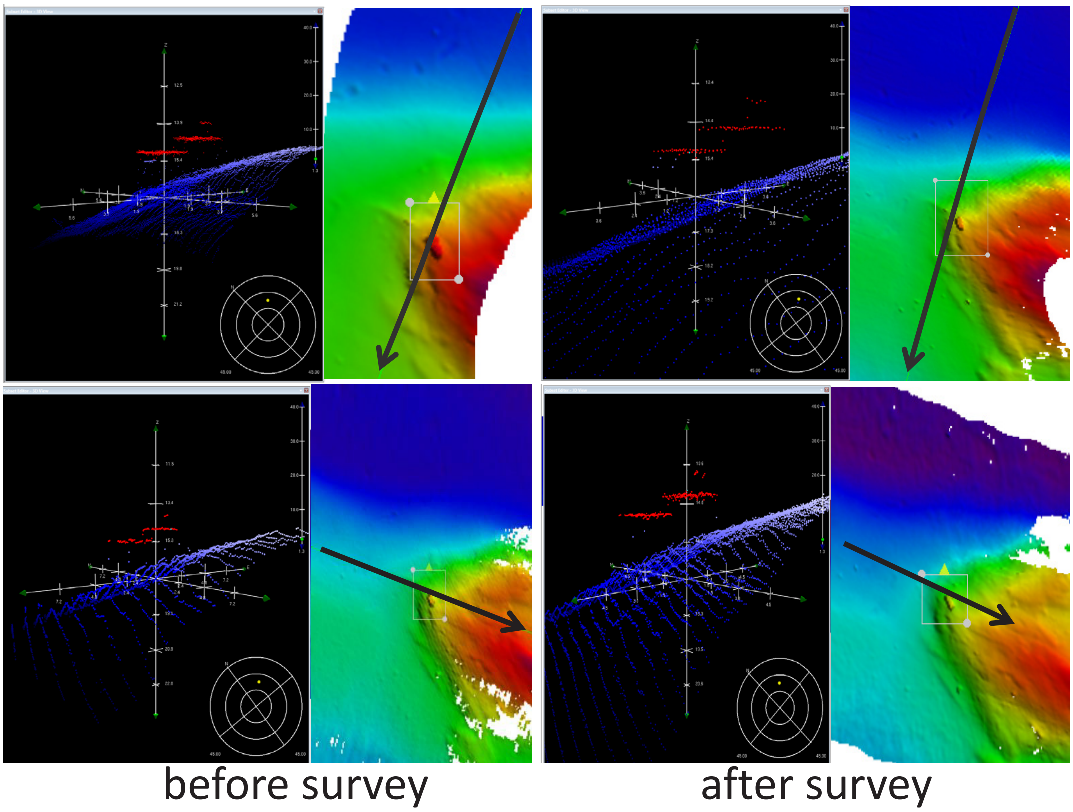


Figure 7. Two perpendicular survey lines over Plattan taken before and after the 2017 survey season together with the 3D view of the pings on the Plattan. Red dots on the 3D view were the soundings used to compare the depths with the reference SjöV value in Figure 5.

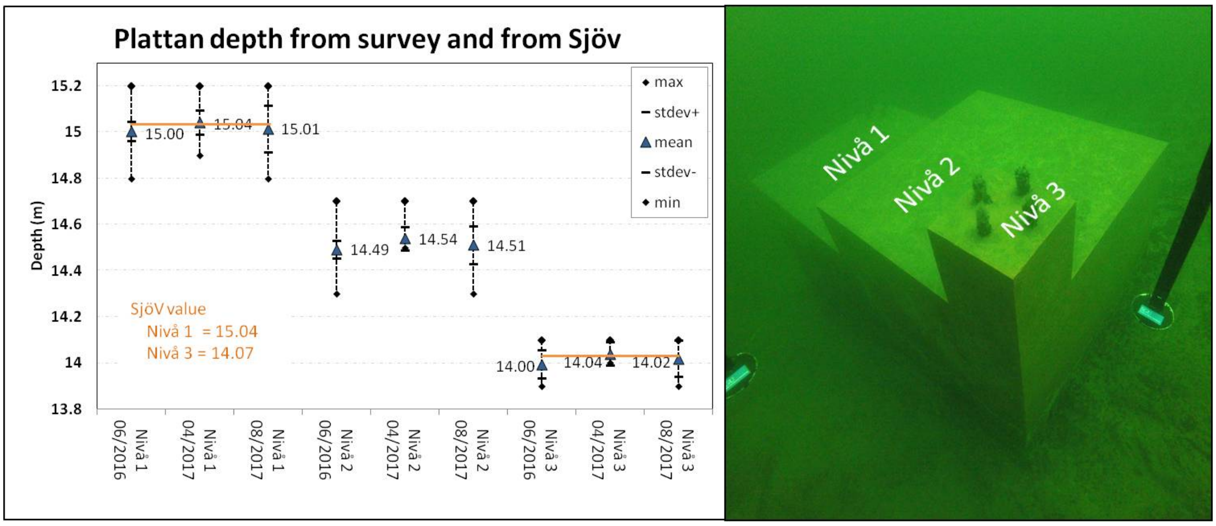


Figure 8. Comparison of depth values from the three different levels of Plattan with the reference depth from SjöV (orange line) from two calibration periods (before and after the survey) in 2017 and from 2016. Average values are shown in the chart.

The mean depth sounding from the test survey differed by approximately ±0.05 m when compared with the reference depth values of Ostplattan. We also found a small difference in our depth measurements collected in autumn 2016, and before and after the survey in 2017 (Fig. 8). These differences (maximum 5 cm) were within the error margins of the survey system.

Object detection capability

The object detection capability of the EM2040 system was also evaluated by counting the number of detections both on top and in total on the object above the seafloor at various beam angles. This was done by computing the number of pings that hit Ostplattan from lines run at three different distances from the Ostplattan; nadir beams –line run directly above the Ostplattan, middle beams line run ~10 m from the Ostplattan, and outer beams – line run ~35 m away from Ostplattan (Fig. 9).

Results from the analysis of the object detection capability of OS EM2040 showed that the three Ostplattan levels (with surface area equal to; level 1 = 1 m^2; level 2 = 0.25 m^2; level 3 = 0.0625 m^2) were detectable based on the number of hits on the target in both the nadir and mid beams. However, the outer beams (i.e., beams from ~45° and above) managed to detect only level 1 and level 2 and only had a few hits on level 3. The position and placement of Ostplattan on a sloping terrain may also have contributed to the difficulty in the detection of Ostplattan surface levels by the outer beams. The number of pings that hit the target on the different levels of Ostplattan exceeded the requirements set by IHO S-44 System Detection Capabilities, (International Hydrographic Organization, 2011).

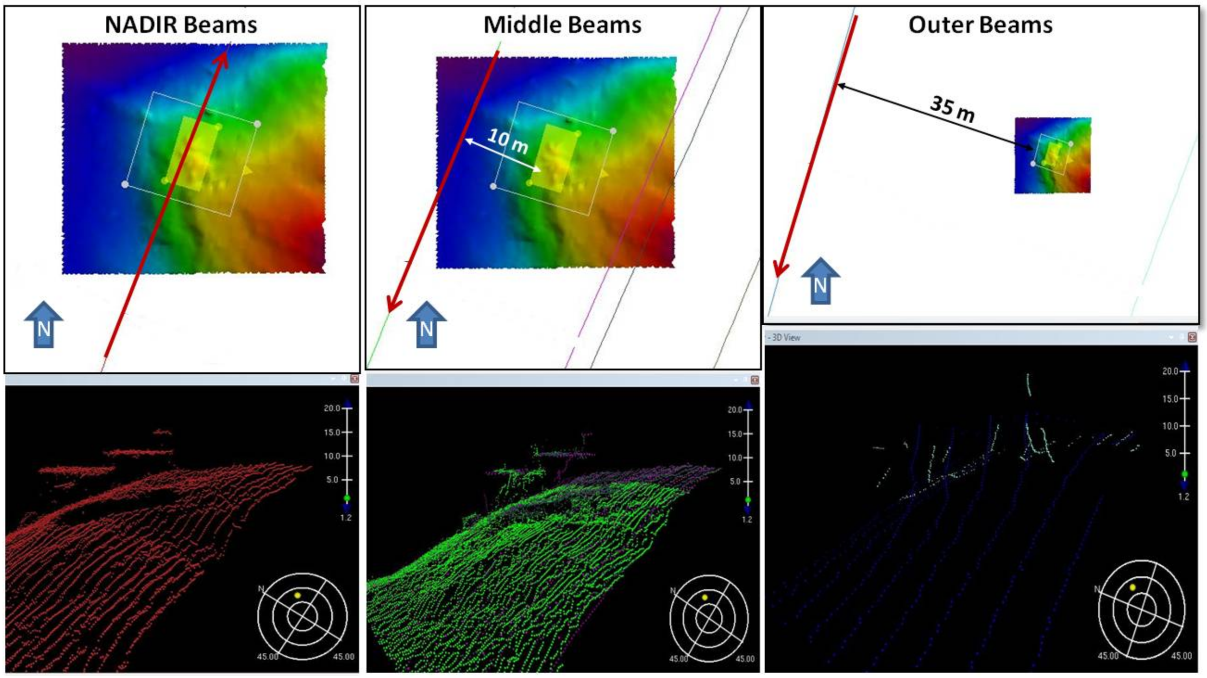


Figure 9. Survey lines over Ostplattan at varying distance to determine the object detection capability of the OS EM2040 at the Nadir, middle and outer beams. Survey lines are shown as red arrows and show the direction of the boat during the survey.

Bathymetric data processing

Data Import

MBES data were acquired using *SIS* to record the raw bathymetry (*.all*) and water column data files (*.wcd*) collected by the EM2040D. The raw bathymetric data (*.all*) files were then imported into *Caris HIPS and SIPS* (10.3.1) for bathymetric data processing using the Caris semi-automated batch import processing tool.

After data conversion, GPS tide was computed (converting to datum model SWEN08_RH2000) and exported as a separate file. This was manually cleaned, filtered and smoothed to remove RTK dropouts and remove wave motion already compensated for by the MRU. The filtered GPS tide was then applied back to the lines to remove some of the artifacts. The lines were then merged and the Total Propagated Uncertainty (TPU) for each sounding was calculated.

Data cleaning and filtering

A data processing workflow was followed to standardize the bathymetric data processing of the project (Fig. 10). Data cleaning and filtering were done using Caris HIPS and SIPS. A bathymetric grid surface was created using the CUBE algorithm (source) to examine the data. Blunders and outliers were rejected, and major systematic errors were identified for further data cleaning and filtering.

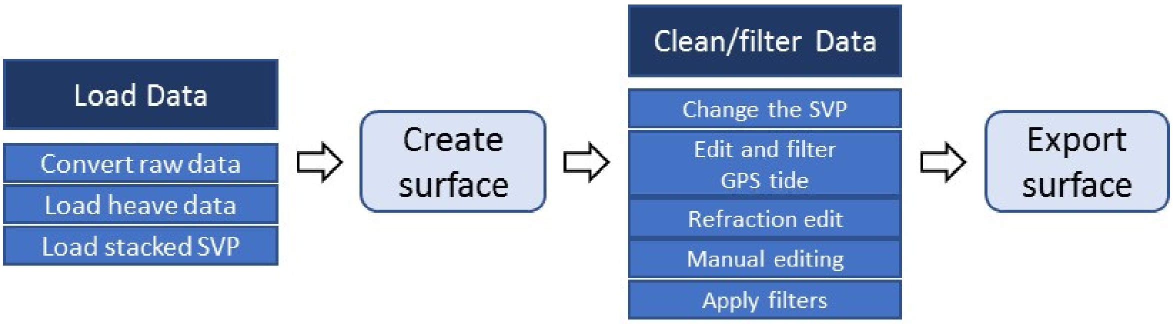


Figure 10. Multibeam data processing workflow using Caris HIPS and SIPS (v. 10.3.1)

The process of cleaning the bathymetry data involved the following steps; 1) manual editing of CUBE hypotheses, 2) replacing SVPs for lines with incorrect SVP values, 3) filtering out spikes in the various auxiliary sensors and 4) correcting lines with poor sound velocity profiles using Refraction Editor. After the cleaning procedures, a surface filter was used on the CUBE surface to reject/flag soundings outside a confidence interval of 95.44% to reduce the amount of manual editing required to produce a clean sounding set. Finally, other artifacts not removed by the aforementioned methods were manually removed using different editor functions (i.e., subset, attitude, swath) in Caris HIPS and SIPs.

After all the cleaning was done, the final surface was exported to a 16-bit floating geotiff format as the final bathymetry product for the project.

Accuracy Analysis

Computation for uncertainty

The error budget of the R/V *Ocean Surveyor* hydrographic survey system was calculated using the template provided by SjöV. The values obtained from the calculation were used to compare the system to the standards specified in the joint Finnish-Swedish implementation of the IHO S-44 standards (i.e., FSIS-44). Results showed that the survey system used in the project met the Special Order requirement for Total Vertical Uncertainty (TVU) and the Order 1a requirement for Total Horizontal Uncertainty (THU) (refer to Table below).

Table 3. Total Propagated Uncertainty estimates of MBES system based on the Swedish-Finish FSIS-44 hydrographic standards.

Ärendenummer:	HOB16	**Ärendenamn:**	SGU - Hoburgs Bank Project
Fartyg:	Fritt fartyg	**MB-system:**	Fritt MB-lod
Positionssystem:	GNSS RTK	**Djup under svängare:**	20 m
RPH-givare:	Seatex MRU-5	**Stråkbredd:**	150°
Gyrokompass:	Fritt gyro	**Höjdbestämning:**	RTK

Anteckning: Hoburgs Bank geophysical survey 2016-2017

Horisontellt

Felkälla		**Uppskattat fel (+/-)**	**Enhet**	**Totalfel (+/-) meter** Mittstråle Långskepps	Mittstråle Tvärskepps	Ytterståle Långskepps	Ytterståle Tvärskepps	**FSIS-44** Order	Meter
Positionssystem		0.15	meter	0.15	0.15	0.15	0.15	Exclusive	2
Gyrokompass		0.04	grader	N/A	N/A	0.027	N/A	Special	2
Rollfel från RPH-givare		0.04	grader	N/A	0.014	N/A	0.014	1a	6
Pitchfel från RPH-givare		0.04	grader	0.014	N/A	0.054	N/A	2	22
Fel i LTH-koordinaterna		0.02	meter	0.02	0.02	0.02	0.02		
Storlek på footprints	1°x0.5°	1	grader	0.175	0.087	0.674	2.608		
Ljudhastighet		Beräknat	meter	N/A	N/A	N/A	0.261		
Position antenn-R.P	27 m	Beräknat	meter	0.019	N/A	N/A	0.019		
			Totalfel:	**0.23**	**0.18**	**0.69**	**2.63**		

Vertikalt

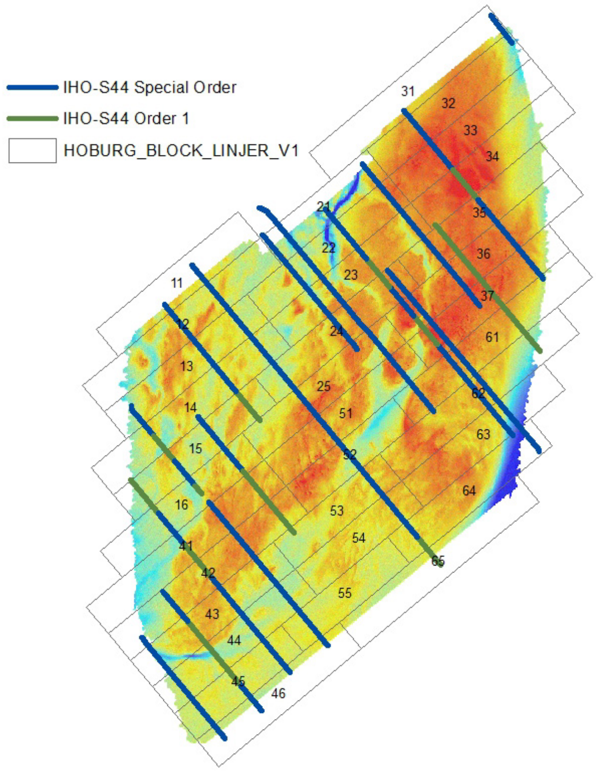
Felkälla		**Uppskattat fel (+/-)**	**Enhet**	**Totalfel (+/- meter)** Mittstråle	Ytterstråle	**FSIS-44** Order	Meter
Upplösning i syst. + Barcheck res.		0.05	meter	0.050	0.050	Exclusive	0.17
Ljudhastighet	0.02 m/s	Beräknat	m/s	0.001	0.002	Special	0.29
Rollfel från RPH-givare		0.04	grader	N/A	0.052	1a	0.56
Heavefel från RPH-givare		0.1	meter	0.100	0.100	2	1.10
Fel i LTH-koordinaterna		0.02	meter	0.020	0.020		
Vattenstånd		0	meter	0.000	0.000		
Squat / sättning		0	meter	0.000	0.000		
Djupgåendeavläsning		0	meter	0.000	0.000		
Positionssytem		0.05	meter	0.050	0.050		
Anslutningsfel, GNSS-höjd - ref.plan		0.05	meter	0.050	0.050		
			Totalfel:	**0.13**	**0.14**		

Crossline analysis

A series of systematic crosslines were conducted throughout the survey area to verify and evaluate the consistency of the survey. The crosslines totaled ~27 kilometers that intersected the main survey lines of each of the blocks (Fig. 11). The crosslines were processed using the same methods and standards as the main lines before statistical comparison, which included loading the SVPs, filtering the tide files and rejecting all erroneous soundings. The Crossline© analysis tool from Fledermaus (v 7.7.6) was used to determine the error limits between the final reference surface and crosslines and then compared to the accuracy requirements of the IHO S-44 standards (International Hydrographic Organization 2011). At least 300 million soundings from the crossline data were used and compared to the final bathymetric grid. The results showed that the entire survey fell within the Special-Order requirement of IHO, although some of the crosslines soundings did not meet the accuracy standards specified in IHO-Special Order (see Fig. 11 and Table 4). For crosslines that did not meet the IHO-Special Order standard it was not verified whether this was due to the quality of the crossline or the survey lines. However, some crosslines where run with the prevailing southeast swell orthogonal to the survey direction, which lowered the quality of the crosslines compared with lines in the more favorable main survey direction.

A comparison of the 2016 and 2017 datasets using the 2017 crosslines showed a consistent and significant difference in the depth values between the two resulting surfaces of 14 cm. This could be a result of the change in the processing method employed for 2017. To fix this, a static offset value (14 cm) was added to the 2016 data to produce the combined (2016 and 2017) final bathymetric surface. Furthermore, many data points in 2016 had been unnecessarily deleted, which were subsequently re-introduced. Because of these issues, we recommend that further data processing/cleaning is necessary if the survey results are to be used for navigation purposes, especially in areas where the new minimum depth is deeper compared with the existing charts.

The bathymetric dataset collected from our 2017 survey, based on the crossline analysis, was considered to fulfill the IHO S-44 Special Order specifications. However, based on the Total Propagated Uncertainty (TPU) estimates of the survey system, the survey fulfills the FSIS-44 Order 1 specifications.



Hoburg Bank Survey Crossline Analysis	
Number of Soundings used in the Comparison	325290717
Data Mean	-25.194427
Reference Mean	-25.193597
Mean	-0.00083
Median	0.002934
Std. Deviation	0.14254
Data Z - Range	-60.27 to -13.39
Ref. Z - Range	-41.04 to -27.39
Diff Z - Range	-10.49 to 6.89
Mean + 2*stddev	0.285915
Median + 2*stddev	0.288019
Ord 1 Error Limit	0.597718
Ord 2 Error Limit	1.155753
Special Order Error Limit	0.313373
Custom Error Limit	0.412028
Ord 1 P-Statistic	0.002205
Ord 2 P-Statistic	0.00007
Special Order P-Statistic	0.038637
Custom P-Statistic	0.014495
num rejected Order 1	717418
num rejected Order 2	22623
num rejected Special Order	12568114
num rejected Custom	4715154
Order 1 Survey status	ACCEPTED
Order 2 Survey status	ACCEPTED
Special Order Survey status	ACCEPTED

Figure 11 and Table 4. Results from crossline analysis showing the lines that meet the required standards for IHO S-44 Special Order (blue) and the lines that meet the Order 1 (green) specifications. In the case where the crosslines did not meet the IHO-Special-Order standard it is not verified if this was due to the quality of the crossline or the survey lines. Table on the right is the output from the Crossline© analysis software (Fledermaus 7.7.6) for the Hoburgs bank survey.

Comparison of survey result with Navigational Chart

The results from the hydrographic survey were compared with the latest versions or editions of the largest scale navigational chart and raster charts that cover the project area. Critical soundings from our data was extracted in Caris and plotted over the existing navigational chart in the area (see Fig. 12). Results show that our depth measurements generally agreed with the existing nautical chart of the area (i.e. SE71). However, the positions of the chart's critical soundings, to some extent, differ with our data. An obvious difference between our data and SE71 nautical chart is the shape and extent of the 20 m contour line particularly in the mid and southeastern part of the survey area (see Fig. 12). Our survey result shows that the shoal area is significantly larger than that depicted in the chart. Another shoal feature that we found that is not indicated in the Chart is the presence of shallow shoal (~12 m depth) located in the mid portion of the survey area which has, because of its shallow depth, the potential to impact the navigation capabilities of large vessels in the area.

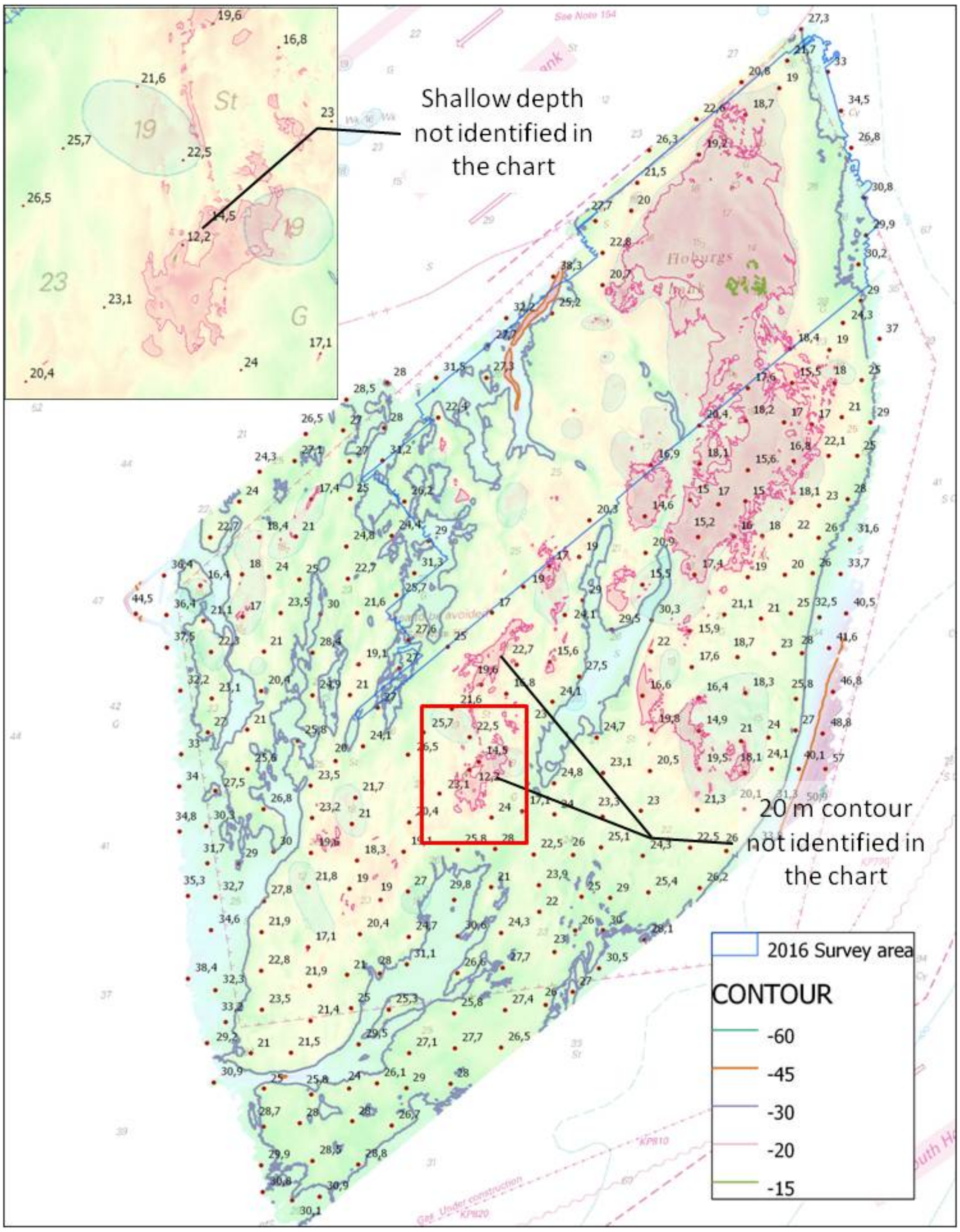


Figure 12. Semi-transparent bathymetric grid from this survey together with contour lines and critical depths (i.e., numbers with points) overlaid on the existing navigation chart (#SE71). Inset is an enlarged image showing difference in extent of the 20 m contour line from the chart (in light blue) and in our data (light red. Also shown here is the shallow depth (~12 m) not identified on the chart.

Issues encountered

Occasional Real-Time Kinematic (RTK) signal dropouts were experienced during the whole survey, which had some limited effect to the accuracy of the position measurements. In addition, computer hardware and software upgrades at the start of the survey season in 2017 resulted in problems with the data processing workflow that caused substantial delays to the project and limited the time available for manual cleaning. These problems were eventually addressed in the latter part of the survey

An incident occurred during the survey where the SIS acquisition settings were unintentionally altered, specifically the 3D scanning feature was turned on, which lead to systematic artifacts in the data. This artifact is described as a sudden small (~10 cm) change in the depth value across the beam at regular intervals in the survey line (Fig. 13). To fix this, we followed Kongsberg's recommendation and slightly changed the pitch and roll vessel file settings in Caris, which removed much of the artifact.

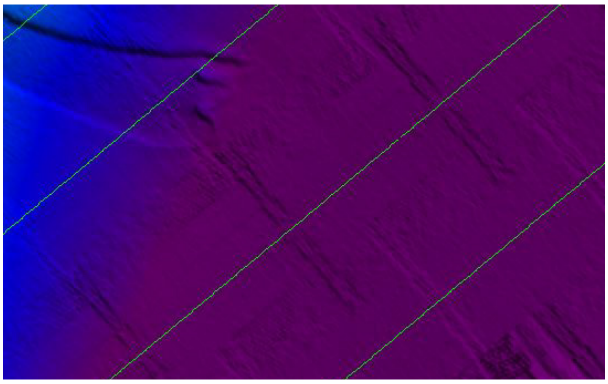


Figure 13. Small area of the survey showing the systematic artifacts that resulted when the 3D line scanning option in the SIS was unintentionally turned on (30x vertical exaggeration). This artifact was removed by applying a modified vessel file.

MBES – DERIVED DATASETS

Backscatter

Backscatter data are collected together with the depth using the same multibeam system. Backscatter is a measure of the intensity of the sound that is reflected back from the seafloor and thus provide information about the characteristics of the seafloor. Studies show that backscatter can be used as a proxy for different seafloor properties namely; the sea floor hardness, surficial sediment characteristics and sea floor roughness (Fonseca & Mayer 2007).

Three main steps were taken during the acquisition of the data (using SIS) to ensure that consistent and comparable backscatter data were collected across the whole survey area.

1. Absorption coefficient: The water column absorption coefficient used in SIS was calibrated daily using information from the CTD cast. Two different methods were used: (1) calculating the full absorption profile for 300 kHz based on the daily CTD cast using Sound Speed Manager, alternative (2) using a fixed mean salinity value in SIS based on the daily CTD cast, during this survey the two methods resulted in comparable results.
2. Settings: All data were collected in CW mode at 300 kHz, pulse length was kept constant (pulse length "medium" used 2016, while pulse length "short" was used 2017 to increase the resolution in shallow water. Some smaller differences were noticed)
3. QC: All survey modes were tested pre-survey and QC checked for backscatter differences. Daily processing of backscatter mosaics in FMGT geocoder ensured backscatter quality during survey.

The backscatter information from the raw MBES (.all) files was processed using the FMGT software (v 7.7.6) from QPS-Fledermaus. The backscatter mosaics were created from the beam time series data type of the EM2040 system (Fig. 14). FMGT uses the same standard geometric and radiometric corrections of Geocoder as described by (Fonseca & Calder, 2005). Most of the parameter in FMGT were kept at their default settings, not only because adequate backscatter mosaics were produced using these settings, but also to favor consistency of the backscatter mosaics from different surveys as shown by Ierodiaconou et al. 2018. A more detailed explanation of the various "default" settings implemented by FMGT is provided in aforementioned article.

The final backscatter mosaics were exported as floating-point geotiff grids at the following resolutions: 50 cm, 1 m and 2 m. Furthermore, backscatter derivative products were produced, namely statistical derivatives of the backscatter and sediment characterization derivatives, using Angular Range Analysis (ARA) from (Fonseca & Mayer 2007) in the same grid format as the mosaics. These sets of map products were used as spatial predictor-variables in the final habitat modelling process.

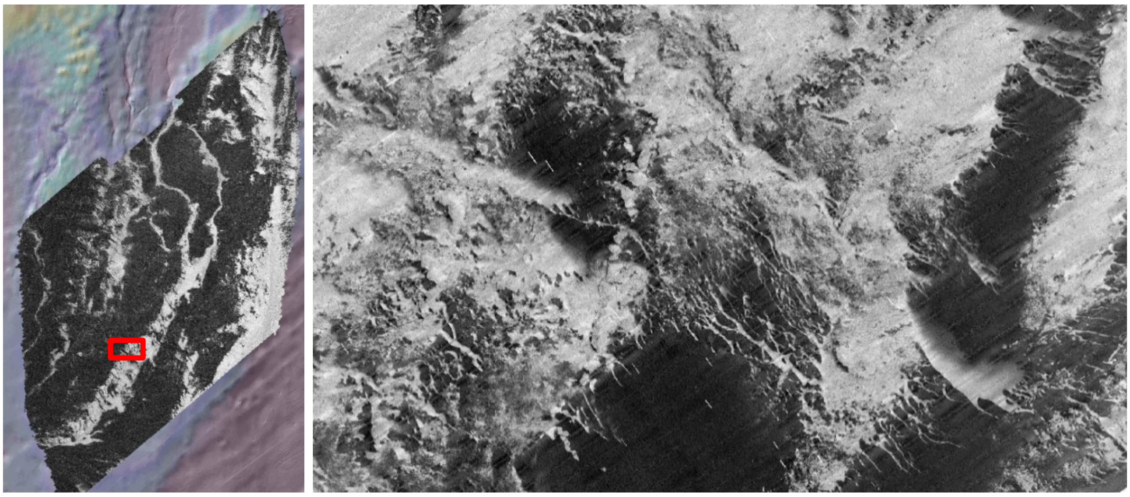


Figure 14. Backscatter mosaic (dark = high backscatter signal) extracted from raw files collected by the EM 2040 MBES system using FMGT software. The backscatter mosaics provide a general idea of the sediment characteristics of the survey area and was used in the habitat modelling.

Water column data

Water column data (e.g. a compressed version of the full backscatter record) was also recorded together with the bathymetry data by the EM2040 MBES system but in another file using another format. Objects within the water column (i.e., above the seafloor) scatter that emitted sound pulse from the MBES are recorded as a *.wcd* file, by the system. The measurement and analysis of this property in the water column is now an area of intense research because of its usefulness in other areas of marine science (Colbo et al. 2014).

In this project, the use of the water column data as a tool for fisheries assessment was undertaken in collaboration with SLU Aqua (Fig. 15). Initial analysis was done by SGU using QPS-FMMidwater software to investigate the potential of water column data from multibeam echosounders designed for hydrographic surveys and do some pilot testing of target detection capabilities, while SLU Aqua was further testing how fish targets and biomass in the water column could be identified (Fredriksson et al. 2017).

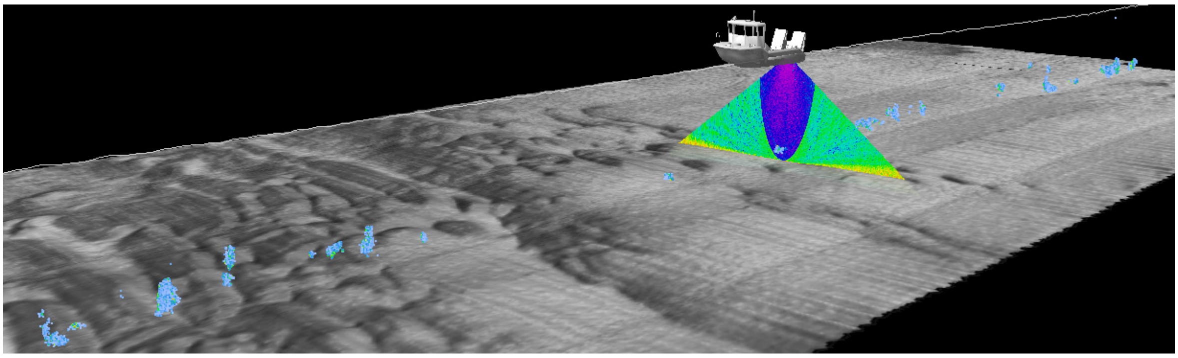


Figure 15. Image of processed water-column data that is believed to be schools of fish found along near the seafloor.

HYDROGRAPHIC SURVEY SUMMARY AND RECOMMENDATIONS

The main objective of the survey was for a 100% seafloor ensonification and that the quality of the sonar data was suitable for high resolution sediment and habitat mapping. Both of these objectives were achieved. The entire Hoburgs bank hydrographic survey produced a total of > 4 billion soundings that covered an area of 1400 km^2, with 340 and 1060 km^2 completed in 2016 and 2017 respectively. These soundings had depths that ranged between 10.45 and 63 m.

The data collected for the 2016–2017 Hoburgs bank geophysical survey produced systematic artifacts that were mainly due to the high beam width and minimal overlap in the main survey lines used during the survey, that were more apparent during the summer when a significant thermocline was present. This contributed to lower sounding accuracy but was necessary to suit time and funding constraints. Despite the lower accuracy, the hydrographic survey achieved and even surpassed its original objective of attaining the IHO S44 Order 1 standard. Finally, the hydrographic survey system installed on R/V *Ocean Surveyor* can attain an even higher survey standard if more conservative settings are used. This was not done in this survey not only because of limited resources but also because it was not deemed necessary for the original objective.

It is also noted that although the data in its current form may not be suitable for direct use in charting applications because limited effort was applied in the data processing, we still recommend that the gridded cube-surface from SGU be used as soon as possible to highlight or correct the areas identified as shallower than the current chart shows.

The backscatter information (i.e., mosaics and backscatter derivatives) produced from the survey will be used in the sediment and habitat predictive modelling part of the project. This was the first survey where we gave high importance to the quality of the backscatter data due to its role in the habitat modelling. Thus, there were times that settings were changed during the course of the survey in order to optimize both the quality of the bathymetry and backscatter data. This led to differences in the backscatter values from those collected during 2016 and 2017. Hopefully, the same problem will not happen again in future surveys especially with the added features of backscatter calibration of the existing system.

OTHER HYDRO-ACOUSTIC SYSTEMS

Sub-bottom profile data as well as hydro-acoustic fishery splitbeam sonar data was collected in conjunction with the hydrographic surveys.

Sub-bottom profiler (SBP)

Vertical profiles of the sediment layers were collected using an SBP 120 system (SBP, Kongsberg). SBPs send sound pulses that travel to the seafloor and the sediment layers at a speed that dependent on the density of the material. The difference in the time for this signal to return and recorded would indicate the characteristics (i.e., type and thickness) of the sediment layers below the seafloor. System settings used during the survey are the following: 10 ms, linear chirp pulse from 1,7 kHz – 5,3 kHz frequency and data was saved using the standard SEGY format. Data were processed in Meridata Processing Software (MDPS) (v5.2, Meridata, Lohja). A subset of the data (every 2nd to 4th line) was manually interpreted and digitized, and the resulting data interpolated to provide a model of estimated postglacial sand depth used as a predictor for modelling (Fig. 16).

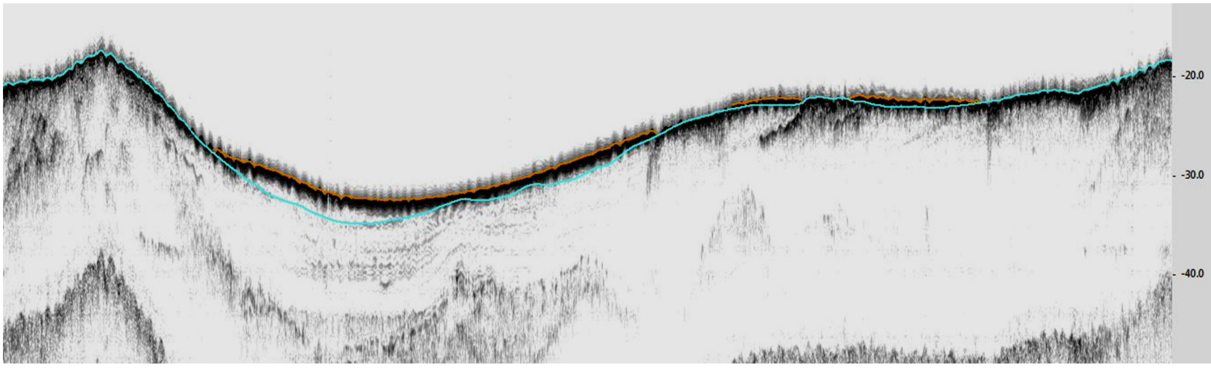


Figure 16. Interpreted SBP data as postglacial sand (using MDPS software) from a sediment profile in Hoburgs bank.

Split-beam echosounder

A Kongsberg EK60 70 kHz splitbeam echo-sounder was deployed (Fig. 17) half-way through the 2017 survey season to collect water column backscatter information that would be used for fish stock assessment. The system was pole-mounted on the starboard side of R/V *Ocean Surveyor* and positioned using RTK. The sonar was pre-calibrated by SLU Aqua at salinity ~5 ppt (salinity at Hoburgs bank varied between 7–8 ppt). The post-processing of the EK60 calibrated water column backscatter data was conducted by SLU Aqua.

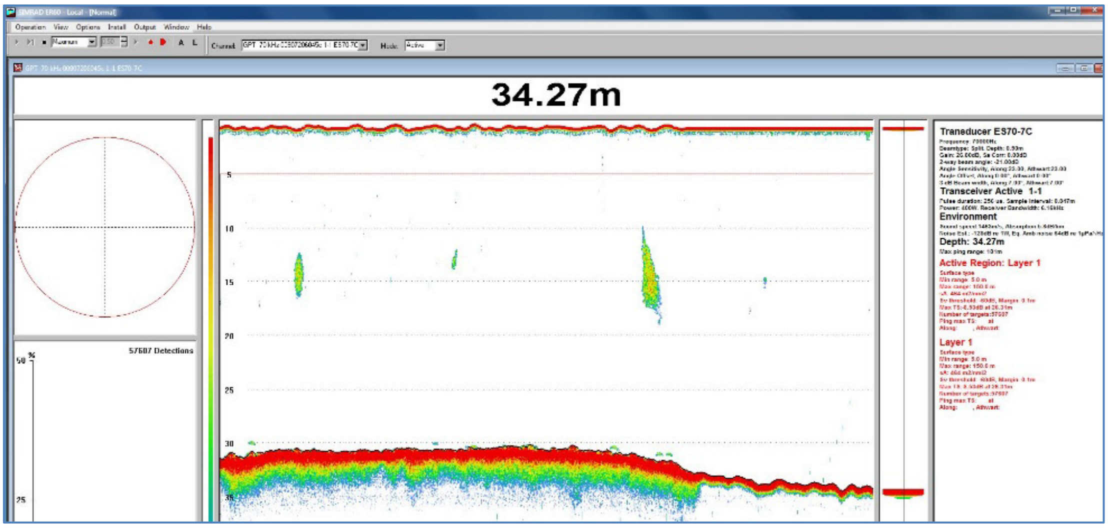


Figure 17. Splitbeam echo-sounder EK60 interface showing fish schools in the water column as well as individual targets close to the seabed on Hoburgs bank.

GROUND TRUTHING

Ground truthing is the collection of actual data that shows the conditions and parameters used to describe the seafloor environment. It is the source of "correct" in-situ information that is used in the interpretation and analysis of remotely sensed (multibeam) data and eventually, to validate the results of the modeled data. A well distributed ground truth dataset is required to properly classify the dataset to minimise bias. The ground truthing strategy on Hoburg's bank were driven by the following requirements.

1. To adequately sample all major benthic habitat types across the bank in a way suitable for mapping, modeling and monitoring.
2. To be able to plan the sampling in a standardized and repeatable fashion using the recently collected sonar data as a guide.

A widely practiced ground truthing strategy, the two-stage methodology as reported by Clements et al. 2010, is used in this project that can answer the requirements set forth in the project. The first stage is a dedicated hydro-acoustic remote-sensing phase that produces the acoustically derived maps (i.e., outputs from the hydrographic survey) of the area. The outputs from this remote sensing phase, that includes the depth, substrate structure and backscatter, are classified using unsupervised classification to quickly identify different ground-types. The different ground types were the basis for identification of sampling sites using a stratified random sampling tool developed by NOAA.

The second stage focused on the direct measurement and observation of the bio-geophysical parameters of the identified ground-types (i.e., the selected sampling sites) using two main general methods; 1). sediment grab sampling, and 2) underwater visual observation. The succeeding paragraphs further elucidates the methods mentioned above.

Sampling design (sampling site selection)

Define the sampling area

To facilitate the ground truthing of Hoburgs bank, the area was divided into 34 blocks (Fig. 2) that is on average 3 x 12 km large, or 36 km^2. The size of the blocks was estimated to be the total amount of area that could be acoustically surveyed in a 24h time period. The sampling was generally conducted on a per block basis, so that a block could be ground-truthed upon completion of the survey. For the smaller blocks it would be possible to complete a block during the night then sample it during the day and then move on to the next survey block. This way transport time between survey and sampling was minimized.

Estimate of the sampling effort

The number of sampling sites was estimated from the time available for sampling and the size of the area. Blocks with less complex habitats was allocated fewer samples, while blocks that are more diverse were allocated more samples. The complexity was determined by the number of unique ground-types identified in the initial classification. The blocks' complexity, together with expert interpretation, was used to determine the sampling effort for each survey blocks. Approximately 550 ground-truthed sites were sampled for the Hoburgs bank which is around 15 sites per survey area or 1 sample per 3 km^2.

Classification Tool

The outputs from the initial acoustic survey was classified to determine the different ground-types. The depth and backscatter data were used to produce a simplified habitat map (shape file with habitat classes) based on depth zones, substrate type and seafloor shape and complexity. This was done using a simple and quick classification analysis developed in ArcGIS model builder. It was tuned to capture the main habitat types found on Hoburgs bank and takes about 20 minutes to run on a survey block (~30 km^2). It used the following input data; depth geotiff 2m and backscatter geotiff 2m and used a minimum mapping unit of 1000 m^2.

Selection of sampling sites

A Sampling Design Tool developed by NOAA (Buja & Menza 2007) was used to do stratified random selection of sampling sites based on the classified draft map on daily basis during the survey. The tool optimizes the ground truthing requirements of the project by using estimates of area-based metrics for each classified ground-types. This allowed us to specify the number of samples for every ground-types that was encountered/produced in the classified map. This tool was incorporated into the in-house developed ArcGIS model builder to hasten the process.

Expert interpretation was used to locate unique or special features from the classified map that the automated classification missed referring to both the high resolution multibeam data as well as information from the sub-bottom profiler data. We also manually made small adjustments on the locations of some samples that were not located on or very close to habitat boundaries although some samples were left on boundaries to capture this common habitat feature as well. Modifications were also made to the number of samples in certain classes that were considered under sampled, extra important or complex using the stratified random sampling tool. Finally, we did a quality check to reduce the risk of over or under sampling certain parts of the survey area.

Bio-geophysical sampling

Sediment sampling

Surface sediments were sampled to examine the quality of the sediments and to compare it with the substrate type identified in the classified hydroacoustic data (depth and backscatter). Occasionally, "in-fauna samples" were also collected, to investigate the presence of animals and plants in the topmost layers of sediment. Two different types of sediment samplers were used, the Van Veen and OPB (Orange Peek Bucket), depending on the predicted substrate type and the data from the live video (from the underwater camera about 20 meters aft). The former was used for sandy bottoms, while the latter was used for all other kinds of bottoms (silt, clay, rocks and boulders).

The samplers were operated from the moonpool and positioned using RTK-GPS using the position of the moonpool. All samples were analyzed/interpreted onboard by a geologist and the interpretations were then encoded on SGU's inhouse geological sampling database. In addition, approximately 25% of the samples were sent to lab for further grain-size analysis.

I.A.1 Van Veen sediment sampler

Two 0.2 m^2 Van Veen samplers were used to take sediment and infauna samples on gravelly, sandy and silty bottoms, penetrating approximately 15 cm into the sediments. SGUs main sampler is made by KC of Denmark (http://www.kc-denmark.dk). It has four lids lined with stainless steel mesh, 4 mm thickness and area of 1000 cm^2, but it turned out to be overly sensitive to small rocks and had to be repaired multiple times. The backup Van Veen from the manufacturer Swedaq (www.swedaq.se) were more robust but smaller and had less penetration depth was used on mixed bottoms (see Fig. 18).

I.A.2 OPB – Orange peel bucket sampler

Orange peel bucket sampler is the bigger option for collecting hard and larger sediments (e.g., silt, clay, rocks and boulders). It is a heavy and sturdy grab with four segments or "peel" that are hinged around a central core which closes as the grab is lifted up thereby trapping the particles underneath. They are better at grabbing hard and uneven large loads, rather than just scooping small finer particles, and is usually used in rocky, or hard clay substrates (see Fig. 18).

I.A.3 Sediment sample processing and interpretation

The retrieved sediment sample was assessed regarding its composition and particle size as well as probable sedimentation environment. Any occurrences of animals and plants were also noted. The sample was photographed and if necessary, a small amount of the sample was brought into the lab for further assessment. Occasionally, circa 2 kg of material was saved in plastic bags for later particle size analysis, which was later used for calibration of the substrate estimates made based on the reflectivity (backscatter) data. In addition to geological parameters such as grain size and origin, a sub selection of the samples was analyzed for infauna characteristics during the 2016 survey season in collaboration with SLU Aqua (Karlsson et al., 2017). These samples were sieved using a 1mm mesh and all the remaining material was sent to the SLU lab for biological analysis (Fig. 18).

Figure 18. Sediment sampling using Van Veen and Orange Peel Bucket (left). The collected sediment samples were described recorded using the in-house sediment database system while at least 25% of the samples were sent for a more detailed grain-size analysis. Selected samples were also analysed for infauna population during the 2016 sampling (right). Photos: SGU.

Underwater observations

The underwater benthic observation and sampling stations were designed to provide direct observation on state of the geology and biology in the study area. The method uses visual classification which involved indexing/identifying each interval point in the photo in terms of its substrate type, biota type, dominant species presence, algae type and presence, etc. as proposed in the report "Visuella Metoder" (draft version from Havs- och vattenmyndigheten 2015).

In this project, SGU aimed to undertake a more standardized data collection method that could be compatible with both the historical sampling methodology of SGU as well as the current versions that are designed to collect data about biotopes and identify flora and fauna. The main system used in this endeavor is the in-house build drop-camera system that was developed from the scratch using basic. The details on the use of this system is as follows:

I.A.4 The drop-camera system

A drop-camera system, designed and built in-house by SGU, was used to collect 360° underwater videos and photos of the seafloor as well as associated oceanographic data. The camera was operated from the A-frame and positioned using RTK-GPS based on the location of the A frame (Fig. 19). The succeeding paragraphs details the various aspects in the operation of the drop-camera system including the processing and analysis of the data.

I.A.4(a) Camera cage dimensions

When perpendicular (0°) to ground the lens of the waterproof housing was 73 cm from the substrate. Camera cage dimensions were 112 x 112 cm with a diagonal distance from foot to foot of 140 cm. The camera cage was deployed via winch from the stern of the vessel.

I.A.4(b) Camera set up

Video footage and images of each site were recorded with a DSLR camera (Canon EOS 6D DSLR) fitted with a lens (Canon EF 28mm f/2.8 IS USM, Φ 58mm). The camera was encased within a waterproof housing that was fixed to a positioner arm (Pan & Tilt; Sidus, San Diego) that was attached to the drop camera cage. The camera set up was equipped with two red laser pointers (brand, type) that were placed on either side of the waterproof housing at a distance of 30 cm from each other. Furthermore, lighting was provided by 2 sets of LED strobe lights placed on either side of the waterproof housing and a GoPro Hero 4 digital camera was affixed to the base (Fig. 20). All parts of the camera set up, except the GoPro camera and laser pointers, were controlled remotely from the deck via computer.

Figure 19. In-house designed, underwater observation drop-camera and sensor system. Associated sensors: **A.** CTD system. **B.** Doppler current sensor. **C.** DO meter (Oxygen optode). Photos: SGU.

I.A.4(c) DSLR camera

Video footage and images recorded by the DSLR camera used a pre-defined script driven pattern that covered 360° in the horizontal direction at several different vertical angles. More specifically a total of 23 images were recorded at six different vertical angles relative to the bottom substrate (Table 5). The scripts were developed during the project and during the 2017 survey season an additional 6 images, photographing the 360° horizon using only natural light, were added to more complex sites. Also, a passive light section sweep was added to the initial video recording. The passive light photos allowed for seafloor observations beyond the range of the camera light system. Add table row with passive light 360° images.

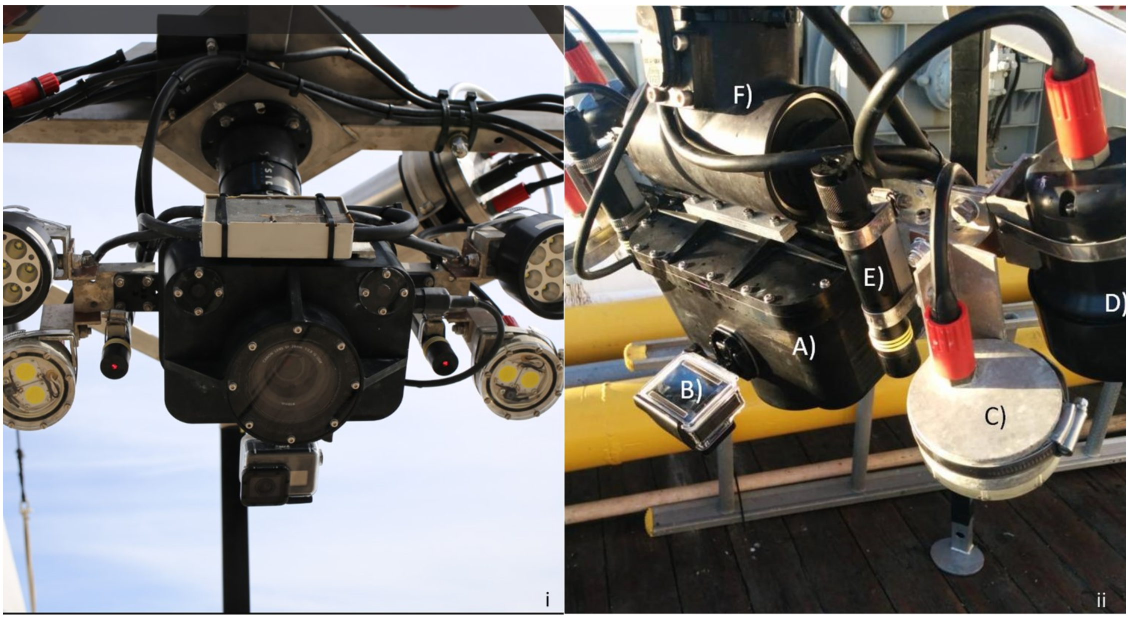


Figure 20. Front **(i)** and back **(ii)** view of the camera set up. **A)** Waterproof housing for DSLR camera. **B)** GoPro Hero 4 digital camera in waterproof housing. **C)** Strobe light. **D)** Strobe light. **E)** Laser pointer. **F)** Motor. Photos: SGU.

Table 5. Vertical and horizontal angles of images recorded by the DLSR camera. Values in bold denote the images used in photo mosaics and analysed according to visual methods.

Vertical (Tilt) angle (°)	Horizontal (Pan) angle (°)	Images (N)
0	**-150**	**1**
27	**-150, -110, -60, -21, 30, 70, 120, 159**	**8**
50	**157, 115, 67, 26, -24, -63, -110, -153**	**8**
62	136, 45, -40, 130	4
70	-140	1
75	-140	1

Of these 23 images 17 (those ≤ 50°) were analysed according to visual methods by Aquabiota AB. These images were also used to produce planar photo mosaics using the photo stitching software Autopano Giga (v4.2.3, Kolor LLC) (Fig. 21). These mosaics were subsequently analysed for substrate and benthic species cover using the freely available annotation software photoQuad (v1.4, Trygonis and Sini 2012). Images recorded at an angle of 50° were cropped to 75% of their vertical size in order to cover the required 5 m^2. This value was calculated in underwater conditions on flat ground.

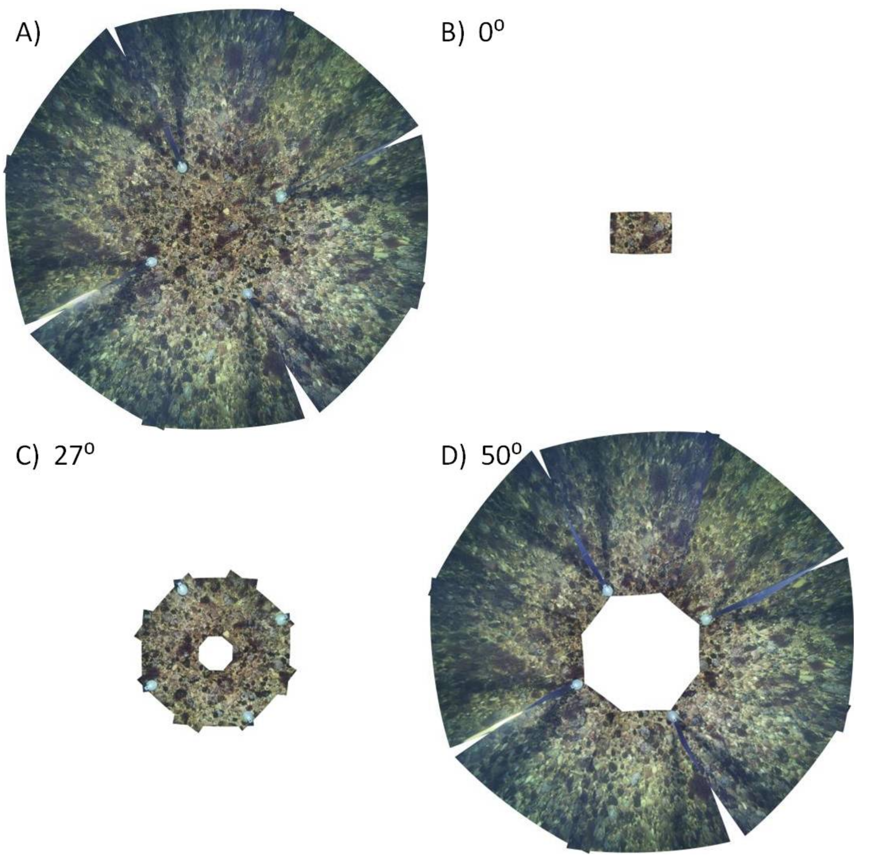


Figure 21. Underwater photo images taken by the DSLR camera and mosaiced using auto stitching in Autopano Giga software. **A)** Photo mosaiced image produced by combining images taken at 0° **(B)**, 27° **(C)**, and 50° **(D)**, **B)** 1 image recorded at vertical angle 0° (inner image), **C)** 8 images recorded at vertical angle 27°, **D)** 8 images recorded at vertical angle 50° (outer part of the image in **A)**.

Video footage was used in analyses to provide a better understanding of the site particularly with regards to substrate identification, which was often difficult with only still images.

I.A.4(d) Accessory cameras

Video footage and images were also recorded using a consumer grade “action camera”. The GoPro camera (Hero 5) recorded video footage throughout the whole deployment as well as still images every 10 seconds via a predetermined setting.

I.A.4(e) Other camera cage equipment

In addition to the camera set up the camera cage was equipped with a variety of sensors to collect different sets of environmental data. The CTD (Valeport Ltd, Devon) to record water temperature and salinity, a Doppler Current Sensor (Aanderaa Data Instruments AS, Bergen) to record water current as well as an Oxygen Optode (Aanderaa Data Instruments AS, Bergen) to record water oxygen concentration (Fig. 19).

Summary specs of the underwater observation system:

- Camera specifications:
  - Power and vdsl over coax in wire.
  - Computer on chip running windows 7.
  - Pan and tilt.
  - Powerful led light.
  - Canon EOS 6 full format camera.
  - Live view onboard OS with scriptable software control.
  - Laser for object size assessment (30 cm distance between points).
- Oxygen sensor.
- Current doppler sensor.
- CTD.
- Fluxgate compass.
- One spare connection for release system, grip etc.

Image processing and interpretation

Collected underwater images were mosaiced using the photoQuad software (Trygonis & Sini 2012) and subsequently used for manual interpretation. A more detailed explanation of the methodology involved in the image processing and interpretation is found in the accompanying report. Outputs from the manual interpretation served as inputs to the habitat modelling process that is explained in the main part of this report.

CONCLUSION

This document is a comprehensive report of the field work undertaken for SGU's Hoburgs bank project. It details our experiences, equipment and methodologies used in the biological and geophysical mapping of Hoburgs bank. This document serves as accompaniment to the habitat map production, modeling and analysis report which is found in another document. This was made as a separate report because it is our intention that other future habitat mapping projects by SGU or other entities can easily refer to this document during the conduct of field surveys. Finally, we believe that this work is not the final product, instead it is but a part of a continuous process where improvements can be added with new technology and better equipment to create better habitat maps for coastal environment.

SURVEY PERSONNEL INVOLVED

Marin miljö och planering

Name	Position
Gustav Kågesten	Project/Survey leader
Ola Hallberg	Survey co-leader
Caroline Bringensparr	Data processor/Geologist
Annika Dahlgren	Data processor/Geologist
Francis Freire	Data processor/Report author
Olof Larsson	Survey personnel/Geologist
Liselott Wilin	Survey personnel/Geologist
Finn Baumgartner	Survey personnel/Geologist
Anna Svensson	Survey personnel/Geologist
Johan Norrlin	Survey personnel/Geologist
Par Nordgren	Survey personnel/Geologist

Marin drift och teknik

Name	Position
Peter Kask	Captain
Mattias Nyström	Captain
Björn Bergman	Survey engineer
Ronnie Grimm	Survey technician
Åke Hellgren	Survey technician
Philip Cederlund	Chief engineer
Veiko Virula	Chief engineer
Per Wittmar	Boat Technician
Olof Bovin	Second mate
Pontus Bergström	Able seaman
Tommy Svensson	Able seaman
Holger Karlsson	Able seaman
Helen Pettersson	Cook

REFERENCES

Buja, K. & Menza, C., 2007: Sampling Design Tool for ArcGIS. National Centers for Coastal Ocean Science. <https://coastalscience.noaa.gov/project/sampling-design-tool-arcgis/> accessed 200908

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*(4), 628–637. <https://doi.org/10.1093/icesjms/fsp280>

Colbo, K., Ross, T., Brown, C. J. & Weber, T., 2014: A review of oceanographic applications of water column data from multibeam echosounders. *Estuarine, Coastal and Shelf Science*, *145*, 41–56. <https://doi.org/10.1016/j.ecss.2014.04.002>

Council of the European Communities., 1992: *Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora.*

Fonseca, L. & Calder, B., 2005: Geocoder: An Efficient Backscatter Map Constructor. *U.S. Hydro 2005 Conference*, 9. <https://doi.org/10.1126/ SCIENCE.249.4972.1017>

Fonseca, L. & Mayer, L., 2007: Remote estimation of surficial seafloor properties through the application Angular Range Analysis to multibeam sonar data. *Marine Geophysical Researches*, *28*(2), 119–126. <https://doi.org/10.1007/s11001-007-9019-4>

Fredriksson, R., Kaljuste, O. & Bergström, U., 2017: Kartering av pelagisk fisk och metodutvärdering vid Hoburgs bank 2016. *SLU ID: SLU.ua. 2016.5.2-374.* Sveriges lantbruksuniversitet.

Hell, B., Broman, B., Jakobsson, L. & Jakobsson, M., 2012: The Use of Bathymetric Data in Society and Science : A Review from the Baltic Sea, (41), 138–150. <https://doi.org/10.1007/s13280-011-0192-y>

Ierodiaconou, D., Schimel, A. C. G., Kennedy, D., Monk, J., Gaylard, G., Young, M., Diesing, M. & Rattray, A., 2018: Combining pixel and object based image analysis of ultra-high resolution multibeam bathymetry and backscatter for habitat mapping in shallow marine waters. *Marine Geophysical Research*, *39*(1–2), 271–288. <https://doi.org/10.1007/s11001-017-9338-z>

International Hydrographic Organization., 2011: *Manual on Hydrography*. *Publication C-13* (Vol. 2005).

Karlsson, E., Fredriksson, R., & Bergström, U., 2017: *Bottenfaunaprovtagning vid Hoburgs bank 2016. SLU ID: SLU.ua. 2016.5.2-374.* Sveriges lantbruksuniversitet.

Kongberg Maritime AS., 2012: *Kongsberg EM2040 Multibeam Echo Sounder Instruction Manual. Kongsberg Maritime AS*. <www.kongsberg.com> accessed 200908

Kongberg Maritime AS., 2013: *SIS - Seafloor Information System. reference Manual for EM 2040. Release 4.0.*

Kurland, J. & Woodby, D., 2008: What Is Marine Habitat Mapping and Why Do Managers Need It? In J. Reynolds & H. G. Greene (Eds.), *Marine Habitat Mapping Technology for Alaska*. Fairbanks, Alaska : Alaska Sea Grant College Program, University of Alaska Fairbanks. <https://doi.org/doi:10.4027/mhmta.2008.02>

Trygonis, V. & Sini, M., 2012: Journal of Experimental Marine Biology and Ecology photoQuad : A dedicated seabed image processing software, and a comparative error analysis of four photoquadrat methods. *Journal of Experimental Marine Biology and Ecology*, *424–425*, 99–108. <https://doi.org/10.1016/j.jembe.2012.04.018>