

BALANCE

Data mining and collation: an overview

BALANCE Conference

25-26 October 2007

Copenhagen, Denmark

Denmark
Estonia
Finland
Germany
Latvia
Lithuania
Norway
Poland
Sweden

SGU

Sveriges geologiska undersökning
Geological Survey of Sweden

Johan Nyberg, SGU



Introduction

Data Collation

- General
- metadata
- spatial reference
- base maps
- formats
- data categories

Conclusions

Data type

Country/Organisation	Nr.	Publisher	Content Title	Data	Info
Denmark					
Danish Forest and Nature Agency		-	-		
National Environmental Research Institute	1	NERI	Bottom fauna samples in Danish waters 1970-2005	Biology	Sampling of soft bottom fauna
	2	NERI	Bottom fauna samples in BALANCE Pilot Area 1	Biology	Dataset contains information on several samples per station in a grid.
	3	NERI	Near Real Time data - NRT	Physical Oceanography	Data from CTD casts (salinity and temperature) and concentrations of total nitrogen (TN) and total phosphorous (TP).
	4	NERI	Water chemistry data from Danish waters	Geochemistry	Water chemistry data from Danish waters including nitrate, phosphorous, TN, TP, silicate and chlorophyll a.
	5	NERI	CTD data from Danish waters	Geochemistry	CTD data are publicly available through the web.
Danish Institute for Fisheries Research		-	-	-	-
DHI - Water and Environment		-	-	-	-

Data type

Country/Organisation	Nr.	Publisher	Content Title	Data	Info
WWF Denmark		-	-	-	-
Geological Survey of Denmark and Greenland	6	GEUS	Marine Landscape	Physical Oceanography/Substrates (geology)	This dataset is the result of a raster calculation on the 3 underlying maps: Sediment, Photic-depth and Salinity.
	7	GEUS	salinity layer	Physical Oceanography	Consists of 6 classes: <5 psu, 5-7.5 psu, 7.5-11 psu, 11-18 psu, 18-30 psu and >30 psu.
	8	GEUS	photic zone	Physical Oceanography	Consist of two classes: Euphotic zone and Non-Photic zone.
	9	GEUS	sediment layer	Substrates (geology)	Consist of 5 classes: Bedrock, Hard bottom complex, Sand, Hard clay and Mud.
	10	GEUS	Topographic Marine Landscape	Substrates (geology)/Physical Oceanography/Shoreline Morphology	4 underlying maps: Sediment, Photic-depth, Salinity and Topography from the Bentic Terrain Model.
Estonia	11	GEUS	5 m dybdekurver	Bathymetry	Digital højdemodel - bestående af polygoner i 5 meter intervaller - over både landjord og havbund.
Estonian Marine Institute, University of Tartu	12	EMI	Estonia - sediment	Substrates (geology)/Biology	Seabed sediment data: digital geological maps and point data from macrozoobenthos grab sampling stations
	13	EMI	Estonia - benthos	Biology	Coastal sea. Biomass of benthic macrophytes. Biomass and abundance of benthic macroinvertebrates.

Data type

Country/Organisation	Nr.	Publisher	Content Title	Data	Info
	14	EMI	Estonia - hydrography	Physical Oceanography/Geochemistry/Biology	Hydrographic (salinity, temperature, nutrient profiles) and plankton (phytoplankton, zooplankton) data from 25 stations
	15	EMI	Estonia - bathymetry	Bathymetry	Digital bathymetric data (depth isolines) of the Estonian coastal sea
Finland					
Geological Survey of Finland	16	GTK	Northern Baltic Sea Sediment Map	Substrates (geology)	BALANCE sediment data from the Northern Baltic Sea:
	17	GTK	Archipelago Sea Sediment Map	Substrates (geology)	Data is based on the marine geological maps which has been reclassified into the BALANCE substrate classes.
WWF Finland		-	-	-	-
Metsähallitus Natural Heritage Service	18	METSA	Archipelago zonation	Natural environment/Shoreline Morphology	Zonation in SW Finland, categorized to three classes (inner, middle and outer archipelago).
	19	METSA	Communication infrastructure	Natural environment/Biology/Socioeconomics	A predictive map of communication infrastructure indicating the pressure induced by anthropogenic influence on marine nature; (1 = lowest, 5 = highest) classes.
	20	METSA	Effects of marine management activities on fishing	Natural environment/Socioeconomics	A predictive map of the probable effects of marine management activities on fishing; classified to five (1 = lowest, 5 = highest) classes.
	21	METSA	Effects of recreational fishing on sensitive habitats	Natural environment/Socioeconomics	A predictive map of effects of recreational fishing on sensitive habitats; (1 = not vulnerable, 2 = moderately vulnerable, 3 = highly vulnerable) classes.
	22	METSA	Habitat heterogeneity	Substrate (geology)/Bathymetry/Marine meteorology/Shoreline Morphology	A predictive map of habitat heterogeneity created using depth, wind exposure and shoreline data; classified to five (1 = lowest, 5 = highest) classes.
	23	METSA	Human influence on coastal lagoons	Natural environment/Biology/Socioeconomics	The map is a prediction on anthropogenic influence on coastal lagoons classified to five classes (1 = lowest, 5 = highest);

Data type

Country/Organisation	Nr.	Publisher	Content Title	Data	Info
	24	METSA	Marine noise disturbance from vessels	Natural environment/Biology/Socioeconomics	A predictive map of noise disturbance in the marine environment induced by vessel traffic; classified to five (1 = lowest, 5 = highest) classes.
	25	METSA	Potential dredging sites	Natural environment/Biology/Socioeconomics	A predictive map of potential dredging sites associated with maritime traffic and recreational boating.
Finnish Environment Institute	26	SYKE	Water turbidity in the Archipelago Sea	Physical Oceanography/Biology/Natural Environment	The turbidity of surface water in the Archipelago Sea is monitored during the open water season using MERIS and Terra/MODIS satellite images.
	27	SYKE (HELCOM)	Harbours	Structures/Natural environment	Dataset contains the ports (with commercial traffic) of the Baltic Sea.
	28	SYKE (HELCOM)	Oil Terminals	Structures/Natural environment	This dataset includes information about the oil terminals along the Baltic Sea coastline which have an annual turnover of more than 3 million tonnes.
	29	SYKE	Shipping lanes	Structures/Natural environment	The dataset contains the main shipping and boat lanes in the Finnish coastal waters (excluding the Åland islands).
	30	SYKE (HELCOM)	Exclusive Economic Zones in the Baltic Sea	Structures/Natural environment	Dataset contains the border lines of the Exclusive Economic Zones in the Baltic Sea.
	31	SYKE (HELCOM)	Baltic Sea Subbasins	Structures/Natural environment	Dataset consists of the sub-divisions of the Baltic Sea sub-basins according the HELCOM's Combine Manual (www.helcom.fi).
	32	SYKE	Bathymetry	Bathymetry	The dataset is a rasterized TIN-model of the bathymetry of the Finnish coastal waters, based on the depth data (shorelines, depth points and depth iso-curves).

Data type

Country/Organisation	Nr.	Publisher	Content Title	Data	Info
	33	SYKE	Secchi depth	Physical Oceanography	Dataset consists of point observations of secchi depth in the Finnish Sea areas from 1995 to 2006. The dataset contains 1647 sites and some 56000 observations.
	34	SYKE	Turbidity and surface water temperature in Finland (1995-2005)	Physical Oceanography/Marine meteorology	The dataset contains surface water (0-1 meters) temperature data from the Finnish coast as well as turbidity data of the area. The data is collected in 1995-2005.
	35	SYKE	Surface water salinity in Finland (1995-2005)	Physical Oceanography/Marine meteorology	The dataset contains surface water (0-1 meters) salinity data from the Finnish coast.
	36	SYKE	Bottom water salinity in Finland (1995-2005)	Physical Oceanography	The dataset contains bottom water salinity data from the Finnish coast.
	37	SYKE	Wave exposure	Physical Oceanography/Marine meteorology	Wave exposure grids created using the method SWM (Isæus) A nested-grids technique was used to ensure long distance effects on the local wave exposure regime.
	38	SYKE (Lantmäteri)	Land-uplift model NKG2005LU	Shoreline Morphology	The estimated apparent land uplift in mm/year in the Scandinavia based on the model NKG2005LU.
	39	SYKE	Average ice-days 1996-2000	Marine meteorology	Dataset contains the average number of days of ice-cover at the Finnish coastal observation stations between 1996-2000 (228 points).
	40	SYKE (HELCOM)	Baltic Sea Protected Areas (BSPAs)	Natural environment/Biology	The dataset contains the Coastal and Marine Protected Areas in the Baltic Sea Region (HELCOM) and the new proposed offshore Baltic Sea Protected Areas.
	41	SYKE (BirdLife International)	Important bird areas (IBAs)	Natural environment/Biology	Dataset contains the important bird areas (IBAs) in the Baltic Sea area.

Data type

Country/Organisation	Nr.	Publisher	Content Title	Data	Info
	42	SYKE	Natura 2000 sites in Finland (lines)	Natural environment/Biology	The dataset contains those Finnish NATURA 2000 sites according to the Finnish government decisions
	43	SYKE	Natura 2000 sites in Finland (polygons)	Natural environment/Biology	The dataset contains the Finnish NATURA 2000 sites according to the Finnish government decisions
	44	SYKE	Natura 2000 sites in Åland	Natural environment/Biology	The dataset contains the NATURA 2000 sites in the Åland Islands according to the Finnish government decisions
	45	SYKE (HELCOM)	Ramsar and EU Bird Directive Areas	Natural environment/Biology/Structures	Dataset consists of Ramsar and EU Bird Directive Areas in the Baltic Sea based on Baltic Pipeline System: Environmental Impact on the Baltic
	46	SYKE (HELCOM)	UNESCO Biosphere Reserves	Natural environment/Biology/Socioeconomics	Dataset contains the Unesco Man and the Biosphere (MAB) Biosphere reserves, promoting solutions to reconcile the conservation of biodiversity.
Finnish Game and Fisheries Research Institute	-	-	-	-	-
Germany					
WWF Germany	-	-	-	-	-
Leibniz-Institute of Marine Sciences	-	-	-	-	-
Latvia					
Institute of Aquatic Ecology, University of Latvia	-	-	-	-	-
Lithuania					
Coastal Research and Planning Institute	47	CORPI	Red algae Furcellaria lumbricalis reefs in the Lithuanian coastal waters	Biology/Substrate (geology)/Shoreline Morphology/Physical Oceanography	Statistical model for prediction of reefs, based on habitat mapping data. Model prediction map considers photic zone of from the coastline to the 20 m depth.

Data type

Country/Organisation	Nr.	Publisher	Content Title	Data	Info
	48	CORPI	EUNIS habitats types in the Lithuanian Exclusive Economic Zone	Biology/Substrate (geology)/Bathymetry/Structures	Map of EUNIS habitat types Coastline vector data set, raster layers of sediment composition and bathymetry were used along with biological variables.
	49	CORPI (Klaipeda University)	Biodiversity study and mapping of marine habitats in the vicinity of the Butinge Oil Terminal.	Biology/Substrate (geology)/Socioeconomics/Bathymetry	Habitat distribution incoastal waters in the Butinge Oil Terminal area, by integrating biological material with the bathymetry data and sediment distribution maps.
	50	CORPI (Geological Survey of Lithuania)	Geological Atlas of the Lithuanian part, the Baltic Sea	Substrate (geology)/Shoreline Morphology/Natural environment	The dataset contains information on sediments and geological geomorphological map and map of the anthropogenic strain (1 : 5 000) and explanatory script.
	51	CORPI (Fisheries Department under the	Baltic Sea, Lithuanian part, fishery map	Substrate (geology)	The map contains information on sediments compiled by integrating marine chart and geological images.
	52	CORPI(Lithuanian Maritime Safety	Baltic Sea middle part, approaches to port Klaipeda	Bathymetry/Substrate (geology)/Structures	The map supplies information on the bathymetry. The map also informs about light characters, colors of lights and buoys, nature of seabed.
	53	CORPI (Geological Survey of Lithuania)	Baltijos jūros Lietuvos krantų geologinis atlasas	Shoreline Morphology/Substrate (geology)/Structures	The dataset contains information on coastline, geological geomorphological map and map of the anthropogenic strain .

Data type

Country/Organisation	Nr.	Publisher	Content Title	Data	Info
The Geological Survey of Norway		-	-	-	-
The Norwegian Institute for Water Research	54	NIVA	Predictions of Laminaria hyperborea at the Norwegian Skagerrak coast	Bathymetry/Substrate (geology)/Biology/Natural environment/Physical Oceanography/Marine meteorology	This map was created using GRASP to model presence of Laminaria hyperborea. Predictors: wave exposure, depth, curvature and light exposure.
Sweden	55	NIVA	Predictions of Nephrops norvegicus at the Swedish Skagerrak coast	Bathymetry/Substrate (geology)/Biology/Shoreline/Physical Oceanography/Marine meteorology	This map shows predictions (probabilities from 0-1) for presence of Nephrops norvegicus, created in GRASP. Data: depth, slope, aspect and substrate.
County Administrative Board of Stockholm		-	-	-	-
Department of Marine Ecology, Gothenburg University		-	-	-	-
National Board of Fisheries	56	National Board of Fisheries	Perch nursery habitat	Biology/Bathymetry/Substrates (geology)/Physical Oceanography	Fish habitat modelling
	57	National Board of Fisheries	Roach nursery habitat	Biology/Bathymetry/Substrates (geology)/Physical Oceanography	Fish habitat modelling
	58	National Board of Fisheries	Perch spawning habitat	Biology/Bathymetry/Substrates (geology)/Physical Oceanography	Fish habitat modelling

Data type

Country/Organisation	Nr.	Publisher	Content Title	Data	Info
	59	National Board of	Pike nursery habitat	Biology/Bathymetry/Substrates (geology)/Physical Oceanography	Fish habitat modelling
	60	National Board of	Sander nursery habitat	Biology/Bathymetry/Substrates (geology)/Physical Oceanography	Fish habitat modelling
Swedish Environmental Protection Agency		-	-	-	-
WWF Sweden		-	-	-	-
Geological Survey of Sweden	61	SGU	Bottom substrate types in the Baltic	Substrate (geology)	Tentative representaiion of bottom substrate types in the Baltic
	62	SGU	Marine Geological maps of the Baltic, Bothnian Sea and Bothnian Bay	Substrate (geology)	Digital marine geological maps
	63	SGU	Marine Geological map of the Skagerrak area	Substrate (geology)	Digital marine geological maps
	64	SGU	Marine Geological Map of the Gävle area	Substrate (geology)	Digital marine geological maps
	65	SGU	Marine Geological map over the Norrtälje area	Substrate (geology)	Digital marine geological maps
	66	SGU	Marine Geological map of the Umeå area	Substrate (geology)	Digital marine geological maps
	67	SGU	Batymetric map of the Bothnian Bay	Substrate (geology)/Bathymetry	Digital bathymetric raster map of 50*50 meters resolution over the Bothnian Bay.
Polen					
Polish Sea Fisheries Institute		-	-		
USA					
University of California		-	-		

Spatial Reference

Datum: WGS 84

Coordinate system for whole Balance area: **UTM zone 34N (extended)**

Coordinate system for detailed pilot areas: **UTM with appropriate zone**

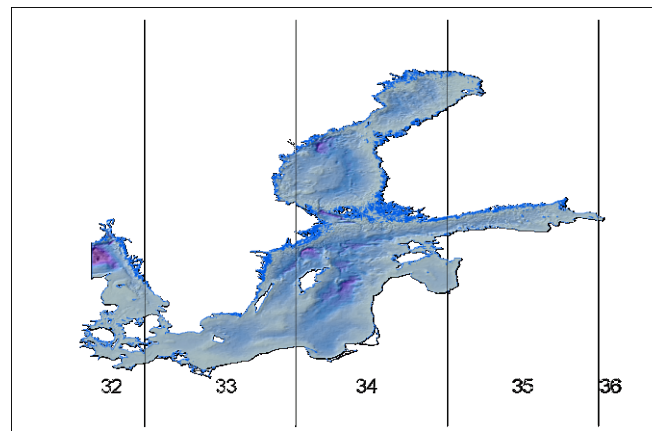
The objectives for choosing spatial reference system were Standard reference system

Easy to use (implemented in most commercial GIS)

Minimal areal distortion

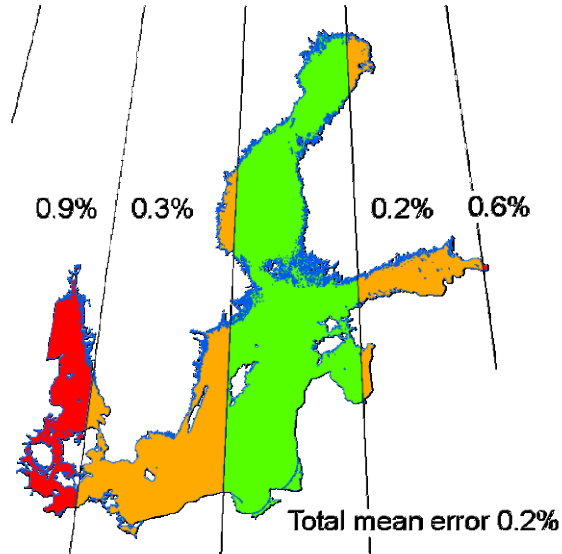
Spatial Reference

The choice of UTM based on WGS 84 and with a single extended zone fulfils the two first objectives. The Balance area covers parts of UTM zones 32N – 36N with its major part within zone 34N



Spatial Reference

Total minimal areal distortion is obtained if zone 34N is chosen as the extended zone for the whole Balance area. The picture below depicts the relative areal errors of the Balance area within the different UTM zones.



Base Maps

The following datasets have, by the courtesy of ESRI Inc., been freely used within the Balance project

Coastlines 1:15 000 000

Excerpt from the *"World Countries 2002"* dataset published in *"ESRI Data & Maps"*. Originator: ESRI Inc.

Coastlines 1:250 000

Excerpt from the *"Europe Countries"* dataset published in *"ESRI Data & Maps"*.

Originator: AND Data Solutions B.V. and ESRI Inc.

Data type

Vector data: Vector data consist of point, line or polygon themes including attribute data to each object

Raster data: This data can be either grid data or digital images.

Grid data: A grid in a GIS is a geographical referenced rectangular array of equally sized, quadratic cells.

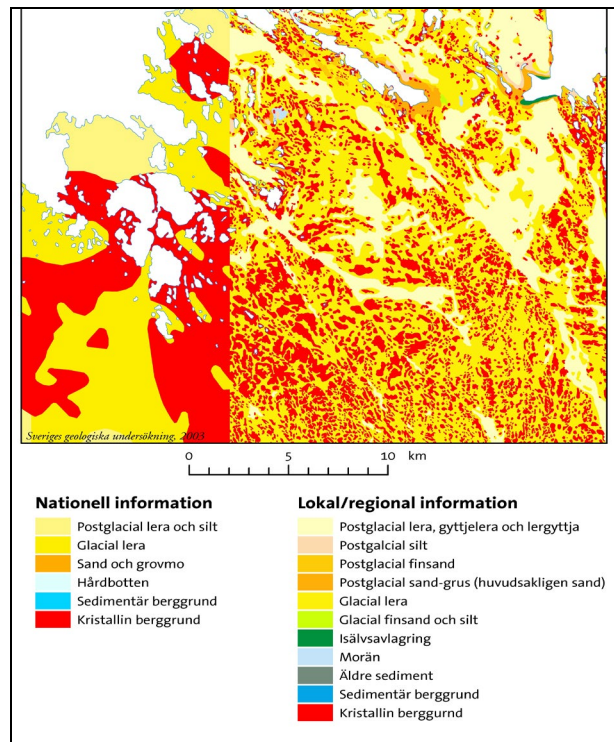
Image data: Scanned images (map) can result in a digital image. Geo-referenced images can be converted to a grid.

ASCII data: ASCII (text) files. This data must be handled in other systems before loading into the GIS

Base Maps

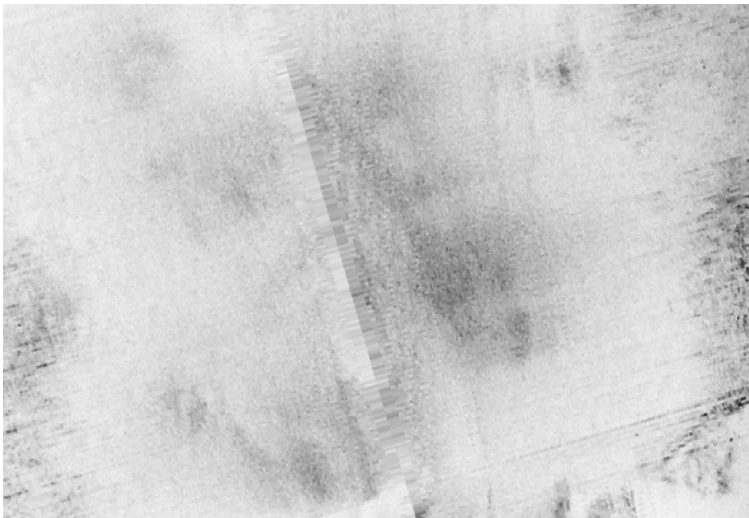
“Nationell” information is based on literature, single-point studies and sea-charts

“Lokal” information is based on survey lines 1 km apart

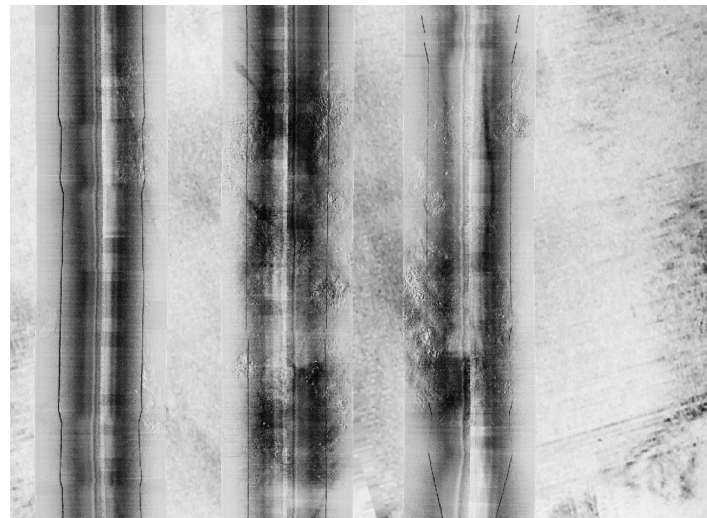


Base Maps

Side-scan sonar image
100 kHz
1000-1500 m coverage
1 meter pixel size



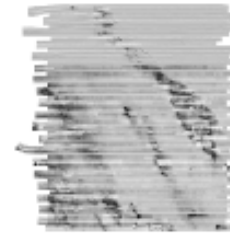
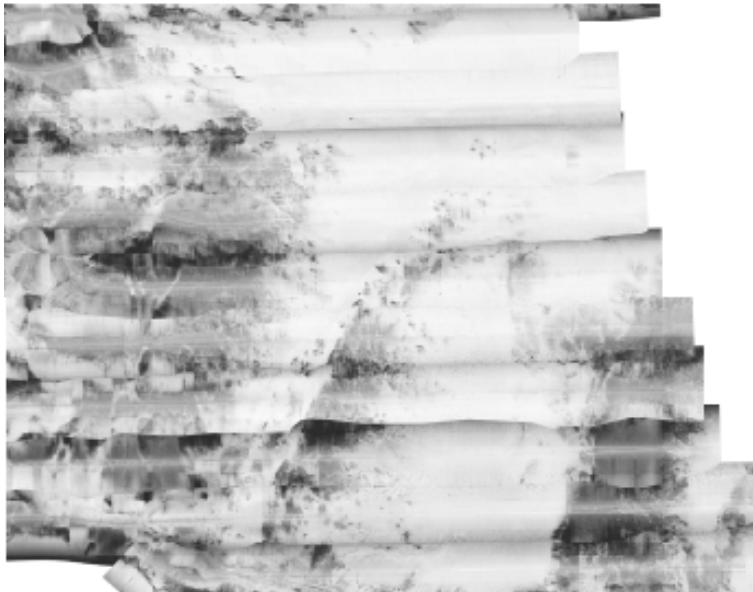
Side-scan sonar images
370 kHz
125-150 meter coverage
20 cm pixel size



Base Maps

Mosaics of Side-scan sonar images, geographical corrected

100 kHz
survey lines 1 km apart



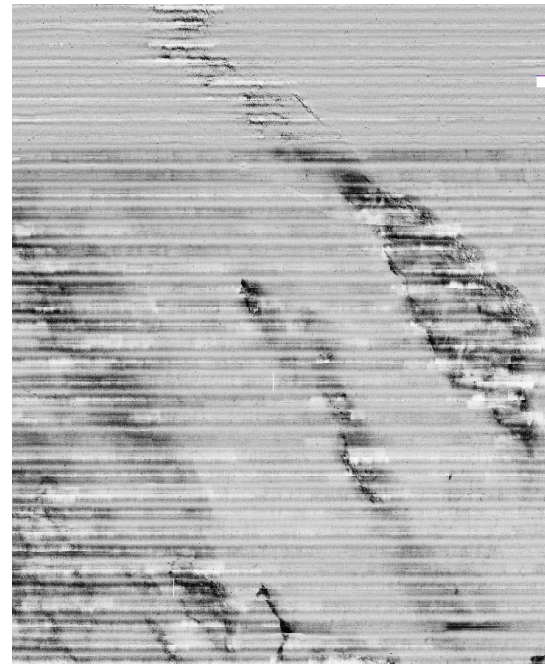
370 kHz
survey lines 100 m apart

Base Maps

100 kHz
survey lines 1 km apart



370 kHz
survey lines 100 m apart



Data

Bathymetry

Data Sets:

Depth contours (Vector elevation curves), Gridded surface (DEM)

Depth areas: Dredged areas, dredged disposal areas

Data collection:

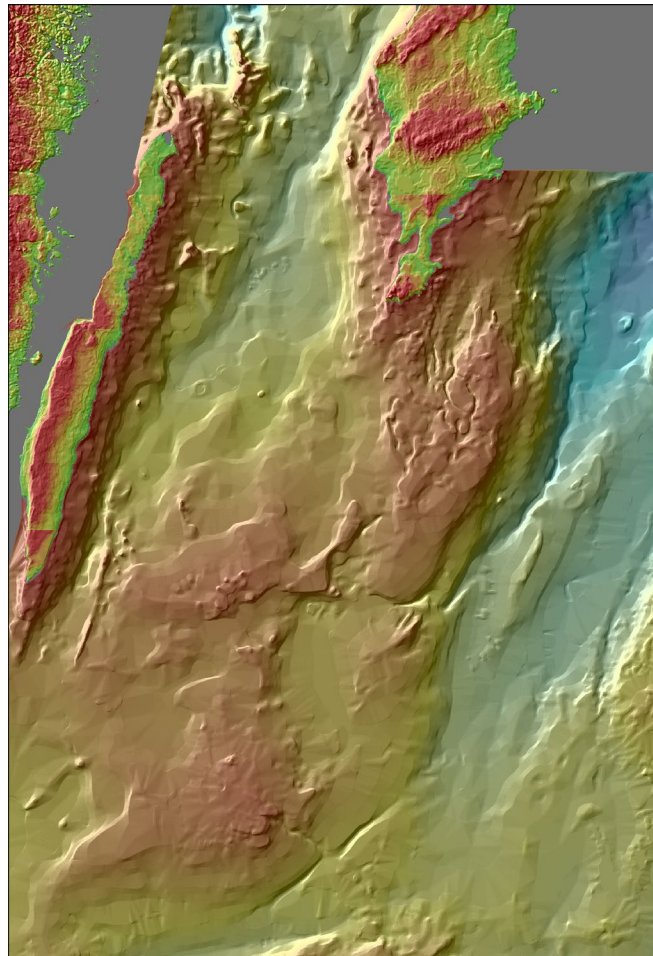
Remote sensing methods: Satellite imagery, Airborne techniques

Hydroacoustic methods: Multibeam echosounder, Echosounder (single-beam AGDS), Interferometry

Data

Bathymetry

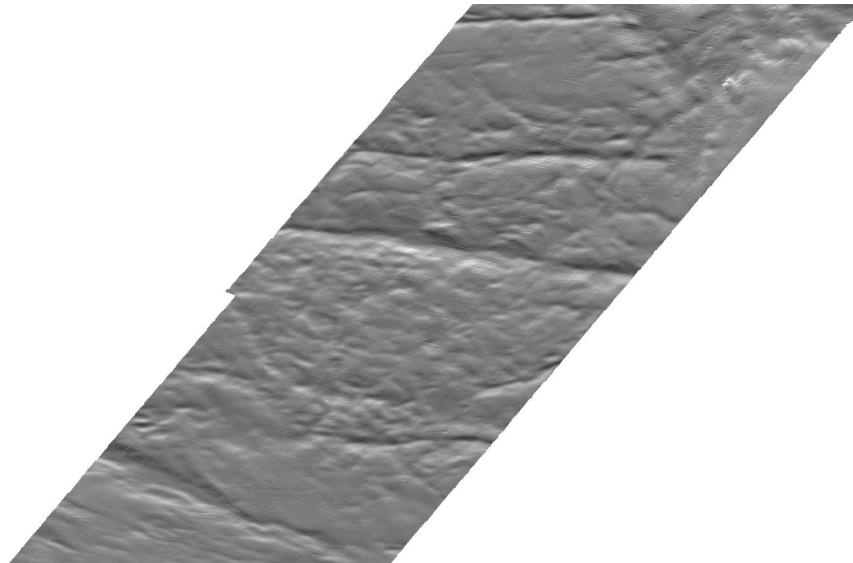
Shaded model
the Baltic



Data

Bathymetry

Multi-beam



Data

Shoreline Morphology

Data Sets: Shorelines, land cover, topographic maps, imagery, unclassified images, annotations

Natural Topography: Sea Cover: Coastal Type, Ice, Bedform

Hydrology: Lakes, Rivers

Land uplift areas

Data

Shoreline Morphology

Gavviksfjärden

“Höga Kusten”



Data

Substrates (geology)

Data Sets:

Seabed Sediments, erosion/sedimentation, Holocene and Pleistocene geology, Bedrock geology, Cores and Samples, Results of analyses, Interpretations, Reports and Maps

Data collection:

Hydroacoustic methods: Multibeam echosounder backscatter, Interferometric sonar, Sidescan sonar, Sub-bottom profiling, seismic

Video and imagery: Camera, remote operated vehicles

In situ sampling: Core sampling, Grab sampling, Particle size analysis of sediments

Data

Geochemistry

Data Sets:

Sample locations and sample analyses of sediment and organic and inorganic components

Physical Oceanography

Data sets:

In situ: Wave measurements, surface and subsurface water temperatures, surface and subsurface currents, surface and subsurface salinity, sea level, secchi depth, tidal amplitudes

Satellite: Sea level, geostrophic currents

Model: Gridded modelled nowcasts and hindcasts of most of the above parameters at a range of resolutions in time and space.

Data

Chemical Oceanography

Data sets

Water Column Chemicals: Organic and inorganic, nutrients from “Smart Buoys”, moorings (temporary and permanent), monitoring stations (temporary and permanent), cruises, radiation Monitoring

Suspended Sediments: From “Smart Buoys”, moorings (temporary and permanent), monitoring stations (temporary and permanent), cruises, remote sensing.

Data

Marine meteorology

Data sets

In situ: Surface wind velocities, surface temperatures (air and sea surface), sea level pressure, relative humidity, wave measurements, wave exposure, ice cover

Satellite: Surface winds and waves, sea surface temperatures, ice cover

Model: Gridded modelled nowcasts and hindcasts of most of the above parameters at a range of resolutions in time and space.

Data

Biology

Data sets

- in situ data sets (+ diving)
- modelled data (+ fish stocks) e.g. habitat modelling data

Distribution of key organisms

Plankton surveys, trawls

Fish (including fisheries related / derived data)

Fish Spawning Areas, Fish Nursery Areas, fish abundance and spatial distribution, Fish behaviour, age, scales

Fisheries: Fish Catches? mean length on age, year class strength, catch per unit effort

Data

Biology

Data sets

Marine Mammals

Populations of e.g. seal; breeding sites

Marine Birds

Populations of waders and wildfowl, seabirds, breeding sites

Benthic Marine Flora/ Fauna

Marine Benthic surveys of species and habitats, images and videos, ROV image data and side scan sonar image data

Data collection: Video and imagery: Towed video sledges, Camera, Sediment profile imagery, Remote operated vehicles

In situ sampling: Diver surveys, Trawls and dredges, Grab sampling, Drop-trap, Beam-trawl, Push-net, white plates and scoops, Juvenile trawl, Low Impact pressure wave

Data

Algal bloom in the Northern Bothnian Sea 2006



Data

Structures

Data sets

Shoreline Constructions

Piers, Pontoons, Slipways, Training Walls

Obstructions

Outfalls, Diffusers, Barriers,

Offshore Installations

Surface Structures: wind turbines, wave energy devices

Sub-surface Structures: marine turbines, pipelines (oil and gas, other), cables (communication, electrical, other)

Navigational Aids

Buoys, Beacons, Light Vessels

Wrecks, Archaeology

Data

Structures



Data

Structures



Data

Human Activities

Data sets

National Limits and boundaries

Administrative Boundaries: Port Limits, Exclusion Zones, Pollution Control Zones, Political and administrative boundaries

Shipping, Transportation

Traffic Separation, Reporting Areas, Anchorage Areas, Vessel Routes, Shipping Density, Transportation / Shipping Routes

Activity and Licence Areas

Military Exercise Areas, Aggregate Extraction Areas, Wind Farm Development Areas, Disposal Sites, Standing Approvals for Dispersants

Aquaculture

Fish and Shellfish Farm Areas,

Fisheries

Fisheries Areas, Sensitive Fish Areas, Fish Shellfish Growing Waters, Closed fishing areas

Data

Natural Environment

Data sets

Reserves with marine components, Natura 2000 sites, BSPA etc., including seal sanctuaries

Bathing Waters, Recreational Waters, Environmental pollution, environmental impact assessment, monitoring environmental risk, Nitrate vulnerable zones, Industrial Discharge Surveys, Bathing Water Quality, Marine Litter Surveys

Other Examples: HELCOM Regional Seas Boundaries, WFD River Basin Districts, WFD Coastal Water-bodies, WFD Transitional Water-bodies, environmental resources, protection and conservation

Conclusions & perspectives

Key messages

models and maps are not better than the input data

international common standards are needed for the data that are used as well as for data collection, collation and management

standards and protocols are needed for intercalibration of data and survey methodology between sectors and nations.

Next steps

Continuation and enhancement of data collection preferably with higher resolution

Perspectives

More and better data is crucial for enhancing knowledge and thus making an improvement of the management of the Baltic Sea

Acknowledgements

The results presented were funded by BALANCE

BALANCE is part-financed by
the European Union (European Regional Development Fund)
within the BSR INTERREG IIIB Programme