Article

A Data Model and Processing Environment for Ocean-Wide Bathymetric Data Compilations

Benjamin Hell and Martin Jakobsson, Department of Geology and Geochemistry, Stockholm University (Sweden)

Abstract

The compilation of ocean-wide digital bathymetric models (DBM) requires specific features of the bathymetric data storage and great flexibility of the data processing chain. In this article a solution based upon a spatial relational database management system and a Geographical Information System front end is introduced, which will eventually serve the compilation of a new DBM of the North Atlantic Ocean. As shown in a preliminary case study, the abundance of sounding data—both single beam and multibeam—available in that area to date bears an extremely high potential to derive a DBM with much greater accuracy and resolution than the DBMs commonly used today.

Résumé

La compilation des modèles bathymétriques numériques (DBM) des océans nécessite des éléments spécifiques du stockage des données bathymétriques et une grande flexibilité de la chaîne de traitement des données. Dans cet article, une solution reposant sur le système de gestion de la base de données relationnelle et un système d'information géographique frontal sont introduits, ce qui servira en fin de compte à la compilation d'un nouveau système DBM de l'océan atlantique nord septentrional. Comme indiqué dans l'étude de cas préliminaire, le grand nombre de données de sondage, à la fois monofaisceau et multifaisceaux, disponible dans cette zone, constitue à ce jour un potentiel très élevé pour la mise au point d'un DBM avec une exactitude et une résolution bien supérieures à celle aujourd'hui des DBMs en service dans ces jours.

Resumen

La compilación de modelos batimétricos digitales oceánicos requiere características específicas de almacenamiento de datos batimétricos y una gran flexibilidad en la cadena de procesamiento de datos. En este artículo se presenta una solución basada en un sistema de administración de una base de datos relacionales espaciales y se introduce un Sistema de Información Geográfica, que servirá finalmente para la compilación de un nuevo Modelo Batimétrico Digital del Océano Atlántico Norte. Tal y como se muestra en un estudio de un caso preliminar, la abundancia de datos de sondeos - tanto multihaz como monohaz - disponibles actualmente en esa zona ofrecen un potencial extremadamente alto para alcanzar un Modelo Batimétrico Digital con mayor precisión y resolución que los modelos utilizados comúnmente hoy en día.

1. Introduction and background

Even in the ages of multibeam echo sounding, ocean-wide digital bathymetric models (DBMs) still rely upon a multitude of historic measurements, mostly single beam echo soundings collected during the climax of ocean mapping activity between the 1960s and 1980s. In certain places even digitised spot soundings or depth contours from paper charts may be the only information available. Only the combination of both historical and contemporary measurements allows for the modern ocean-wide DBMs, such as the International Bathymetric Chart of the Arctic Ocean (IBCAO) (Jakobsson et al. 2008) or the 1-minute grid derived from the General Bathymetric Chart of the Oceans (GEBCO) (IOC et al. 2003).

Raw bathymetric sounding data used for the compilation of ocean-wide DBMs come from a variety of measurement methods and data sources. Because of the high resolution and spatially good seafloor coverage of contemporary multibeam surveys, these should be included in ocean-wide compilations wherever available and possible. Nevertheless, to date only a small percentage of the world's oceans is covered by multibeam data, focusing along the coast lines and within the exclusive economic zones. Hall (2006) considers that approximately 90% of the ocean is unmapped with multibeam, and the GOMaP initiative (Vogt et al. 2000) estimates 215 ship years of mapping activity required in order to obtain full coverage of the world ocean below the 500m isobath. In the Arctic Ocean, less than 6% of the ocean's area is mapped with multibeam (Jakobsson et al. 2008). Therefore in many regions, particularly of the open oceans, one has to rely upon older single beam echo soundings and interpolate between the data points. Apart from single beam measurements, and to a lesser extent, spot soundings can contribute valuable information. Digitised data from paper charts (contours or points) may be needed in places without better alternatives and where the underlying raw data are not publicly available.

Using a diversity of such heterogeneous data sets and data sources, however, poses questions to be considered for the compilation process: (1) How to quantify and assign data uncertainty bounds to the various data sets? (2) How to account for greatly varying data uncertainty in the compilation process? (3) How to estimate the uncertainty of the final gridded product (error propagation)? (4) How to handle huge data amounts for efficient processing?

Most of the answers to these questions can only be found if the bathymetric data used for the compilation (soundings, digitized contours etc.) are accompanied with a proper description, henceforth referred to as metadata. The metadata should include basic information regarding how, when and for which purpose the data were produced, with positioning and sounding methods being of particular importance. More specifically what the metadata should include will be discussed in this present work. To take metadata into account during the DBM compilation process is far from trivial. But considering the data's source and its uncertainty all the way from the beginning of the compilation process to the end products bears great potential for realistic uncertainty and reliability estimations of the final DBM.

The compilation of an ocean-wide DBM from both single beam and multibeam measurements involves processing of millions of raw data points — soundings and depths extracted from processed grids of multibeam surveys: The NGDC GEODAS repository alone contains more than 2.5 million single beam soundings, and multibeam mapping commonly produces at least the same order of magnitude of processed data points. This means that due to the sheer amount of soundings available, in order to be able to take advantage of the metadata, efficient solutions for data integration, storage, handling and processing need to be developed.

In this article we present a flexible data model, data management and processing environment, designed to resolve these problems. This solution is based on an underlying spatial database management system, which through the use of predefined queries (views) can be used with a Geographical Information System (GIS) front end (Fig. 1).

As will be shown in the last section of this article, the approach presented here allows for easy tracing of DBM problems to the underlying errors in the raw data. This greatly facilitates the data processing when compiling ocean-wide DBMs. Eventually the database and processing chain introduced here will be deployed for a compilation of a new DBM of the North Atlantic Ocean.

Several hydrographic offices world-wide have implemented spatial database technology in their archiving and processing of bathymetric data. The solution presented in this study is particularly designed to be used for compilations of DBMs.

2. Errors and uncertainty of bathymetric data

2.1. Systematic errors

As shown by Smith (1993) through cross-track analyses on more than 2000 — mostly deep-water — single beam surveys, a few percent contain systematic depth errors. Most of these errors originate from unknown or erroneous time to depth conversion and travel time readings when digitising paper roll records.

Travel time to depth conversion errors result in depth offsets proportional to the water depth, errors which were present in about 4% of the single beam surveys in Smith's original study. The most prominent case is a mix-up between the sound velocities of 1500m/s and 800fm/s, leading to a systematic 2.5% error of the measured depth. Jakobsson et al. (2008) describe a striking case of time to depth conversion confusion with submarine soundings from the Arctic Ocean, which lead to extensive systematic errors in previous versions of the IBCAO DBM. These errors were due to a lack of metadata, and once detected, they could be corrected.

Obtaining digital data from analog echo sounder records on paper rolls can result in systematic offsets of the depth measurements by a constant, if misinterpretations occur when the stylus moves off a paper edge. This commonly results in travel time offsets by multiples of full seconds, and is therefore referenced to as 400-fathom errors (the equivalent of one second two way travel time). Smith (1993) detected sections with obvious errors of this kind in 1.7% of all analyzed single beam surveys. Similar horizontal errors can occur with radio-based positioning systems: For example when offsets of the detected pulses occur, so called lane jumps can sometimes happen with the LORAN system.

Uncertainty bounds (typically the standard deviation), which can be assigned to data depending on measurement quality information, are strictly valid only for random errors and usually do not consider systematic anomalies. As systematic errors in the underlying soundings easily result in errors in the final DBM (refer to the last section of this article for examples), it is vital to both detect and reduce these systematic errors in the underlying (mostly) historic data as well as possible. One possibility for this is the crossover error approach by Smith (1993).

2.2. Random uncertainty

Random raw data uncertainty will also result in uncertainty of the compilation product. Particularly positioning with legacy systems such as LORAN, OMEGA or celestial navigation is prone to much larger random uncertainty bounds than modern (D)GPS navigation. Depending on the positioning and sounding systems used, uncertainty bounds can be assigned to the data sets.

In a similar manner as with bathymetric measurements, uncertainty bounds can be assigned to information digitized from paper charts (Jakobsson et al. 2005).

3. Metadata: Critical information for error tracking and uncertainty estimation

The most promising way to assess random uncertainty is by looking at metadata information, while crossover errors at crossing ship tracks can be used to detect systematic errors. It becomes clear that the raw sounding data are only complete with a comprehensive metadata record. Although in some cases it would be rather easy to detect errors manually, for example by looking at track line intersections, processing several thousand data sets in this way is tedious and calls for more automated routines.

To obtain an uncertainty estimate of the DBM, raw data uncertainty has to be taken into account during the compilation process and an uncertainty propagation must be performed. The uncertainty of the final product will then be a result of two factors: (1) the uncertainty of the underlying raw data and (2) the error propagation during the data compilation. A possible approach to the first factor could, for example, be the multibeam error model by Hare et al. (1995), whereas methods such as Monte Carlo simulation (Jakobsson et al. 2002) could be applied for the error propagation. Both approaches are essentially based on metadata.

The data storage structure should therefore incorporate all available metadata, with the most important records readily available for automated processing routines: measurement time, equipment (including positioning method, sounding method and uncertainty estimates), platform etc.

Two important standards exist dealing with metadata of geographical data: the United States FGDC Content Standard for Digital Geospatial Metadata

(CSDGM) (FGDC 1998) and the international ISO/ FDIS 19115 (ISO/TC 211 2003). Both the metadata information required by either of the two and the implementation structure of the two standards are very similar and it is easily possible to translate metadata records following one standard into the other. The data model presented here implements fully FGDC CSDGM compliant metadata records in XML format.

4. Data storage and analysis capabilities

For the capability to store multibeam data, single beam soundings, digitized contours and spot soundings, the data model needs to be able to handle two-dimensional data sets (multibeam) as well as data profiles and single data points. Preferably there should be no principal difference in dealing with these different types of data from a user's perspective. This means that the underlying data structure should generalize the differences inherent in the raw data.
Ideally intermediate products of the compilation work, i. e. preliminary processed data sets, should be handled in the same manner and with the same interface towards the user (Fig. 1). This means that adding processing related flags to soundings and surveys should be simple, and handling processed DBM grids should be possible.

4.1. A multi-dimensional data structure

Central to bathymetric data is a survey or, more generally, a data set comprised of many soundings, which share common metadata. Working on the basis of surveys or data sets, instead of soundings, while structuring a database results in one big advantage: For many operations the actual sounding data may not be needed. For example soundings can be replaced by the outline of the survey or a simplified ship track, when the data extent is analysed. Not having to handle the bulk of the sounding points at all times can simplify the processing to a great extent.

When a number of such data sets of the same kind are accompanied by metadata containing the same classes of information, the metadata can be seen as a part of the overall data structure. In this case the data can be seen as a multidimensional cube with the metadata constituting additional, often discrete, data dimensions (Fig. 2a). In the case of sounding data, these dimensional entities can, for example, be navigation or sounding instrumentation, measurement time, data source or measurement platform, just to name the most important ones. Often the metadata dimensions are discrete and in many cases they can be generalized and ordered into groups (aggregation), e.g. single-beam echo sounders and multibeam equipment. The dimension entities have one-to-many relationships with the data sets: One navigation instrument, for example, is used for the measurement of several data sets. Even the set of sounding points for a survey can be seen as a dimension, although with a many-to-one relationship to the data sets: One data set contains many soundings.
In a relational database, data models of this structure lend themselves to a particular database sche-

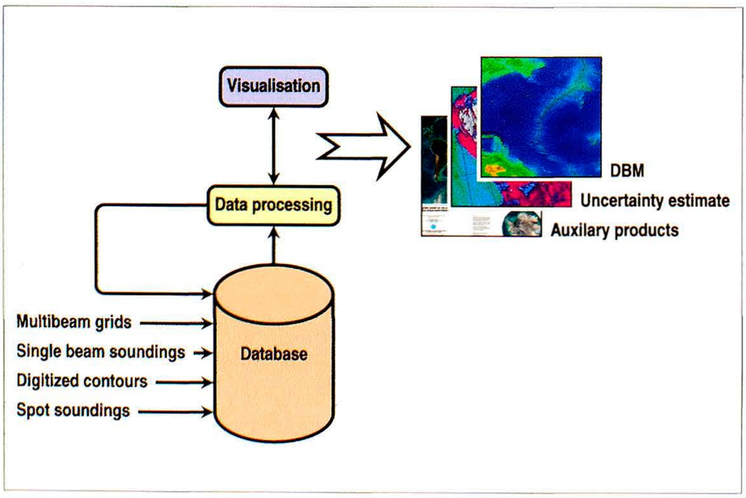

Figure 1: The underlying database management system stores and handles both the data which are used for the DBM compilation and compilation products, such as DBMs of test areas or gridding test runs. A GIS and visualization software based environment is used for data analysis and processing towards the end products: DBM, uncertainty estimation of the DBM and other products, such as maps.

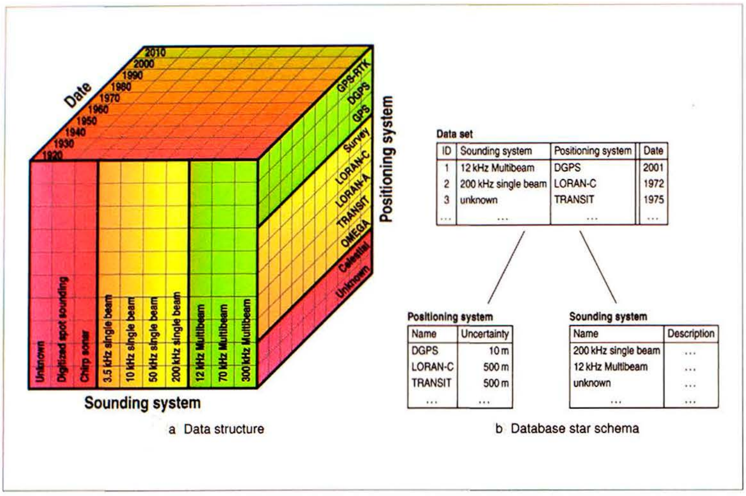

Figure 2: Each data set has, according to its metadata, a specific place in a multidimensional data cube (Fig. 2a). The dimensions of this cube are the various metadata entities, such as measurement date, positioning equipment or sounding system. Often the metadata dimensions can be generalized or aggregated, e. g. sounding systems into single beam and multibeam. Such a data structure lends itself to a star-shaped schema (Fig. 2b) with dimension tables (positioning & sounding system) ordered around a fact table (data set) in a relational database. Simplified illustration with three dimensions only; dimensions without further attributes can be merged into the fact table (date).

ma layout (Fig. 2b), where the metadata dimensions are tables ordered around a central fact table in a star-shaped manner (Kimball and Ross 2002). This database schema allows for easy extraction of data subsets along a common metadata axis, e. g. of all data measured with a certain positioning equipment.

4.2. Relational database storage

For an ocean-wide DBM compilation the amount of surveys and raw soundings is high: Potentially, several thousand single beam and some hundred multibeam surveys need to be considered. It should be noted that for the compilation of ocean-wide DBMs processed high-resolution grids from multibeam surveys constitute the input data, not raw multibeam soundings. The system presented here is not designed to store or process the latter. The total number of single beam soundings will soon reach some millions. The amount of grid points in a single high-resolution multibeam survey can be even larger. Since the final resolution of a regional DBM will be much lower than the resolution provided in most multibeam data, subsampling of multibeam surveys prior to their inclusion in the database can lead to substantial space saving.

Consider the storage of a few hundred bytes per point — the data types commonly used for storing geometries in spatial databases are unfortunately rather space intensive —, hundreds of millions of points will lead to a database with space requirements in the order of hundreds of gigabytes.

The amount of data dealt with for DBM compilations is usually increasing with time, as the compilation project moves on and new data gets available and is added to the compilation. Good scalability should therefore be considered an important advantage of the chosen storage system.

The only off-the-shelf storage solution, which satisfies these needs, is a spatial relational database management system (DBMS). To use a relational database for storing bathymetric data yields another great benefit: The query-based data retrieval from spatial relational databases offers basic analysis functions directly on the raw data level. For example finding depth differences at crossing measurement tracks (cross over errors) simplifies into a single spatial SQL query, the tools for which are already built into the spatial extension of the DBMS.

As some of the metadata information needed varies between raw soundings and processed data sets, a certain flexibility of the database schema concerning these metadata records has to be accomplished.

4.3. Data to be stored

For the purpose of regional DBM compilations, the information shown in Table 1 is considered to be relevant and should be stored in the database.

Depending on whether information is common for several data sets, within a data set or varies from sounding point to sounding point, the required database records can be split into information per sounding, per survey and more general metadata, e. g. per vessel or measuring system. The corresponding entities and attributes can therefore be spread over several database tables accordingly.

	Entity	Attribute	Description	Database table/column
Per sounding	Sounding	Sounding ID	Unique point identifier	SOUNDING.PID
		Position	Including information about coordinate system and geodetic datum used	SOUNDING.Geometry
		Depth	Observed depth	SOUNDING.Depth
		Measurement time	Time stamp	SOUNDING.Time
		Travel time	Two way travel time, if available	SOUNDING.TWT
		Time/depth conversion	Basic travel time to depth conversion information (can vary within a single data set, e. g. with Carter's tables)	SOUNDING.ReductionFlags
		Processing flags	Possibility for data flagging during the compilation process	SOUNDING.ProcessingFlags
Per data set	Dataset	Data set ID	Unique data set identifier	DATASET.DatasetID
		Bathymetric datum	Vertical bathymetric datum	DATASET.BathDatum
		Bounding box	Minimal bounding rectangle	DATASET.BoundingBox
		Simplified geometry	Track line (single beam) or coverage polygon/G-Polygon (multibeam, grid data)	DATASET.SimpleGeometry
		Measurement time range	Survey start	RAW_DATA.BegTime
			Survey end	RAW_DATA.EndTime
		Data transfer date	Alternatively processing date	RAW_DATA.GotTime
		Metadata record	Comprehensive FGDC CSDGM compliant metadata record (in XML format)	DATASET.FGDCMetadata
General metadata	Measuring platform	Ship name		PLATFORM.PlatformName
		Platform type	Surface ship, submarine, drifting station etc.	PLATFORM.PlatformType
		Home port country		PLATFORM.CountryCode
	Navigation system/ Horizontal uncertainty	System name		NAVIGATION_SYSTEM.NavigationSystemID
		System class	GPS, DGPS, LORAN, etc.	NAVIGATION_SYSTEM.NavigationSystemClass
		Navigation uncertainty	In case of raw data	NAVIGATION_SYSTEM.NavigationSystemAccuracy
		Resolution	In case of processed data	NAVIGATION_SYSTEM.Resolution
	Sounding system/ vertical uncertainty	System name		SOUNDING_SYSTEM.SoundingSystemName
		System type	Single beam/multibeam	SOUNDING_SYSTEM.SoundingSystemType
		Depth uncertainty	Generalized for one sounding system, percent of water depth	SOUNDING_SYSTEM.SoundingSystemAccuracy
		Footprint	Generalized, percent of water depth	SOUNDING_SYSTEM.Footprint
	Data source	Originator name	Who did the measuring/processing?	ORIGINATOR.OriginatorName
		Originator country		ORIGINATOR.CountryCode
		Source repository	Where does the data come from?	DATA_SOURCE.DataSourceName
		Underlying data sets	What raw data is this data set based on?	BASED_ON.BaseDataID

Table 1: Data attributes valuable for regional DBM compilations. Certain information needs to be stored per sounding point, other attributes only vary from data set to data set or are common for several data sets. See Fig. 3 for the referenced database tables and columns.

4.4. Database schema

When we apply the multidimensional data structure to the sounding data considered here, it is most apparent to first focus on the single data set or survey. The survey/data set entity then takes the role as the central table of the star-shaped database schema. Because the attributes needed differ a bit between raw surveys and processed data sets, two extra tables are required to store these special attributes (RAW_DATA and BASED_ON in Fig. 3). The different metadata will then represent the dimensions: Navigation system, sounding system, measurement platform, originator and data source. All of these can be modeled as dimension tables around the data set fact table. The soundings, which belong to each data set, can be seen as a dimension as well, with the exception that there are many sounding tuples per data set tuple. Strictly, the (less relevant) tables COUNTRY and BASED_ON are not part of the star structure.

Fig. 3 shows the resulting database schema in standard Unified Markup Language (UML) notation.

5. Database-GIS coupling

Many Geographic Information Systems (GISs) available can make use of georeferenced data stored in external DBMSs, following the widely accepted OpenGIS standards (OGC 1999). Unfortunately, common GIS software only supports flat, two-dimensional data tables. To our knowledge is not possible to make use of complex database schemas with interlinked tables as the layout described above. This problem can be solved through the use of predefined SQL queries, so-called views.

As all SQL queries have their output ordered as single data tables, views can take a role to convey data from several tables joined together into large indi-

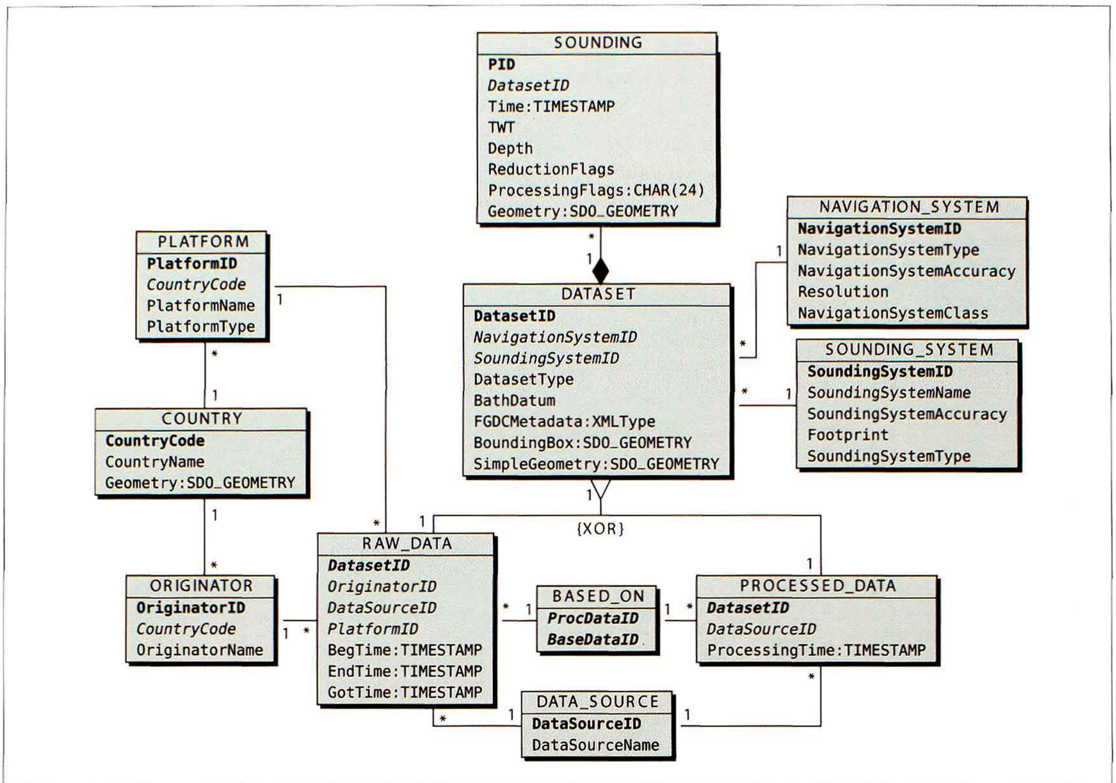

*Figure 3: Database schema in UML notation. Shown are all tables with their primary keys (bold font), foreign keys (italic) and regular attributes (some special data types are denoted after colons). *—1 denotes zero-or-more to one relationships, the diamond composition, and the triangle generalization: One navigation system can, for example, be used by many data sets, whereas a single data set is linked to exactly one navigation system.*

vidual tables, which can then be used by a GIS. The SQL query underlying a view is executed every time data are read from the results table. Therefore the computational cost of using views as a data front end is rather high — especially when complex join operations are necessary, as with most data warehousing schemas. To overcome this drawback, some DBMSs implement materialized views, where the query result is stored in a temporary table, which is updated automatically when the underlying data change (see e.g. Chaudhuri and Dayal (1997) for details). The major disadvantage of materialised views is the extra database hard disk capacity needed for the redundant storage, an increase of about a factor of two in total disk space.

6. Software implementation

The data model and processing environment described in the previous sections was implemented based on the Oracle 10g relational DBMS, Oracle Spatial for handling geospatial data and the Intergraph GeoMedia Professional GIS in combination with IVS 3D Fledermaus visualisation.

Oracle Spatial is Oracle's database extension for storing and analyzing spatial data. To date it is by far the most comprehensive product of its kind and therefore the industry standard in GIS business. Oracle Spatial features the data types, indexes and methods necessary to store and query both vector and raster data. Furthermore, a fairly complete set of GIS-like processing and analysis functionality as SQL functions is implemented in the DBMS. However, the powerful visualisation and interactive analysis features of a full-blown GIS are lacking.

Of the full-featured GISs available, Intergraph's GeoMedia Professional is the one with the most advanced database coupling. Internally, GeoMedia handles all data sources according to the relational data model, which makes data conversion between different sources a rather straight-forward task. Geo-

Media Professional is a vector based GIS. Nevertheless, raster data can be displayed, and there are extensions available for raster data analysis. GeoMedia Professional features powerful 2D visualisation functionality but lacks 3D visualization abilities completely. Therefore we combine GeoMedia Professional with the extremely powerful three-dimensional visualization tools of the IVS 3D Fledermaus software package.

The general data structure and processing approaches presented here are not limited to this specific combination of software and could generally be implemented with a different DBMS or GIS (although the number of functioning DBMS/GIS combinations is rather limited).

7. Application and outlook: A Digital Bathymetric Model of the North Atlantic

The North Atlantic Ocean is arguably the best mapped ocean in the world, covered to a large extent with an abundance of sounding data, which feature a tremendous variability in accuracy, resolution and density (Fig. 4).

For this reason it is an ideal test bed for data compilation techniques. The data handling and processing introduced in this article will be used for the compilation of a new DBM of the North Atlantic. The proposed International Bathymetric Chart of the North Atlantic (IBCNA) (Macnab and Travin 2007) is an undertaking to assemble and to rationalize all available bathymetric observations from the Atlantic Ocean and adjacent seas north of the Equator into a coherent DBM.

Neither of today's most commonly-used large scale bathymetric models — GEBCO (IOC et al. 2003), based upon digitized contours derived from single beam echo sounding measurements, and ETOPO2v2 (U.S. Department of Commerce et al. 2006), satellite altimetry constrained by single beam echo

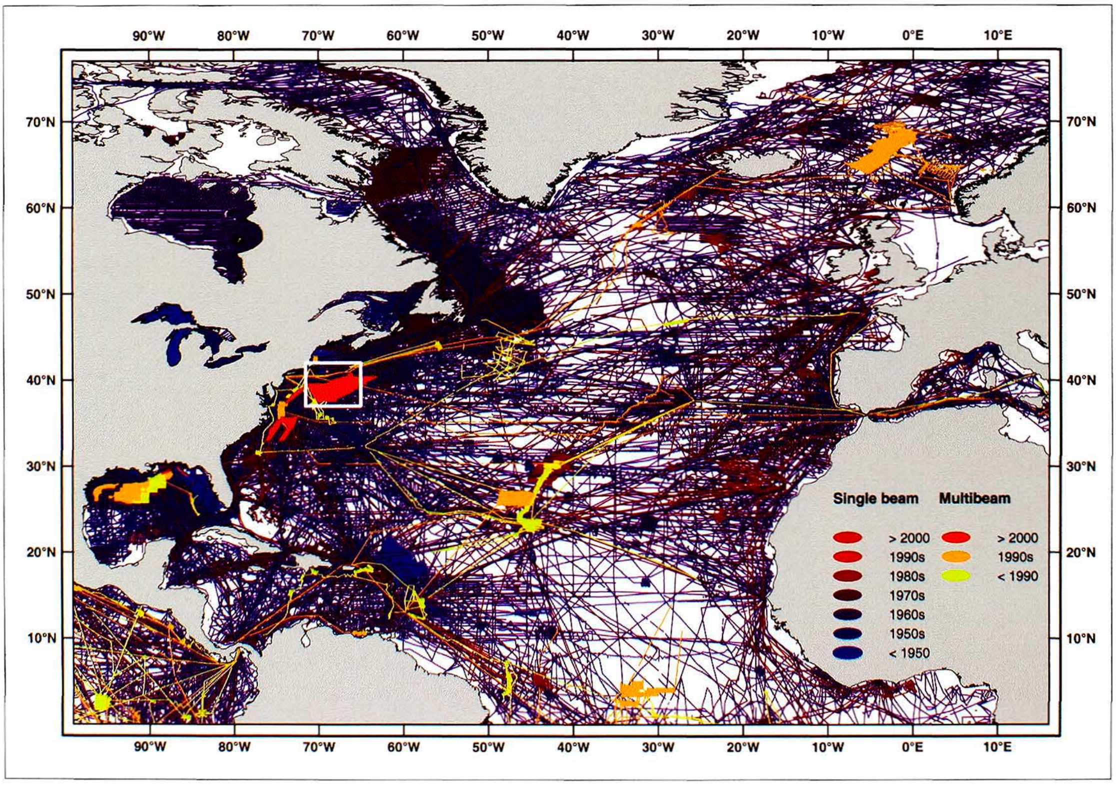

Figure 4: Readily available bathymetric surveys in the North Atlantic region (mostly NGDC, NOS, USGS), color coded by age. Single beam surveys are shown in blue to red colors, multibeam in yellow to orange. Younger surveys and multibeam are plotted on top of older ones and single beam. The lion's share of the data comprises single beam measurements from the 1960s to 1980s. A small but growing part of the North Atlantic is covered with multibeam data. The white box outlines the area shown in Fig. 5.

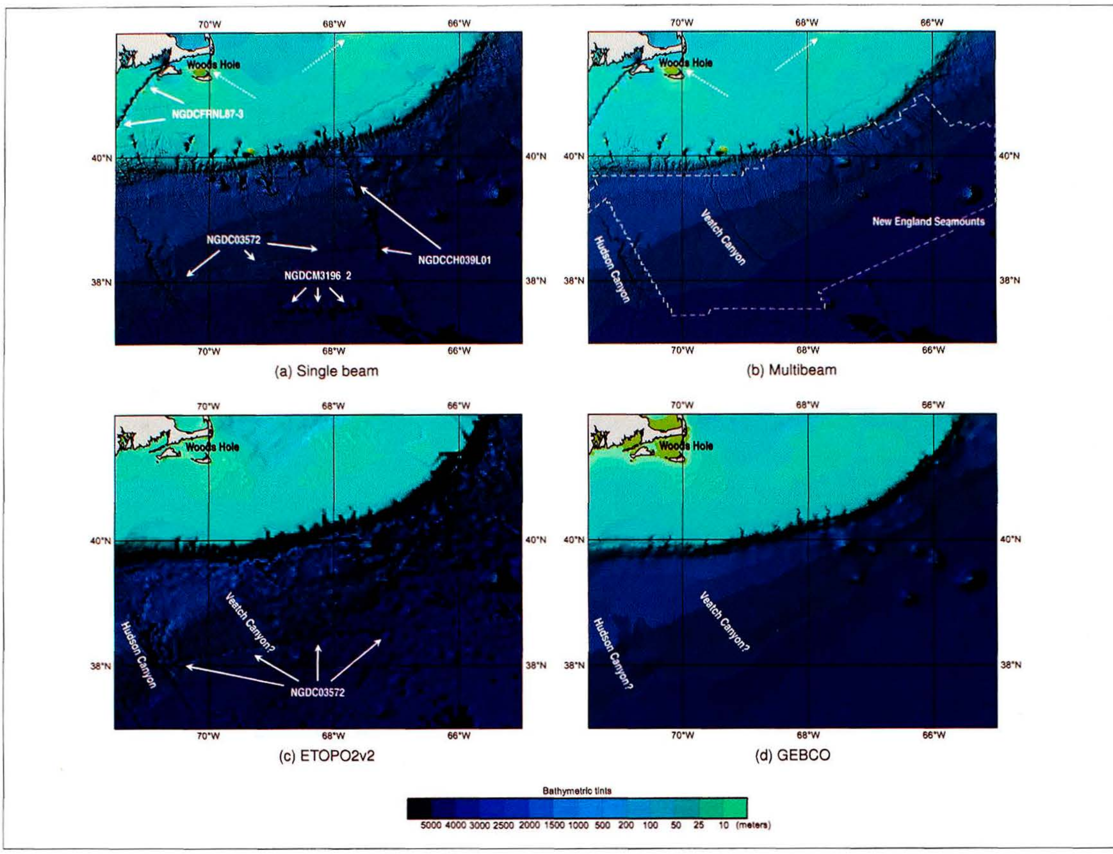

Figure 5: The comparison of DBMs based on purely single beam data (Fig. 5a), and single beam and multibeam data (5b, outline of multibeam survey dashed), with the ETOPO2v2 (5c) and GEBCO (5d) DBMs shows the potential of using an up-to-date bathymetric database for DBM compilation in the North Atlantic.

sounding track lines (Smith and Sandwell 1997) — incorporates the large amount of recent multibeam echo sounding data. Therefore, a new model based on all existing multibeam information has the potential to provide a significantly enhanced portrayal of the North Atlantic's seafloor.

Predicted seafloor topography using satellite altimetry, no doubt, provides the key instrument for the most poorly mapped regions of the world's oceans. But in the case of the North Atlantic, the existing data allows for a high-resolution DBM using echo sounding measurements exclusively, as will be shown in the following preliminary case study.

Figure 5 shows an area of the U. S. East coast (location marked in Fig. 4) and some of the potential of bathymetric data readily available to date. The DBM shown in Fig. 5a is purely based on single beam tracks from the NGDC GEODAS repository. In Fig. 5b additional recent multibeam data from the U. S. UNCLOS mapping program (Gardner 2004) were included in form of processed grids with a lateral resolution of 100 m.

Both DBMs were gridded with the approach used for the IBCAO compilation (Jakobsson et al. 2008), to a resolution of one-half arc minute. In short this involves filtering the data with a block-median filter and gridding using surface splines in tension (Smith and Wessel 1990). However, the single beam data were not at all cleaned for outliers or processed prior to the gridding, as one of the purposes of the examples is to show how easily some of the most obvious error sources can be traced down by means of the database and GIS. A comprehensive uncertainty analysis will be performed in future work.

For comparison purposes the two most commonly used DBMs of the North Atlantic region are shown: ETOPO2v2 (Fig. 5c) and GEBCO (Fig. 5d) in their native resolutions of two arc minutes and one arc minute, respectively.

A qualitative comparison of the DBMs shows some of the potential of an up-to-date bathymetric database for a regional DBM.

7.1. Tracing DBM problems to their origins

Apart from some gridding artifacts due to the lack of coastline constraints or sparse data close to the grid border (dotted arrows), there are several obvious track line artifacts visible in the single beam based DBM. By means of the GIS and database, these artifacts could easily be traced down to their originating surveys and—without a comprehensive analysis carried out so far—plausible error sources could be found, although in three of the four highlighted cases the metadata supplied with the data sets is unfortunately very sparse. NGDCFRNL87-3 is a Woods Hole based survey from 1987, where presumably data was recorded directly after leaving port and included in the data set before proper measuring activity started later and further off the coast.

In the case of survey NGDCM3196_2 the problem affects only a small part of the track line and looks like a seafloor tracking problem of the Simrad EA500 echo sounder used. NGDCCH039L01 is a survey from 1963 and shows the typical 400-fathom errors from analog recording described above. For the fourth survey, NGDC03572, the error is rather small and constant and a travel time to depth scaling error seems most likely. Comparing the depth values to crossing surveys reveals differences of around 2.5 %, which is the discrepancy between the sound velocities of 1500 m/s and 800 fm/s.

7.2. The power of an up-to-date sounding database

Adding extensive and high-resolution processed multibeam grids to large parts of the compilation area (Fig. 5c, dashed outline) further improves the seafloor image in these areas drastically. Single beam track line artifacts are overridden by the sheer abundance of multibeam grid points. The complete coverage and high resolution of the multibeam data reveals seafloor features, which are not visible in the single beam based DBM (e. g. Veatch Canyon). The small-scale structure of other features, such as the New England Seamounts, is further refined.
The (qualitative) comparison with ETOPO2v2 and GEBCO shows the great potential of data that became available at a later time and could therefore not be considered for either of these two DBMs. In areas with a thick sediment cover, such as on continental shelves, the capabilities to accurately reproduce seafloor topography from satellite altimetry are limited. Therefore it should be mentioned that this comparison is certainly not to the advantage of ETOPO2v2.

When looking at a small segment of the global DBMs ETOPO2v2 and GEBCO, they reveal a very different character. GEBCO shows a rather smooth bathymetry and rich details in areas with good single beam data coverage (e. g. along the continental slope or the New England Seamounts). Some other features (e. g. the Veatch Canyon) are missing completely. ETOPO2v2, on the other hand, shows all large features but their structure is far less clear than in GEBCO (partly due to the lower grid resolution). Flat areas, such as abyssal plains, display a characteristic orange peel like surface pattern. Interestingly, ETOPO2v2 shows artifacts originating from survey NGDC03572 as well.

Taken into consideration that no proper data preprocessing was performed, already a good coverage with single beam measurements easily results in a more detailed DBM than either of ETOPO2v2 or GEBCO, with the potential of achieving a higher grid resolution. This can be seen e. g. at the Hudson Canyon, the small-scale canyons along the continental slope or the structure of the New England Seamounts. To get rid of track line error related grid artifacts, appropriate data pre-processing has to be carried out.
As one would expect, including multibeam data in the compilation further increases the overall seafloor image quality drastically.

8. Conclusions

A new data storage model and processing environment was developed, targeting the compilation of ocean-wide DBMs. The GIS and spatial database approach developed facilitates the tracing of data errors due to the implementation of readily available metadata, which accompany all sounding information. Compared to the DBMs commonly used today, for a well-mapped region it was shown that using an up-to-date data base, including multibeam information, can lead to a significantly improved DBM seafloor portrayal. The implemented methods will be used for the compilation of a new DBM of the North Atlantic Ocean.

Acknowledgements

We appreciate the constructive comments by R. Hare and another anonymous reviewer, which helped to significantly improve the manuscript.

References

- Chaudhuri, S. and Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65–74. ISSN: 0163-5808. DOI: 10.1145/248603.248616.
- Gardner, J. V. (2004). Cruise Report USNS Henson: U. S. Law of the Sea cruise to map the foot of the slope and 2500-m isobath of the Northeast US Atlantic continental margin. Tech. rep. Durham, N. H.: Center for Coastal and Ocean Mapping/Joint Hydrographic Center, University of New Hampshire. URL: http://www.ccom.unh.edu/unclos/reports_pdf/report_HEN04-1.pdf.
- Hall, J. K. (2006). Worldwide High-resolution Bathymetry. *Hydro INTERNATIONAL*, May 2006, 57.
- Hare, R., Godin, A., and Mayer, L. A. (1995). *Accuracy estimation of Canadian swath (Multi-beam) and sweep (Multi-Transducer) sounding systems*. Tech. rep. Ottawa: Canadian Hydrographic Service.
- Intergovernmental Oceanographic Commission, International Hydrographic Organization, and British Oceanographic Data Centre (2003). *Centenary Edition of the GEBCO Digital Atlas*. Published on CD-ROM on behalf of the Intergovernmental Oceanographic Commission and the International Hydrographic Organization as part of the General Bathymetric Chart of the Oceans. British Oceanographic Data Centre. Liverpool, U. K. URL: http://www.ngdc.noaa.gov/mgg/gebco/.
- ISO Technical Committee 211 (2003). *Geographic information—Metadata*. International Standard ISO/FDIS 19115:2003.
- Jakobsson, M., Calder, B., and Mayer, L. (2002). On the effect of random errors in gridded bathymetric compilations. *Journal of Geophysical Research*, 107(B10), 2358–2368.
- Jakobsson, M., Armstrong, A., Calder, B., Huff, L., Mayer, L., and Ward, L. (2005). On the Use of Historical Bathymetric Data to Determine Changes in Bathymetry. *International Hydrographic Review*, 6(3), 25–41.
- Jakobsson, M., Macnab, R., Mayer, L., Anderson, R., Edwards, M., Hatzky, J., Schenke, H. W., and Johnson, P. (2008). An improved bathymetric portrayal of the Arctic Ocean: Implications for ocean modeling and geological, geophysical and oceanographic analyses. *Geophysical Research Letters*, 35, L07602. DOI: 10.1029/2008GL033520.
- Kimball, R. and Ross, M. (2002). The data warehouse toolkit: *the complete guide to dimensional modeling*. 2nd ed. New York: Wiley Computer Publishing. ISBN: 0-471-20024-7.
- Macnab, R. and Travin, D. (2007). International Bathymetric Charts—A Framework for Compiling and Rationalising Acoustic Soundings. *Hydro INTERNATIONAL*, 11(9), 14–17.
- Metadata Ad Hoc Working Group (1998). *Content Standard for Digital Geospatial Metadata*. Tech. rep. FGDC-STD-001-1998. Federal Geographic Data Committee. URL: http://www.fgdc.gov/metadata.
- Open GIS Consortium (1999). *OpenGIS Simple Features Specification for SQL*. Version 1.1. URL: http://www.opengeospatial.org/standards/sfs.
- Smith, W. H. F. and Wessel, P. (1990). *Gridding with continuous curvature splines in tension. Geophysics*, 55(3), 293–305.
- Smith, W. H. F. (1993). On the Accuracy of Digital Bathymetric Data. *Journal of Geophysical Research*, 98(B6), 9591–9603.
- Smith, W. H. F. and Sandwell, D. T. (1997). Global Sea Floor Topography from Satellite Altimetry and Ship Depth Soundings. *Science*, 277, 1956–1962.
- U. S. Department of Commerce, National Oceanic and Atmospheric Administration, and National Geophysical Data Center (2006). *2-minute Gridded Global Relief Data (ETOPO2v2)*. URL: http://www.ngdc.noaa.gov/mgg/fliers/06mgg01.html.
- Vogt, P. R., Jung, W.-Y., and Nagel, D. J. (2000). GOMaP: A Matchless Resolution to Start the New Millennium. *EOS Transactions*, 81(23), 254, 258.

Benjamin Hell, e-mail: *benjamin.hell@geo.su.se*