

ICES CM 2007/Q:19

Not to be cited without prior reference to the author

Facing the past – recovering historic echodata

Daniel Stepputtis¹, Uwe Böttcher¹, Jörg Voskamp², Eberhard Götze³

¹ Federal Research Centre for Fisheries / Institute for Baltic Sea Fisheries Rostock, Alter Hafen Süd 2, D-18069 Rostock, Germany

² Fraunhofer Institute for Computer Graphics (IGD), Joachim-Jungius-Strasse 11, D-18059 Rostock, Germany

³ Federal Research Centre for Fisheries, Palmaille 9, D-22767 Hamburg, Germany

email: science@stepputtis.net

Abstract

Since decades, a vast quantity of data is being collected in the field of fishery science. A valuable source of data about the distribution and abundance of living resources within aquatic environments are collected with echosounders. This type of data is a valuable source of information which is used for the assessment of fish stocks, as well as for ecological research. Nowadays, echosounders store their data electronically, which is the basis for fast access and easy analyzes of these data. Until the beginning of 1990s only integrated data were stored, but are often inaccessible since they are archived in paper form solely. This results in acoustic survey time series, which are often truncated before 1990. One example is the Baltic spring acoustic survey, which was carried out by the GDR and the USSR in the late 1970s and 1980s, a period prior to the regime shift from cod domination to sprat domination. Nevertheless, these data are needed to extend the actual international acoustic spring survey to the past and to have a reference of a period prior to the regime shift in the Baltic Sea. The recovery of these survey data includes the digitisation of all paper stored information (meta data, biological data and integrated echo data). Additionally the paper echograms were scanned and electronically archived. Furthermore the vertical distribution of fish species was extracted from echograms using a software package, developed in the German GLOBEC-project. The steps of recovery, as well as exemplary results will be presented.

Keywords: data recovery, hydroacoustic survey, Baltic Sea, regime-shift

Introduction

Nowadays, a vast quantity of marine science and fisheries data are being collected year by year. These data are mostly recorded electronically and stored in databases, which enhance their accessibility (hopefully). However, data collection in marine and fisheries science started more than hundred years ago. These historic datasets often offers valuable time-series, which could answer a variety of key-questions of today's marine science, such as the change of marine ecosystems and fisheries under different environmental regimes (Goal 1 of ICES Strategic Plan) or the understanding and quantification of human impact (Goal 2 of ICES Strategic Plan). Unfortunately, most data and publications from former times are solely available as (often unique) paper hard copies or on data media which are not compatible with today's technology. It is alarming, that these hard-copies are often unique and therefore very vulnerable to breakdown.

The way of storage heavily limits the availability of these important data. Therefore, most of former data series and knowledge are lost from scientific memory.

See Zeller *et al.* (2005) for a good argumentation why it is meaningful to recover historic datasets.

A valuable source of data about the distribution and abundance of living resources within aquatic environments are collected with echosounders. This type of data is a valuable source of information which is used for the assessment of fish stocks, as well as for ecological research. Nowadays, echosounders store their data electronically, which is the basis for fast access and easy analyzes of these data. Until the beginning of 1990s only integrated data were stored on mile basis, i.e. one echo-value per nautical mile (representing all echoes within the water column), usually stored on paper-printouts. This type of storage makes it very difficult to access and analyse these data. Additionally, these integrated data does not contain information about the vertical distribution of echoes.

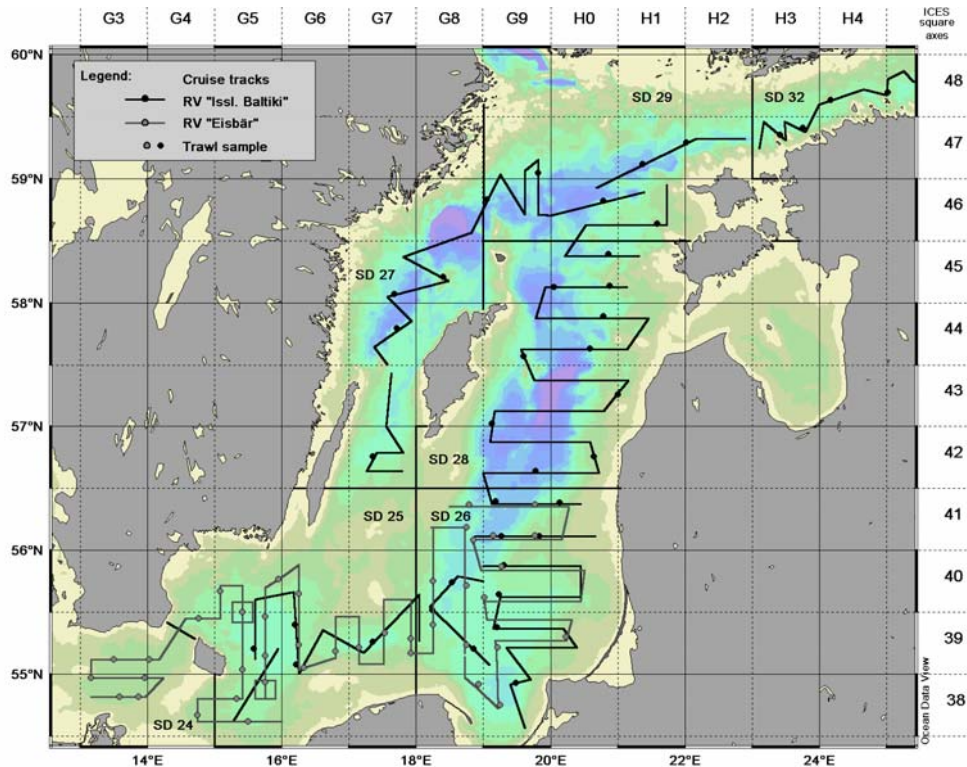


Fig. 1 Example for the spatial coverage of spring acoustic surveys during the 1980s. Hydroacoustic transects and trawl stations for RV «Eisbär» (GDR) and RV «Issl.Baltici» (USSR) are shown for 1986.

In the late 1970s and in the 1980s regularly acoustic surveys were carried out within the Baltic Sea during spring time (see Fig. 1). These surveys were carried out by the former GDR and the former USSR and targeted the commercially and ecologically important clupeid species sprat (*Sprattus sprattus*) during spawning time (May/June). Additionally, herring (*Clupea harengus*) was partly covered by these surveys.

The surveys provide information on a period prior to the regime shift in the Baltic Sea during the early 1990s (Köster *et al.* 2003, Fig. 2). Some fifteen years ago, the pelagic ecosystem changed from a cod-dominated to a sprat-dominated system (Kornilovs *et al.* 2001; Alheit *et al.* 2005). A combination of reduced reproduction success and increased fishing pressure led to a rapid decrease of the stock size of Baltic cod. Simultaneously the sprat stock size increased rapidly. Stabilizing effects are assumed for both regime states, due to predatory control. During the cod-dominated phase, cod controlled the sprat stock, while at present, sprat strongly influence cod reproduction success preying upon cod eggs. Therefore, the spatial and temporal overlap of both species is essential to understand the dynamics of the Baltic Sea pelagic ecosystem. Additionally, the surveys are precursors to the recent International Baltic Acoustic Spring Survey (BASS).

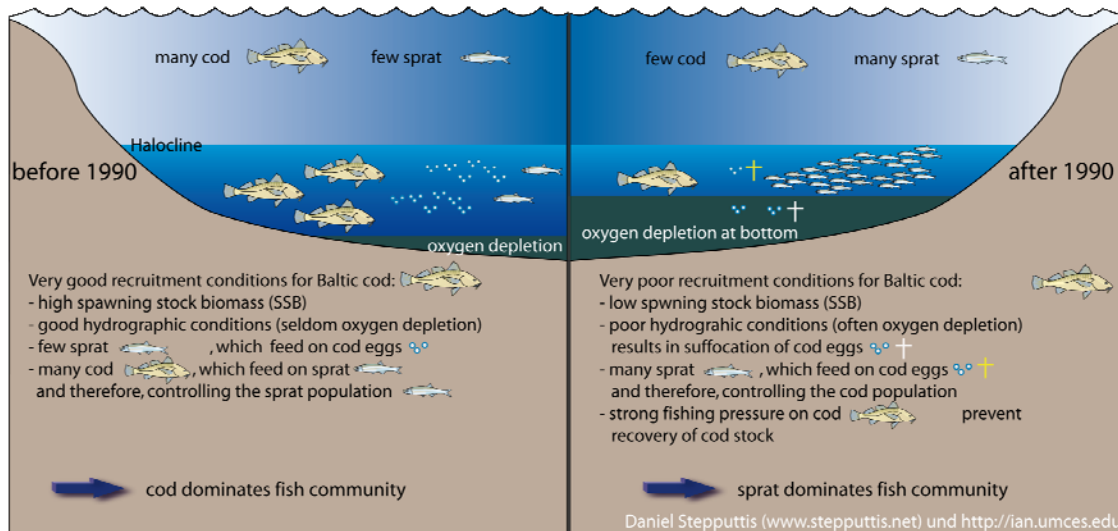


Fig. 2: schematic representation of the regime shift in the Baltic Sea

The unique survey data are currently stored at the Institute for Baltic Sea Fisheries in Rostock (GDR data) and the Latvian Fish Resources Agency (USSR data). The datasets include hydroacoustic data (integrated echo recordings per EDSU and raw-data available as printed echograms, which will be used to extract vertical distributions), as well as catch data (catch composition, length, weight, age and maturity data).

Aim of the project was to digitize, quality-check and deliver the data accompanied by relevant metadata (this process is still ongoing). The proposed data will a) extend the time series of recent surveys (and improve the use of the series as index for the assessment of Baltic sprat), b) enhance the understanding of the Baltic ecosystem (especially spatio-temporal mechanisms connected to past and future ecosystem changes) and c) help to develop the basis to incorporate the clupeid distribution into advanced up-to-date modelling tasks, such as area disaggregated MSVPA.

This manuscript will focus on the recovery of hydroacoustic data and will present a methodology to digitize paper-printed echograms. Additionally first results of the German data will be presented exemplary.

Material and Methods

The recovery of historic, paper archived hydroacoustic data include:

- a) input of integrated echo values (as nautical area backscattering coefficient – NASC) per horizontal sampling unit (EDSU, usually 1 nautical mile (nmi)) and
- b) scanning and analysis of paper printed echograms

Whereas task a) is straightforward, specialized software was needed to analyse scanned paper echograms. This software was developed at the Fraunhofer Institute for Computer Graphics Rostock and offers a high degree of automation. This software and the process of the analysis of paper-printed echograms will be described below.

Requirements of the software

The main task of the developed software is to allow the semi-automatic preparation (optimal rotation of the scanned echogram, detection of ROIs (regions of interest) and horizontal markers (i.e. nautical miles), calculation of thresholds, and others) and analysis (bottom detection, ship noise detection, identification of fish echoes, and deletion of unwanted pixels (e.g. depth markers)) of scanned echograms. To optimize and if necessary to correct automatic analysis, manual manipulation of the results of the automatic algorithms should be possible.

Additionally, the system should have a comfortable graphical user interface to analyze and manipulate the automatically calculated results. It had to be usable on different computer platforms. Furthermore all analyzed data had to be stored in a well-known, manually editable format like XML for textual data and TIF for image data to ensure a platform and vendor independent way of data storage.

Another requirement was to use existing basic imaging technologies being able to easily support image files in the size of 2700 x 3700 gray-scale pixels in a fast way.

Implementation technologies

After doing a user requirements analysis it was decided to use Java as implementation technology as it is available for different operating system platforms and with the Java Advanced Imaging Library (JAI) there exists a robust library which allows handling large image data in a very fast way.

Using the Java Swing Technology it was able to quickly implement a graphical user interface for the presentation and the manipulation of the images and the analysis results.

The processing of scanned echograms follows a four-step approach:

1. Pre-Processing
2. Automatic Analyzing
3. Manual correction of masked pixels
4. Export of relative vertical echo distribution

After some preliminary tests, it was decided to separate the pre-processing from the remaining tasks, as pre-processing showed to very time-consuming. Therefore, it is possible to run the pre-processing on all scanned image files in a separate job over night without having the need of an interaction with the software.

Implemented Algorithms

In the following, we will shortly describe some of the algorithms used during pre-processing and analysis of the scanned echogram images. As scanned documents are grey-scale images (Fig. 3) all image processing steps are based on 256 grey-scale colours or black and white depending on the used algorithms and processing step.

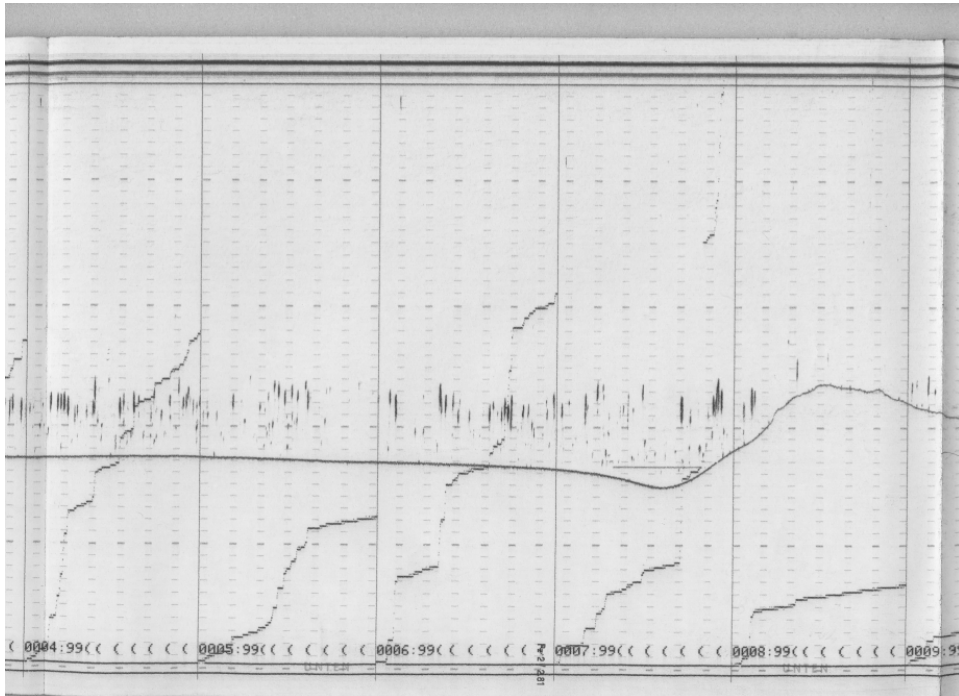


Fig. 3: Example for a paper-printed echogram, which contains five complete nautical miles (grey-scale picture prior to preprocess (e.g. thresholding process, rotation etc.).

Pre-processing

The aim of the Pre-processing step is to prepare the scanned echogram images for further analysis. This task is divided into following subtasks:

1. *Propose an Orientation of the scanned image* - Images are often scanned in wrong orientation. This algorithm tries to find the right orientation of the image.

2. *Proposing initial region of interest (ROI) for the scanned image* - Scanned images contain dark borders around the original sheet of paper that are not interesting for the analysis and could disturb further algorithms. The region of interest focuses onto the echogram itself.

3. *Background / Foreground separation* - Separation of background and foreground separation (black and white threshold) is performed using Otsu's method (Otsu 1979) which is part of the JAI.

4. *Calculating optimal rotation of image* - To easily separate the scanned image into nautical mile images the marking lines should be oriented exactly vertically. The optimal rotation of the image is calculated using an iterative combination of horizontal and vertical projection of the images and the Hough transformation (Hough 1962) which is a special case (Deans 1981) of the Radon Transform.

5. *Finding and proposing nautical mile separators in images* - Using the data from Hough transformation vertical lines are proposed to be markers for nautical mile separation. As there are other vertical miles like folding marks (echograms are typically printed on continuous forms paper) the width of the vertical line is taken into account to identify lines, which separate horizontal sampling units (typically 1 nautical mile).

Automatic pre-processing stops here. A graphical user interface allows the correction of the pre-processing results, especially of found nautical miles and their numbers. Each horizontal sampling unit (1 nautical mile) is stored in a separate image file for further analysis. Preprocessing parameters are stored in .xml-files (see below)

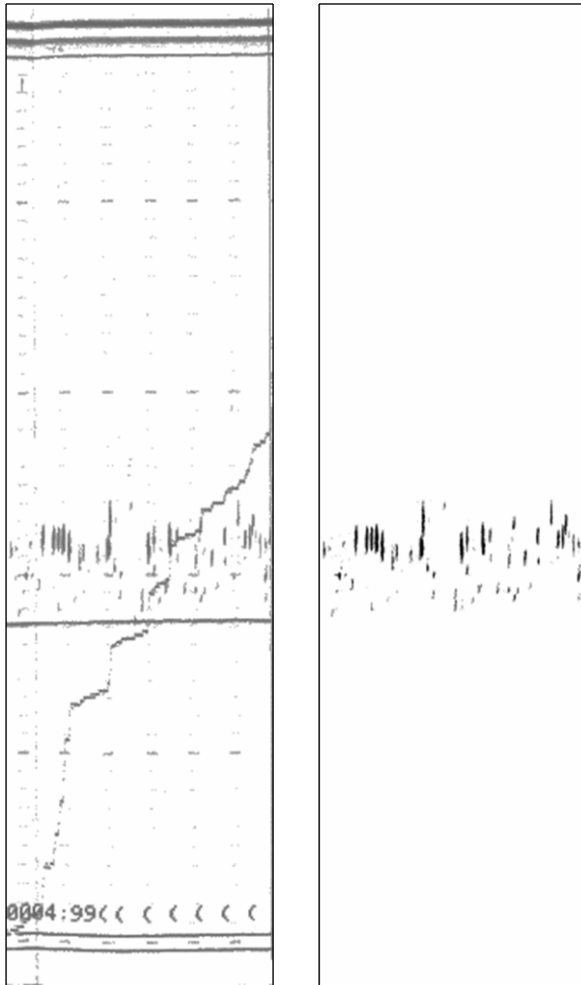


Fig. 4: Example for the automatic algorithm, which separates fish echoes from remaining black pixels in the echogram, such as bottom line, depth markers, integration line, annotation etc.. Left: Original scanned echogram of a single nautical mile in May 1986 saved as .tif-file. Right: corresponding .tif file, which contains solely automatic separated fish echoes, which are used for later analysis.

Analyzing each nautical mile

The software analyzes each nautical mile image and tries to separate fish echoes from the remaining black pixels in the image (Fig. 4).

1. *Erosion and dilatation* - A combination of erosion and dilatation processes clears the mile image from unwanted smaller pixels and connects pixels being of interest for further processing like depth markers, possible bottom lines and fish echoes.

2. *Cleaning mile images from vertical lines* – The next step in the automatic analysis of echograms is to remove the vertical lines in the echogram (nautical mile - separation lines and folding lines of continuous forms paper) from further processing. It was found that all lines

covering more the 10% of image height are not of interest and can be removed including adjoining lines covering 4% of image height. This is done by calculating vertical projection of the mile image and removing the vertical pixel line.

3. *Finding and cleaning the ship noise line* - It was assumed, that the noise line (just below the transducer) is a straight horizontal line in the upper third of the image. Using a horizontal projection of the mile image the third main line crossing the nautical image is assumed as the noise line. This assumption can be corrected using the graphical user interface.

4. *Finding and removing markers* - Every paper printed echogram contains several markers to identify e.g. the different depth intervals. These markers are short lines distributed over the echogram. A combination of horizontal and vertical projections identifies the markers at the crosspoints. As result, these identified markers will be removed from further analysis.

5. *Finding bottom line proposals* – Typically, echograms contain a bottom echo. This echo should be removed from further analysis. This is done by the definition of a bottom line. To find proposals for the bottom line all remaining pixels under the bottom line are collected into separated regions and classified into ground line, depth markers, fish and other black pixels. Information from previous miles is used to decide which bottom line is going to be used for the current nautical mile.

All analyzed results are stored in separate files. The user can correct the results of the automatic analysis algorithms using the graphical user interface (Fig. 5)

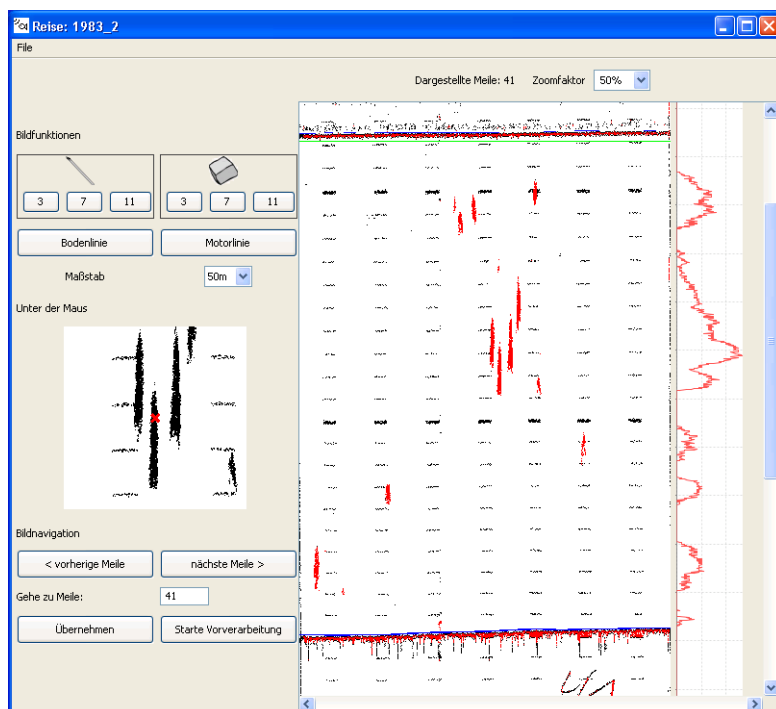


Fig. 5: Graphical user interface for the manual refinement of the results of the automatic analysis.

Result storage

Intermediate and final results are stored in different files. A central .xml-project file describes the current state of the analyzing process of the journey. It contains information on state of the pre-processing of the image files.

For every pre-processed image (usually multi-nautical-mile) file exist an .ini-file, formatted in XML-syntax which contains pre-processing results like threshold parameter describing foreground/background separation, the optimal rotation of the image, proposed region of interest, pixel number for mile separation of the image file, and numbers of the nautical miles this image shows. This file stores intermediate data needed to separate the image files into nautical miles image files. It is produced during pre-processing allowing a timely separation between pre-processing (which is quite time consuming), and analyzing of nautical miles.

After pre-processing and hence mile-separation, three different files for every nautical mile are stored in a sub-folder.

a) One image file per nautical mile which contains the thresholded raw echogram of that specific horizontal sampling unit (nautical mile, Fig. 4 left).

b) The results of the analysis step in another separated .ini-file which is again formatted in XML-syntax. This file contains the results of the analysis steps, such as data on the vertical scaling factor of this nautical mile (i.e. the depth range adjusted in the echosounder during recording: 50m, 100m, 250m), noise and bottom line coordinates, as well as minimum, maximum and average bottom depth. Depth information is calculated by dividing the maximal pixel height of a scanned echogram image (2340) by the vertical scaling factor of the mile.

c) Another black/white image file showing the marked pixels (pixels, which are detected as fish-echoes) as black pixel (Fig. 4 right). This file is modified by manual correction of fish echo detection and is later used to export vertical echo distribution.

After pre-processing, analysis and manual correction of all scanned images and nautical miles, the user can export data into a .csv-file. This file contains data on horizontal mile width (in Pixel), range information (vertical scaling factor), minimum, maximum, and average depths (in m), as well as a sum of black and of white pixels taking into account the noise and the bottom line for each depth meter of each nautical mile. Applications like Microsoft Excel, OpenOffice Calc, SPSS or R can import the exported data.

Of course, these exported vertical distributions are relative measures only. Nevertheless, as mentioned above, a fundamental step of historic hydroacoustic survey recovery is the digitization of integrated echo values per horizontal sampling unit (nautical mile). With the exported relative vertical distribution of echoes from analysed paper echograms, these integrated values can be divided into depth layers.

Results

After finishing the development of the software package, first results of data recovery are available (Fig. 6 and Fig. 7). Since we are still in the process of data recovery, a final presentation of results is not possible so far.

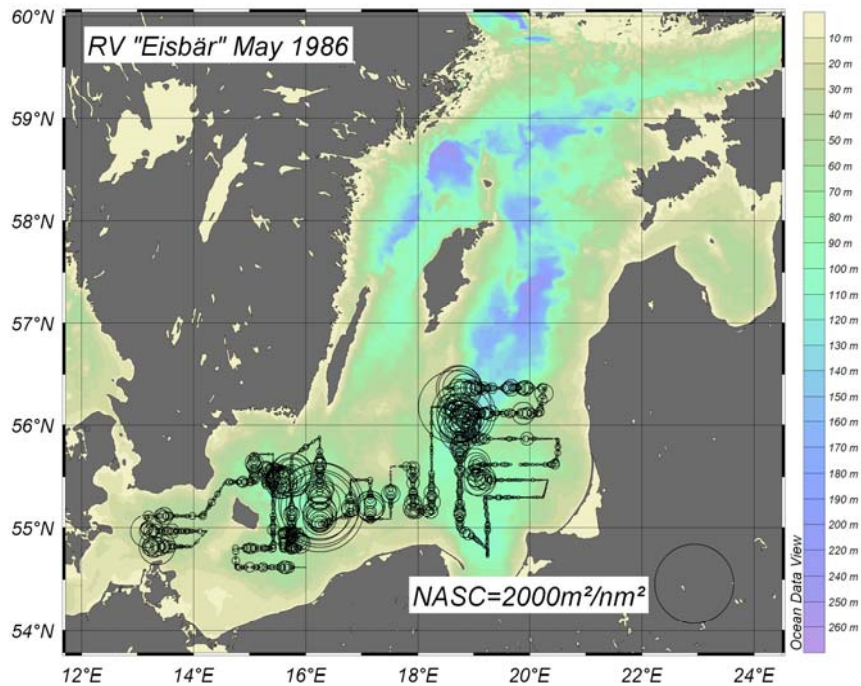


Fig. 6: Example for restored echodata from the 1980s. Shown are horizontal integrated echovalues (nautical area backscattering coefficient – NASC in m^2/nm^2) in May 1986.

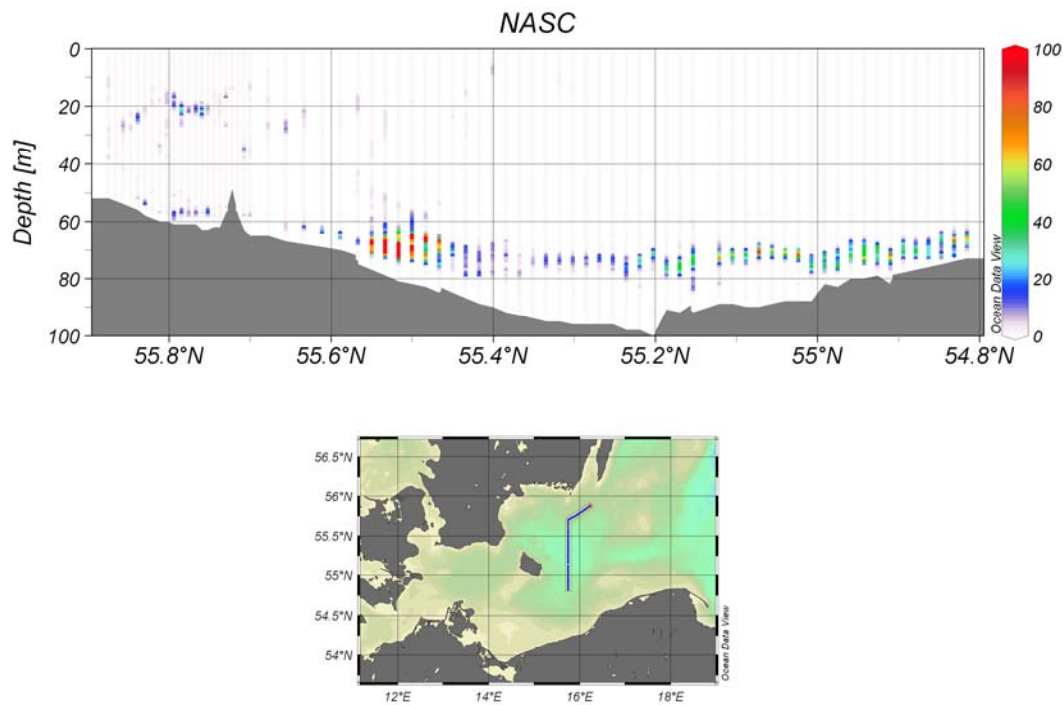


Fig. 7: Example for restored vertical distribution of clupeids along a hydroacoustic-transect through the Bornholm Basin (central Baltic Sea) in May 1986. The basis for this graphic are digitized and analysed paper-printed echograms.

Discussion

The presented method to restore historic hydro-acoustic data, allows to transfer old analogue data which are not accessible for analysis into digital archives and databases. A fundamental prerequisite for a long-term storing is the use of standard file formats for all processing steps (.tif-files for scanned “raw”-echograms and analysed echograms, as well as text-files with XML-syntax for all results and processing parameters).

There is one limitation for the extraction of the vertical fish distribution from the paper-printed echograms: Whereas nowadays echograms store single echo measurements at an intensity level (i.e. the colour encode the intensity of the echo), the analysed paper-printed echograms are limited to black and white colour encoding (i.e. echo (black pixel) or not echo (white pixel)). Therefore the vertical distribution of fish, as extracted from the historic echograms is simply based on the relation of black and white pixel per depth interval without any information about the echo intensity within these black pixels (i.e. echoes which represent fish).

Nevertheless, with these vertical resolved echo data from the mid 1980s we can compare the vertical distribution of clupeids (mainly sprat) echoes before and after the regime shift in the pelagic fish fauna of the Baltic Sea (Köster *et al.* 2003). For that purpose, the models presented at ICES ASC 2007 by Stepputtis *et al.* will be a helpful tool.

We would be happy to be contacted from colleagues facing similar problems. In that way we could offer to share our knowledge and if necessary the software.

References

Alheit J, Möllmann C, Dutz J, Kornilovs G, Loewe P, Mohrholz V, Wasmund N (2005) Synchronous ecological regime shifts in the central Baltic and the North Sea in the late 1980s. ICES Journ Mar Sci 62:1205–1215

Deans SR (1981) Hough Transform from the Radon Transform. IEEE Trans Pattern Analy And Mach Intellig 3(2):185–188

Hough PVC (1962) Method and means for recognizing complex patterns. U.S. Patent 3,069,654, Dec. 18

Köster FW, Möllmann C, Neuenfeldt S, Vinther M, St. John MA, Tomkiewicz J, Voss R, Hinrichsen H-H, Kraus G, Schnack D (2003) Fish stock development in the central Baltic Sea (1976-2000) in relation to variability in the environment. ICES J Mar Sci 219:294–306

Kornilovs G, Sidrevics L, Dippner JW (2001) Fish and zooplankton interaction in the Central Baltic Sea. ICES J Mar Sci 58(3):579–588

Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9(1):62–66

Schlitzer R (2004) Ocean Data View (ODV); <http://www.awi-bremerhaven.de/GEO/ODV>

Steffens M (2005) Refbase; <http://www.refbase.net>

Stepputtis D, Böttcher U, Neumann T, Andersen KH (2007) The distribution of Baltic Sprat (*Sprattus sprattus* L.) and its relation to physical features – observations and models. 15 pp

Zeller D, Froese R, Pauly D (2005) On losing and recovering fisheries and marine science data. *Mar Pol* 29:69–73

Acknowledgements

This study has been carried out with financial support from the German Ministry for Education and Research (BMBF 03F0320D; project: GLOBEC-Germany) and from EUR-OCEANS (<http://www.eur-oceans.eu>) within the project HISBASS (HISToric Baltic Acoustic Spring Survey). Additionally, we like to thank R. Schlitzer for “Ocean Data View” (Schlitzer, 2004), which was used for several graphics and Matthias Steffens for the excellent open-source reference manager “refbase” (Steffens, 2005).

The symbols for Fig. 2 courtesy of the Integration and Application network (<http://ian.umces.edu/symbols>), University of Maryland Center for Environmental Science.