Processing of the Hydrographic Data

1. Download, reformatting, and data selection

CTD, BTL (water samples), and PFL (float) measurements of temperature, salinity, and pressure/depth for the region 47 to 65 °N, 15 °W to 15 °E were downloaded from different data centres: NODC (National Oceanographic Data Center, U.S.A.), ICES (International Council for the Exploration of the Sea, Denmark), and DOD (Deutsches Ozeanographisches Datenzentrum, Germany). Most of the data were from the World Ocean Database 2009 (WOD09; Boyer et al., 2009) of the NODC. All of the data were reformatted and checked for duplicate stations with the same date and position (dlat<0.05°, dlon<0.05°), which were rejected. The temporal distribution of the measurements is shown in Figure 1.

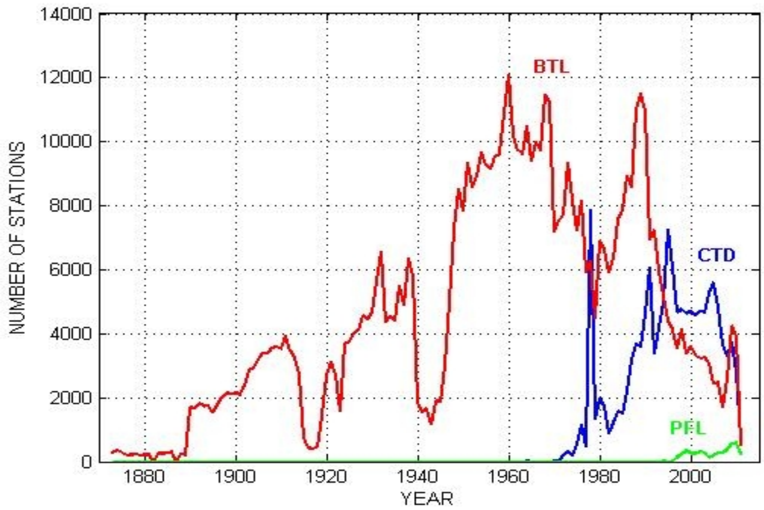


Fig.1: Annual number of stations/profiles with CTD, BTL, and PFL measurements of temperature and/or salinity

The profiles of temperature and salinity were vertically interpolated (after Reiniger und Ross, 1968). Larger data gaps were maintained if the distance between two measurements (>A) and the distance between the interpolation point and the neighbouring measurement (>B) were too large. The following table shows the depth dependant values used for A and B (in m; Z: depth):

Z	A	B
<10	30	3
<100	50	4
<200	75	6
<300	100	8
<500	150	10
<800	200	12
<1200	270	15
<2000	550	20
<5000	1100	30

Data from the following 179 depth levels were selected:

0	to	10 m	every	2 m,
10	to	200 m	every	5 m,
200	to	500 m	every	10 m,
500	to	1000 m	every	20 m,
1000	to	5000 m	every	50 m.

The near-surface levels of 0, 2, and 4 m were filled if there was a measurement in the upper 5 m.

This yielded a data base of 734,415 stations/profiles in the period 1890 to 2011 with about 13.3 Mio. temperature and 12.8 Mio. salinity data. 62 percent of the stations originated from WOD09, 9 percent from ICES, and 29 percent from DOD.

The stations were then sorted in 0.25° x 0.5° latitude-longitude boxes. Figure 2 shows the temporal coverage of the region by filled boxes.

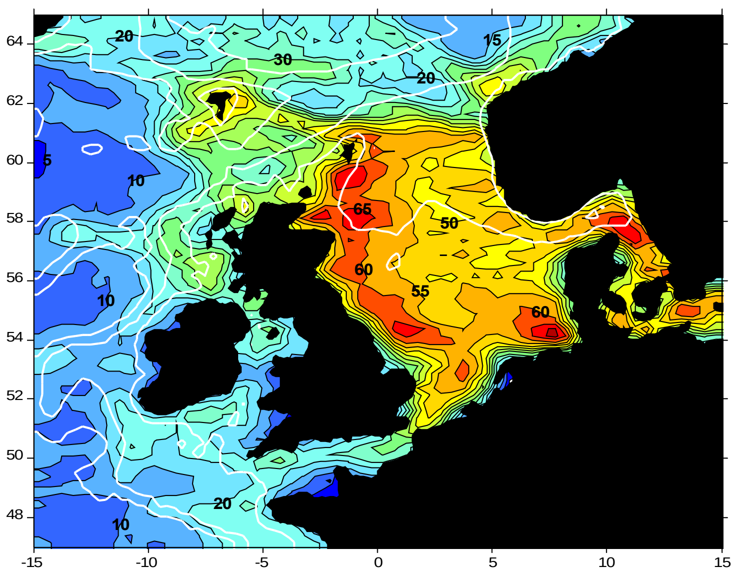


Fig. 2: Temporal coverage by filled boxes with measurements: percent of years with measurements in the period 1900 to 2010. The white lines show the bottom contours of 100, 500, and 2000 m.

For the whole period (122 years) a filled box contains on the average 235 measurements at the sea surface, 77 at 100 m depth, 40 at 500 m, 24 at 1000 m depth, 8 at 2000 m depth, and 5 at 4000 m depth.

2. Quality control

The temperature and salinity data of each box on each depth level were controled for erroneous measurements, using several quality criteria. Bad data were flagged and ignored in the following steps of the quality control.

Data were flagged as bad

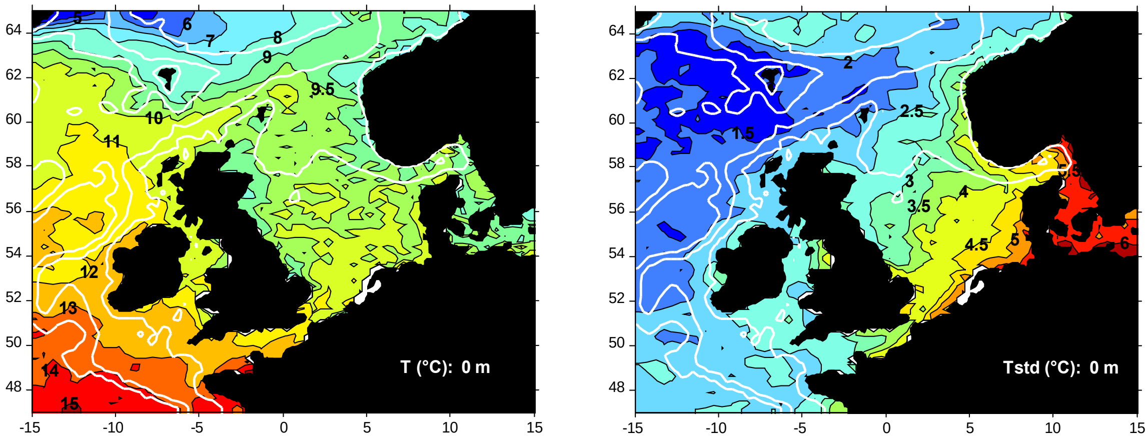
- for total stations in inshore boxes,
- below the maximum water depth of the box,
- for total stations containing more than two thirds of bad data.

At each depth level data in a box were flagged as bad
- outside a realistic overall parameter range separated for the Baltic and North seas,
- outside a specified range of the standard deviation for the total period, for 20-years intervals, and for all months (upper 200 m),
- outside a specified local parameter range (upper 200 m).

Emphasis was put on a sufficient number of data for the computation of the standard deviation in each box, allowing the inclusion of neighbouring boxes if necessary, and on weighting of spatial and temporal inhomogeneities and the distances of the boxes. For each box on a depth level a number of 600 data in the upper 200 m, 400 data up to 500 m depth, and 300 data below were envisaged, whereat the maximum number of surrounding boxes included were set to 11 x 11. Several runs were performed per one range of the standard deviation as well as additional runs with increasing quality standards (reductions of the standard-deviation range, the adequate number of data, and the maximum number of included boxes). In the upper 200 m weighted, left- and right-sided standard-deviation ranges were used if the data of a box show one-sided scattering. Objective of all runs was to reject a maximum number of erroneous data and to keep as much as correct data.

Despite these extensive quality checks the data set still contained erroneous measurements, which cannot be eliminated by statistical methods without loosing a lot of good data. Thus, a visual control of horizontal and vertical data distributions was performed. Finally for each box of a depth level, from the remaining good data the all-data mean and standard deviation for each month (in the upper 200 m) and for the total period were calculated, taking care for the temporal inhomogeneity of the data. Conspicious means and standard deviations were corrected by horizontal and vertical neighbouring data. Then, the optimized means and standard deviations (Figs. 3 to 6) were used to control the data set once more.

Overall, about 2.4 % of the temperature data and 7.9 % of the salinity data were marked as erroneous, leaving about 12.9 Mio. temperature and 11.8 Mio. salinity data in the period 1890 to 2011 for further calculations.



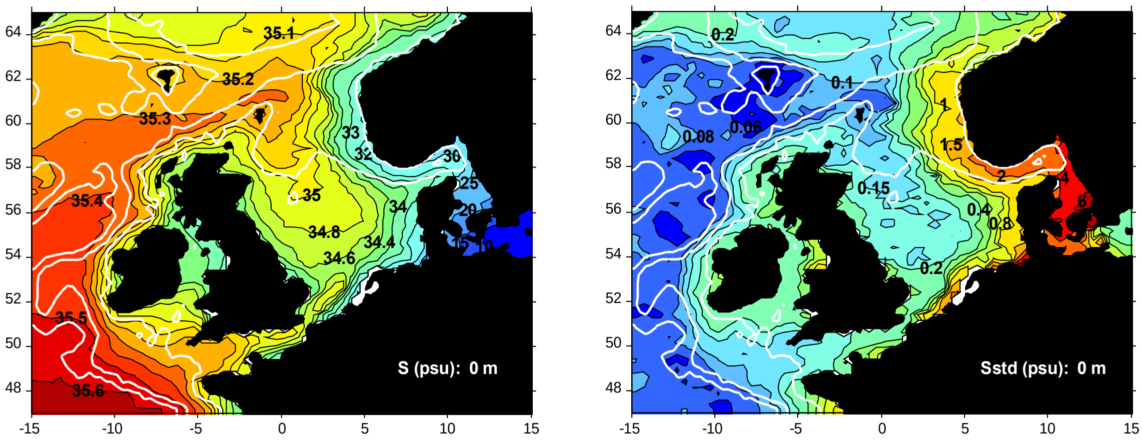


Fig. 3: Mean and standard deviation of all good temperature and salinity data at the sea surface in the period 1890 to 2011.

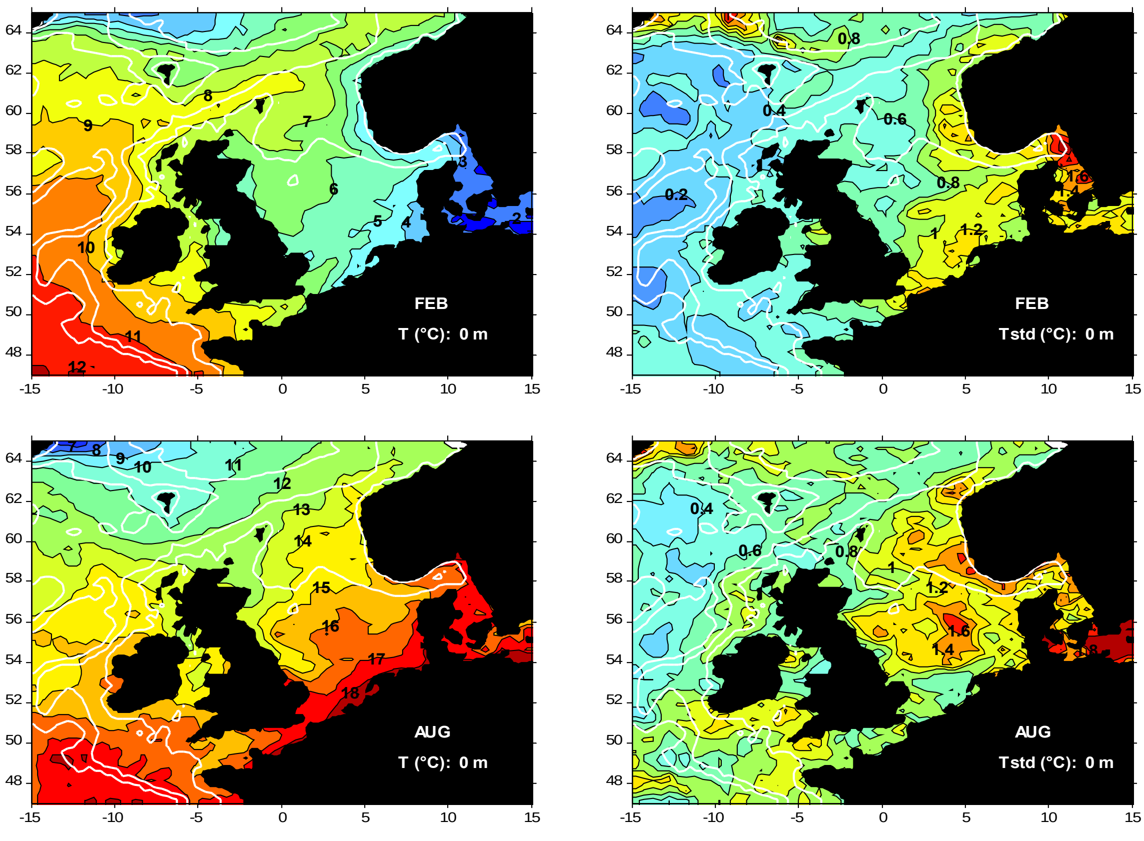


Fig. 4: Mean and standard deviation of all good temperature data at the sea surface for February and August in the period 1890 to 2011.

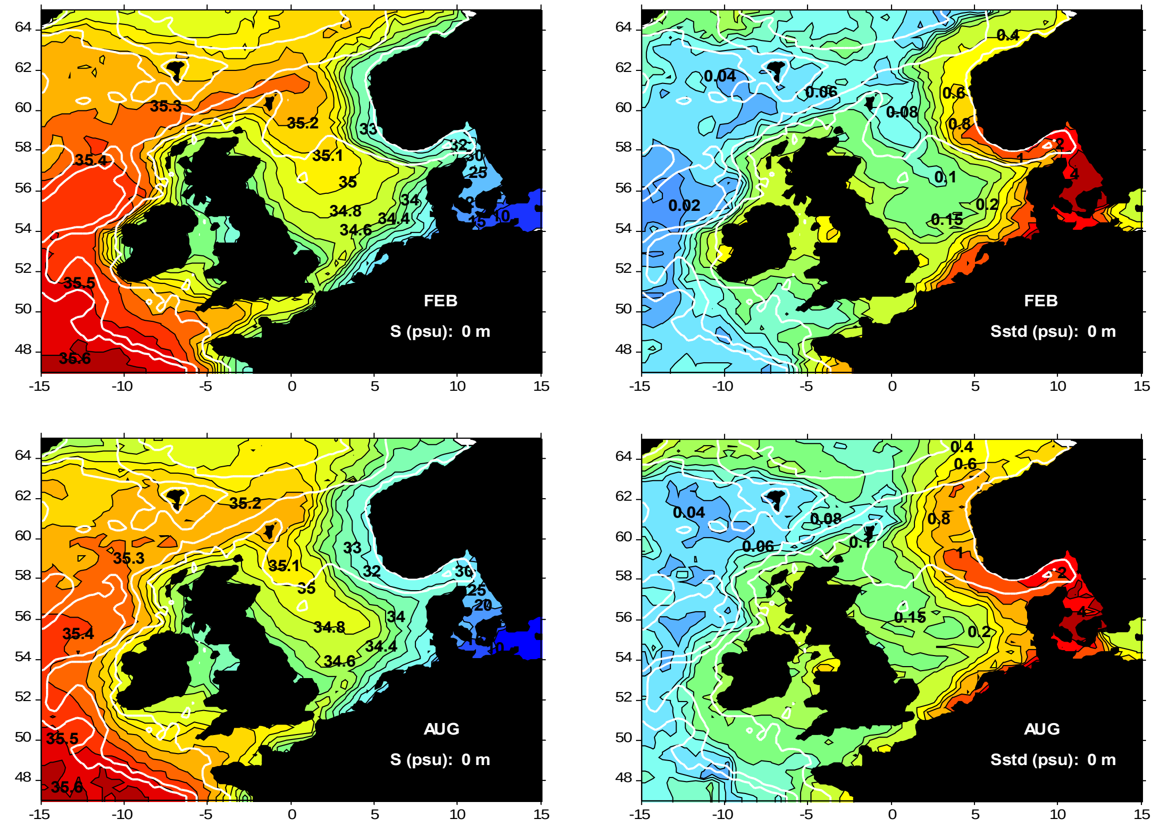


Fig. 5: Mean and standard deviation of all good salinity data at the sea surface for February and August in the period 1890 to 2011.

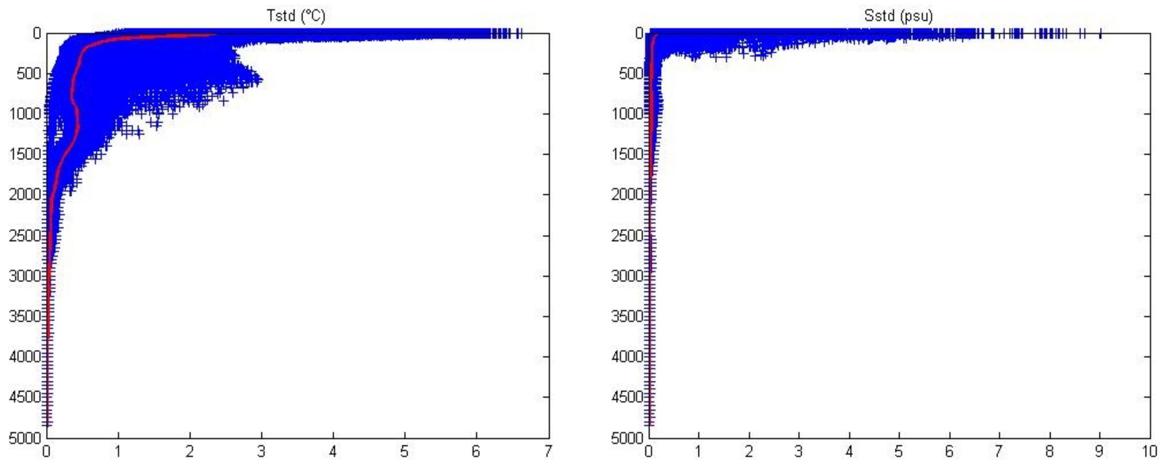


Fig. 6: Vertical distribution (0 to 5000 m depth) of the standard deviations of temperature and salinity for all boxes, using all good data in the period 1890 to 2011 in each box. The median value is shown by the red line.

3. Monthly and yearly means

Before calculating the mean for each year, a polynomial fit of 11th degree for the mean seasonal variations was subtracted from the temperature and salinity data, respectively, of each box on a depth level in the upper 200 m, using surrounding boxes for a better coverage of the seasonal variation if necessary. The same fit was used to eliminate the mean intra-monthly variation. By this method a maximum coverage of the region for the monthly and yearly means was reached.

Finally, monthly and annual means for each year in the period 1890 to 2011 were calculated for each box (0.25° latitude x 0.5° longitude; 72 x 60 boxes) on a depth level (179 levels). The means were ascribed to the box centres:
Latitude= 47.125, 47.375, …, 64.625, 64.875 °N
Longitude= -14.75, -14.25, …, 14.25, 14.75 °E.
After the correction of some conspicuous means the horizontal distributions were smoothed weakly.

Despite the extensive data quality control, some remaining erroneous data could have let to inaccurate means. Additionally, residuals of the annual and monthly variations could have also contributed to some inaccuracies. Thus, especially on small spatial scales the climatology should be used with care. An error estimation for a spatial scale of 0.5° latitude x 1° longitude yielded an uncertainty of 0.15 (0.45) °C for the monthly (yearly) mean temperature and 0.11 (0.16) psu for the monthly (yearly) mean salinity at the sea surface, of 0.13 (0.27) °C and 0.04 (0.06) psu, respectively, at 50 m depth, and of 0.12 (0.22) °C and 0.02 (0.03) psu, respectively, at 100 m depth.

The new climatology was compared coarsely with existing climatologies as HadISST1 (Rayner et al., 2003) and that of the Japan Marine Science and Technology Center (2005), showing no obvious erroneous deviations.

References

Boyer, T. P., J. I. Antonov, O. K. Baranova, H. E. Garcia, D. R. Johnson, R. A. Locarnini, A. V. Mishonov, T. D. O'Brien, D. Seidov, I. V. Smolyar, M. M. Zweng (2009): World Ocean Database 2009, Chapter 1: Introduction, NOAA Atlas NESDIS 66, Ed. S. Levitus, U.S. Gov. Printing Office, Wash., D.C., 216 pp, DVD.

Frontier Research System for Global Change, Japan Marine Science and Technology Center, Japan (2005): Subsurface Temperature And Salinity Analyses by Ishii et al., Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, http://rda.ucar.edu/datasets/ds285.3.

Rayner, N.A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, A. Kaplan (2003): Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res., 108 (D14), 4407, doi: 10.1029/2002JD002670.

Reiniger, R. F. and C. K. Ross (1968): A method of interpolation with application to oceanographic data. Deep-Sea Res., 15, 185-193.