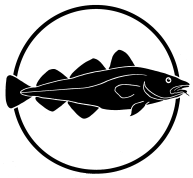
# The Faroese Fisheries Laboratory

**Fiskirannsóknarstovan**



# The NISE Dataset

**By**

**Jan Even Øie Nilsen, Hjálmar Hátún, Kjell Arne Mork,
and Hedinn Valdimarsson**

**Technical Report No.: 08-01**

**Tórshavn** **February 2008**

# The NISE Dataset

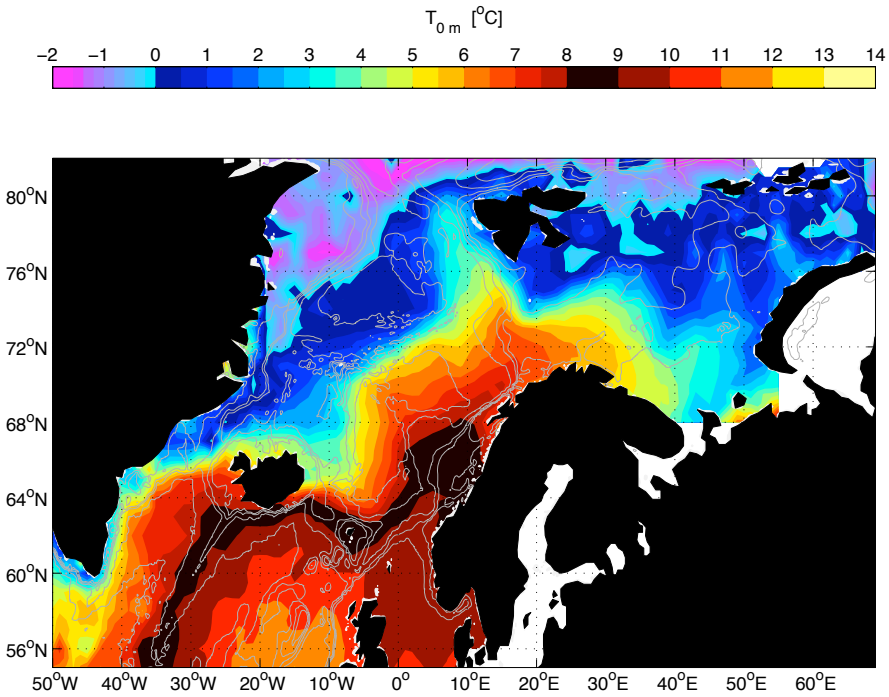
J. Even Ø. Nilsen[1] , Hjálmar Hátún[2] , Kjell Arne Mork[3] , Héðinn Valdimarsson[4]

[1] *Nansen Environmental and Remote Sensing Center, Bergen, Norway*

[2] *Faroese Fisheries Laboratory, Tórshavn, Faroe Islands*

[3] *Institute for Marine Research, Bergen, Norway*

[4] *Marine Research Institute, Reykjavik, Iceland*




Faroese Fisheries Laboratory Technical Report 08-01

FISKIRANNSÓKNARSTOVAN
FAROESE FISHERIES LABORATORY

NERSC

HAVFORSKNINGSINSTITUTTET
INSTITUTE OF MARINE RESEARCH

MARINE RESEARCH INSTITUTE

# Contents

# Chapter 1

# Introduction

Project NISE (Norwegian Iceland Seas Experiment) is a cooperation between the three fisheries (and marine) research institutes in Iceland (MRI), Faroes (IFR), and Norway (IMR), Geophysical Institute (GFI) at the University of Bergen (Norway), and the Nansen Environmental and Remote Sensing Center (NERSC). Project NISE aims at revealing important climate-ecology links in the study region, clarifying the impacts of climatic variation and change on the physical conditions and thereby the fish stocks.

The NISE region includes the subpolar North Atlantic Ocean (east of 55°W and north of 50°N), the Greenland–Scotland ridge, the Iceland and Norwegian Seas. The relatively cold and fresh subpolar gyre and the northward flow of warm and saline waters characterize the region south of the ridge. In the southeastern part of the Nordic Seas warm and saline water from the North Atlantic flow northward while in the western part of the Nordic Seas cold and relatively fresh water flow southwards. The interaction of water masses and their mixing product is essential for the climate and living conditions in these areas.

The observed hydrography will be combined with a "state of the art" numerical ocean model in order to evaluate and develop a tool which can be used by physical and biological oceanographers to describe and potentially predict the marine climate and its impact on biochemical processes, primary and secondary productivity, and migration patterns.

This report concerns the hydrographic dataset in NISE. The purpose is to provide the best possible set of hydrographical data for the NISE project. It is not intended to be *the* database for the Nordic Seas, neither wrt. quality (see Section 2.3) nor public access (see Section 3.1). By adding the collaborating institutes' own data as well as data from some other sources, to a public database (ICES) we aim at providing an optimal dataset for the NISE project participants.

## NISE Contact

Héðinn Valdimarsson, Marine Research Institute, Skulagata 4, 121 Reykjavik, Iceland. Tel.: +354 575 2000, email: hv@hafro.is.

## Institutions

**FFL** Faroese Fisheries Laboratory, Nóatún 1, FO-110 Tórshavn, Faroe Islands (*www.frs.fo*)

**GFI** Geophysical Institute, University of Bergen, Allégt. 70, N-5007 Bergen, Norway (*www.gfi.uib.no*)

**IMR** Institute of Marine Research, PB 1870 Nordnes, 5817 Bergen, Norway (*www.imr.no*)

**MRI** Marine Research Institute, Skulagata 4, 121 Reykjavik, Iceland (*www.hafro.is*)

**NERSC** Nansen Environmental and Remote Sensing Center, Thormøhlensgt. 47, N-5006 Bergen, Norway (*www.nersc.no*)

# Chapter 2

# The Dataset

## 2.1 Data Sources

### 2.1.1 Platforms

The hydrographic variables included in this version (V3) of the NISE dataset[1] are *temperature*, and *salinity*. The measurement of these variables are either done by collecting water samples in (Nansen/Niskin) bottles for later analysis and reading reversible thermometers, or electronically using a CTD (Conductivity Temperature Depth) sonde. In the latter, depth is given by the pressure and from the three variables the salinity can be calculated. The CTD is a relatively modern instrument, but bottle sampling are still used especially for calibration of the electronic data.

The concept "station" refers to a geographical position and time where and when hydrographic measurements are made. Each measurement made at a given station is referred to as a "sample", and a vertical sequence of samples is termed a "profile". A horizontal sequence of profiles constitute a "section". Another type of instrument used in the collection of the data used here, is the float. These are automated CTD-sondes floating at a pre-set depth or density level, surfacing at regular intervals, taking hydrographic profiles during their ascent. These profiles are in this report also referred to as "stations".

### 2.1.2 Contributors

The base of the NISE dataset is the public database maintained by ICES[2]. The ICES database is chosen since it is the most comprehensive for our regions, it has undergone thorough and documented quality control, and it is easily accessed via the WWW.

This base is then supplemented by data from the institutions in NISE: The Faroese Fisheries Laboratory, Institute for Marine Research (Norway), the Marine Research Institute (Iceland), the Geophysical Institute (GFI) at the University of Bergen (Norway), and AARI[3] and NERSC in collaboration (Johannessen et al., 2000; Alekseev et al., 2001).

When found suitable and accessible, data from other sources have been included. These include some sections in the North Atlantic from WOCE[4] and data from ARGO-floats[5] in both the Nordic Seas and North Atlantic.

The number of stations and samples from different sources are listed in Table 2.1.

[1] There was a set of some chemical properties in version 1, but not in later versions. Future inclusion of chemical and other variables will require demand as well as effort from the users.

[2] International Council for the Exploration of the Sea (ICES), H.C. Andersens Boulevard 44–46, DK-1553 Copenhagen V, Denmark (*www.ices.dk*).

[3] Arctic and Antarctic Research Institute (AARI), Department of Ocean–Atmosphere Interaction, 38 Bering str., St. Petersburg, Russian Federation, 199397 (*www.aari.nw.ru*).

[4] The World Ocean Circulation Experiment (*www.woce.org*).

[5] Data from these floats are publicly available from the International Argo Project, which is a pilot program of the Global Ocean Observing System (*www.usgodae.org*).

Table 2.1: Data contributors, data amounts, and cruise labels for identification in the database (See Section 3.3).

| Source | Stations | Samples | Cruise label (first 2 characters) |
|---|---|---|---|
| ICES | 450 180 | 7 842 158 | IC |
| AARI | 34 734 | 539 584 | aa |
| FFL | 8 045 | 805 628 | FA |
| IMR | 1 944 | 67 050 | NO |
| MRI | 2 186 | 71 563 | IS |
| ARGO | 9 269 | 608 701 | RG |
| GFI | 2 779 | 104 027 | BS, FS, GS, NS, t9 |
| WOCE | 488 | 9 051 | 06, A0, A1, A2, AR, ME |
| Total | 509 625 | 10 047 762 | |

## 2.2 Data Coverage

### 2.2.1 Horizontal Coverage

The region covered by the NISE dataset is the northern North Atlantic, the Nordic Seas, and the Barents Sea. The southwestern corner of the domain is at 50°N,55°W, and the northeastern corner at 82.5°N,70°E. The British Isles (10–0°W) and North Sea (0–14°E) are cut in the south at 55°N and 56°N, respectively, due to low NISE focus and large data concentrations. The Baltic Sea, White Sea and Kara Sea are not included either.

The horizontal distribution of salinity samples are shown in Figure 2.1. There are in general more temperature than salinity samples, so these are conservative estimates of the data coverage. The distribution at both 10 m and 200 m are shown to give an indication of the data coverage at different depths, and the latter is chosen for its ecological importance. See next subsection about vertical resolution of the profiles. The density of data varies considerably between different regions of the domain and there is a clear bias towards better coverage in summer (compare middle and lower panels in Figure 2.1). These graphs are presented to give a first impression of the spatial coverage. With Ocean Data View (ODV, see Section 3.3), the user can easily explore the data availability further, e.g. for other depth levels.

### 2.2.2 Vertical Resolution

The NISE dataset consists of both CTD-data and bottle data. CTD-data are usually given at 1 m intervals, while bottles are usually taken at coarser depth resolution, often at what is called "standard depths" (see below). To avoid an unmanageable amount of samples in the database, the CTD-stations are decimated by resampling them at coarser depth intervals. The intervals are finer than standard depths to preserve some of the improved vertical resolution given by CTD. The original CTD-data are available as a separate dataset (see Section 3.2).

To give an impression of the different depth resolutions used, the vertical distribution of samples for each contributor is shown in Figure 2.2. The depth resolution differs for each source. These can be categorized as *regular* and *irregular* depth intervals:

***Regular***

**FFL** Samples are selected from CTD data at 5 m depth intervals

**IMR** No data are available from this source between 1960 and 1980. Prior to 1960 all data are sampled by bottles at the following depths: 0, 10, 25, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1200, 1500, 1800, 2000, 2500, 3000, 3400 m[6]. After 1980, samples are selected from CTD data at 5 m intervals.

**GFI** Data before 1976 are sampled by bottles at the following depths: 0, 10, 25, 50, 75, 100, 150, 200, 300, 400, 600, 800, 1000, 1200, 1500, 2000 m[6]. After 1976 data are interpolated from CTD data onto 10 m intervals (0–500 m), 50 m intervals (500–1500 m), 100 m intervals (1500–2200 m) and 400 m intervals (2400–4000 m).

[6]too few to be visible in Figure 2.2b/c

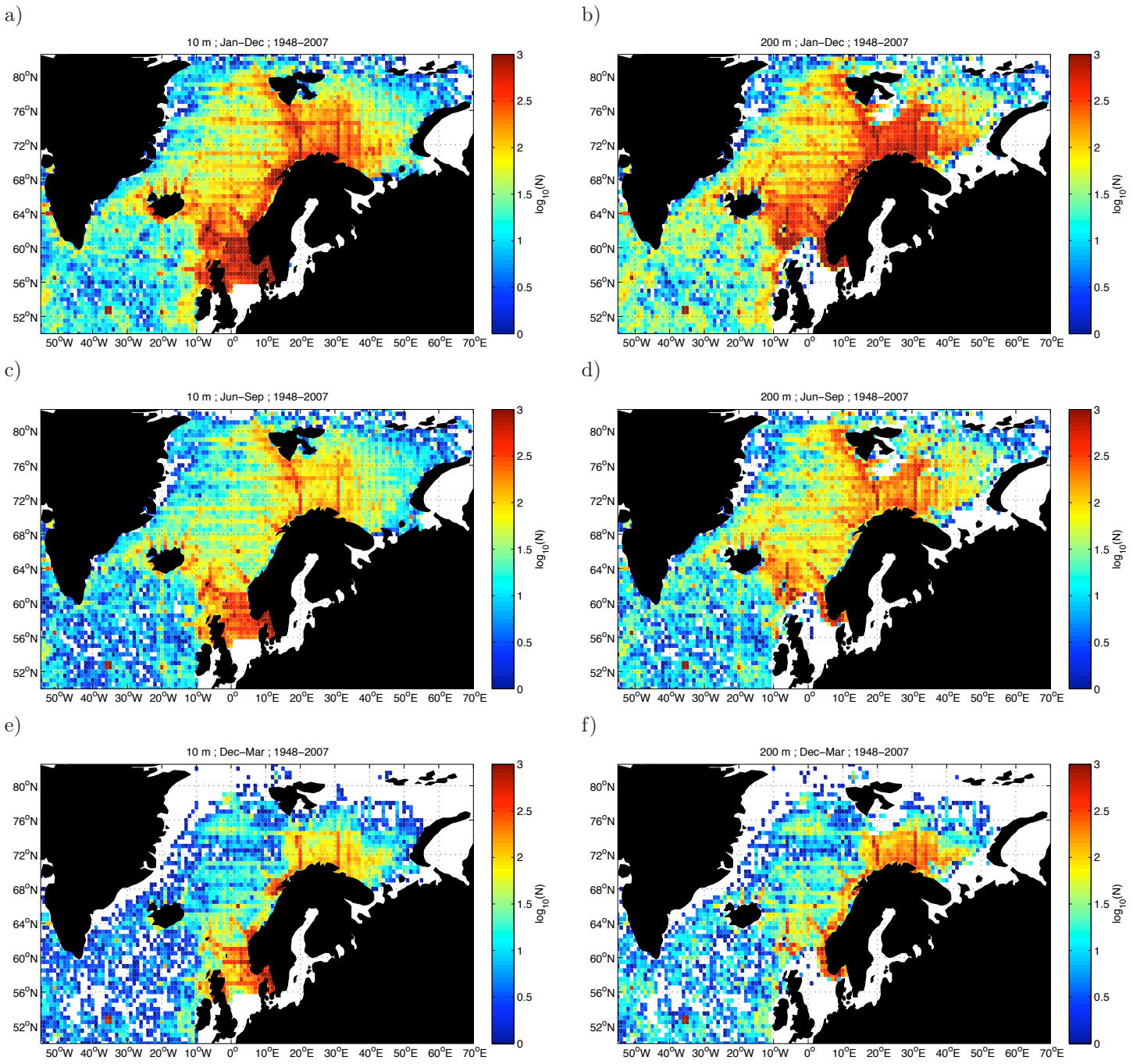



Figure 2.1: Number of *salinity* samples ($N$) at 10 m (left panels) and 200 m (right panels) for all (a,b), summer (c,d), and winter data (e,f). Colour (logarithmic scale) indicate number of data inside the 0.5° latitude by 1° longitude squares. White areas have no data.

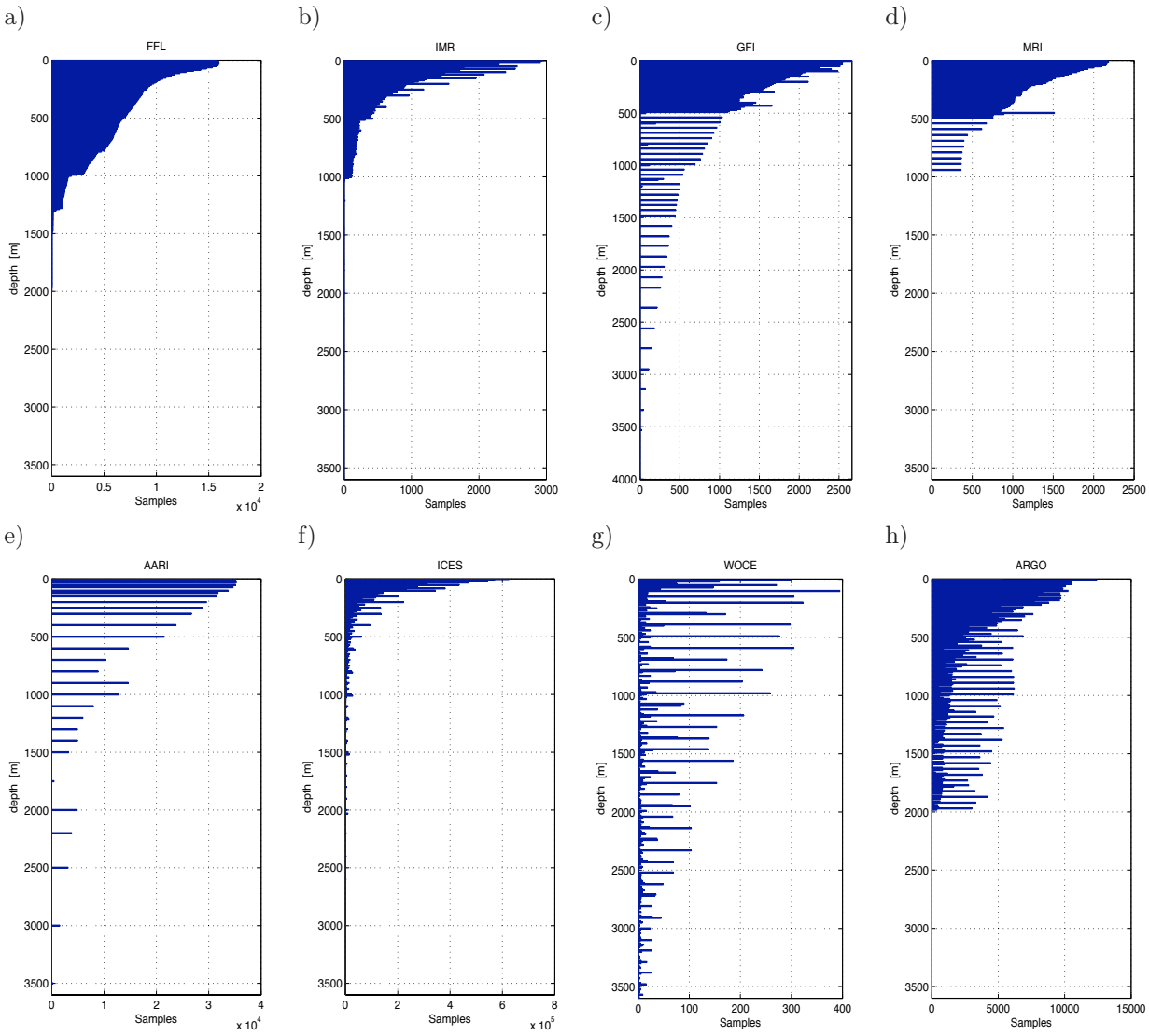



Figure 2.2: Distribution of sample depths (salinity and/or temperature) from the different contributors.

**MRI** Only data after 1994 are delivered directly from Iceland. These are interpolated from CTD data onto the same depths as was done for the GFI data (above).

**AARI** These data are delivered on Standard (Levitus) depths: 0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2200, 2500, 3000, and 3500 m.

***Irregular***

**ICES** Data are delivered both as bottle data and decimated CTD-data. As a coarse rule the vertical depth intervals are 20–50 m in the upper layers (0–400 m) and 100–200 m below 400 m.

**WOCE** Data are delivered sub-sampled from CTD data. As a coarse rule the vertical depth intervals are 20–200 m in the upper layers (0–400 m) and 200–300 m below 400 m.

**Argo** Data are delivered sub-sampled from high-resolution profiles. As a coarse rule the vertical depth intervals are 10 m in the surface layers (0–100 m), 20–25 m in intermediate layers (100–400 m) and 50 m below 400 m.

Depths in the NISE-dataset are given in meters. When depths have been measured and given in decibars in the sources, they have been converted.

### 2.2.3 Temporal Coverage

The time frame for this dataset is from 1900 to 2006. The distribution of stations over the years are shown in Figure 2.3. Naturally, data is scarce in the first 50 years, and due to the usual data sharing restrictions also in the latest few years. The two database sources used in the NISE-dataset (ICES and AARI) contain the early period data, while the supplementary contributions ensure higher concentrations in the latest decades.

The monthly distribution for the NISE dataset is shown in Figure 2.4. As already indicated in Figure 2.1, there is a bias towards more data during summer than winter. Twice as many stations have been taken in the calmer April–September period, than during October–March. This is the general trend for most of the different data contributions (some shown in Figure 2.5).

## 2.3 Quality Control

### 2.3.1 Data Quality

The contributing institutions all have their own routines for quality checking prior to delivery. However, documentation for this is, in addition to being difficult to access, beyond the scope of this report. Users of the NISE dataset will have to contact the relevant institutions for details. All data from the ICES-database has gone through ICES' own quality checking routines (See *www.ices.dk/ocean*), in addition to their sources' own control. For AARI, the other data collection used in the NISE-dataset, details of quality control is unknown. The additional data from WOCE are documented at *www.woce.org*, while the Argo are documented at $www.coriolis.eu.org//cdc/argo/argo-dm-user-manual.pdf$. Samples flagged as "questionable" or "bad" are removed before inclusion into the NISE-dataset.

The NISE data group have performed no thorough quality assurance, due to its limited resources and large size of the dataset. However, a simple outlier test and visual inspection have been done, in order to eliminate clearly erroneous data, using the following ranges:

**Depth** 0–5000 m

**Temperature** -2°C–25°C

**Salinity** 1–36.1

The lower salinity limit have removed some near-shore data, but these are few and outside the NISE focus. A further, but still coarse, visual inspection in T-S space have been performed to eliminate more outliers. Closer inspection for specific regions can be performed by the user, e.g. using ODV (see Section 3.3).

| All final quality checking prior to publication, is ultimately the responsibility of the user! |
|---|

### 2.3.2 Redundancy

When adding supplementary data to such a comprehensive database as ICES', some stations are likely to exist in both datasets. These are called duplicates, or redundant data. ODV checks data on import against the data existing in the collection, but unless all metadata match (Cruise, Station, time, position, etc.), duplicates will be imported. Since the identification labels often differ between the different contributors, this is clearly the case for the NISE-dataset.

Removing these duplicates in the NISE-dataset is done using our own programmed routines (using MATLAB) after collection is completed and exported to text file from ODV. The duplicate-removal procedure is as follows:

**Detection** Duplicate *stations* are detected when time difference is less than 1 day and longitude and latitude differences are less than 0.01°.

**Identification** Cruise labels of the stations in each group of duplicates are checked to identify their contributors (see Table 2.1).

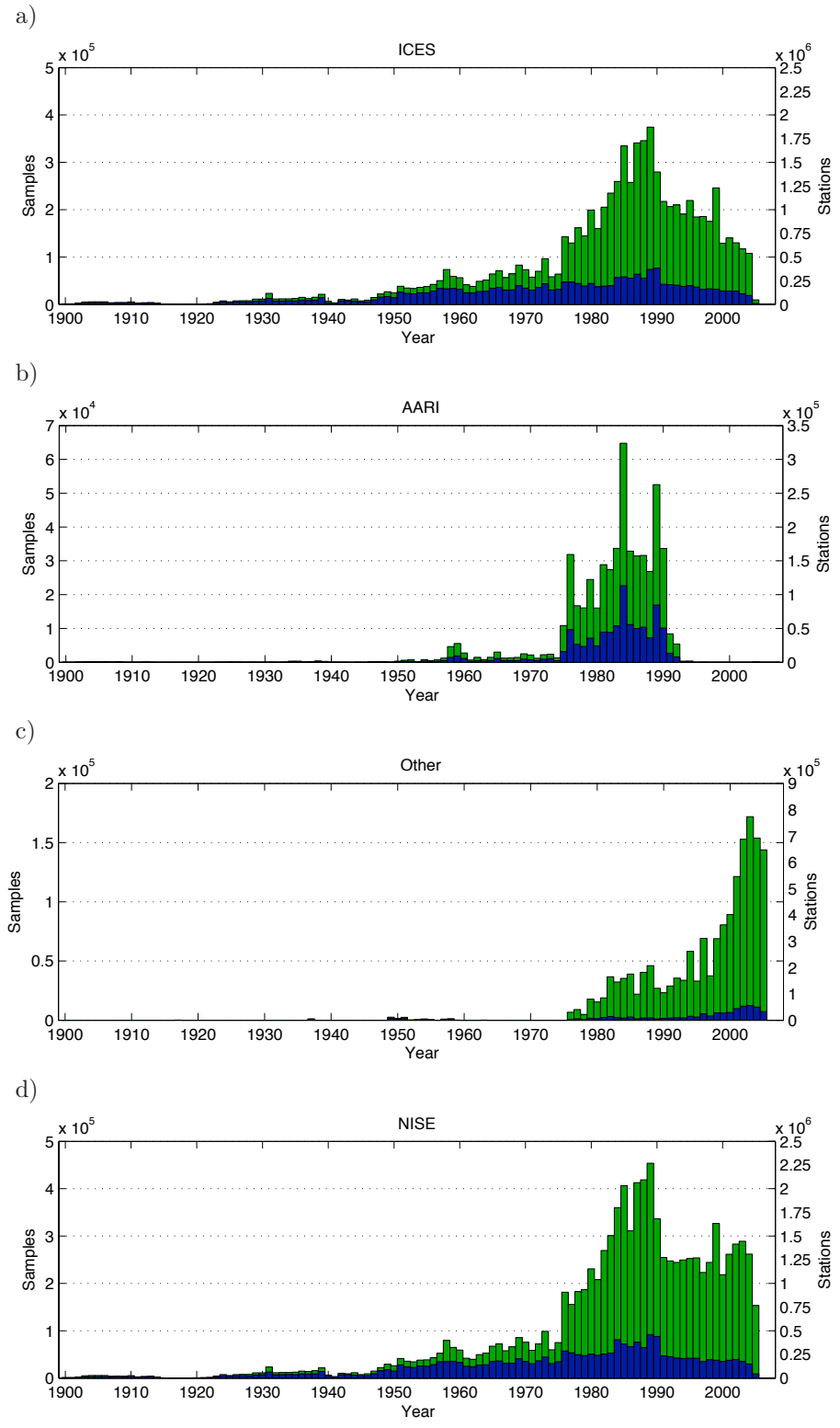



Figure 2.3: Distribution over years of number of samples (higher green bars, left axis) and stations (lower blue bars, right axis) for the data from ICES (a), AARI (b), other sources (c), and the NISE-dataset in total (d).

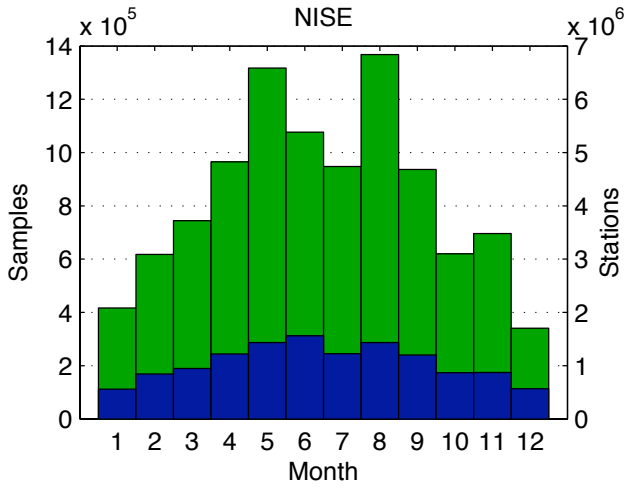



Figure 2.4: Distribution of number of samples and stations in months of year for the NISE-dataset in total. Colours as in Figure 2.3.

**Selection** Stations from only one source are kept, according to the following prioritisation: ICES, FFL, MRI, IMR, GFI, WOCE, ARGO, and AARI.

**Elimination** The other duplicates are not flagged and kept, but are simply eliminated from the dataset.

The list of priority is set up by the following considerations: i) ICES is the chosen base for the NISE dataset; ii) The NISE participants and GFI are first hand sources for additional data; iii) WOCE and ARGO are external add-ons; iv) AARI is a collection in itself, thus an indirect source for any duplicates of the above data.

There were about 150 000 duplicate stations found, mostly because of stations aready being present in the ICES-database. The same goes for AARI, which contained duplicates from all sources apart from MRI, WOCE, and ARGO. There were also a relatively high degree of redundancy (44 000) in the IMR-data, reflecting their high degree of input to ICES. WOCE had some 100 duplicates in ICES, while ARGO had no duplicates at all.

Duplicates from the same contributor are kept, since these may be multiple stations from the same cruise on the same day. Thus, our detection criteria of 1 day is not likely to have caused removal of useful shorter term time series.

The stations are defined as series of samples having identical cruise-, station-,and type-labels (bottle or CTD), as well as monotonically increasing depth (ODV sorts exported data by these parameters). This means that samples occurring at same depth are considered part of the same station, and thus not treated as possible duplicates. It is judged as unlikely that these are duplicate samples, and it would in any case have little effect on the dataset.

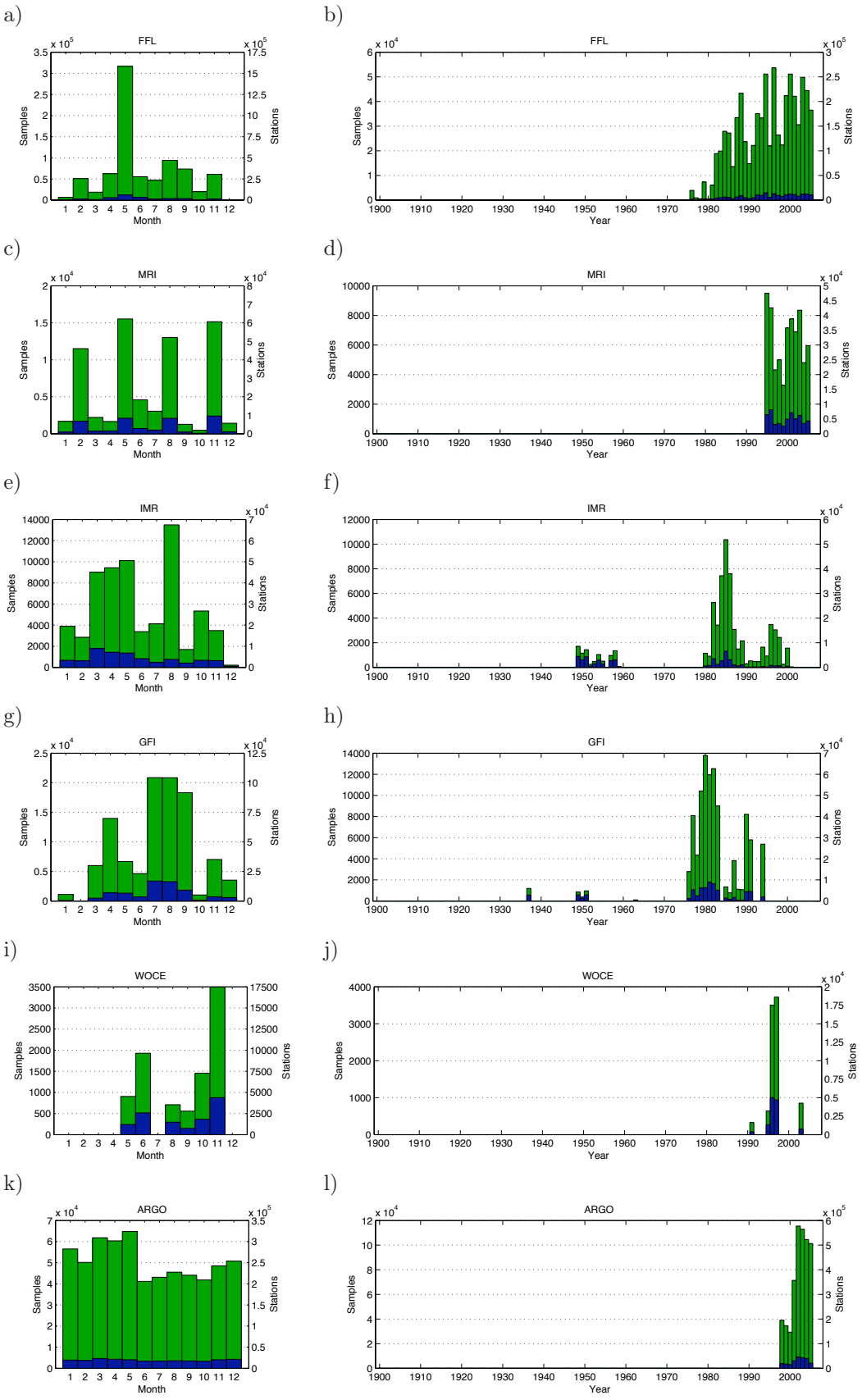



Figure 2.5: Temporal distribution of data from the separate (other) contributors. Colours as in Figure 2.3.

# Chapter 3

# Data Usage

## 3.1 Terms for use

This dataset is created for use within the NISE project. To fulfill the goals of the project, it will be necessary to have access to as much as possible of the data acquired through observations. This includes data from ICES and other data banks, but also from sources where access may be restricted. In order to secure the widest possible access without infringing upon originator rights, the following guidelines apply to all data, shared within the project:

- Data exchanged within the project must not be applied for commercial use.
- NISE data are only open to scientists involved in NISE.
- If a publication relies heavily on a dataset from a specific originator, then the data originator should be included in the author list.
- Any use of data for publication, etc., should bear the following acknowledgement:

  > *"Data were provided by: The Marine Research Institute, Iceland; Institute of Marine Research, Norway; the Faroese Fisheries Laboratory; the Arctic and Antarctic Research Institute, Russia, and Geophysical Institute, University of Bergen, Norway, through the NISE project."*

The reference to use for technical background is the present report:

> Nilsen, J.E.Ø., H. Hátún, K. A. Mork, H. Valdimarsson (2008). The NISE Dataset. *Technical Report 08-01, Faroese Fisheries Laboratory, Box 3051, Tórshavn, Faroe Islands.*

## 3.2 Data Organisation

The main dataset is the "bottle" dataset. This set contains all data, bottle- and CTD-samples, but with the latter decimated to "standard" or "bottle" depths (see Section 2.2.2). This makes up a comprehensive and at the same time manageable dataset on most computer systems. As an extra service for more detailed studies, a CTD-only base is provided.

The data is stored as the following ODV-collections (see Section 3.3) from which they can be easily exported to other formats if necessary:

`nise_bottle_v3` The main NISE dataset

`nise_ctd_v3` Supplementary CTD-data (compiled from ICES, FFL, MRI, and GFI)

NISE participants can obtain the data by contacting the authors of this report (i.e. the NISE data group) who will provide a CD/DVD of the current version of the dataset (v3).

## 3.3 About using Ocean Data View

The Ocean Data View (ODV) software can be downloaded from *http://odv.awi.de*. If you are not a registered user, you have to register first. You are then supplied with a user name and password. Then click on the ODV version, which you want to download and you will get access to several files to download, e.g. text files with installation info and a zip file containing the ODV installation.

ODV is a software package for the interactive exploration, analysis and visualization of oceanographic and other geo-referenced profile or sequence data. It gives a nice data overview from where data subsets are easily selcted and exported to ASCII (text) for import to other software for further analysis. Furthermore, in ODV you can examine your data coverage by maps and simple graphics. Subsets of data can be selected[1] by time periods and geographical domains (incl. polygons), as well as cruise and station labels (see Table 2.1).

Since the NISE dataset is both created, relies heavily upon, and is distributed using ODV, all users of the NISE dataset should reference Ocean Data View in your publications as follows:

*Schlitzer, R., Ocean Data View, http://odv.awi.de, 2006.*

[1]Note: In the selection criteria dialog of ODV, the station label lower range needs to be set to * for all data to show.

# Acknowledgments

Data were provided by: The Marine Research Institute, Iceland; Institute of Marine Research, Norway; the Faroese Fisheries Laboratory; the Arctic and Antarctic Research Institute, Russia, and Geophysical Institute, University of Bergen, Norway, through the NISE project. ICES, WOCE, and Global Ocean Observing System is thanked for making data publicly available. Thanks to Schlitzer (2006) for the ODV software. Financial support from NAF (Nordic Working Group for Fishery Research within the Nordic Council of Ministers) is gratefully acknowledged.

# Appendix A

# Climatology of the Nordic and Barents Seas

To give an impression of the contents and possibilities of the NISE dataset, horizontal mean fields of temperature and salinity at some standard depths are presented.

To create this climatology, the NISE-data have been divided into bins of 2° longitude by 1° latitude, and vertically into bins centered around standard (Levitus) depths. In every bin, mean values unbiased by the seasonal varying amounts of data (e.g. Figure 2.4) are created by first calculating monthly means and averaging these. For presentation purposes, the coarse fields are interpolated onto a 1° longitude by 0.5° latitude grid (Figures A.1 and A.2). See Nilsen (2003) for details on the bin mean averaging procedure used.

The horizontal fields show the typical situation with the warm and saline Atlantic Water spreading into the western part of the area, and the positions of the main fronts are revealed.

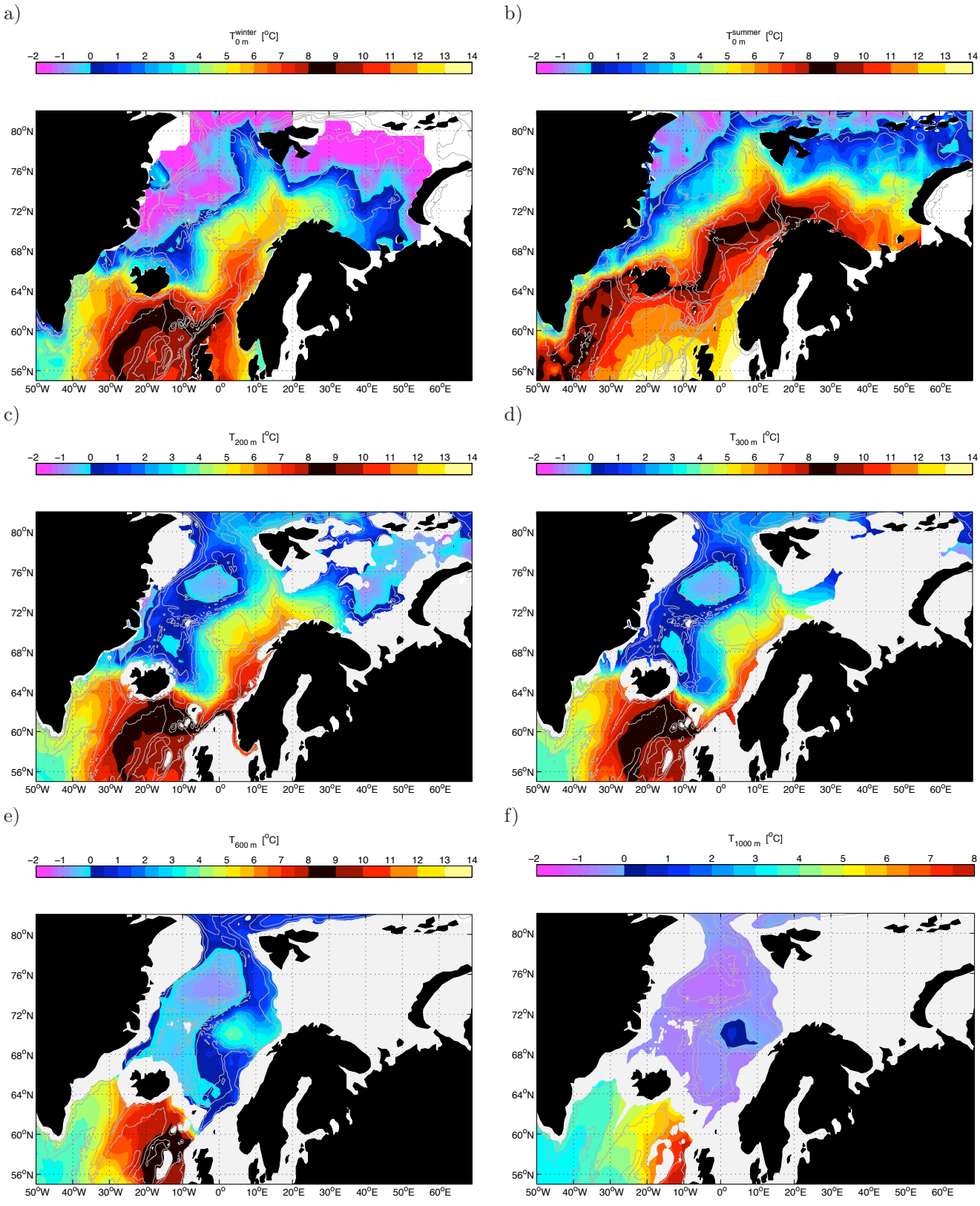



Figure A.1: Mean temperature at selected depths. The seasonally influenced surface layer is represented by both winter (a; Dec–Apr) and summer (b; Jun–Sept) fields, while deeper layers are represented (c–f) by their full-year means.Isobaths are drawn in grey for 250 m, 500 m, and every 1000 m.

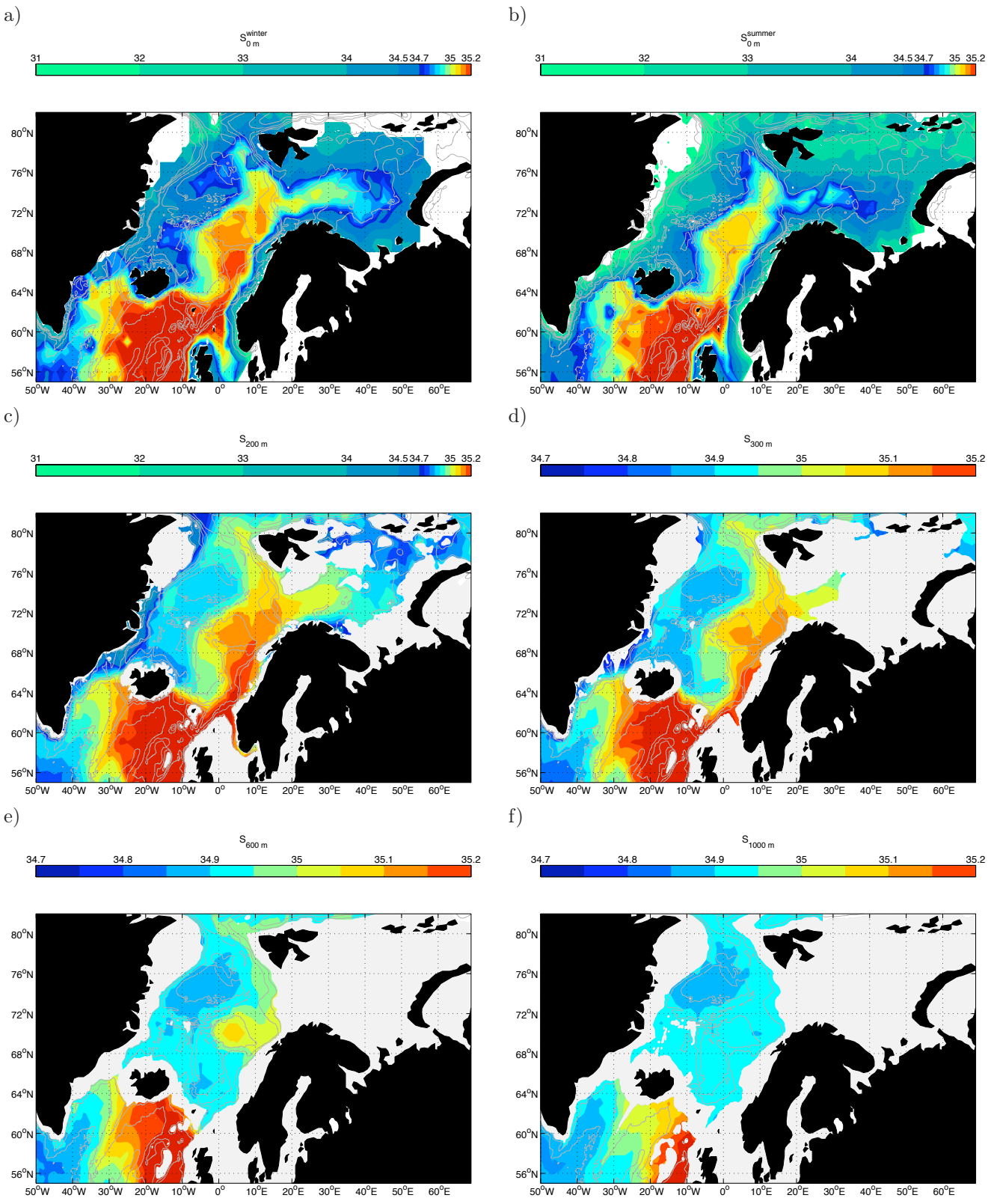



Figure A.2: Mean salinity at selected depths. The seasonally influenced surface layer is represented by both winter (a; Dec–Apr) and summer (b; Jun–Sept) fields, while deeper layers are represented (c–f) by their full-year means.Isobaths are drawn in grey for 250 m, 500 m, and every 1000 m.

# Bibliography

Alekseev, G. V., Johannessen, O. M., Korablev, A. A., Ivanov, V. V., and Kovalevski, D. V. (2001). Interannual variability of water mass in the Greenland Sea and the adjacent areas. *Polar Research*, 20(2), 201–208.

Johannessen, O. M., Alekseev, G., Ivanov, V., Korablev, A., Kovalevsky, D., Myakoshim, O., and Zakharov, V. (2000). *Detection and Modelling of Greenhouse Warming in the Arctic and Sub-Arctic. Oceanographic Data Analysis: Greenland Sea*. Technical Report Task 3, INTAS 97-1277, Nansen Environmental and Remote Sensing Center/Arctic and Antarctic Research Institute, Thormøhlensgt.47, 5006 Bergen, Norway.

Nilsen, J. E. Ø. (2003). *Aspects of the Atlantic Flow through the Norwegian Sea*. Dr. scient. thesis in physical oceanography, Geophysical Institute, University of Bergen, Allégt. 70, 5007 Bergen, Norway. Reports in Meteorology and Oceanography, University of Bergen, Report No. 3-2003 (www.nersc.no/∼even/doc/thesis).

Schlitzer, R. (2006). Ocean data view. http://odv.awi.de.