PRECEDING PAGE BLANK NOT FILMED

N93-24931

Data Archaeology at ICES

H.D. Dooley

Introduction

This paper provides a brief overview of the function of the International Council for the Exploration of the Sea (ICES), both past and present, in particular in the context of its interest in compiling oceanographic data sets. Details are provided of the procedures it adopted to ensure adequate internationally collaborative marine investigations during the first part of the century, such as how it provided a forum for action by its member states, how it coordinated and published the results of scientific programmes, and how it provided a foundation, through scientists employed in the ICES Office, for the establishment of the original oceanographic marine databases and associated products, and the scientific interpretation of the results. The growth and expansion of this area of ICES activity is then traced, taking into account the changing conditions for oceanographic data management resulting from the establishment of the National Data Centres, as well as the World Data Centres for Oceanography, which were created to meet the needs of the International Geophysical Year (IGY). Finally, there is a discussion of the way in which the very existence of ICES has proved to be a valuable source of old data, some of which have not yet been digitized, but which can be readily retrieved because they have been very carefully documented throughout the years. Lessons from this activity are noted, and suggestions made on how the past experiences of ICES can be utilized to ensure the availability of marine data to present and future generations of scientists.

The International Council for the Exploration of the Sea

ICES was formed in 1902 with a primary remit to provide scientific advice on the state of fish stocks in the Northeast Atlantic, in particular the fish stocks of the North Sea and the Baltic. It was recognized that the accuracy of this advice depended very much on a detailed knowledge of the state of the marine ecosystem. Hence the initial ICES activities leaned heavily towards the establishment of an international data set of physical, chemical, and biological oceanographic parameters which would form the basis of products to be made available for the use of fishery biologists.

The basic structure of ICES was then as it is today, that is, it had a number of committees which covered most marine disciplines (but not marine geology), whose members were chosen experts in relevant fields. There were eight Member Countries in 1902, which have now grown to 17, all coastal states of the North

Atlantic, the North Sea, and the Baltic. At the time of its formation, an Office to coordinate these activities was set up in Copenhagen, which was manned by scientific staff who provided services to the ICES community, not least of which was the provision of internationally consistent data sets. Hydrography (i.e., physical and chemical oceanography) was a very important component of the early ICES, with one of the two senior staff being a physical oceanographer (Martin Knudsen), who was in charge of the "Service Hydrographique" for more than 40 years.

The present-day ICES is basically similar to the one that was created in 1902. It still has an office in Copenhagen, now called the ICES Secretariat which has a staff of 25 professional and secretarial staff. The number of ICES Committees has expanded to cover most areas of relevance to marine living resources. In addition to these "science" committees, ICES has in recent decades created two advisory committees, one concerned with the management of fish stocks, the other concerned with the management of the marine environment, which was created as a result of the increasing threat posed by man's activities. Although science-based discussion and research remain a vital part of ICES, mainly via its other committees these two committees now form the basis and incentive for most of ICES' applied activities. Below these committees are the working and study groups which at present number about 100. These groups provide the expert analysis and opinion on which ICES advice and initiatives are based. They make use of various databases that are maintained at ICES, including those on fishery biology, fishery statistics, marine contamination, and oceanography. Every effort is made, as before, to ensure the consistency of the international data in these databases by intensive quality assurance of the data, backed up by the rigorous application of standards and frequent intercalibration exercises.

Early Coordination of Oceanographic Programmes

Following the approval of a programme for the hydrographic and biological work in the northern parts of the Atlantic Ocean, the North Sea, and the Baltic Sea in 1901, an International Council and an International Laboratory to provide Standard Seawater were established. The main function of this Council was to sustain international collaboration in marine research, and to steer and publish the work. The nature of this research underwent continuous scrutiny, and there is no doubt that the data now available to the marine community were determined by the recommendations made by ICES. The progress in understanding and the evolution of scientific programmes were carefully documented in ICES reports (specifically the ICES "Publications de Circomstance"), which summarize on a grand scale the management of what can be seen as a 50-year-long oceanographic project.

A typical example of the direction which ICES provided can be found in the 1924 annual report of the Council, which stated:

> The programme outlined in 1922 is recommended for continuation with certain changes to be introduced as a result of the progress of research: for instance the collection of water samples on the English light vessels to be carried out every fourth day instead of daily, and the saving attained to be allotted to the institution of a new Atlantic Route.

ICES was also attentive to the need for new techniques for measuring the ocean, and many new programmes had their infancy in ICES resolutions. For example, also in 1922, ICES was to recommend

> Experiments to be carried out by England with drift indicating instruments, which provided the foundation of many years of research on the water circulation of the North Atlantic.

Publication of Oceanographic Data

In 1902 ICES, in accordance with the recommendations of the International Conference that led to its formation, commenced publication of the Bulletin Hydrographique, a responsibility delegated to the Hydrographic Department of the ICES Bureau, which was to become known as the ICES "Service Hydrographique" in 1926. This publication, which appeared annually until 1956 — data were published retrospectively for the war years — lists data from hydrographic stations in the North Sea, Baltic Sea, and the Northeast Atlantic, including waters around Greenland. In addition to the profile data, many hundreds of thousands of entries on ship track temperature and salinity were also listed. These data include the so- called "Route" data, a series of ship routes across which observations were regularly taken. In many cases ICES received the original thermograph charts from these cruises, from which the data were listed and published along with the actual charts. ICES placed much emphasis on the need to "rescue" and publish data in order to make them available as soon as possible, a policy which it pursued with considerable energy and single-mindedness, as depicted in the following extract from the report of the 1935 meeting of the ICES Hydrographical Committee:

> The "Dana" was commissioned from June 6th to 22nd on which date she was rammed by a German trawler and foundered near the Dogger Bank. Most of the hydrographical data were saved and will appear in the Bulletin Hydrographique

Figure 1 shows the extent to which the "route" and light vessel data have been digitized, based on the data sets available at NODC and ICES. Clearly there are serious shortfalls that must be rectified as soon as possible.

From the beginning, these data lists were accompanied by details of methods and analysis, as well as track charts and distribution maps. In addition the staff of the Service Hydrographique, under the leadership of Professor Martin Knudsen, would utilize and interpret these data in a way that could be of use to fishery scientists in particular. Atlases were prepared and products relating to time series of anomalies of temperature and salinity were regularly produced, as well as accounts of the scientific analysis of the data.

Although the primary data concerned temperature and salinity, the Bulletin Hydrographique also contained listings of chemical and biological (plankton) data, as well as descriptive details of the method of collection, for example. In the beginning all these data were listed by country and date, but from 1936 on the data were geographically sorted by date of measurement following the introduction of the Card Index (see below).

Although most of the hydrographic data collected in the ICES area up to 1956 were published in the Bulletin Hydrographique, this was not always the case. The alternative location for publication was always noted by ICES in its annual reports, thus simplifying the task of ensuring that these data have remained available to the scientific community. For example, in its 1924 report, ICES noted:

> A *Bulletin Hydrographique* containing the material for 1924 will be published on the same lines as the preceding four years, and to contain also the English Channel surface observations of temperature and salinities, if these observations are not published in the report of the Committee of the Atlantic Slope.

Indeed, the reports of the Atlantic Slope Committee, which were produced during the 1920s and 1930s, contain a wealth of hydrographic data for this period for the continental shelf and deep water areas to the south and west of the UK, Ireland, France, Spain, and Portugal. Although the interests of the Atlantic Slope Committee were primarily related to fisheries (in particular mackerel) questions, this series of reports provides additional detailed scientific interpretation of the hydrographic data published therein.

In the mid-1930s ICES agreed to an initiative of its then Administrative Secretary, Cdr Nellemose, to "automate" its hydrographic and biological data, by creating a Central Card Index of all surface and profile hydrographic data that had been published in its Bulletin Hydrographique and other ICES publications such as the Atlantic Slope Committee reports and also of data published at the National level. The hydrographic element of this index is still preserved in the ICES archives, with each card carefully cross-referenced to its source. In 1956, when this activity was stopped along with the cessation of the publication of the Bulletin Hydrographique because there was now too much material being produced each year for the Service Hydrographique to handle, there were almost 200,000 cards which were arranged by position, year, and month. Until this date the Card Index had served as an aid to the arranging of the entries in Bulletin Hydrographique, but more importantly it provided a "database" which was to form the basis of a large number of scientific publications and products describing oceanographic conditions in the ICES area.

From 1957 on, hydrographic data sets submitted to ICES were prepared on punch cards, using the ICES punch card (hydro-chemistry) format which was designed to be similar to the NODC format with whom regular exchange of data was anticipated. Both of these formats are still in use today. For a while the punch card data were published in a new series called the ICES Oceanographic Data Lists (IODL), but this publication survived only six years because of the diminishing interest of the scientific community in using printed data lists. The demise of the IODL was also partly due to the establishment of the World Data Centres, which were originally created to receive data collected during the International Geophysical Year (IGY) in 1958. The establishment of the WDCs implied that all future data would be made available to the scientific community via National Data Centres, which would perform quality control and reformatting of the data. ICES had, however, a continuing need for data in order to meet its commitments with regard to products, and still relied on its Member Countries to supply data, presumably through their National Data Centres. In practice, however, since product preparation also required quality control of data from a perspective not available at the National Centres, e.g., comparisons of multi- national data sets, and because only a very small number of Data Centres equipped to handle oceanographic station data actually came into being in the ICES area, existing arrangements have to a large extent remained in force. Unfortunately, the loss of a commitment to submit data to ICES by institutes in Member Countries has resulted in only a very small percentage of data collected in recent decades being put into the public domain.

This abandonment of data lists in the early 1960s meant that there was now a clear gap in information about who was collecting research cruise data, and where. This situation was remedied in 1967 by the Intergovernmental Oceanographic Commission (IOC) which introduced the first edition of a cruise inventory called "Report on Scientific Cruises and Oceanographic Programmes (ROSCOP) which has subsequently passed through to revisions and is now called the Cruise Summary report (ROSCOP III). In order to identify the data required by ICES scientists, and for the preparation of products, ICES immediately took on the task of coordinating the return of completed forms from its member countries, and published edited versions of them in the successor to the IODL series, ICES Oceanographic Data Lists and Inventories (IODLI). These were published in manuscript until 1975, but a rapid expansion in their use led to the necessity of publishing them in microfiche form until 1983. The handling of these forms was further streamlined in 1984 by their digitization. Now portable computer software is available to allow searches of these forms, which currently number in excess of 11000. This database now serves as the primary means of searching for data sets that have not been made public, and also serves as a catalogue of cruise data sets which have been submitted to ICES.

Completeness of Digitized Data Sets Available at ICES

Given the combination of the thorough published documentation of the past and the relative completeness of its ROSCOP database of the past 25 years, ICES is in a pretty good position to assess what data collected over the past Century have not been released by their originators for use by the international community.

ICES itself was not involved in the huge undertaking of digitizing the data following the establishment of the NODC (Washington) in the early 1960s. Most of these data were of ICES origin, published in ICES publications, but it was only very recently that ICES was able to assess how successful the digitization of data in the early 1960s was in terms of the data it originated. Indeed, ICES had available only the temperature and salinity component of the Bulletin Hydrographique data set, since it possessed no chemical oceanography data prior to 1957. In addition it did not hold any of the data generated by its Atlantic Slope Committee in the 1920s and 1930s, which was a most worrying state of affairs. However, it is now clear that at least 80% of these data are digitized, and why these data could not be accounted for at ICES until recently cannot be readily explained.

It is apparent, thanks to the efforts of NODC, that the majority of the "old" ICES data are digitized. Inevitably errors must have crept into such a huge operation and these are coming to light during the cross-checking being undertaken by ICES at present. Most of the errors arise not from too little data, but too much, as many cruise data sets have found their way into the data set not once, or twice, but on occasion as much as three times. Most of these duplications and triplications have arisen from the incorrect assignment of ship codes, but are relatively easy to locate and therefore eliminate. Systematic cross- checking also reveals a number of omissions. For example the Danish data sets for 1924 and 1926 have been missed, and much of the data and errata published in various appendices to the Bulletin Hydrographique have been missed, for example the 132-station cruise of the Norwegian ship "Heimland" to the Denmark Straits in 1932. The precise documentation of this period means that the identification of missing data for this period is relatively straightforward, but the presence of some data at ICES which must have been received at some time from NODC, but which is no longer at NODC, is indeed puzzling.

Examples of Time Series Data at ICES

In recent years ICES has been expending effort on digitizing Russian data sets collected mainly as part of Soviet section and Weather Ship programmes. These data had been delivered to the World Data Centre A (WDCA) and, in spite of their limited accuracy (salinity is accurate to about 0.05), it was felt that they would provide a valuable contribution to our understanding of time series processes in the North Atlantic. A total of about 10000 deep water stations were digitized as part of this cooperation with WDCA. Data from Ocean Weather Ship "Charlie", which was situated about 200 miles south of Iceland in the Central North Atlantic, was believed to be a particularly valuable data set. The scale of effort was immense, with hydro-chemistry stations being worked six times daily, every day from mid-1975 to the end of February 1990. For ten years prior to 1975, this station was occupied somewhat less frequently by the USA. Thus as a time series, the data from OWS "Charlie" represent one of the most intensive that is likely to become available, and providing a unique opportunity to evaluate the time scales of variability in this part of the Atlantic. Figure 2 is presented as an example of this. It shows monthly averaged data for the period 1975-1985 at Station Charlie, and variability at various time scales is clearly apparent. It is also clear that these scales of variability would not necessarily be apparent from less intensive sampling programmes. The most dominant scale in these data is one at ten-year periodicity, which occurs in phase throughout the water column at Charlie. This scale is confirmed by the full Charlie data set from 1965 to 1989. Throughout this period, salinity and temperature were at their minimum in mid-decade, a fact which influenced the interpretation of the mechanisms giving rise to the mid-1970s low-salinity anomaly that had major implications for the ecology of the Northeast Atlantic at that time. The fact that the mid-1980s minimum in temperature and salinity at Charlie was even greater in the mid-1980s implies that factors that resulted in the mid 1970s anomaly have not yet been fully explained. Figure 3, compiled from data stored at ICES, shows the magnitude of the mid-1970s anomaly in the Faroe—Shetland Channel, and illustrates the unique nature of this event.

Data sets such as those available for OWS "Charlie" will almost certainly prove useful in determining the intensity of sampling required in order to establish without ambiguity the natural scale and range of oceanic processes. From a data centre point of view, knowledge of such ranges is important for assessing outliers in data sets. In the English Channel, for example, data being currently reported to ICES show salinity levels more that two standard deviations in excess of the values collected throughout the twentieth century. The problem is that the intensity of sampling in this area is so poor that the statistical basis for these data is unreliable. However, the data received are confirmed from two sources, one of which was also from a cruise that worked other areas where comparisons with a third ship were possible. This led to the conclusion that the salinity outliers of 1991 in the English Channel are indeed valid data, and that values of 35.55 in salinity are some 0.2 higher than observed before. The poor overall data coverage, however, leads us to draw back from the conclusion that these values are "exceptional".

In locations of suppressed high-frequency variability, relatively sparse data sets are probably sufficient for filtering out the effects of ocean change. For example, the deep water in the Skagerrak, an area in the North Sea lying between Norway, Sweden, and Denmark, is a fairly good indicator of changing conditions in the North Sea, and responds well to changing density as a result of changes in temperature and salinity, especially in winter. At this location one value a year is probably enough to give confidence in the distributions as shown in Figure 4, which shows the annual mean values of temperature and salinity at 600 m depth in the Skagerrak. Both the temperature and salinity distributions tell a story that can be readily linked to climatic events during the century. For example, the well-documented cold winters in the northern North Sea are clear, as well as the effects of the sequence of four exceptionally warm winters that are occurring up to the present. Similarly, the salinity picture shows a gradual decline in peak salinity values since 1978. This year is an important one from the point of view of conditions in the Baltic Sea, as it was the last year in which there was a substantial inflow of "new" water to the Baltic. As a result the Baltic is gradually stagnating, with large areas of hydrogen sulphide forming, much to the detriment of the commercial fisheries in the area. The Skagerrak is the "open sea" end of the Baltic Sea, and conditions there may have a major impact on Baltic exchange processes.

Current Acquisition of Oceanographic Profile Data - Successes and Problems

At the present time the ICES Oceanographic Data Centre has received and quality-controlled more than 6000 profiles of CTD and nutrient data for the year 1990 alone. More than half of these are for the North Sea. This suggests that current data exchange is in a fairly healthy state. However, there is no room for complacency as the true situation is one about which we should be extremely concerned. In particular the following should be noted:

a) Many of the data were accrued because the submitter was obliged to contribute to specific projects for which ICES is the project data centre, viz. SKAGEX (Skagerrak Experiment), NANSEN (North Atlantic—Norwegian Sea Exchange), the NSTF (North Sea Task Force) and the IYFS (International Young Fish Survey). These data are at presently available only to participants in these projects.

b) Exclusive of these projects, cruise summary reports indicate that less than 10% of collected data have been submitted for this year.

c) Less than 10% of the data were submitted via the approved routes of the International Ocean Data Exchange (IODE).

d) No internationally agreed formats, apart from the ones approved by ICES, were used.

e) But, apart from submissions from four institutes, the formats used by submitters mostly demonstrated a serious lack of understanding of how data sets should be structured. The mode of submission generally bordered on anarchy in spite of the detailed specifications given to participants. Many of these data sets had to be submitted more than once because of ambiguous structures. A total of 34 different formats (structures) were used to submit data, the majority on DOS diskettes.

f) Quality control procedures are being dominated by the need to check the conversion from the user formats. A particular problem lies in checking users' undefined "missing data" fields, which have been defined in as many as five different ways within a data set.

Thus, although relatively large amounts of station data profile data have been made available to ICES in a timely way, the acquisition of these data sets is fraught with problems, which implies that the present successes cannot be sustained indefinitely. Too many data sets are still missing, a situation that has been common throughout the last two decades.

Efforts must be made to wrest these data sets from institutes, but options for doing this are limited. One option open to ICES is to withhold service from sources which are requesters but reluctant suppliers, and to exercise this authority. Other possibilities must be considered, as this one option is not sufficient to lead to the release of most of the data sets. In addition to the classical CTD/bottle station data, efforts should be made immediately to retrieve similar data from other instrument platforms, for example seasoar- and batfish-type data. It should also be recalled that the area of activity to which our predecessors in the first part of the century applied so much energy, was the acquisition of surface temperature and salinity data obtained whilst ships were on passage. We should learn from this example, and attempt to acquire the huge volumes of data collected by thermo-salinographs which are routinely obtained and calibrated on most research-vessel cruises. This instrument has been used routinely for almost 20 years, yet not one data point has been acquired by data centres from what must be several million miles of data.

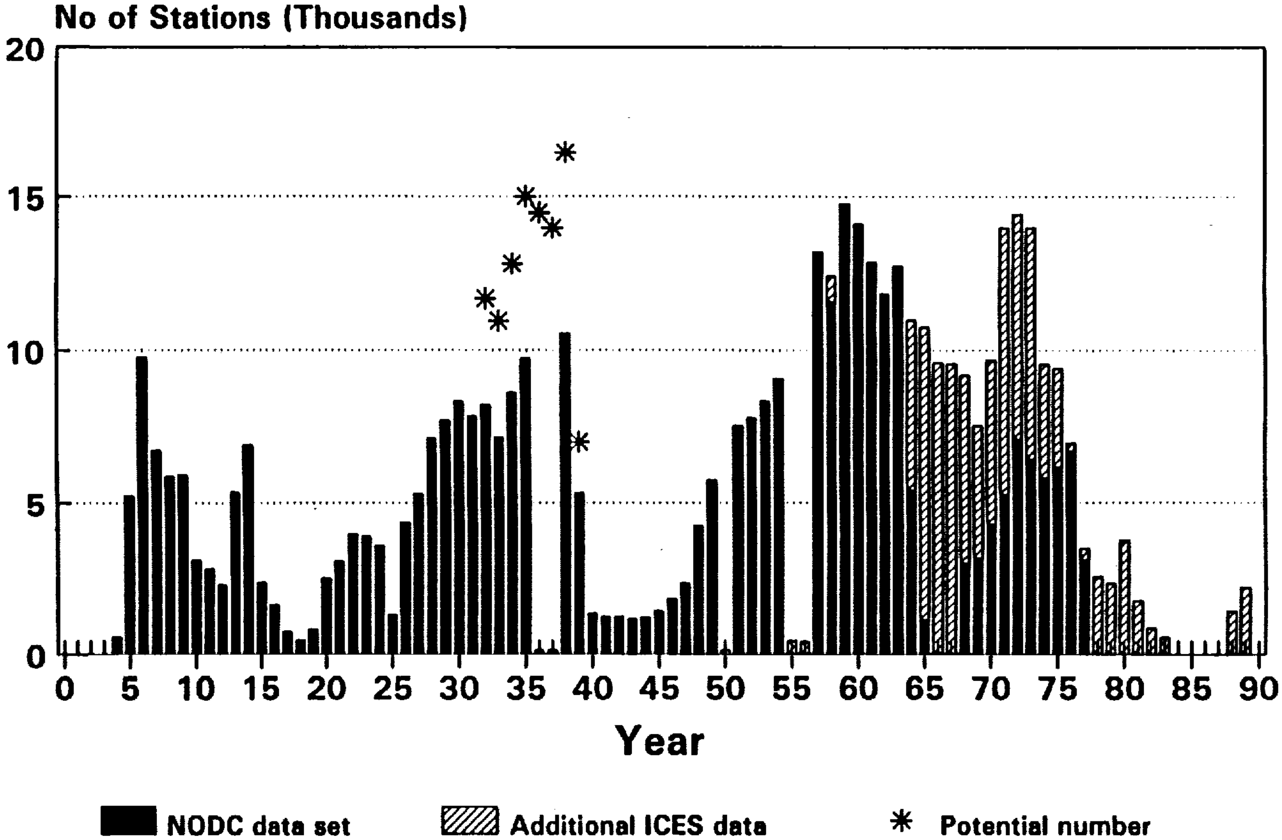

Fig. 1. Number of surface temperature and salinity observations in the NODC and ICES surface data files, 1900—1990. The potential numbers during the 1930s refer to the number of observations listed in the Bulletin Hydrographique. (Similar numbers not published in other decades).

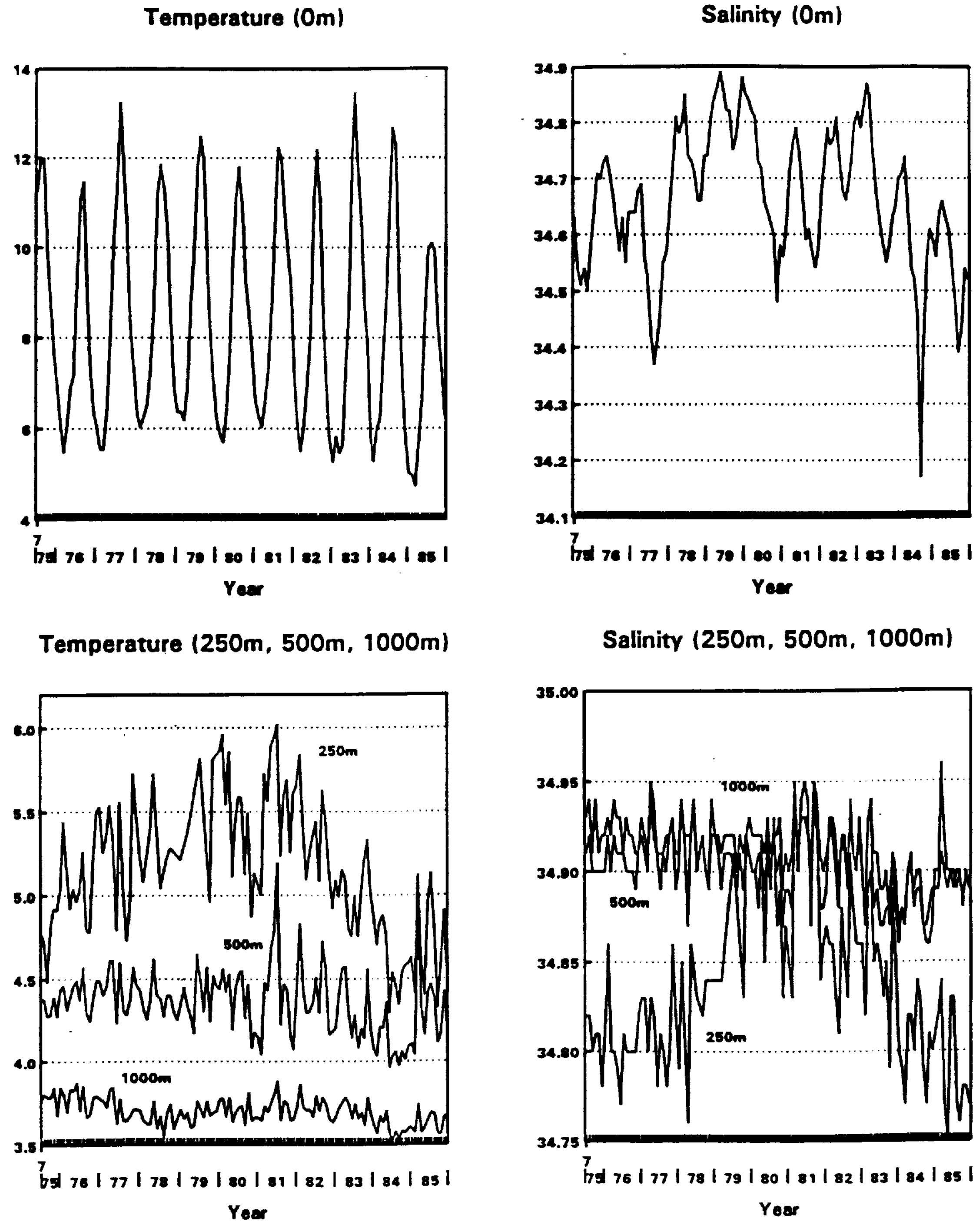

Fig. 2. Time series at OWS "Charlie", 1975—1985.

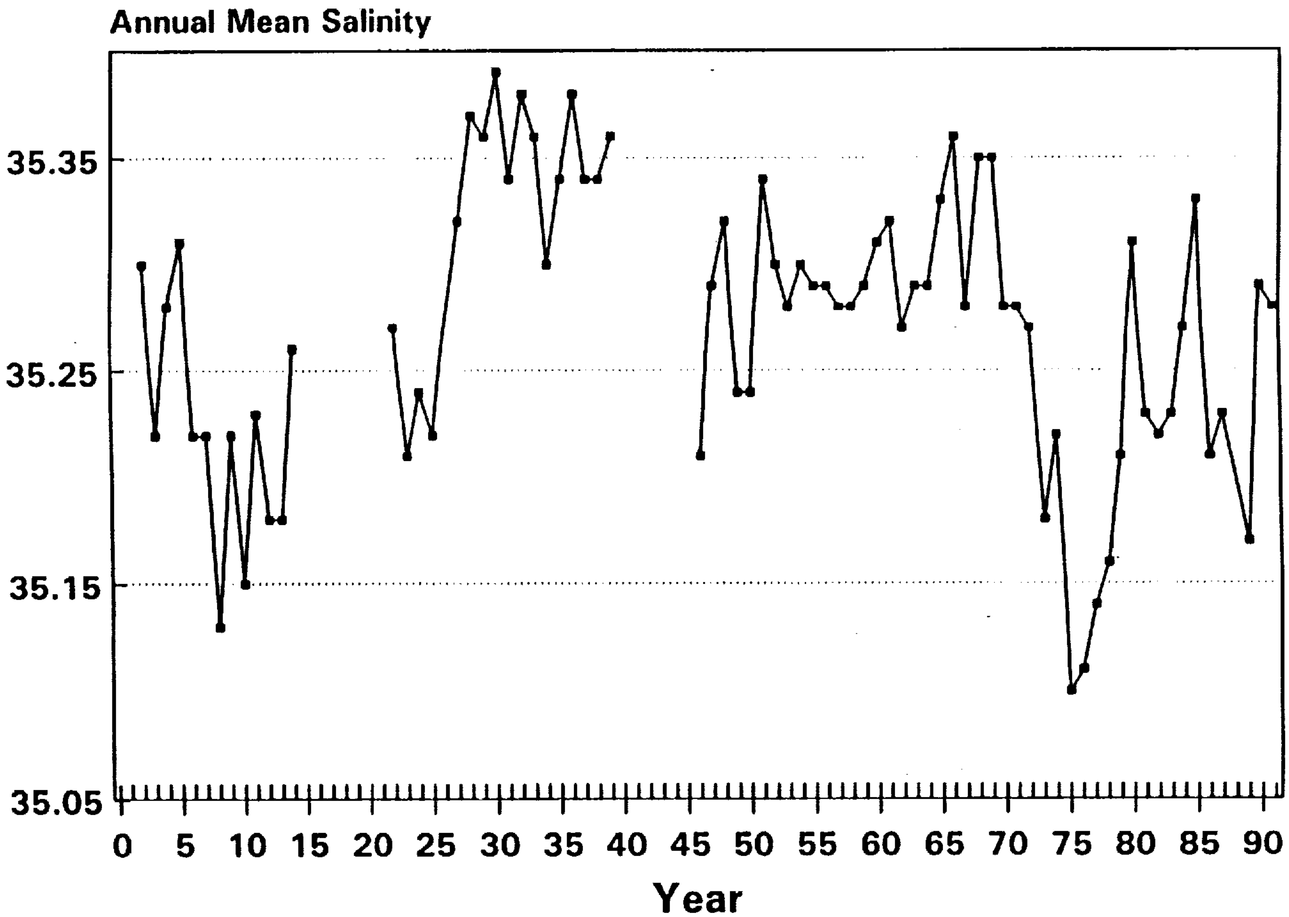

Fig. 3. Annual mean salinity at 0-300 m in the eastern Faroe—Shetland Channel (59°30'N 7°W to 63°00'N 1°E, ±20 nm), 1902—1991.

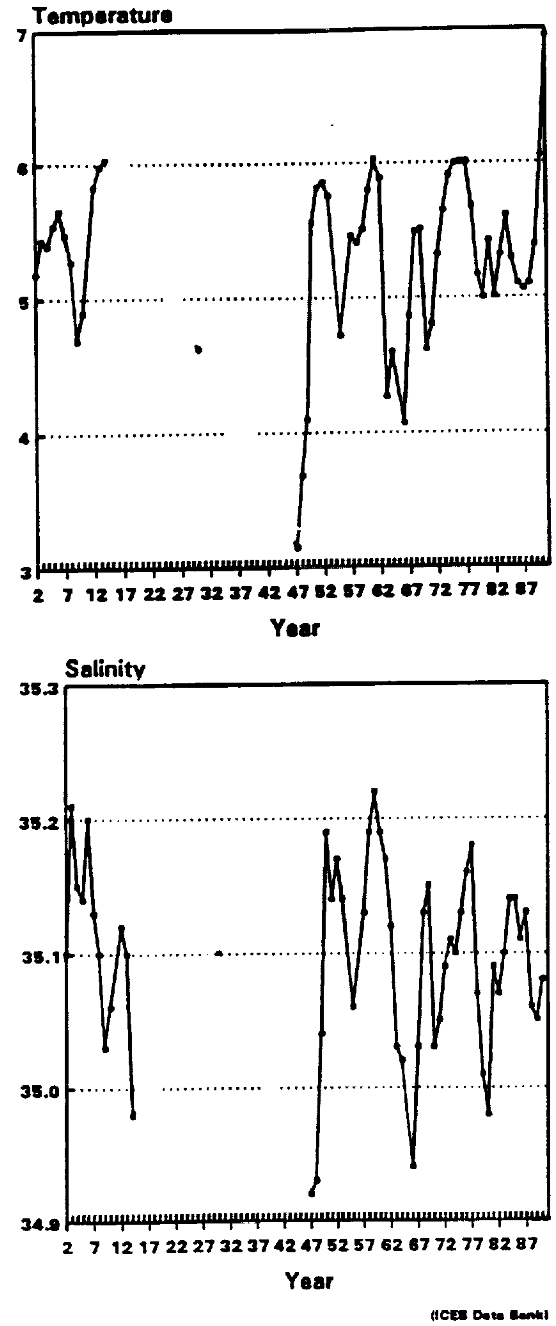

Fig. 4. Annual mean temperature and salinity at 600m in the Skagerrak, 1902—1991.