Pergamon

Prog. Oceanog. Vol. 33, pp. 303-346, 1994

0079 - 6611/94 $26.00

0079-6611(94)00006-9

Surface salinity of the North Atlantic: Can we reconstruct its fluctuations over the last one hundred years?

GILLES REVERDIN[1,2], D. CAYAN[3], H.D. DOOLEY[4], D.J. ELLETT[5], S. LEVITUS[6], Y. DU PENHOAT[7] and A. DESSIER[8]

[1]*Lamont-Doherty Earth Observatory of Columbia University, Palisades, N.Y. 10964, USA*
[2]*Laboratoire d'Océanographie Dynamique et de Climatologie, Université Pierre et Marie Curie, Paris, France*
[3]*Scripps Institution of Oceanography, La Jolla, CA 92093, USA*
[4]*International Council for the Exploration of the Sea, Palaegade 4, Copenhagen, Denmark*
[5]*Dunstaffnage Marine Laboratory, Oban, Scotland*
[6]*National Oceanographic Data Center, Universal South Building, Washington, DC 20235, USA*
[7]*Centre ORSTOM, Nouméa, Nouvelle-Calédonie*
[8]*Centre ORSTOM, IFREMER, Plouzané, France*

Abstract – Surface samples have been collected in the North Atlantic in the past one hundred years for determining the ocean salinity and its temperature. A large share of the data we have used were collected by merchant vessels or weather ships of European countries and to a large extent are listed in reports, in particular in the *Bulletin Hydrographique*. We investigate whether these data are relevant for determining low frequency fluctuations of the sea surface salinity. We find many crossings in the 1920s for which salinity is anomalously high compared with the climatology or with other crossings collected on the same ship line. These anomalies are indicative of a contamination of the sample. By examining hydrographic data, reports and recent experience in collection and storage in sea water, we can attribute these large errors to unclean buckets where salt crystals dissolve into the sample and to breathing of the samples during the storage. Each of these stages contributes in estimating a too large salinity and adds to the scatter of the measurements.

To further investigate these errors we compare the surface salinity and temperature for each monitoring program with nearby hydrographic casts, mostly in the eastern Atlantic. We find large differences between the various monitoring programs of different periods, and we use comparisons to empirically correct the data. Unfortunately, the number of comparisons is often too small resulting in a large uncertainty in these corrections, in particular before 1914 and for the UK and German monitoring programs before 1939 which exhibit the largest average bias in the 1920s. Despite this, we find that surface samples provide a useful complement to the hydrographic station data for investigating low-frequency variability of upper ocean waters. In the two areas where we did construct these time series: the Faeroe-Shetland Channel and the eastern Atlantic near 50°N, the surface data critically reduce the aliasing caused by insufficient sampling by the hydrographic casts. Both areas present minimum salinities around 1910 and in the late 1970s.

CONTENTS

1. INTRODUCTION

Our interest is in reconstructing past low-frequency fluctuations of the North Atlantic water masses. In the shallow seas and shelves bordering the Atlantic Ocean, large data sets of hydrographic stations have been collected which illustrate well-defined low-frequency interannual variability (SCHOTT, 1966; DICKSON, 1971; and TAYLOR, REID, MARSH, STEPHENS and JONAS, 1983, for the European shelves and shallow seas; MANNING, 1991 and PETRIE, LODER, LAZIER and AKENHEAD, 1992 for the North American shelves). In the deep-ocean sectors before 1948, however, only certain areas, mostly adjacent to the continental shelves have been regularly sampled. The hydrographic time series in the Faeroe-Shetland Channel (DOOLEY, MARTIN and ELLETT, 1984; TURRELL and SHELTON, 1993), for example, illustrate low-frequency fluctuations of the water-mass properties, suggestive of changes in the ocean climate on inter-annual to inter-decadal time scales.

Most of the hydrographic knowledge of the ocean and its low-frequency variability has been gathered since 1948 when the number of subsurface observations increased considerably. In the upper ocean, they have illustrated low-frequency variations which are coherent over large areas. One striking example is the low salinity anomaly which has been found in various parts of the European polar seas, the North Atlantic subarctic gyre and the north-eastern Atlantic between 1965 and 1983 (DICKSON, MEINCKE, MALMBERG and LEE, 1988). This salinity lowering contributed as much to the modification of surface ocean density as did temperature and, therefore, played a major role in the formation of subsurface waters. The amplitude of these signals is typically of the order of 0.10psu, but larger signals are observed in polar currents (up to 0.5psu in the Icelandic Sea or in the Labrador current). We also expect larger amplitudes where spatial gradients are large. However, the sampling by hydrographic casts does barely resolve these structures, even for the best sampled periods (LEVITUS, 1989a,b).

We will discuss in this paper whether one can extend and supplement the deep-sea records of upper ocean salinity. Sampling of the surface waters has been carried out more extensively by surface monitoring than by hydrographic stations; approximately 400,000 surface samples being collected in comparison with 200,000 casts for the deep ocean. The seasonal cycle of surface salinity is large (SMED, 1993; LEVITUS, 1986) but interannual signals have a comparable amplitude so that one can hope to detect them in sea surface records. The regular collection of surface salinity samples was advocated by PETTERSSON (1894), both for the North Atlantic and adjacent shelf seas. In the North Sea, steamers were first used in 1890 to provide regular collection, and in 1896 this monitoring was extended to the North Atlantic (although Pettersson already used a cross-Atlantic vessel on at least one occasion in 1892). Surface sampling is still continuing for chemical, biological, physical and fishery research and one of the largest tasks is to assemble the different data sets. Recent use of the surface data is illustrated by ELLETT (1982) and by TAYLOR and STEPHENS (1980) for the North Atlantic, by DONGUY and DESSIER (1994) for the tropical Atlantic and by DELCROIX and HÉNIN (1991) for the equatorial Pacific. Usually, this involves some careful testing of the data associated with *a priori* judgements by an experienced oceanographer have removed erroneous data or spurious effects related to the spatial or temporal inhomogeneity of the data.

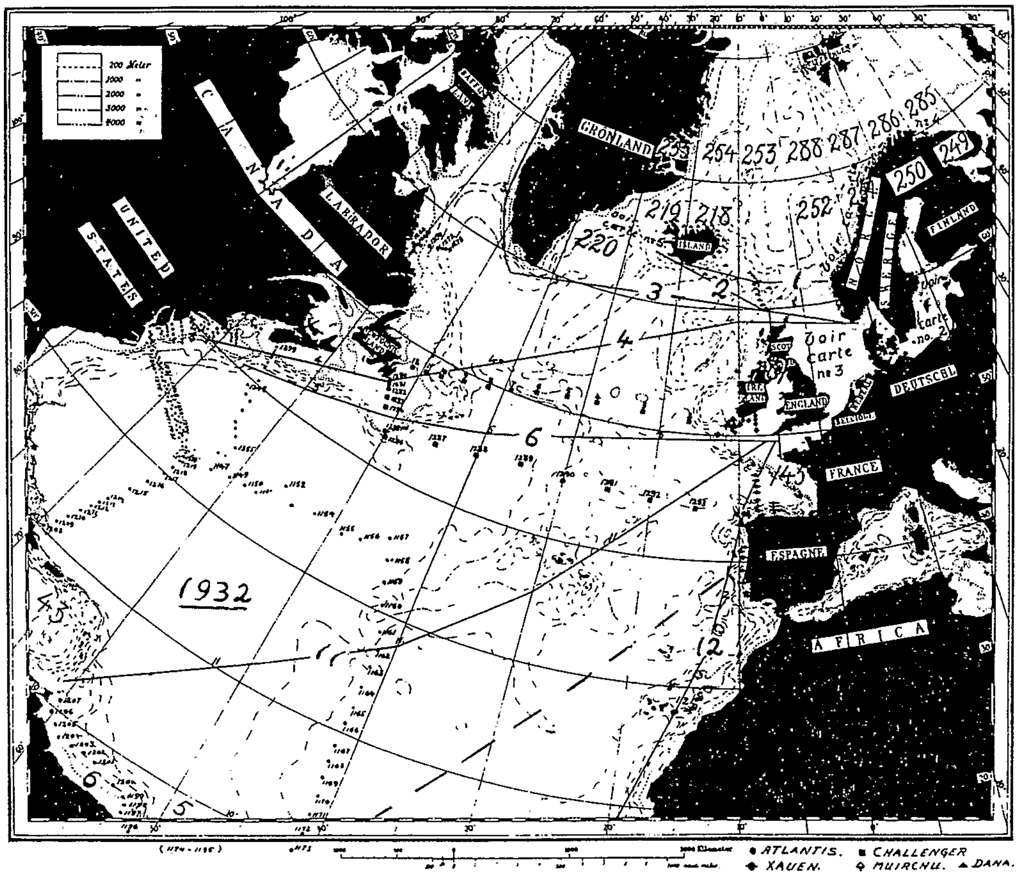
Most of the attempts to investigate surface low-frequency variability have involved a single data source, an exception being the analysis of the ICES (International Council for the Exploration of the Sea) surface data and hydrographic stations of the 1902-1939 period by SMED (1943). Smed identified a widespread increase of surface salinity in the North Atlantic between the 1910s and the late 1920s and 1930s. At sites in the Faeroe-Shetland Channel, this is coherent with the change at 200m. We should, however, comment that surface salinity data can have many problems. A large portion of the samples was collected without scientists on board by a poorly-trained crew. The sea water was then stored in capped bottles and analyzed after the elapse of time varying between a few hours and over 6 months. Two general sources of errors are possible. First, an improper collection technique may have modified samples, for example, if the collection bucket was not clean. Second, the sample may have changed in the bottle, because either the bottle was not clean or dissolved in the sample, or because of evaporation through the cap. To measure the salinity, the older samples were titrated, more recent samples have their conductivity measured. In many cases after 1901, the work was performed by institutions provided with a standard water (initially provided by ICES and recently by IAPSO) and the estimation of a chlorinity (or a conductivity) is made relative to this water with a standard analysis error of the order of 0.02psu (salinity is the Practical Salinity having the dimensionless unit psu). However, this was not always so, in particular in the UK between the 1920s and the late 1950s, when the analysis was done at the Office of the Government chemist in London with a lesser accuracy. Between 1895 and 1901, when the use of a calibrated standard water was not so widespread, these errors can exceed 0.05psu (REVERDIN, 1993).

The difficulty of mixing different data sets was recently illustrated by the Rockall Trough time series between December 1988 and December 1991, during which in the eleven months with both weather ship and research ship data, eight of the monthly maxima were from weather ship samples (ELLETT and TURRELL, 1992). The mean difference between the weather ship and research ship monthly salinity maxima is 0.025psu, with a maximum difference of 0.076psu. In both sets of data, water was collected from intakes 3-4m below the surface, and the samples stored for later analysis onshore. Out of 78 weather ship observations for this period, five salinity values which appeared to be too high by 0.2 to 0.5psu were rejected, the discrepancy being attributed to leaky or contaminated sample bottle caps. To correct this problem, the weather ship now uses bottles which have a separate plastic seal in addition to the cap.

The question that we have to address is how the different data sets can be mixed to form long term records of surface salinity and temperature. For this, we will concentrate on the North Atlantic, for which the longest time series of data is available and for which a coordinated effort by ICES assembled the different sets between 1905 and the 1960s. How the samples were collected and kept and the spatial area covered by these data have all varied in time. But this information is not always available from written records, which greatly adds to the difficulty of combining the different data sets. In one case, the titration book (Norwegian sampling for 1932) lets us know how long the samples were stored before they were analyzed.

The prime issue is to identify the sources of errors and how they influence the accuracy of the different data sets. We mostly investigate salinity, but the examination of co-sampled surface temperature (SST) often helps to explain sea surface salinity behaviour and its accuracy may also bear on that of the salinity. A summary of the studies on the past SST measurements is found in PARKER and FOLLAND (1991). After presenting the data sources in section 2, we discuss how the water was collected. In section 4, we present evidence for the water modification during the collection. Then, in section 5, we investigate how the samples were stored before being analyzed. Following this, we discuss the presence of extreme outliers and we compare the distribution of surface salinity from these sets relative to collocated contemporaneous hydrographic stations or sets of surface data of higher accuracy. In section 8, we compare time series constructed for the same area with different data sets, and discuss in section 9 how the surface data collected in the last one hundred years may contribute in defining past oceanic surface conditions.

a



2. THE DATA

The *Bulletin Hydrographique*, a publication of ICES, is the largest source of surface data (other large sets of surface data are discussed in Appendix A and their data also included on Fig.2). A large portion of these ICES data was digitized by the National Oceanographic Data Center (NODC, Washington, USA). Unfortunately, the codes for the country, the vessel and the shipping lines were not retained by NODC before 1954. We attempted to acquire most other data sets for which a reference was available. Except for a few Danish, French and American data sets which have apparently been lost, and some German and UK surface data which we have yet to collect, our attempts at locating the data were mostly successful. Some data sets were communicated on magnetic tapes from various hydrographic offices and ICES. Cruise reports and various bulletins provide complementary surface data. Finally, some data are listed in manuscript archives which we only partially digitised and used.

Between 1905 and 1958, the bulk of the surface observations is listed in the *Bulletin Hydrographique*. The *Bulletin* indicates which countries and institutions were in charge of the monitoring, and in some cases the vessels' names are indicated. Fig.1a for a typical year shows how the sampling was conducted, and Table 1a mentions the main participating institutions and the areas

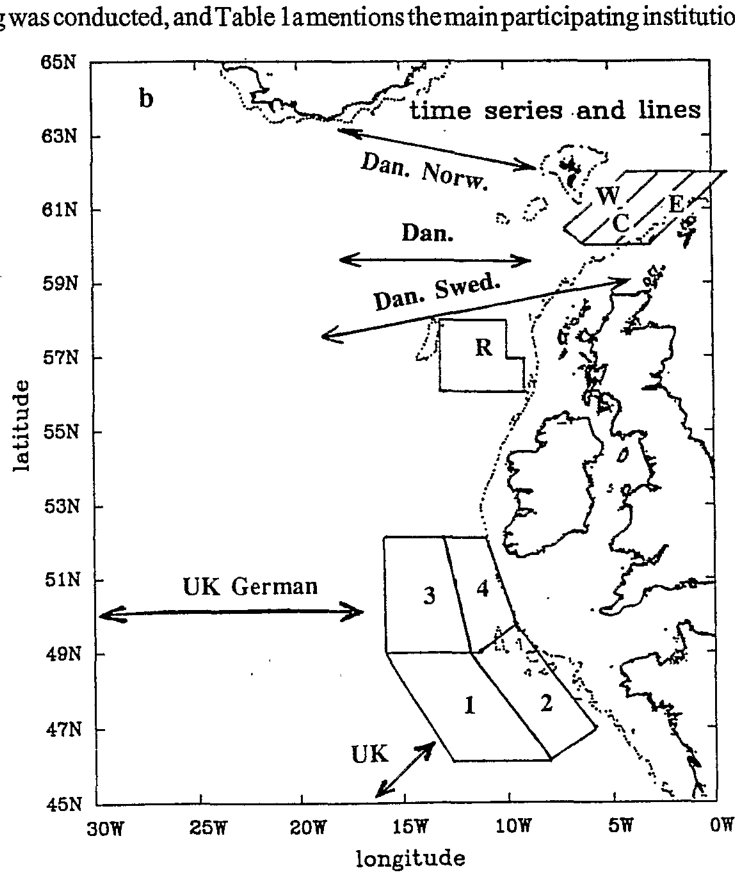


Fig.1. (a) (left) *Bulletin Hydrographique* map indicating the North Atlantic lines sampled in 1932 for ICES. Line 1 was monitored by Norway, lines 2,3 4 by Denmark, line 6 by Germany, and lines 11 and 12 by England. (b) (above) Map identifying the eastern Atlantic boxes where time series are constructed as well as some of the lines; the dotted contour corresponds to the 150m isobath.

which they sampled. As is clear from Fig.2, more sampling was conducted in the shallow European Seas, because many institutions participating in the surface sampling were involved in fishery research. Actually, more than half of the ICES samples are collected over shelves and shallow seas (Table 1b), and these are summarized in the ICES atlas of surface temperature and salinity (ICES, 1962). For various reasons, the sampling was conducted differently in inshore areas than over deep water from cross-Atlantic liners and steamers. Unfortunately the shallow seas data are often of a higher quality than the deep-sea samples. We also consider surface data collected during research vessel oceanographic cruises or for some German institutions before 1940, for which an experienced observer was present, and which should provide more accurate data (a list of the cruises which contribute to 11580 data between 1905 and 1939 is provided by REVERDIN, 1993).

After World War II, information about the sampling is often available from the investigators or from the ICES data inventories. In many cases, it is difficult to find where the observations are stored or archived, and even some recent data have been lost. In recent years a problem has been the increasing use of thermo-salinometers or undulating CTDs (SeaSoar or SeaRover) towed behind the vessel. Data from thermo-salinometers are only useful if the instrument has been properly calibrated, and if the water is pumped from a suitable site on the vessel. Unfortunately, information is not always available on how these instruments have been calibrated. These data, if uncalibrated, may have errors larger than other surface data, and should not be used as a reference in a comparison.

The surface data distribution is shown by decade on Fig.2 where most of the samples are associated with a temperature and salinity. Data collected prior to 1900 present special challenges because 'standard water' had not yet been developed, and have been described in an unpublished report (REVERDIN, 1993). There is very little data for the periods 1890-95, 1899-1903, 1916-19 and 1940-45. Coverage for the other years has also enormously varied in time with specific ship routes being either interrupted for long periods or diverted. The most regularly sampled route until 1960 was between the Shetland Islands and south-west Greenland, which was maintained by the Danish Hydrografisk Laboratorium. Altogether, there are 697607 samples available in the Atlantic ocean and adjacent seas for the period 1890-1990. Most of them were collected north of 10°S, including 350643 samples from the North Sea and Baltic Sea. There are also 67685 calibrated thermo-salinometer reduced data in the Atlantic Ocean, each of which approximately corresponds to 1 nautical mile of ship route.

We also use near-surface samples from hydrographic stations for comparison with the surface data. These originate from 517565 casts (mainly Nansen casts), including 170014 in the Baltic and North Seas and roughly 150000 on the North American continental shelves. Usually, the samples have been collected and stored by trained personnel with more care than for the ships-of-opportunity programs. Some times, the analysis was done on-board (in particular, on many German cruises), but in other instances, the samples were not analyzed until the ship is back to port (in particular, for early Norwegian and Danish oceanographic cruises). In those cases (for example, the Norwegian *Michael Sars* cruises from 1901 to 1910), the samples were generally analyzed within 3 months of collection, but often more than a year elapsed before analysis (for example, the Danish *Dana* oceanographic cruise in 1921-1922). Furthermore, the near-surface water at a hydrographic station was often drawn with a bucket, which may introduce similar biases to those of the surface data set. Therefore, we also need to investigate the error in these surface-level data from Nansen casts. In the early decades, the sampling was mostly on shelves (the European shelves, and also the North American shelves and Grand Banks of Newfoundland after 1912). The total number of stations and of surface data is indicated by decades in Table 1b. In the deep ocean, there are fewer hydrographic stations than surface samples available until 1940, with comparable numbers in the 1950s to late 1970s. In the latter period, the distribution is such that there are more surface samples in the north-eastern Atlantic and more station data in the western Atlantic.

Table 1: Summary of the contribution of different countries to the ICES surface sampling in the North Atlantic. (a) area and time sampled by each country (number of samples collected in a day indicated in parenthesis); (b) total number of surface samples available by decade, for the North Sea, the Kattegatt and Baltic Sea, and other areas of the Atlantic Ocean (excludes all ICES data from 1925, 1936 and 1937).

Table 1a:

Denmark	1897-1915, 1920-1940, 1945-1954-1960, Shetland to Iceland (6/day before 1914 and 2/day after) 1897-1906, 1921-1939, 1945-1960, Scotland to western Greenland (6/day before 1914 and 2/day after) 1904-1915; 1922-1936, North of Scotland to New York. (4/day)
England	1904-1909, 1913-1915, 1922-1939, English Channel to South America and West Indies. (2/day) 1904-1910, 1913-1916, England to New York and Canada. (2/day) 1948-present, Ocean Weather ships
France	1956-1970, Ocean Weather ships
Germany	1928-1939, English Channel to New York. (2/day before 1935, 6/day after)
Netherlands	1913-1914, English Channel to Surinam. (3/day) 1948-1983, Ocean Weather ships
Norway	1931-1939, Bergen to Iceland. (12-24/day)
Portugal	1928-1931, Lisbon to Madeira and the Azores. (4/day)
Sweden	1920-1922, North of Scotland to New York (4/day)

Table 1b

	Surface samples			Hydrographic stations		
period	**Baltic Kattegatt**	**North Sea**	**Atlantic**	**Baltic Kattegatt**	**North Sea**	**Atlantic**
1890-99	602	2654	8820			
1900-09	4898	12544	27531	7892	5241	12966
1910-1919	1546	10875	18077	4226	2162	3927
1920-29	7420	23637	22783	10212	3049	8145
1930-39	15752	39960	34797	17693	7900	15806
1940-49	492	11272	9968	13494	3530	10299
1950-1959	10287	55210	50103	16266	17980	46536
1960-69	3421	73753	36198	16171	19911	86244
1970-79	2677	65045	61446	10171	10014	97814
1980-89	17	8581	*144826	3932	941	65108

*includes 77141 surface samples and 67685 data from thermo-salinometer recordings (usually averaged over 5 minutes)

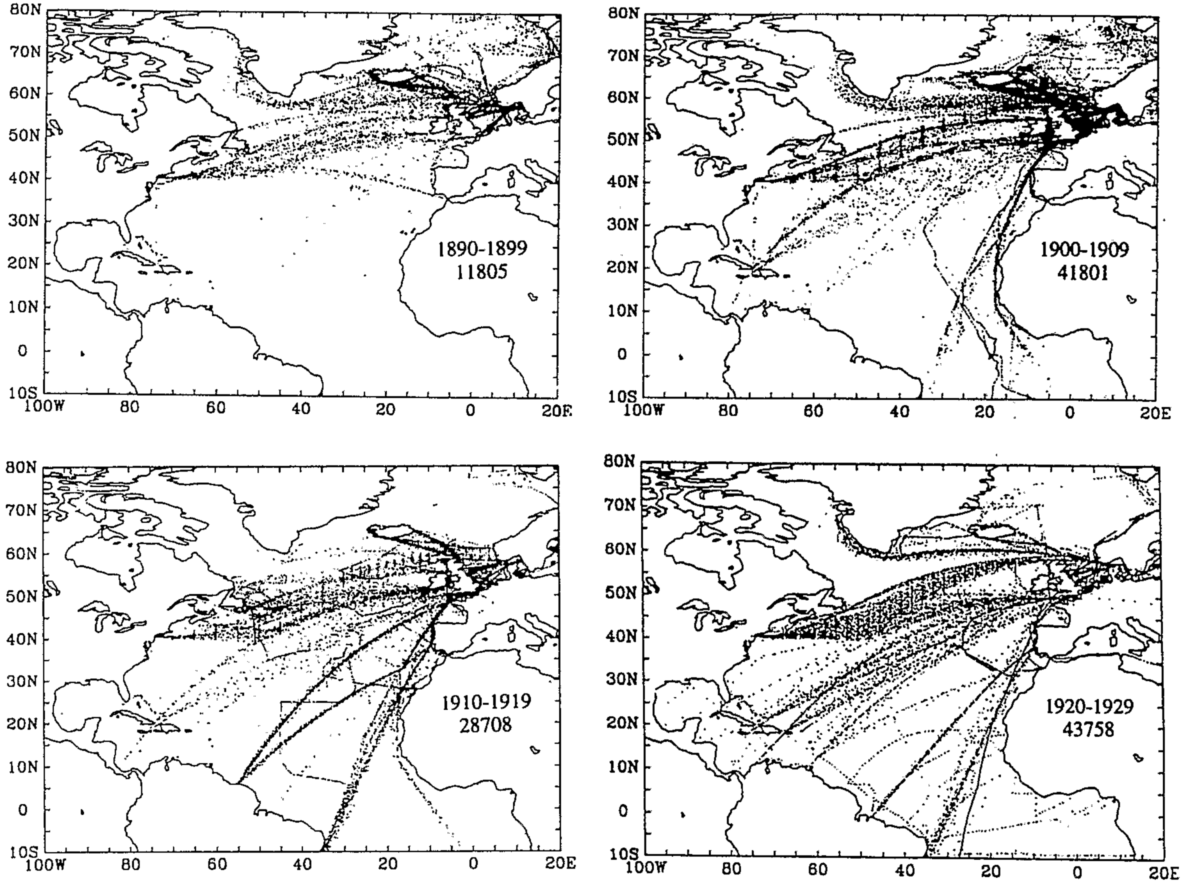
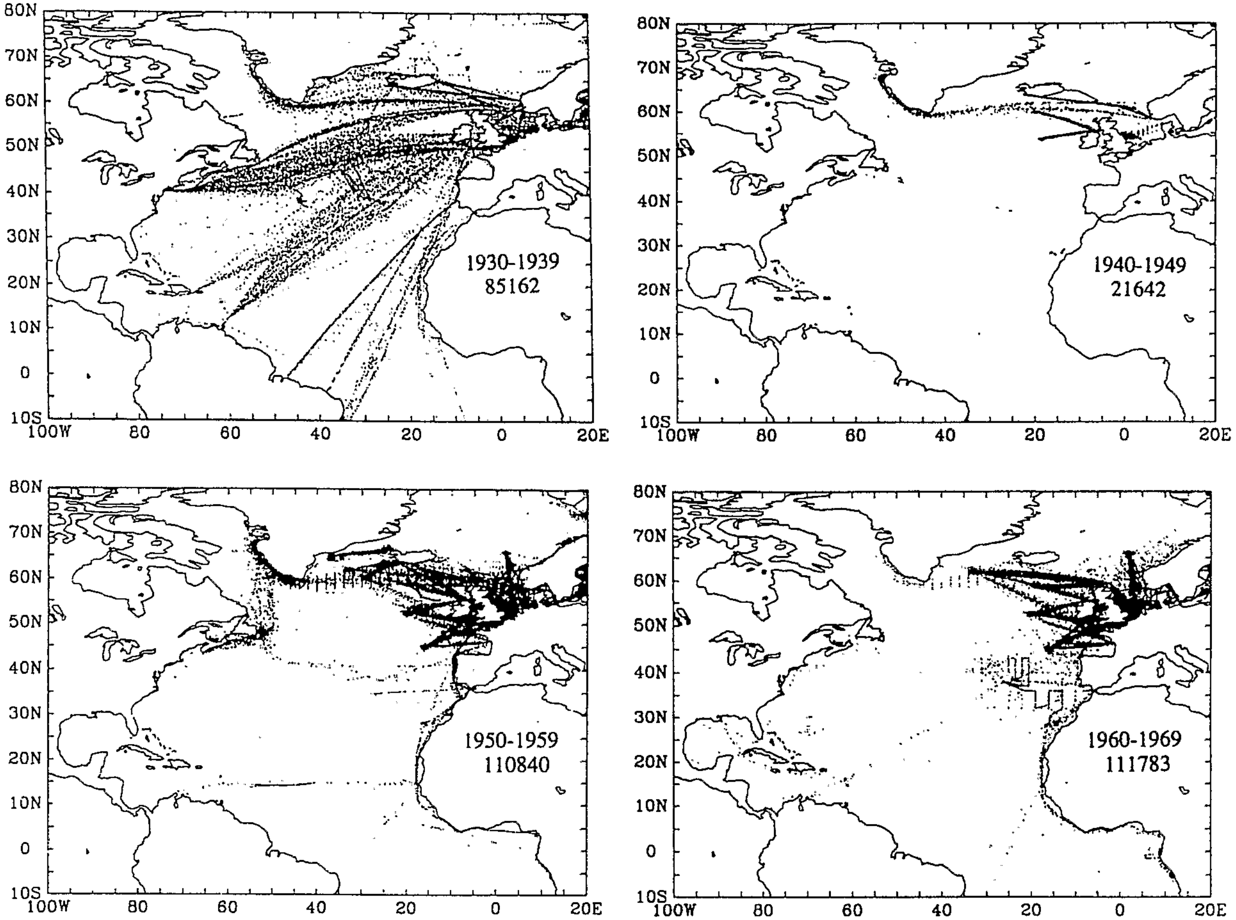
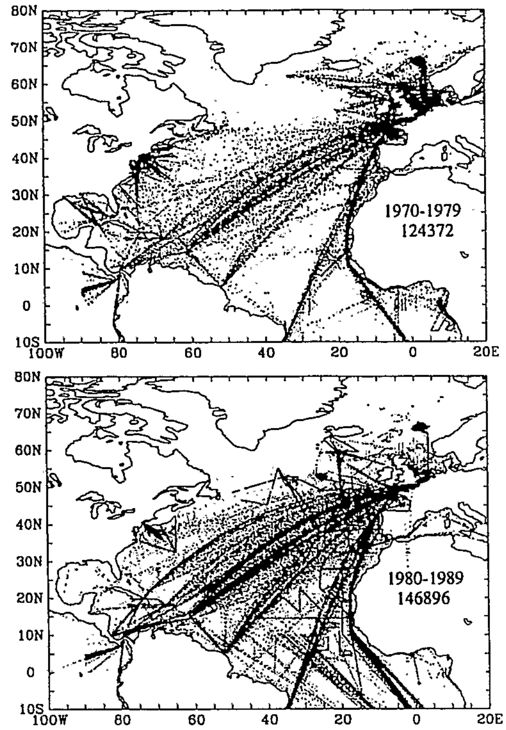


Fig.2. Data distribution by decade of surface samples in the Atlantic Ocean north of 10°S. The total number of data included on the maps is indicated, most of which are associated with a salinity estimate, but notice that some are in the Mediterranean Sea or the Pacific Ocean (the summary on Table 1b includes only data in the Atlantic Ocean). Data on land have not been screened out.



1930-1939
85162
1940-1949
21642
1950-1959
110840
1960-1969
111783
80N
70N
60N
50N
40N
30N
20N
10N
0
10S
100W
80
60
40
20
0
20E



3. DATA COLLECTION

Since the early sampling programs, various ways have been used to collect sea water. Samples have been drawn from a continuous flow of water pumped at subsurface or have been drawn from a bucket. In the case of the continuous flow, the water salinity is unlikely to have been modified by collection. In the instance of a bucket, however, water modification can occur by (1) the dissolution of salt crystals deposited on the bucket, an effect mentioned for the canvas buckets by KNUDSEN (1899); (2) contamination from unclean buckets (see LUMBY, 1927, for comments on the English sampling after 1920); (3) evaporation (WISSER, 1938) which causes both temperature and salinity errors, and is particularly important for the canvas bucket, but also occurs from other buckets. The amount of evaporation depends on the time the bucket is left on the deck before reading the temperature and taking the water. The resulting decrease in the temperature from a canvas bucket was estimated as 0.35°C from the *Snellius* expedition data in Indonesian waters (WISSER, 1938). Using a model of a bucket (FOLLAND and PARKER, 1990) and *in situ* testing (FOLLAND, 1991), a -0.4°C error in global SST was estimated for the period 1911-1941 by PARKER and FOLLAND (1991). However, this error has a strong spatial and seasonal dependency according to this study. Maximum values occur during winter east of the North American seaboard where the negative bias reaches -1°C. This cooling results from different effects, an important one being evaporation from the walls of the bucket (FOLLAND, 1991). One can also expect that the salt gradient in the canvas will diffuse in the bucket water. Even if there is no initial salt deposit on the canvas, an evaporation which causes a 0.4°C cooling would yield a 0.02psu increase if redistributed within the bucket water.

The design of bucket, how well it is rinsed, whether a thermometer was inserted, and how long it took to read the temperature and collect the water sample from the bucket are all factors which can lead to errors in measuring salinity of water samples.

Buckets were in many materials, the most commonly used being canvas (usually, the larger buckets, which commonly had a wooden base, were used to haul a few liters of water on the deck of the vessel) or cast-iron or brass. There are examples of wooden buckets, for example one was used in June 1891 for a section across the British Channel in June 1891 (DICKSON, 1901). Scientists had diverse opinions about the buckets to be recommended. Canvas buckets were used on board the *Deutschland* in 1911 (BRENNECKE, 1921) and on board the *Snellius* in 1929-1930 (WISSER, 1938), but metal buckets were used for most other hydrographic cruises (the *Ingolf* in 1895-1896) for example).

The efficiency of rinsing depends on the time of immersion of the bucket in the sea which has certainly varied, if one trusts the recommendations which range from 20 seconds to 5 minutes. The bucket was then hauled on deck and suspended to read the water's temperature. One technique was to plunge a thermometer into the bucket after its arrival on deck, and leave it there for some time before reading it. An early example of the first method (thermometer placed on the deck) is the 1895-1896 *Ingolf* expedition (KNUDSEN, 1899). Although in this instance the thermometer was not left for a long time, it is clear from reports that in some instances it might have been left for up to 5 minutes. We suspect that in many cases the water sample was drawn from the same bucket after the temperature was measured (this was already done in 1896-97 according to DICKSON, 1901), although the recommendations for the English sampling program after 1904 were that 2 to 3 casts of the bucket be made to rinse the bucket before collecting the water. During certain oceanographic cruises, the bucket was plunged a second time after measuring the temperature before a sample was drawn, and this is also likely to have been done when a scientist was embarked on board to collect the samples.

Issues of sample contamination and temperature errors with this approach were identified by LUMBY (1927). He devised a surface sampler equipped with sample bottle and thermometer which was towed in the water alongside the ship to allow thorough rinsing of the bottles and avoiding direct human intervention to fill the bottle (LUMBY, 1927, 1928). It also resulted in a much smaller temperature error, because reading could be done directly after hauling the sampler to the deck. This sampler was used for UK sampling in the North Sea and English Channel from early 1925 on some lines, and after October 1928 on all, until the adoption of engine intake sampling in the early 1950s. It was probably also adopted by other countries participating in the sampling of the North Sea after 1928. It was possibly also used on the British weather ships between 1948 and the early 1950s. However, SUND (1931) commented that it was not always clear that this rather cumbersome sampler was used properly, and he devised a simpler sampler, consisting of a bucket with an inserted thermometer, which would be read shortly after hauling and a tap at the bottom for filling the bottle. This sampler not only reduced the likelihood of contamination of the sample, but also provided a more accurate temperature reading.

Other buckets, lacking a tap for drawing the sample but with an inserted thermometer, were designed at other times with the same objective of reading temperature quickly (for example, during the *Deutschland* 1911 Expedition, BRENNECKE, 1921). Some commercial designs were probably quite common: an early brass example is depicted in MARINI (1912), which strongly resembles German buckets with inserted thermometers, or the French bucket used after the second World War; a rather similar model is also used at the Woods Hole Institution; a different model elongated and in rubber was designed after the war by the UK Meteorological Office to further reduce the temperature error (FOLLAND, 1991).

In some instances, it is possible that the water collected for measuring its salinity was not the one in which temperature was measured. Our objective is to document how the sampling was carried out. Unfortunately, this is not always possible, in particular before World War II, when the information is often not reported, and we rely on the following sketchy information.

For example, the temperatures from the Norwegian sampling between Norway and Iceland in 1931-1939 were certainly thermographs records from the intakes, and probably also in 1945-1950, whereas the salinities are from samples collected with the Norwegian bucket (see SUND, 1931). The Danish temperatures and salinities on the lines to Iceland and Greenland were mostly from canvas bucket samples. Canvas buckets were in service until 1980, although in the early years the sampling was supervised by M. Knudsen, who disapproved of the use of canvas buckets. In some years after 1945, it is possible that the temperatures were taken, not from the bucket, but from a hull-placed sensor or, after 1960, a towed thermometer (JENS SMED, personal communication).

Water was drawn from a pump for a Swedish program in 1898-1899 (CLEVE, EKMAN and PETTERSSON, 1901). Other examples of continuous water flow include UK OWS samples drawn since the early 1950s from an intake located less than 4m from the sea surface. Some French samples in the 1960s and early 1970s were also drawn from the ship's intake. For the large UK sampling between 1904 and 1940, it was recommended that the participant vessels use a wooden or galvanized iron bucket (personal communication of D. MATTHEWS to J. LUMBY, 1935). However, Lumby suspected that a small canvas bucket was occasionally substituted and a photograph of 1938 shows a canvas bucket used on a French liner for collecting the sea water. It is also likely that canvas buckets were used by the English merchant vessels in the 1896-97 sampling program monitored by DICKSON (1901). Although Lumby planned to substitute these buckets with his samplers, this is unlikely to have taken place on liners. A small iron bucket was probably used on German liners between 1928 and 1938 (a Norwegian model was used in 1932 on the *Westfalen* (SCHUMACHER, 1933). A small iron bucket was used for the important Dutch OWS set, at least between 1962 and 1978 (*Cirrus* and *Cumulus*) and the bucket recommended by the 'Météorologie Nationale' was used for the bulk of the French data since the mid-1960s and until 1991 (J.R. DONGUY, personal communication).

4. WATER MODIFICATION DURING COLLECTION

4.1 Direct estimation

For some oceanographic cruises, we can compare simultaneous samples collected by different means, usually a bucket and a bottle sample from an hydrographic cast or water pumped from a level close to the sea surface. In the absence of rain, there is no evidence of significant differences in salinity between the upper 50cm from which a bucket draws the water and a hydrographic cast sample from a depth of 1-3m. Although the concentration of certain salts could be larger at the air-sea interface because of evaporation. For example, in early July 1993 by sea state 4 to 6 between Iceland and Newfoundland, the difference between nearly-simultaneous samples from a bucket and a ship's intake at a depth of 5m is less than 0.005psu in 15 out of 20 cases (with an average difference statistically non-different from 0).

The French bucket was tested during oceanographic cruises in the western equatorial Pacific in December 1989 and in August 1992 (SURTROPAC cruises 13, 14, 15 and 17) and in the tropical Atlantic in January-February 1993 (CYTHER-1 cruise). During each of these cruises, both during stations and while underway, the bucket was dropped in the sea and rinsed with sea water, although because of weather conditions, both the interval between successive drops of the bucket and the place where the bucket was stored, changed from cruise to cruise. The sample was drawn from the bucket shortly after hauling it on the deck (2-3m above the sea level) and analyzed within a few days and then compared with data from other samplers (pumped water or hydrographic cast). Results vary from cruise to cruise: in 3 Pacific cruises (SURTROPAC 13, 14 and 17), the bucket samples had an average salinity higher than the sea surface (0.10, 0.10 and 0.06psu respectively), but in one (SURTROPAC 15), it was unexpectedly negative (-0.05psu). The standard error on these average differences varies from cruise to cruise between 0.01 to 0.02psu with a large standard deviation within each set of the order of 0.10psu. This contrasts with the comparison made during CYTHER-1, where the average difference was just 0.007psu with a standard deviation of 0.006psu. These differences are likely to have originated from slightly different designs of the French bucket or from the thoroughness rinsing of the bucket. For example, during CYTHER-1, salinities collected from a non-rinsed French bucket were on average higher by 0.013psu than from a rinsed bucket, and had a larger scatter (0.016psu compared to 0.006psu).

The elongated English rubber insulated bucket (without its cover and inserted thermometer, as has been used on French vessels since 1991) was also tested during SURTROPAC-17 and CYTHER-1. If properly rinsed, its salinities were higher by 0.004 and 0.005 psu than measured by a thermosalinometer at a depth of 4m (or CTD casts) for the two cruises respectively. The root-mean-square (*rms*) difference for CYTHER-1 is 0.007psu (30 comparisons) and is slightly larger for SURTROPAC-17 (*rms* of 0.021psu for 17 comparisons), but in both cruises the *rms* is less than for the French bucket. During CYTHER-1, the salinity of the bucket water was higher by 0.010psu, when the English bucket was not rinsed properly.

How water of a canvas bucket with a wooden base is modified when left for a while on deck before the sample is drawn was investigated in early July 1993 between Iceland and Newfoundland in sea state 4 to 6. In this summer situation when the bucket was left in a position protected from wind there was little evaporation and after the bucket had been on deck for 3 minutes only a small insignificant increase occurred averaging 0.003psu (the *rms* scatter is 0.022 for 30 comparisons) and no cooling of the bucket had occurred. Comparison of the samples drawn from the well rinsed bucket just after its arrival on deck with those from intake samples showed a small average difference (bucket higher by 0.003psu), which was determined by a few outliers.

These examples illustrate that there have been large bucket-to-bucket differences which have depended on the weather conditions and the way in which the bucket is used. The comparisons we presented above are mostly from the tropics and only cover a small fraction of the buckets used in the past. To complement these comparisons, it is useful to consider early hydrographic stations.

4.2 Surface samples from hydrographic stations

During hydrographic casts, it was common to use an iron bucket to collect the top sample. Very often, the vertical density profile presents an inversion between the surface and the upper subsurface level of the cast. At many stations, this level was often close enough to the surface for the water to have been isothermal and isohaline with the surface water. During those cruises, we use the profiles for which there is a safe assumption that the layer is well-mixed, to estimate the bias of the bucket sample. If temperature is measured using reversing thermometers mounted on bottles at both levels, we select the stations for which the two temperaturs are similar to within 0.01°C. Near-surface temperatures measured in the bucket are also subject to error, and we have assumed that the water is well mixed when the surface is colder than the next subsurface measurement and report this temperature inversion.

The determination of the temperature error is imprecise because the surface bucket temperatures were usually reported with one decimal, instead of 2 as is more common for subsurface hydrographic measurements (the *Ingolf* 1895-1896 oceanographic cruise report is an exception, where two decimals accuracy was adopted at the surface only, KNUDSEN, 1899). Reports often mentioned that the thermometers used for bucket temperatures were less accurate than the ones used at deeper levels.

In Table 2 we present a comparison between the surface samples collected by iron bucket and the shallowest bottle sample for a selection of oceanographic cruises by several countries. We have omitted a few outliers larger than 0.50psu, because such a large discrepancy probably resulted either from contamination or from a reporting error. The station data are mostly from the winter season covering a broad time-span (1896-1964), and are extracted from oceanographic cruise reports and the *Bulletin Hydrographique*. For the 1939 *Atlantis* cruise in the tropical Atlantic, we considered only those surface temperatures reported to just one decimal point as having derived from a bucket sample. For this particular cruise the comparison may overstimate the positive salinity bias in the bucket samples because salinity in the area studied often decreases below the surface.

Although the observed density inversion could include a contribution from random errors of the measurements there is a clear bias towards surface temperatures being too low and salinity too high. The negative temperature bias for the selection of oceanographic cruises in Table 2 ranges from about 0.05°C to 0.27°C. This is at the lower end of the errors presented in FOLLAND and PARKER (1990) for the wintertime. The associated salinity errors are in the range 0.01psu to 0.07psu, mostly around 0.04psu, which are too large to arise solely from the evaporation associated with this cooling. A possibility which would reconcile the two would be if the surface layer sampled by the bucket is a little warmer and saltier than the subsurface depth. This is, however, unlikely in most cases, considering the season and area sampled. It is more likely that either the buckets used were better insulated or that they were left on deck for less time during these cruises than in the study of FOLLAND and PARKER (1990). Part of the salinity bias is likely to have originated through dissolution of salt crystals or brine deposits on the bucket.

Oceanographic cruises with small salinity biases, mostly before 1914, also correspond to a small temperature inversion. In those instances, the distribution of the salinity biases has a large peak between -0.01 and +0.01psu, values which are indistinguishable from 0 with the titrimetric chlorinity determination in use at this time. On these cruises the bucket may have been better protected from the wind or temperature and samples were collected sooner once the bucket was on deck than for other sets, but this is unknown. Hence it is disturbing to discover that the average cruise bias in the same season, same area, and by the same vessel varies from year to year; for example, the *Muirchu* stations exhibit large biases in 1929 and 1930, but not in 1925 or after 1930.

Many *Bulletin Hydrographique* Norwegian and Danish surface data are from hydrographic stations drawn from water-bottles and were not subject to the biases described above. However, for the Scottish stations in 1904 and 1905 (ROBERTSON, 1907), there is a strong temperature inversion at the surface, and for other cruises there are a few anomalous surface positive salinity deviations of between 0.05 and 0.10psu (for example for the *Explorer* data in 1938 and cruises in 1903). The mix of cruises from which the set of hydrographic casts were derived results in the proportion of stations with surface bucket values varying from area to area. The proportion decreased in the 1950s, when bottles became more commonly used for the surface sample. This will lead to artefacts in the trends in surface salinity based on surface data from hydrographic casts if these biases are removed.

Table 2. Bias on the surface level of hydrographic stations, based on the comparison with the first subsurface level in instances where the upper layer is likely to be well-mixed. We present the average differences as well as the *rms* of the individual comparisons and the sample size both for T and S. Missing information implies that the surface data were not collected with a bucket. Most of the data are for the late winter and spring season (the *Anton Dohrn* cruise in 1955 north of 65°N, is, however, in November-December, and some of the Scottish data and *Pourquoi pas* stations are from the summer season.

	ΔT	σ_T	n	ΔS	σ_S	n	area
Denmark							
Ingolf 1896	-0.10	0.12	14	0.021	0.023	14	vicinity Iceland
Norway							
Isachsen 1910	-0.17	0.24	7	0.025	0.021	8	Spitsbergen
Michael Sars Feb 1901	-0.04	0.08	27	0.011	0.022	27	Norwegian Sea
Michael Sars May 1902	-0.09	0.09	13	0.050	0.045	13	Norwegian Sea
Brategg 1947-1948	-0.07	0.11	43	0.017	0.020	43	Antarctic
Station M Oct 1948-Oct 1949	-0.13	0.14	9	0.037	0.028	10	Norwegian Sea
Anton Dohrn Nov-Dec 1955	-0.27	0.21	23	0.056	0.061	23	North 70°N
Scotland							
Cruises 1904-1905	-0.20	0.21	188	0.014	0.030	204	Faeroe-Shetland and North Sea
England							
MBA (Plymouth) 1903	-0.05	0.06	43	0.021	0.022	43	English Channel
MBA (Plymouth) 1904-1905	-0.09	0.08	93	0.022	0.018	92	English Channel
Salpa 1922	-0.14	0.15	16	0.033	0.028	16	English Channel
Salpa 1925	-0.13	0.13	13	0.041	0.036	13	English Channel
Salpa 1929	-0.06	0.05	8	0.025	0.022	8	English Channel
France							
Pourquoi Pas 2nd cruise 1921				0.077	0.062	20	Eastern Atlantic
Pourquoi Pas 1st cruise 1922				0.071	0.037	9	Eastern Atlantic
Ireland							
Helga 1920-1921	-0.10	0.10	38	0.036	0.035	42	South Ireland
Muirchu 1925	-0.05	0.05	30	0.012	0.015	30	South Ireland
Muirchu 1929 and Feb 1930				0.050	0.070	33	South Ireland
Muirchu 1931-1933				0.014	0.080	83	South Ireland
United States							
Atlantis 1939	-0.03	0.03	23	0.062	0.051	20	West Atlantic 10-40°N
Moroccan Fisheries 1964	-0.13	0.19	18	0.043	0.39	31	Off Morocco

5. STORAGE OF THE SAMPLES

There is no storage bias in the use of thermo-salinometer in intakes, nor is the storage of major importance if the sample is analysed on board the ship within a few days of collection. However, the samples were not analysed on most merchant vessels nor during some oceanographic cruises, but were kept for a relatively long time before being analysed. Most samples were stored in glass flasks of 150-250cc. Prior to 1902, a variety of flasks was used, most of which were sealed with a cork soaked in liquid paraffin as recommended by PETTERSSON (1894). Knudsen criticised the technique because drops of paraffin dropping into the water could result in titration errors, but this problem was not encountered by DICKSON (1901). After 1904, most of the bottles had a porcelain stopper with a rubber washer and were secured by a wire spring clip. This model was still widely in use in the 1960s and still is in some places. In recent years, various other bottles have been in use, some being quite small (100cc) and therefore more sensitive to evaporation than others. Plastic bottles may even have been in use in a few cases, which tend to be more prone to 'breathing' and therefore to greater evaporation losses.

For the small (100cc) corked bottles, HELLAND-HANSEN and NANSEN (1909) reported a positive error of 0.07psu after 5 months with good corks, and of 0.08psu with poor corks. Even after shorter storage times, they suspected that the samples collected in 1900 on *Michael Sars* had a positive error of about 0.01 to 0.02psu, whereas for the larger corked bottles used during the 1901 *Gjøa* oceanographic cruise in the Barents and Greenland seas, they found no significant evaporation had taken place through the cork. LUMBY (1935) investigated the evolution of replicate samples stored in the glass bottles with porcelain stoppers used in England. The scatter between the samples was larger for periods over 6 months, but unfortunately because the statistics are not fully documented we do not know what the average bias was. Recently, a comparison between different bottles, including the glass bottles with swing-top ceramic stoppers and rubber seals (the 'ICES' bottle) has been carried out for ICES (ICES C.M. 1987/C:21, unpublished report by D.S. KIRKWOOD and A.R. FOLKARD, 1987). For ICES bottles an average salinity increase of 0.008psu in 2 months was reported. Even in perfectly well-sealed bottles samples show a conductivity increase, corresponding to a practical salinity increase of 0.005psu in 4 months (SY and HINRICHSEN, 1986) as a result of dissolution of the glass which should not modify the sater chlorinity and therefore not affect chlorinity-derived salinity.

Errors can also arise when sealing and opening samples. Unless the rim is wiped dry before sealing, salt crystals can form around the top of the bottle, and these may fall into the sample when the cap is removed for analysis. In many institutions it has been the practice since the early 1960s to dry the bottle tops and caps with a disposable tissue before sealing, but this practice is probably not often followed on board ships-of-opportunity. Where sealing relies upon a rubber washer, such as in swing-stoppered bottles, it is important that these are in good condition and replaced regularly, since rubber in contact with sea water perishes rather rapidly. Whatever the storage, there is likely to be a percentage of samples in which salinity has spuriously increased by evaporation because of a fault in the bottle or the cap. Sometimes, this effect is so large that it is readily detected during routine work: for example, out of a batch of six 250cc bottles filled with the same sea water in the spring of 1992 at Lamont and analyzed after 3 months, 5 retained their original salinity to within 0.004, but in one it had increased by 0.077. The longer a sample is stored, the more likely it is to suffer a salinity increase.

The duration of sample storage before analysis has changed. For the 1896-97 monitoring by DICKSON (1901), samples were analyzed within 3 months of collection with the following exceptions: whalers in the northern oceans which were gone for over 5 months, and one ship, the *Para*, sampling the tropics and subtropics. In 1896 salinities from the *Para* were often too high by at least 0.20psu, as shown in the vicinity of the Azores by simultaneous collection by *Princesse Alice* and surface samples collected by other vessels in Dickson's set. This error could have originated either from the long storage time between collection and analysis (sometimes more than 5 months) or from the use of 'unclean' buckets or from the bottles not being properly sealed. We suspect that the sampling delay of a few months remained typical until the 1920s. However, LUMBY (1935) mentions that it had increased in "recent years" for the English sampling to 5-8 months for samples collected on transAtlantic ships monitored in the Lowestoft program. These were titrated in London at the Laboratory of the Government Chemist until the introduction in 1958 of a conductivity salinometer at Lowestoft. This delay implies a systematic error larger than 0.02psu.

Delays were probably less in other countries. The titration book for Norwegian data in 1932 and 1933, kept at the Fiskeridirektoratets in Bergen, Norway, provides the following information. This set of 2091 data indicates that 57.8% of the samples were analyzed within 2 months of collection, 23.4%, 12.7%, 5.7% within the third, the fourth and fifth months respectively; 8 samples were analyzed during the sixth month. Assuming that the increase within time of salinity is linear at 0.004psu in a month and that all bottles had been carefully closed and checked for cracks, the average salinity increase resulting from 'aging' is 0.008psu for this set. Delay for the recent sampling by ORSTOM is also known to be a few months.

6. OUTLIERS

According to Table 2, the positive salinity error (negative temperature bias) one expects in surface samples during oceanographic cruises is larger than 0.02psu (0.10°C). The conditions in which a bucket is used during oceanographic cruises is likely to be better than on ships-of-opportunity, where not only are the personnel less likely to be adequately trained, but also the bucket is used less frequently. The bias is probably different between different sampling programs because the instructions on how to collect and store samples varied and were dependent on the weather encountered. We will discuss in this section the presence of anomalous outliers for the ships-of-opportunity, before estimating in section 7 the statistical distribution of the error for different programs.

The presence of outliers was noted by MATTHEWS (1907) for the UK sampling program in 1904-1905. For instances, whole crossings by the *Port Antonio* are obviously biased, sometimes by more than 0.4psu. For the pre-1905 Danish effort, the brig *Peru* in May 1897, the mail steamer *Laura* in June 1899 and a crossing on 22-27 August 1904 brought samples which were often contaminated. LUMBY (1935) also mentioned (without citing a name) a vessel which for a number of years had consistently reported poor data, nearly all of which had been published before being detected. Unfortunately, this refers to data which are in the computerized set of data, for which ship codes and country codes are not provided.

We checked suspicious outliers in the *Bulletin Hydrographique* issues for 1925, 1929 and 1937 by line and country operating it (see Table 1). When such an anomaly is found, the other data collected during the same crossing are checked for their consistency with other crossings at the same locations. However, it is easier to attribute positive outliers to erroneous measurements than negative ones, because large positive surface anomalies would more readily mix vertically (saline water is denser). This procedure is mainly effective in areas where there are weak spatial gradients, but most lines cross such areas in the eastern Atlantic. We find few obviously erroneous Norwegian data in the *Bulletin Hydrographique*. Outliers are also uncommon for Danish lines, although a few crossings have a large percentage of erroneous data (for example, 10-19 July 1922 on the New

York Copenhagen line). The high quality of the early Danish data is also suggested by the similarity between 1904-1905 salinity time series in inbound and outbound crossings of the Danish line to New York close to Scotland presented in MATTHEWS (1907). On the other hand, data collected along the English and German cross-Atlantic lines do not appear to be as accurate as suggested by their inspection in the eastern Atlantic. Many data on these lines exhibit large positive salinity deviations (over 0.50psu) from neighbouring data and from the climatology. Often a crossing contains so many dubious values that the whole crossing is suspect. An example of this unfortunate situation is illustrated in Fig.3. In 1925, this was the case for at least two crossings in January and March among 16 crossings by the English line between the English Channel and the West Indies. In 1929, many such crossings can be identified, in particular in May, September and late December on the German Cuxhaven to New York line, and in August on the English line from the English Channel to the West Indies. Even in 1937, when data seem to be cleaner, two crossings out of 12 in February and March on the English Channel to South America line have such anomalous salinities that one suspects that either the samples were mislabelled or were not reported correctly.

No doubt is expressed about any of these data in the *Bulletin Hydrographique*, and quite often their titration was double-checked. In these issues of the *Bulletin Hydrographique*, we find that such suspect data (both isolated anomalies, as well as whole crossings) amount to 5.5% (76 out of 1390) in 1925 and 5% (43 out of 851) in 1937 of the data collected along these UK or German cross-Atlantic lines (years in which 4 or 5 different lines were monitored). In 1929 at least 5% (129 of 2503) were suspect, but the figure could well be over 20%. This is a large proportion which would at least contribute to a 0.01psu positive error assuming a 5% proportion of positive outliers is valid. We will remove these suspicious data from the set. To find how effective this is in improving the data set, one needs to know the shape of the error distribution. In the previous sections, we have documented various errors resulting from the sampling or storing of the samples. We can expect a non-normal distribution, probably not centered on zero but skewed towards positive errors because of a tail of positive outliers. In the next section, we will investigate

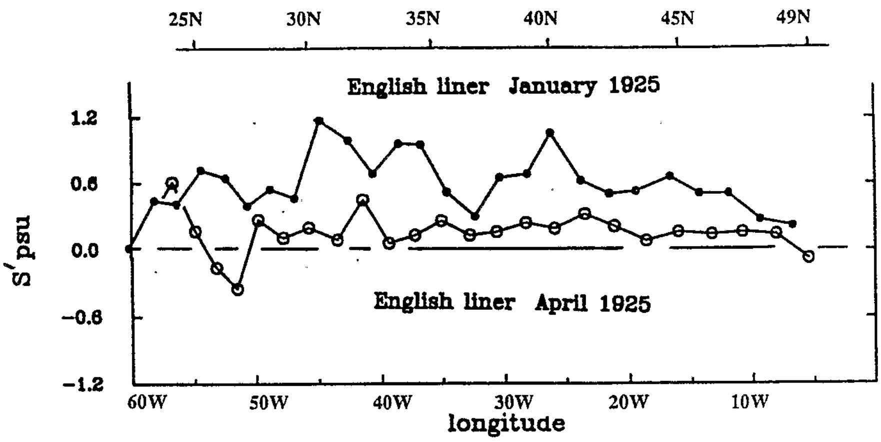


Fig.3. Examples of deviations from the climatological seasonal cycle for crossings along the UK line 11 (Fig.1) between England and the West Indies (Trinidad). Usually, two samples were collected each day, one in the early morning and one in the late afternoon. The two crossings (full dots in late January and open dots in late April) nearly overlap in positions and the seasonal cycle can hardly account for a difference of 0.05psu between the two. The largest value during the January 1925 crossing, 37.93psu, is reported at 30°53'N 44°54'W.

a few examples of the error distribution.

7. STATISTICAL DISTRIBUTION OF THE ERRORS

In general, we do not have a direct validation of the surface samples. To estimate partially the distribution of errors, we compare the ship-of-opportunity data with (a) hydrographic station data, and (b) surface data collected during oceanographic cruises for which we expect the bias to be small. We select samples which are close in space and time to data from the other set, and establish the distribution of the differences between the two. In Appendix A, we complement this with comparisons of a more extensive group of data sets of surface samples, including data collected before 1950, but for which statistics are often more ambiguous.

We first consider two surface sets where we expect that systematic errors are small (Fig.4). These sets are (a) the Dutch sampling of the weather ships on their way to station Mike (Norwegian Sea at 66°N, 2°E) and the North-East Atlantic weather ships during 1968-1983, and (b) samples from the weather ships west of 5°W in 1960-64 and the UK weather ships from 1965-1976 (they collected water on intake 3-4m below the sea surface, and therefore are free from bucket-error). The comparison is with the closest Nansen cast (within 10 days and 15km). These pairs of samples were often collected by the same ship. The distributions of the differences from the hydrographic station salinities have a small non-zero mean (of the order of 0.01psu) and are nearly symmetrically distributed[1]. Interestingly, these distributions prove to be non-Gaussian with a probability density near zero, larger than for a Gaussian distribution with the same standard deviation. A schematic representation of the distributions as a sum of two Gaussian distributions captures most of the distribution. Errors in the titrations or impurities in the bottle certainly contribute to the scatter. However, the non-Gaussian character of the distributions is probably not only related to errors, but to the variability over the distance separating the measurements (this is suggested by the spatial variability in thermo-salinometer records along ship-tracks in the eastern Atlantic). We test the uncertainty on the mean and median value with Monte-Carlo sampling of the distribution. The standard errors are small because of the large number of samples included in the distribution (the largest one for the mean of Fig.4b is 0.005psu), and are one third smaller for the median than for the mean. This results from the non-Gaussian distribution so that a sample median converges faster than the mean.

Two other sets for the Northeast Atlantic are considered, where enough comparisons with nearby hydrographic stations are available to construct the distribution of the differences (Fig.5). The maximum probability is close to 0 in Fig.5b, but shows a positive deviation in Fig.5a. For both sets, the distributions are clearly non-symmetrically distributed with more large positive than negative deviations (Fig.5). In particular, almost all the deviations larger than 0.5psu are positive. We also find this for surface samples collected by the National Marine Fisheries Service (NMFS) in the western Atlantic despite the fact that the inter-station comparison is done in areas with high spatial variability. On Fig.5, the median is smaller than the mean with a difference between mean and median of 0.01 to 0.02psu (as for the other distributions of Fig.4, the uncertainty on the median is less than the uncertainty on the mean). The simplest explanation for the distributions on Fig.5a and 5b is a contamination of the distribution for oceanic small scale variability, either during the collection or from evaporation during storage. The probability distribution of the contaminated samples is the convolution of the error-free distribution with the probability distribution of the error.

[1]We are not assuming a complete lack of error, but that the errors are not skewed. This would happen if the distribution of the error is Gaussian with a mean value close to zero. It is possible that the slightly positive means of Fig.4 are not the result of errors; notice however that the median is closer to zero.

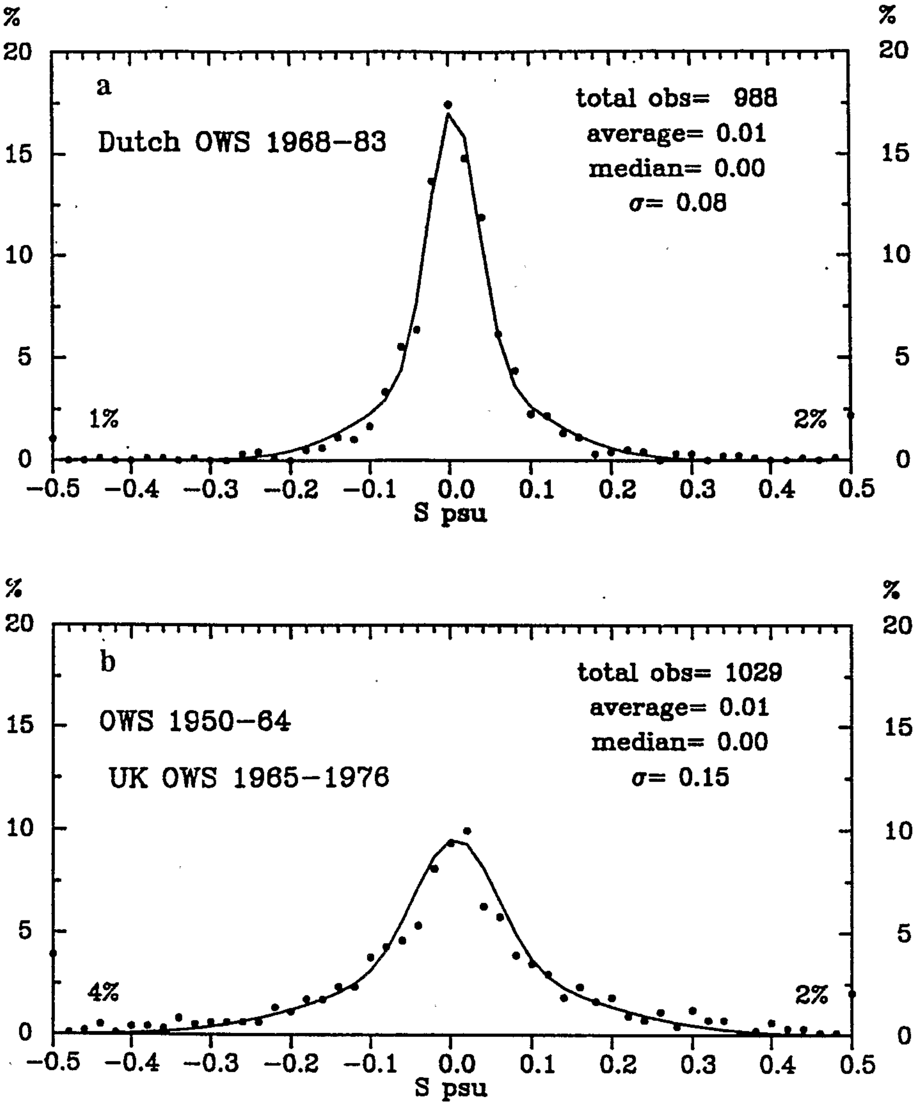


Fig.4. Distribution frequency of salinity differences between surface samples and hydrographic station data. A pair is included if separated by less than 20 days and 15km. The distribution frequencies averaged by classes of 0.02psu are presented by dots and a model curve is fitted to the distribution. Here it is the sum of two Gaussians centered on δS=0.006psu with *rms* σ_1 and σ_2 and respective contributions to the integrated histogram of r_1 and r_2. The summary statistics (mean, median, *rms* for the differences in the [-0.5, 0.5]psu range; the total number of points) are given as well as the percentage of points outside the range. (a) for surface samples of the Dutch program for the years 1968 to 1983 (r_1=0.5 with σ_1=0.03psu, and r_2=0.5 with σ_2=0.10psu); (b) for samples from Ocean weather ships during 1960 to 1964 and English Weather Ships for 1965 to 1976 west of 5°W (r_1=0.4 with σ_1=0.05psu, and r_2=0.6 with σ_2=0.15psu).

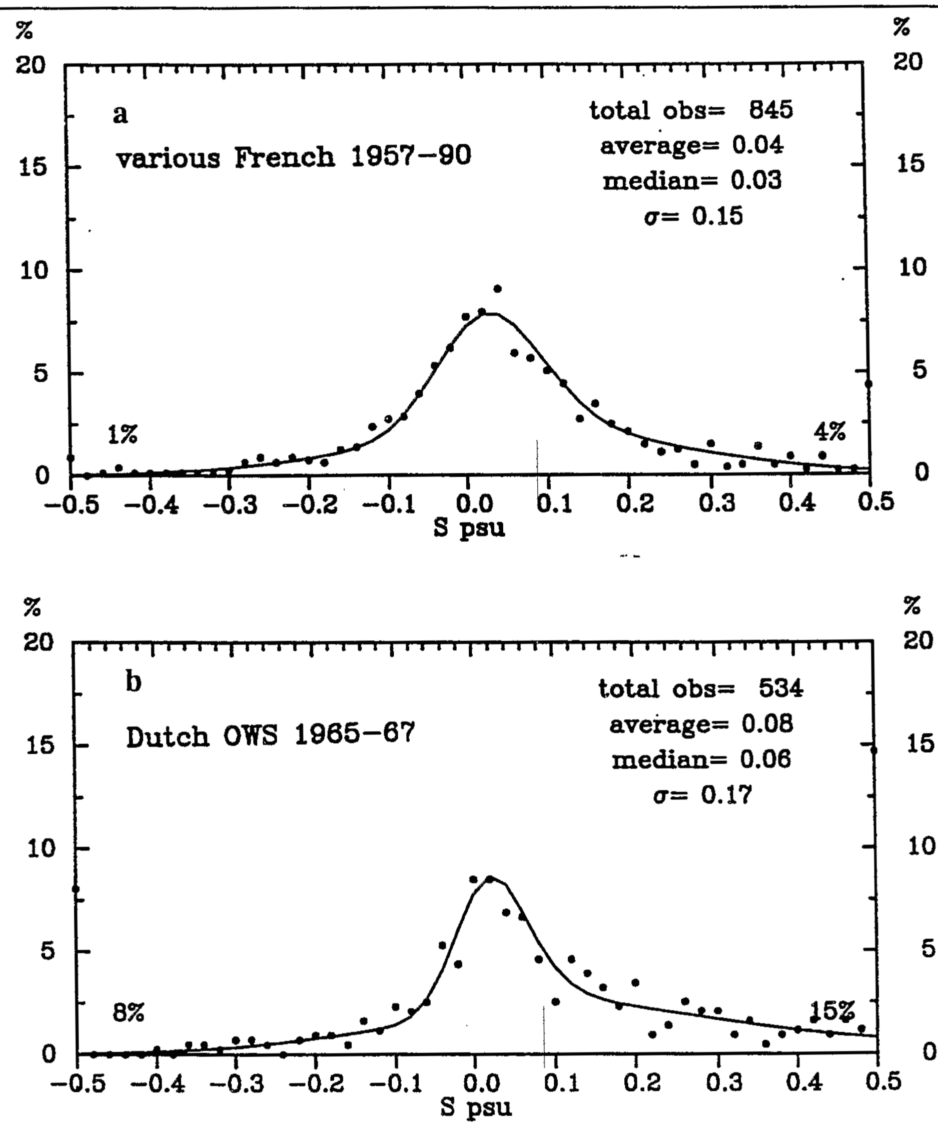


Fig.5. Same as Fig.4, but the model curve is a convolution of the distribution of Fig.4b with a nil distribution for negative δs and a decaying exponential exp(-δs/Δs) for positive δs. (a) for French data (FR set) north of 40°N and west of 5°W between 1957 and 1990 (Δs=0.035psu). (b) for Dutch Ocean Weather ship samples between 1965 and 1967 north of 40°N. Model curve is the combination of two exponentials for the noise: one with amplitude 0.7 and Δs=0.025psu and one with amplitude 0.3 and Δs=0.30psu.

For Fig.5, for example, we have assumed an error probability distribution in exp(-δs/Δs) for δs>0, and 0 for δs<0 (δs is the error). The convolution with a Gaussian ($\exp(-\delta s^2/(2\sigma^2))$) results in the distribution exp(-δs/Δs)*(1+erf(T)) where $T=\delta s^2/\sigma^2+1/(2\Delta s)$. The curve fitted to Fig.5a corresponds to the convolution of the curve on Fig.4 (an empirical combination of two Gaussians) by the exponential distribution with Δs=0.035psu. In the instance of Fig.5b, we also added another exponential error distribution with Δs=0.20psu. Of course, because we are mixing different subsets, this is only a gross indication of the error distributions, and it does not reproduce the frequency of extreme outliers.

No comparably detailed statistics can be derived for the surface sampling from the liners in the Atlantic Ocean and to Iceland or Greenland prior to 1950, because there are insufficient hydrographic stations on which to base a comparison. As just commented, the average error and the standard deviation of the distribution provides an incomplete representation of the errors. However, for lack of other information, in most cases we can only derive these simple statistics which are discussed in Appendix A and summarized in Table 3. We have first excluded from the sets samples from those crossings which were obviously poor, in particular, before 1930 on the UK and German lines south of 52°N. The comparisons illustrate that the errors vary in time and from one monitoring program to another, changes which unfortunately have no obvious explanation. A few of these conclusions are reported here. The Norwegian salinity data in 1931 to 1939 were probably less contaminated than the other surface data (it is possible that they were collected from an intake, as their temperature are reported from a thermograph and have a positive bias). The Danish data of line 2 also exhibit little scatter in the comparison with hydrographic data, but some large biases are found in some years (1949-1952 for example) which almost co-occur with changes in the temperature bias. The German data seem to be better after 1930 than between 1928 and 1930, and the UK data also present smaller average differences in the latter period. The lists in the *Bulletin Hydrographique* also contain data from cruises and special projects, which we suspect to be less susceptible to systematic errors, as are most data before 1914. For example, two vessels, one German and the other Finnish, sailing together between Norway and Spitsbergen in August 1906 simultaneously collected two sets of 8 samples. These two sets have the same mean salinity (temperature) with a *rms* deviation of 0.09psu (0.11°C).

8. TIME SERIES

Our long-term goal is to construct past records of the surface water characteristics which, because most of the data are scattered (with the exception of the weather ships data after 1948), requires a spatial-temporal analysis. There are many techniques whereby the individual anomalies can be averaged with respect to the climatology over some time-space scale. Here, in a first effort in assessing the impact of errors, we gather the individual deviations from the seasonal cycle into time-space boxes without consideration of the spatial-temporal coherence of the signal. The median of the individual deviations from the average seasonal cycle is estimated, which should provide a more stable estimate than the mean if the data distributions encountered are similar to those of Fig.4.

To minimise the effect of the aliasing of the high frequencies as a result of insufficient sampling, we will only consider the low frequencies of the time series of the deviations. We will consider two areas where there is enough data from both hydrographic stations and surface sampling in order to try to compare time series constructed separately from both sets. First, uncorrected surface data are used, then we roughly correct the surface data according to the comparisons presented in Section 7 and in Appendix A.

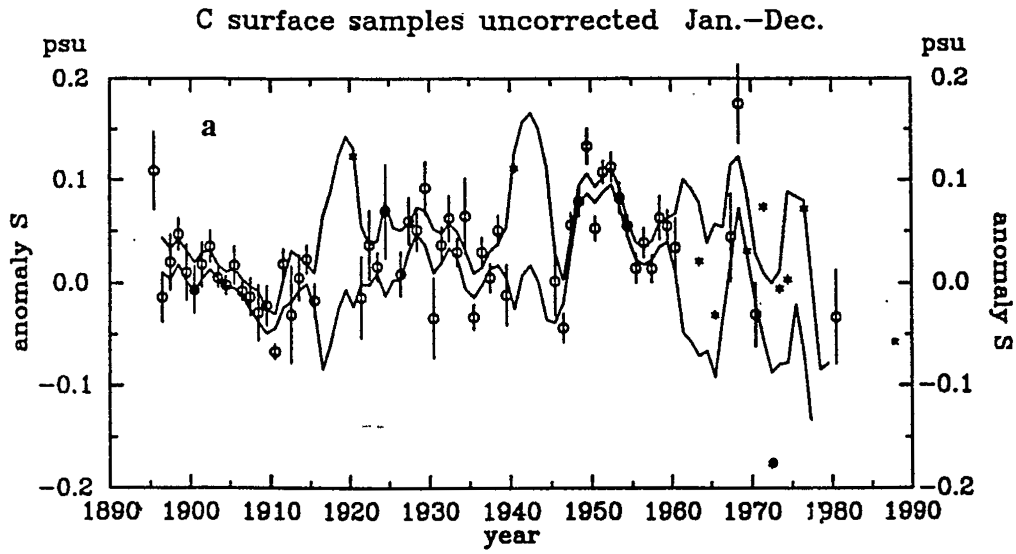
The areas examined are in the Faeroe-Shetland Channel (FS) and in the Northeast Atlantic close to the continental shelf (Fig. 1b). The two areas are intersected by most lines, with the exception of the Portuguese line and the Danish-Swedish line from Northern Scotland to New York. In both areas, there are sufficient data from post-1948 to estimate reliably an average seasonal cycle. Whenever possible, we have separated the data into two seasons: a cold season from December to April and a warm season from June to October. *A priori*, we expected salinity data for the cold season to be less noisy because the mixed layer is deeper and better mixed as a result of the rougher weather. Temperature anomalies are more coherent for the months December to April (RODEWALD, 1972). On the other hand, there are fewer data for this season, resulting in a larger uncertainty for the average anomaly. First, gross outliers were eliminated, i.e. values with deviations from the average seasonal cycle larger than 1.0psu or 5°C. Then for each season we estimated the median of the deviations from the average seasonal cycle. These seasonal medians are fitted with a cubic-spline taking into account the seasonal uncertainties estimated assuming that all measurements are independent. A binomial filter is then applied. This estimate of the low-frequency based on the limited sample is likely to be nearly normally distributed, even if it results from samples distributed as on Fig.5. On Figs 6-9 we present the likely one-standard deviation range of this low-passed time series (based on the errors for the seasonal anomalies).

8.1 Faeroe-Shetland Channel

The spatial variability in the channel results from the different water masses with water from the slope current entering from the south-west along the Scottish shelf, North Atlantic water also entering from the south-west and modified North Atlantic water from the North-west along the Faeroese shelf. We first consider a swath 67km wide along the axis of the Channel between 60°N and 62°N (central Channel) which roughly corresponds to the area considered in DOOLEY *et al* (1984) and where North Atlantic waters dominate. An average correction is applied for each oceanographic cruise by assuming that the Norwegian Sea deep water salinity has not changed and is equal to 34.915 (all samples below 600m and with a temperature lower than -0.5°C are considered part of this water mass). This water is not freshly ventilated, so it is unlikely that its salinity has changed by more than 0.01psu during the whole historical record. The corrections applied sometimes exceed 0.10psu, for example, in recent Russian oceanographic cruises and various Scottish cruises in particular in 1949, 1950 and 1953. These errors probably originate from the use of a sub-standard sea water to calibrate the measurements instead of the Copenhagen standard water. For cruises during which the sampling did not reach the deep waters, we have adopted the correction established for the closest cruise of the same vessel. We also lower by -0.02psu the surface salinity of stations before 1914 which were sampled by bucket (Table 2).

Variability between nearby samples in this domain has a *rms* of around 0.06psu. Because there are few samples, errors related to high frequency, small-scale variability are often quite large and the low-frequencies are not well defined, except during 1948-1960 for the surface data (1107 samples) and the 1948-1990 period for the hydrographic station data (1079 casts) (Fig.6a,b). In recent years, the time series constructed from hydrographic station samples at the World Data Centre is less complete than time series presented in DOOLEY *et al* (1984) and TURRELL and SHELTON (1993). The sampling uncertainty is often large, but some of the differences in the low-frequency salinity anomalies between the ships-of-opportunity and hydrographic casts samples (Fig.6a,b) are large and are also found if seasons are analysed individually, for instance, the maximum salinity is in 1950 for the ships-of-opportunity surface data, and around 1960 and 1968 in station data. The higher surface salinity around 1950 corresponds to the anomalous deviations presented in Table 3. Once these corrections are applied, the time series from station and surface data become statistically identical. The time series based on samples at 10m depth (Fig.6c) closely mimics the surface series, but with slightly lower values in the late 1930s and 1950s. The differences are larger at a depth of 100m (Fig.6d) with a more pronounced peak near 1930 and a more pronounced minimum in 1910.

Hydrographic sampling barely resolves the low frequencies signal, so it is interesting to combine



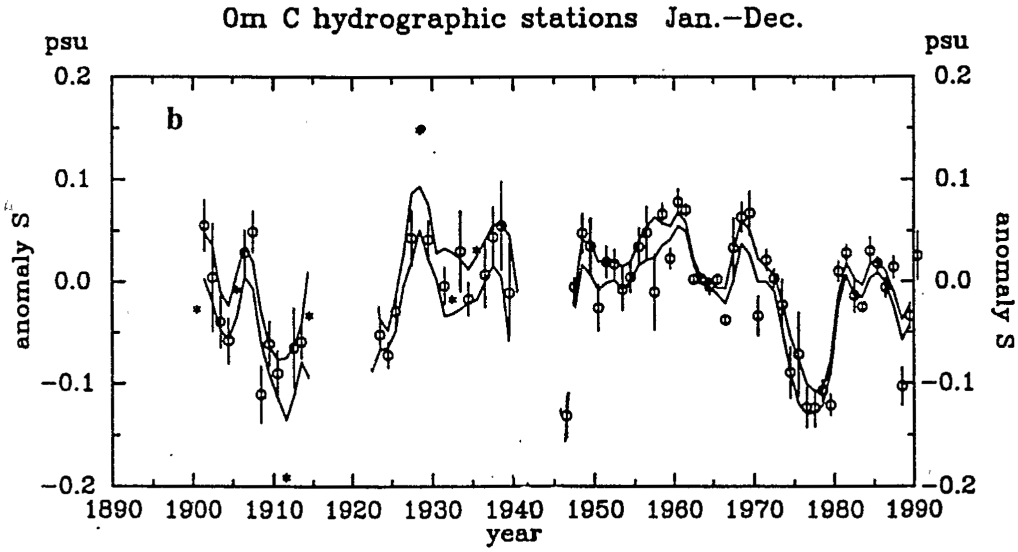
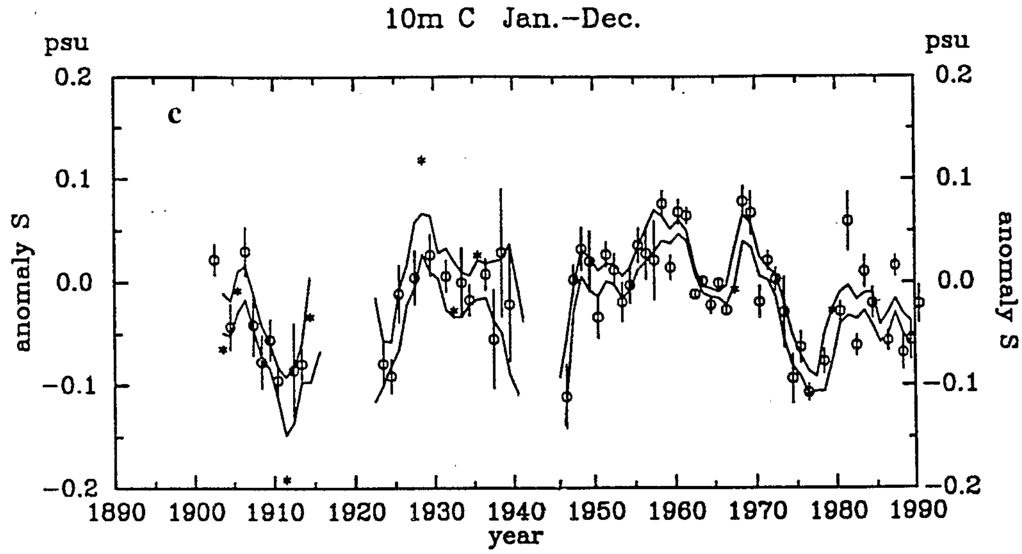
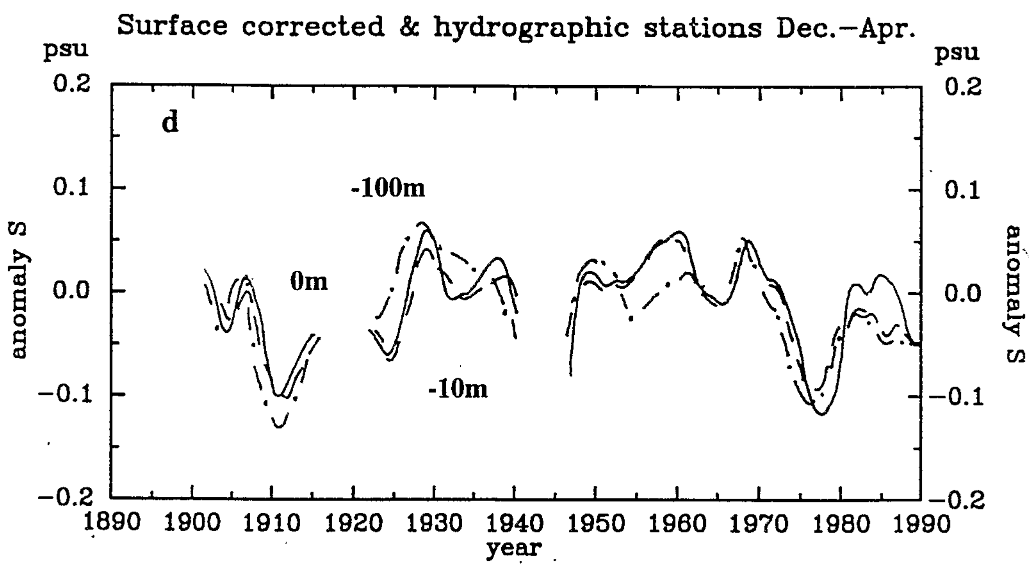


Fig.6. Seasonal average salinity anomalies from an average seasonal cycle in the central Faeroe-Shetland Channel (C on Fig.1b). (a) for 1107 uncorrected surface samples; (b) (right) for 1079 surface values of hydrographic casts; (c) for 823 salinities between -5m and -15m in hydrographic casts. Dots represent the median of the anomalies for a year period and the bars ±σ of its error based on the *rms* value, assuming all data independent and a Gaussian distribution (when there are less than 3 samples, a star is shown without error bars). The curves enclose the ±σ likely range of the binomial filter of a spline-fit to the annually binned data. Low-frequency time series from hydrographic casts at 0m, -10m and -100m are overlaid on Fig.6d (at each depth, we estimate a different seasonal cycle).


10m C Jan.–Dec.
psu
psu
0.2
0.1
0.0
−0.1
−0.2
c
anomaly S
1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990
year


Surface corrected & hydrographic stations Dec.–Apr.
psu
psu
0.2
0.1
0.0
−0.1
−0.2
d
-100m
0m
-10m
anomaly S
1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990
year

the hydrographic data with the surface samples, corrected as suggested by Table 3. We find similar time series (Fig.7) for the combined data sets (surface corrected and hydrographic stations) for neighbouring domains in the eastern and western Faeroe-Shetland Channel (Fig.1b), which confirm what was concluded from analysing the hydrographic casts only. To investigate the seasonal dependence of the signal, we have combined these 3 domains which extend from shelf to shelf (full line on Fig.7 for the annual average). In both cold (Fig.8b) and warm season (Fig.8d) salinity time series, the waters are freshest in the late 1970s and show maxima in the late 1920s, around 1950, 1960 and 1970. The large difference for the period 1910-1925 corresponds to a more irregular sampling; which was poor for the cold seasons of 1911 and 1922, and during the warm seasons of 1915 and 1921. Note that there is no data between the end of 1915 and 1920. Based on the similarity between the two seasonal time series, the time series on Fig.6d which uses all season data should provide a more reliable estimate of the low frequencies than the series for individual seasons. The errors related to the sampling are unlikely to be fully removed by integrating vertically over the water column, so it is not surprising that we find differences with other time series of the upper ocean salinity in different areas of the FS Channel. However, these time series, i.e. in DOOLEY *et al* (1984) for the upper 200m in the central area, and in ELLETT and TURRELL (1992) for the vertically integrated water column over the Shetland shelf edge (slope current), all show the minimum near 1910 and in the late 1970s. TURRELL and SHELTON (1993) time series for the North Atlantic water present higher salinities in 1946-1950. Not surprisingly, the differences since 1970, when the data set we used is less extensive than in other studies (ELLETT and BLINDHEIM 1992; ELLETT and TURRELL, 1992), are also very large.

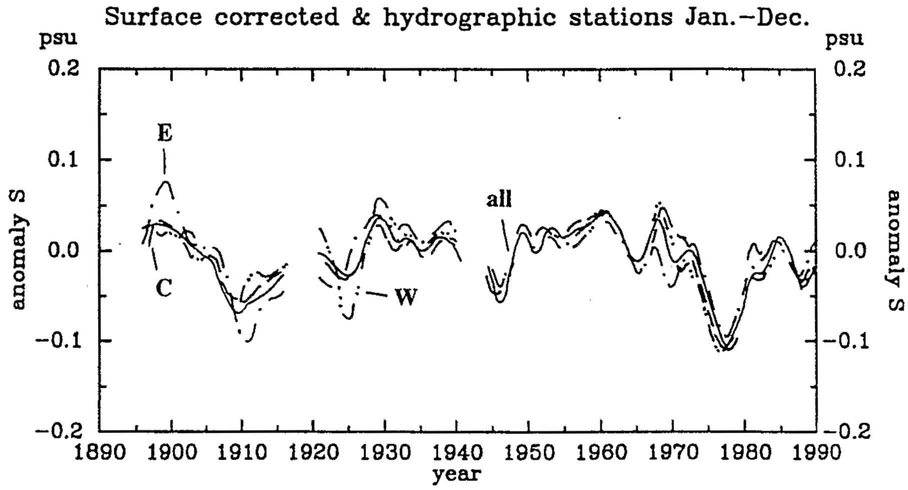


Fig.7. Time series of the low frequency salinity anomalies combining the hydrographic stations and the corrected surface samples for various subareas (W, C and E on Fig.1b) of the Faeroe-Shetland Channel. 'All' refers to the three areas combined (5954 samples).

The temperature anomalies are also shown for the two seasons. As expected, the cold season anomalies are usually weaker, remaining between -0.5°C and 0.5°C for the whole record with maximum values around 1960 and lowest anomalies around 1905. The total increase between the two periods is however less than for the Faeroe coastal stations (HANSEN and MEINCKE, 1984) where it exceeds 1°C. This is probably a real difference related to the coastal sites which are separated by a wide shelf from the deep areas we investigate.

Table 3. Comparison of surface samples with nearby data collected within 20 days and 14km. Average and *rms* difference are indicated both for T and S. Differences larger than 0.5psu or 2°C are removed (except for the NMFS data, where differences larger than 1psu are removed) and the average of the differences is given as well as the *rms* deviation and the number of different surface samples included. The title gives the two sets (1 and 2, where the differences are taken as 1-2) where ICES corresponds to the surface data from ICES, *Dutch, UK, Danish,* and *Norwegian* refer to the country in charge of the monitoring (subsets of the ICES file). After 1948, *UK* and *Dutch* refer to the weather ships; for the earlier period they correspond to liners. *ORSTOM* refers to the French program monitored by ORSTOM (the agency responsible for overseas research) and *FR* to all other French surface data; *hydro* includes all hydrographic casts; *salino* corresponds to data from thermosalinometer and near surface undulating CTD. The range of years is indicated, as well as the average difference, the standard deviation and the number of samples.

Title	Years	ΔT (°C)	σ_T (°C)	ΔS (psu)	σ_S (psu)	n	area and comments
Ocean Weather Ships surface data							
Dutch-hydro	50,55,56	0.05	0.63	0.006	0.128	462	Northeast Atlantic
ICES-hydro	60-64	0.09	0.69	0.003	0.123	218	Northeast Atlantic
Dutch-hydro	65-67**	-0.19	0.79	0.080	0.171	439	(also in North Sea)
Dutch-hydro	68-76	-0.24	0.55	0.020	0.071	323	Northeast Atlantic
UK-hydro	64-76	-0.02	0.70	0.003	0.126	149	Northeast Atlantic
Dutch-UK	55-63	-0.32	0.69	0.010	0.084	443	Northeast Atlantic
Dutch-UK	55-63	-0.15	0.71	0.006	0.150	1944	North Sea
Dutch-UK	64-67	-0.30	0.76	0.003	0.153	710	North Sea
Dutch-UK	68-70	-0.21	0.79	0.001	0.190	626	North Sea
French surface data							
ORSTOM-hydro	77-78	0.06	0.90	0.095	0.160	66	Atlantic south of 50°N
ORSTOM-salino	83-89	0.04	0.83	0.091	0.138	21	Atlantic south of 50°N
ORSTOM-ICES	77-90	-0.25	0.71	0.069	0.251	52	Atlantic near 50°N
FR-hydro	57-90	0.01	0.71	0.041	0.148	805	mostly near France and
Weather Ships							
National Marine Fisheries Service surface data							
NMFS-hydro	71-90	0.00	0.99	0.078	0.371	499	Western Atlantic
ICES liners							
Danish-hydro	05-14	-0.05	0.75	-0.008	0.060	32	Faeroe-Scotland
Danish-hydro	20-30	0.11	0.50	0.014	0.087	10	Faeroe-Shetland
Danish-hydro	31-39	0.05	0.71	0.067	0.103	62	Faeroe-Shetland
Norwegian-hydro	31-39	0.37	0.55	-0.001	0.061	15	Faeroe-Shetland
Danish-Norwegian	31-39*	-0.42	0.70	0.050	0.093	93	Faeroe-Shetland
Danish-hydro	05-39	-0.24	0.79	0.042	0.069	24	52-64°N, 7-35°W
Danish-hydro	45-47	-0.22	0.72	0.020	0.074	28	Faeroe-Shetland
Danish-hydro	48	0.07	0.57	0.053	0.068	30	Faeroe-Shetland
Danish-hydro	49	0.03	0.62	0.130	0.075	26	Faeroe-Shetland
Danish-hydro	50	0.27	0.53	0.095	0.070	63	Faeroe-Shetland
Danish-hydro	51	0.29	0.59	0.089	0.058	56	Faeroe-Shetland
Danish-hydro	52	0.26	0.57	0.137	0.065	70	Faeroe-Shetland
Danish-hydro	53	0.10	0.62	0.069	0.055	63	Faeroe-Shetland
Danish-hydro	54	0.03	0.39	0.062	0.057	21	Faeroe-Shetland
Danish-UK	55-60	-0.55	0.62	0.020	0.068	23	North Atlantic
Danish-Dutch	55-60*	-0.22	0.70	0.035	0.099	112	North Atlantic
ICES-hydro	05-17	0.23	0.83	0.028	0.185	20	40-50°N, 8-35°W
German-hydro	28-29	-0.04	0.70	0.045	0.171	24	40-50°N, 8-35°W
UK-hydro	20-29	0.02	0.86	0.112	0.150	14	40-50°N, 8-35°W
UK-hydro	30-39	0.01	0.99	0.061	0.178	16	40-50°N, 8-35°W

*The data suggest that the temperature difference is larger in winter than in summer.

**The comparisons for 1965-67 between the Dutch and UK surface samples suggest a bias different from the one for the comparison with the station data, suggesting some inhomogeneity within this set or in the station data.

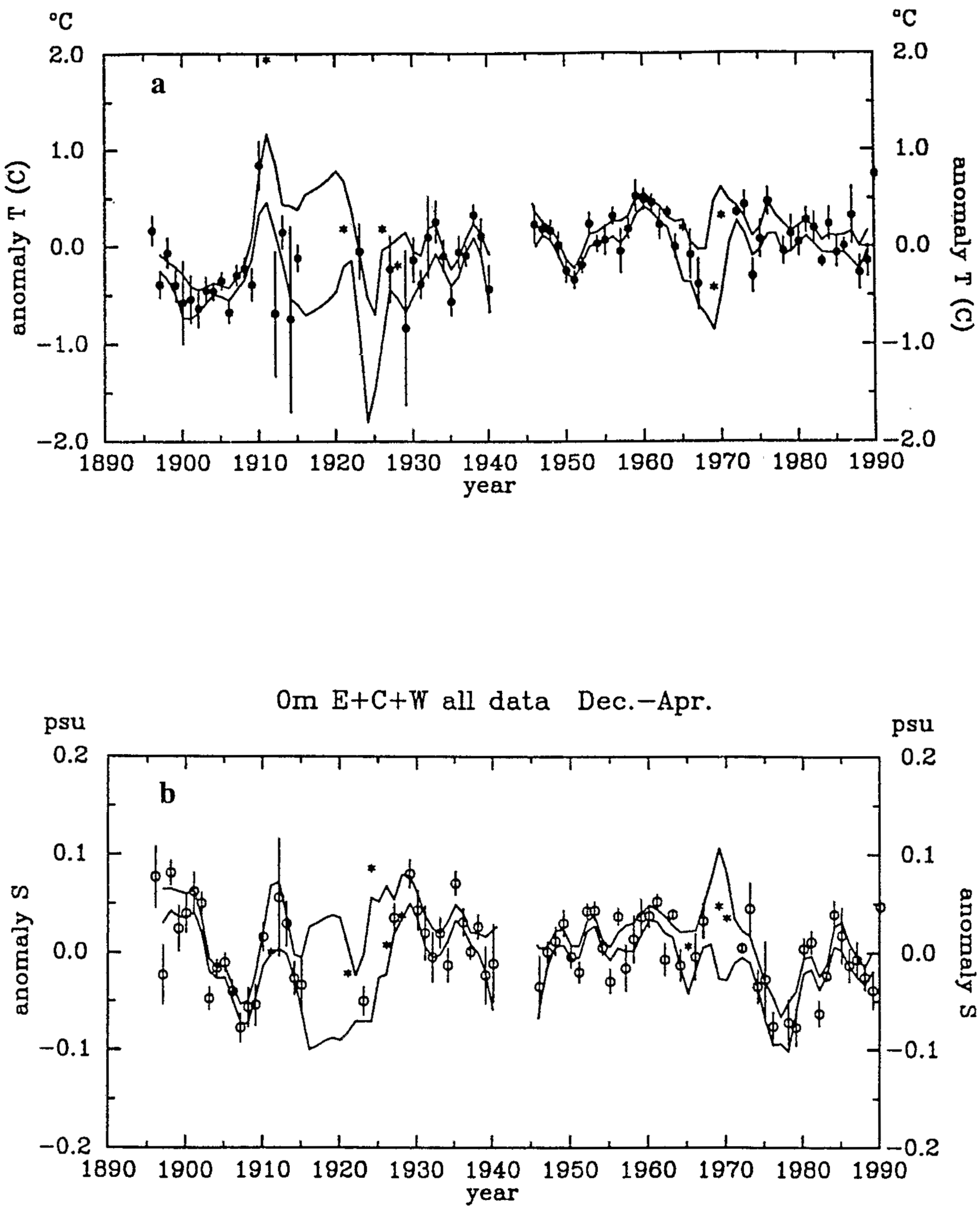
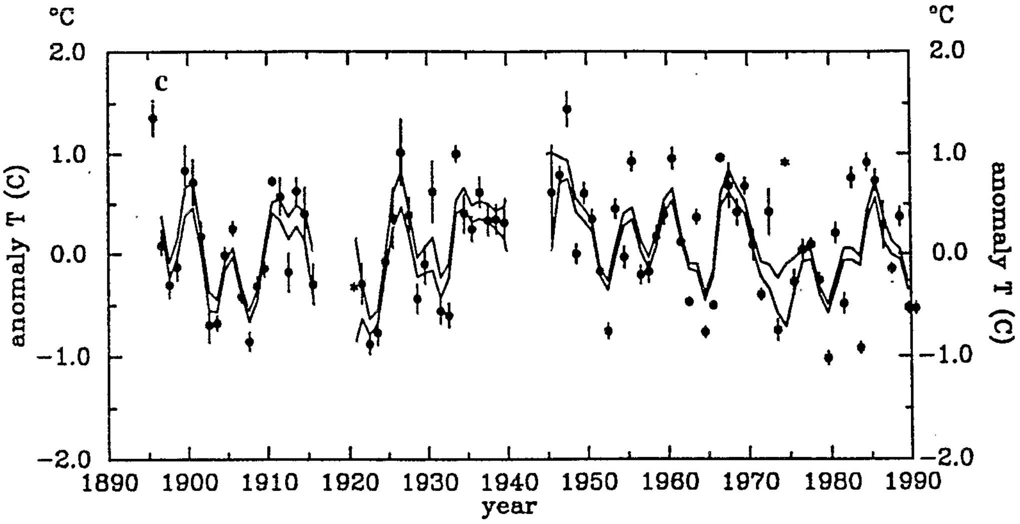
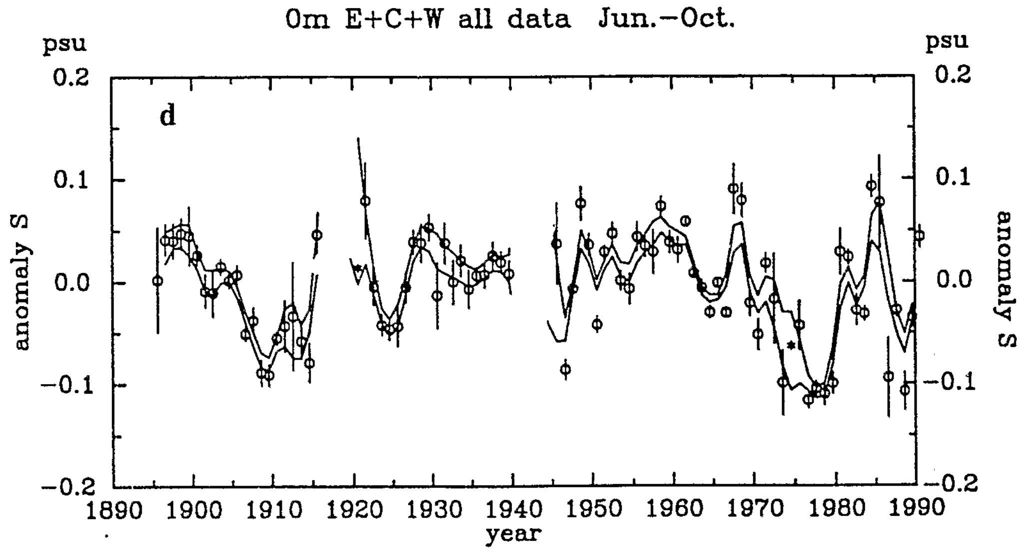


Fig.8. Same as Fig.6 but combining the hydrographic stations and the corrected surface samples for the whole Faeroe-Shetland Channel (areas W, E and C together) and selecting a season. Both temperature and salinity anomalies are presented in Fig.8a (T) and 8b (S) for the December through April season (1463 samples); Fig.8c (T) and 8d (S) for the June through October season (3276 samples).


°C
°C
2.0
1.0
0.0
−1.0
−2.0
c
anomaly T (C)
1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990
year


0m E+C+W all data Jun.−Oct.
psu
psu
0.2
0.1
0.0
−0.1
−0.2
d
anomaly S
1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990
year

8.2 Eastern Atlantic

The other region crossed by different ICES lines and where hydrographic stations were collected regularly is in the eastern Atlantic south-west of Ireland and west of the English Channel (Fig.1b) (48°N - 50°N, 8°W - 12°W). In the vicinity of the shelfbreak is an intermittent poleward flow (slope current) with seasonal and interannual variability (PINGREE and LECANN, 1990); the ocean interior is more quiescent and often has deep mixed layers in winter. Before 1945 hydrographic observations were carried out mostly on the shelf or on the slope. We have excluded most of the shelf area because surface sampling there is very different from the open ocean, with many samples collected during oceanographic cruises and along short lines between Ireland, England and France. The box-division of this region is designed to separate the shelf break from the open ocean.

Unfortunately, the *rms* variability in surface salinity is larger than in the Faeroe-Shetland area, and the time series constructed have larger uncertainties, comparable to the magnitude of the interannual signal (Fig.9). Still, the time series are very different from the ones for surface data without the corrections of the order of 0.10psu which were based on Table 3. The corrections were important for reducing the differences between surface data and hydrographic casts time series. However, these time series suffer from some problems. The larger *rms* variability results in the salinity seasonal cycle being less certain in the area with errors probably exceeding 0.02psu and 0.2°C for individual months. Another caveat is that we did not correct the hydrographic stations as we did in section 7a. Furthermore, the large average corrections that we adopt for the 1920s and 1930s from Table 3 have uncertainties of the order of 0.05psu and this also indicates a likely large scatter of the error from sample to sample.

The data are however appropriate for investigating whether the low-frequency anomalies vary with season and whether they are spatially coherent. The seasonal time series for slope box 2 (Fig.10) are typical of the differences between seasons. The errors on the low-passed time series are comparable for the two seasons and are illustrated on Fig.9 for the cold season (there are more data in the warm season but *rms* deviations are larger). The salinity time series post 1948 suggest deviations are higher by 0.02psu in the warm season than in the cold season, which may result from an error in the average seasonal cycle we used. However, between 1925 and 1935 the summer deviations are more negative. This minimum in the warm season time series is featured both in the hydrographic stations and in the corrected surface data. Large positive deviations in early 1928 and 1929 (Fig.9), and negative deviations in mid-1931 contribute to the difference, but even without those years the cold season deviations are still 0.05psu higher than in the warm season. The cold season minimum in 1935-37 is a well-sampled feature which occurred elsewhere, but not in the hydrographic stations data, whereas a minimum around 1910 is found in both time series. Earlier on, there are large positive deviations which are dubious according to a study of the surface monitoring prior to 1900 (REVERDIN, 1993). However, it is reassuring that similar deviations also occurred during the cold seasons in 1904-1906 when the monitoring was reinstated and recent anomalies also reach a comparable level. The temperature time series at this location for the two seasons does not present any similarities, although most of the differences are not significant according to the estimated errors. The often negative cold season deviations in 1905-1925 which are also found in other areas, have no clear counterpart during the warm season.

The seasonal differences in salinity illustrated for this area are typical of other locations and are probably significant, with the caveat that our corrections for 1920-1939 are based mostly on summer data and might not apply as well to winter data. However, there is the additional issue of spatial homogeneity in the deviations which we have combined for creating time series. The spatial variability of the deviations may be large for the warm season, larger than is the case for

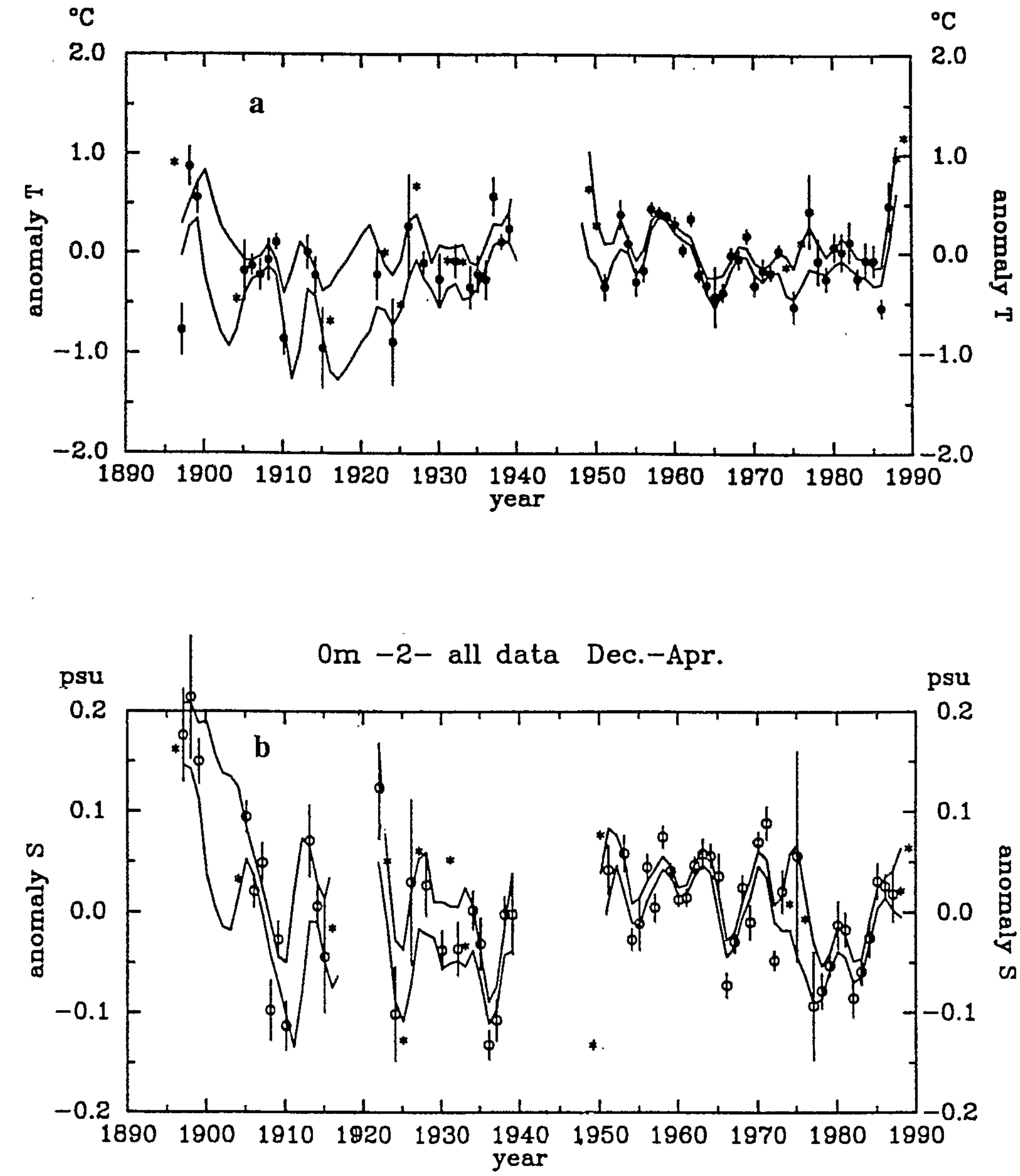


Fig.9. Same as Fig.6 combining the hydrographic stations and the corrected surface samples for area 2 on Fig.1b. Fig.9a (T) and 9b (S) for December through April season (1129 samples).

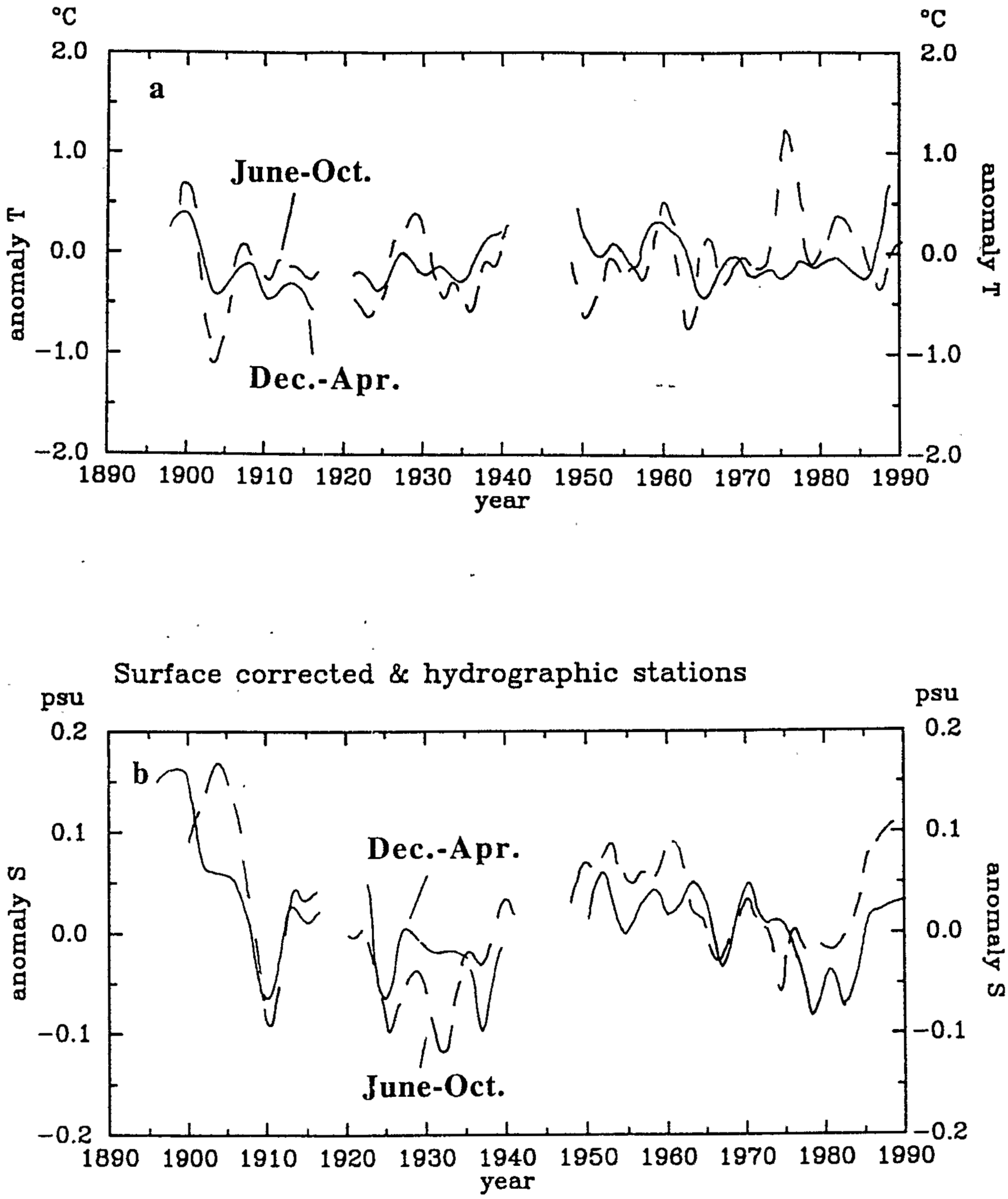


Fig.10. The low-passed time series combining the hydrographic stations and the corrected surface samples for area 2 on Fig.1b. The cold and warm season curves are overlaid for (10a) T and (10b) S (1129 and 2120 samples for the two seasons respectively).

the cold season, which we illustrate in Fig.11. Most of the prominent features of the time series are present in at least three out of the four time series. For salinity, this is the case for the low salinities in 1908-1910, near 1915, 1925, 1935, in the mid-1960s and in the mid-1970s. This is also the case for the positive salinity deviations around 1900, in the late 1930s, late 1950s and in the mid-1980s. The low salinities in the mid-1950s are only found for the two northern boxes. The usually lower temperatures for the period 1905-1920 than for more recent years, and the positive temperature deviations near 1930 and in the late 1950s are also shared by most time series. This suggests that averaging over larger areas, as was done in SMED (1943) with 5° by 5° boxes might not be a major source of uncertainty in the analysis of low frequency variability, at least for the cold season.

The minimum around 1910 in Fig.11 is also consistent with hydrographic data from north of 50°N in the summer of 1910 when there was a broad area of fresher waters (DICKSON, MALMBERG, JONES and LEE, 1984). The sharp increase of salinity between 1935 and 1939 occurred in all time series. The low salinity period in 1910-1920 was also found in the western English Channel (DICKSON, 1921). The low values in the late 1970s and the following increase to nearly unprecedented levels in recent years are also widespread features in the eastern Atlantic and on the continental shelf. For instance, the series from the Seven Stones Light Vessel, west of Cornwall (50°04'N, 06°04'W) exhibits a salinity minimum in 1977-1978. The analysis of hydrographic conditions a little further west in the late 1970s by POLLARD and PU (1985) showed that the 1977 anomaly in the upper 400m extended north of 40°N and they contrasted conditions with those observed in 1957-1958. We also found that positive anomalies in 1957-1958 which switched to being negative in the late 1970s (Fig.9 and 11). However, as DICKSON *et al* (1988) commented, salinities around 1970 were also high, and the transition between the two periods was far from smooth as was assumed by POLLARD and PU (1985), and there was another minimum around 1965-67 both in surface and hydrographic data.

9. DISCUSSION

The question we pose is whether we can combine the surface observations with the hydrographic station data in order to infer the low frequency variations of the upper ocean over an extended period. This issue is important, because hydrographic stations alone barely provide the spatial coverage necessary to resolve the spatial scales of the low-frequency hydrographic signal. Actually, with the exception of the maps by LEVITUS (1989a,b) where the assumption is made that the low-frequency signal emerges from the 'noise' provided by season-to-season variability, most information on the variability is based on time series at a few locations. These time series include the *Panulirus* station close to Bermuda (JENKINS, 1982; TALLEY and RAYMER, 1982; LEVITUS, ANTONOV, ZENGSI, DOOLEY, TSERESCHENKOV, GULEV and MICHAELS, 1992), the surface Rockall Trough (ELLETT, 1982), station *Mike* at 66°N 2°E (GAMMELSRØD and HOLM, 1984), station *Charlie* at 52.75°N 34.5°W (LEVITUS *et al*, 1992) and the other weather ships (TAYLOR and STEPHENS, 1980). These are valuable in providing an estimate of the relative importance of the low frequencies with respect to the higher frequencies and on how deviations from the climatology in surface and subsurface conditions relate. Where water masses are homogeneous, such as in the Rockall Trough, the winter season salinity time series (Fig.12) are very regular and correlations between successive winters are large (0.64 correlation coefficient). However, these various records for hydrography are so far apart spatially that they are difficult to combine, even at low frequencies, so that the analysis of the variability has to some extent to be based on intuition, even for an event as large as the 'Great Salinity Anomaly' of the 1970s (DICKSON *et al*, 1988).

Potentially, it should be possible to complement the hydrographic data with the surface data

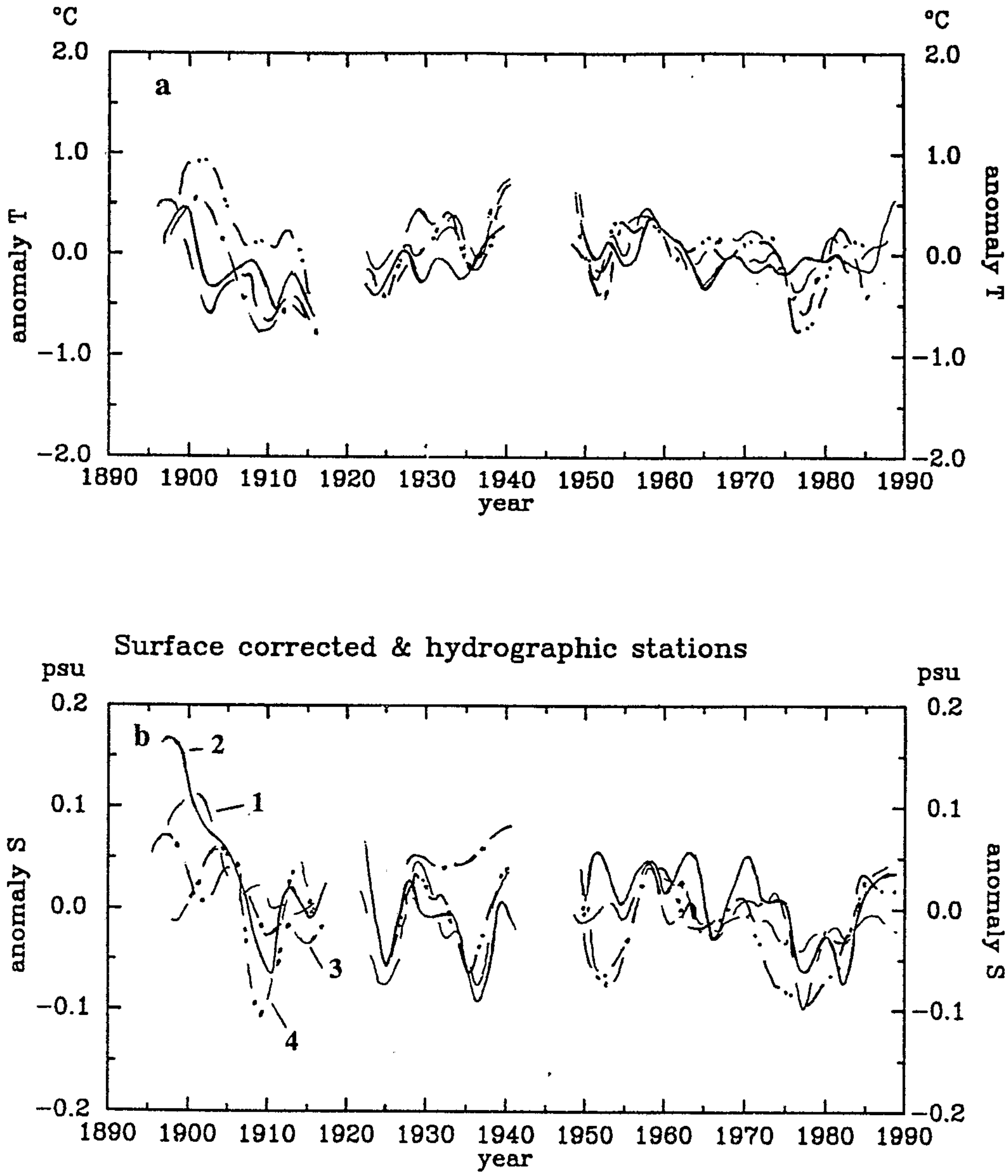


Fig.11. The low-passed time series combining the hydrographic stations and the corrected surface samples for 4 areas in the eastern Atlantic are overlaid for the cold (December-April) season. Notice that there is no data for 1915-1925 for the two northern areas. Fig.11a (T) and 11b (S).

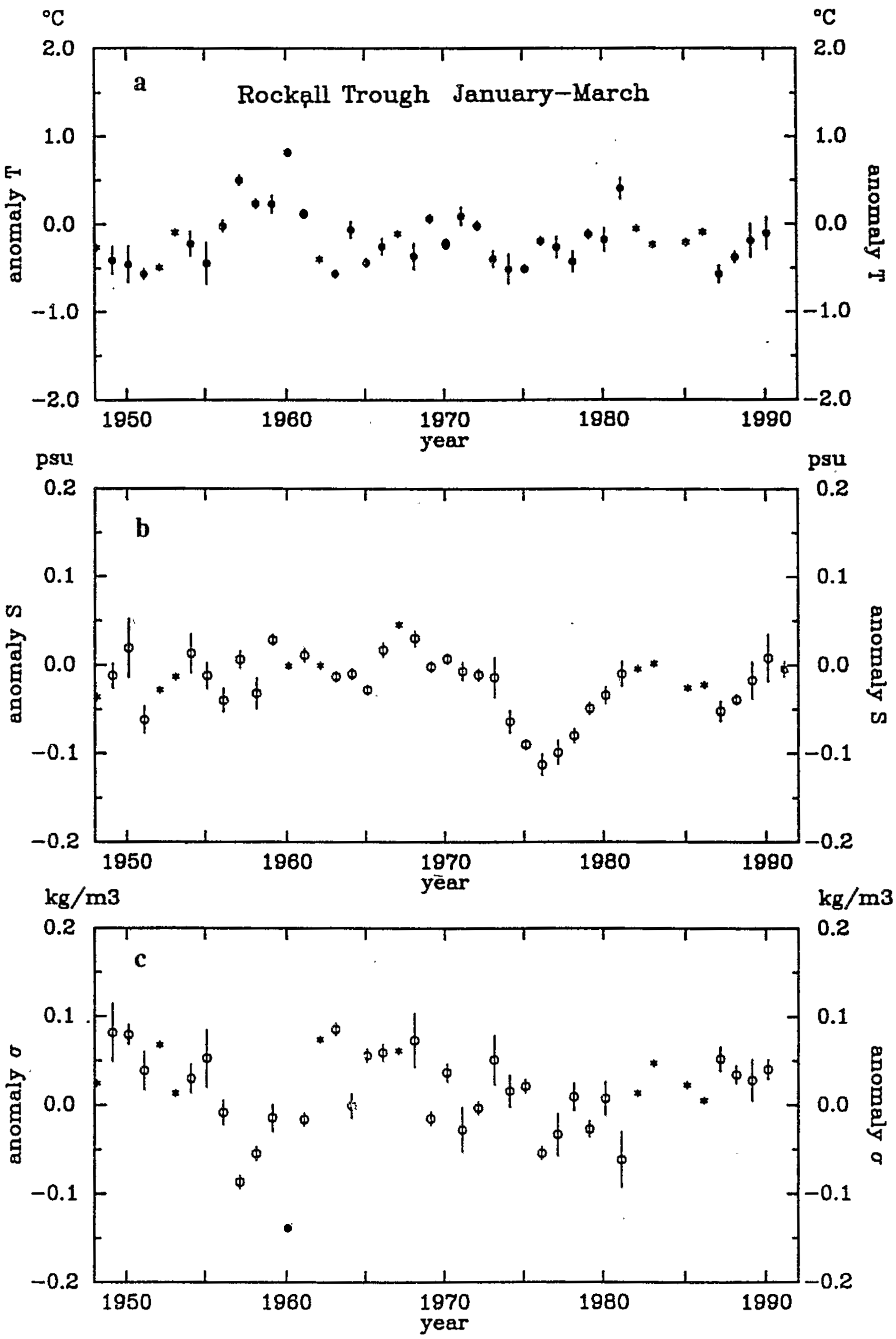


Fig. 12. Time series for the winter seasons (January through March) from 1948 to 1991 in the Rockall Trough area (details of the area and processing in ELLETT and MCDOUGALL, 1983). The average anomalies of, temperature (12a), salinity (12b) and density (12c) are presented as well as error bars based on the month-to-month variability in the winter season.

collected more or less regularly along shipping lines or during cruises over the last 100 years, although Fig.2 shows large areas remain devoid of any data, even on time scales of decades. One argument against analyzing the hydrographic changes from the surface data, is that the seasonal cycle of surface salinity is poorly resolved and that the surface variables are more 'noisy'. However, the picture that has emerged from the present analysis is somewhat more optimistic. To illustrate the signal over noise problem resulting from the seasonal cycle, we will compare the low frequency signal with the peak-to-peak change in the average seasonal cycle for the two domains discussed above. In the Faeroe-Shetland domain, we find a peak-to-peak seasonal signal of 0.10psu with a maximum in early June and a minimum in early October. Encouragingly, the same timing of the extrema was found by KNUDSEN (1905) from the 1897-1903 sampling when averaging his estimates between 3°W and 6°W, with a peak-to-peak amplitude of 0.12psu. We have also found that the interannual variability does not have a strong seasonal modulation. This assumption is shown by ELLETT and TURRELL (1992) to be reasonable for decadal averages of the surface salinity in the Rockall Trough. South-west of Ireland, the peak-to-peak seasonal cycle varies across the shelf-edge. On average it is 0.08psu with a maximum in May and a minimum in September. In both areas, the seasonal cycle contains less variance than the low frequency variability which has peak-to-peak signals of at least 0.15psu. Based on those examples, it is quite important to have a correct sampling of the interannual variability in order to have a good estimate of the average seasonal cycle of salinity. On the other hand, an inadequate knowledge of the seasonal cycle will not be a major hindrance for the investigation of the low-frequency variability, primarily during the cold season when the month-to-month change in the seasonal cycle is small.

These conclusions are based on a limited set of data for the north-east Atlantic, and may not be valid for other areas where the amplitude of the seasonal cycle is much larger, in particular in Arctic waters, in the western Atlantic or in tropical waters (LEVITUS, 1986; SMED, 1943). However, the low-frequency interannual anomalies are probably also larger there. For example, at ocean weather station *Charlie* (52°45'N 35°30'W), the peak-to-peak seasonal cycle is 0.16psu, but the lowering of the surface salinity between the winters of 1962 to 1968 and 1969 to 1973 exceeds 0.2psu (TAYLOR and STEPHENS, 1980). Similar conclusions hold at the *Panulirus* site near Bermuda (LEVITUS *et al*, 1992).

Another inconvenience in using surface data is that the high frequency variability is larger at the surface than a few tens of meters below the surface, because of the surface air-sea exchange of water. However, the weather ship time series of surface salinity (often daily data) illustrate that the subseasonal variability is often not overwhelming for these sites where the winter mixed layer is deep or very deep. Typical values of the intra-seasonal *rms* variability in the north-eastern Atlantic during the winter season are in the range 0.04psu to 0.10psu (in winter time, the distribution of the monthly deviations is adequately modelled by a Gaussian distribution). In summer time, the standard deviation is often larger, but the distribution is less Gaussian with a long tail for the negative anomalies. Because so many of the surface data are loosely distributed, one would like to know what sampling frequency is needed to reconstruct the low frequency time series. We can illustrate this question by sampling the nearly monthly time series of the Rockall Trough by a few 'data' distributed randomly through the record, which are assumed to have a noise corresponding to the sample high frequency statistics (*rms* 0.05psu for S and 0.40°C for SST). For the January-March season, we find that with 20 (40) data, the standard error on the low-frequency signal of 0.017 (0.009)psu compared to a *rms* low frequency signal of 0.029psu. The situation there is less favourable for temperature or density, because of higher season-to-season variability (with 40 samples the *rms* error for temperature is 0.13°C for a signal *rms* of 0.10°C and for density 0.027kg m^{-3} compared to a signal of 0.028kg m^{-3}).

Although this suggests that surface samples, even loosely distributed should contribute to defining the past low frequency variability, this study illustrates the difficulty of validating the very inhomogeneous set which one has to use. This problem was recognized by LUMBY (1935) who commented: "For instance, a case arose in which two ships running over the same route (not in the English Channel) took alternate series of samples, and it was noticed that the two sets were inconsistent. The discrepancy alone was not sufficient to point definitely to inaccuracy in either one set or the other, but subsequent inquiry elicited the information that the observer in one of the ships was not well inclined towards the work...".

We have documented the possibility of errors related to the collection of the water or to the storage if not rinsed properly. For example, a French bucket with a rubber rim can strongly contaminate a sample of sea water. Bottles with loose caps can suffer from extensive evaporation during the storage. Most of these errors will result in an increase of a sample's salinity. The other errors, either reporting errors (position, time or value), or errors from the titration (originating from the uncertainty in the end-point or of the chlorinity of the standard water) are more likely to be normally distributed with a zero mean. We find that the combination of these various effects often results in a very skewed distribution of the errors toward the positive deviations. These errors imply that a salinity spatial average from surface samples sometimes exceed by about 0.10psu the average obtained from hydrographic stations. The median is a slightly less erroneous estimate, but is also biased towards positive values. Many surface samples at hydrographic stations were also drawn by bucket and are therefore subject to similar contamination, as shown by comparisons with the first subsurface bottle sample. For hydrographic stations, at least in winter, it is often possible to replace the value from the surface sample with that from the first subsurface bottle sample.

The problem with the systematic errors, in particular those associated with use of buckets and uncalibrated thermo-salinometers, remains a concern. It is interesting to refer to the studies by PARKER and FOLLAND (1991) and FOLLAND (1991) of the temperature bias with canvas buckets which has a complicated spatial pattern and is also seasonally modulated. It could be that the contamination of samples collected with this bucket would also have a spatial pattern. Unfortunately, the validations are mostly for the north-east Atlantic and information from elsewhere is scant. In the western Atlantic, a comparison based on 30 surface samples of mixed origin between 1930 and 1939, mainly along the edge of the Grand Bank, suggests that the average bias of the surface samples does not exceed 0.09psu at the 95% confidence level. This is encouraging, because interannual signals there seem to have a large amplitude (PETRIE *et al*, 1992). In many cases we have no hint of why the errors have a specific magnitude in a given set and why they appear to change at certain times. It may be that undocumented modifications were introduced as investigators became aware of various problems; for example, we are aware of three such changes in the last 20 years of ORSTOM data (see Appendix A).

In this connection, it is important to assemble all available information on the sampling methods used before they are lost forever. We have shown that such details as the type of bucket or sampler, the type of sample bottle and the method of salinity determination have important bearings upon the reliability of the data, and these with any comments by the originator of the series should form an essential part of the 'data archaeology' projects now proposed. Experience shows that it is sometimes difficult to obtain this background information even for current programs.

Based on what we have presented in this paper, the interdecadal signal presented by SMED (1943) mainly from uncorrected surface samples is questionable south of 50°N for Smed's areas E to I. We should, however, mention that although the inconclusive comparisons we were able to carry out on the 1920s Danish samples suggested a possible positive bias (0.06psu based on 9 samples), such a bias is not large enough to explain the late 1920s maximum on SMED's (1943) low-passed curves in the subarctic gyre (areas A to C). The errors in surface samples are large enough to have strongly distorted the low-pass time series of area averaged salinity, as illustrated for two areas in the north-east Atlantic. In these two cases, corrections suggested from direct comparisons with neighbouring stations resulted in time series, which, if not fully satisfying, are at least realistic and complement earlier analyses from hydrography. Actually, the Faeroe-Shetland Channel time series clearly shows that the 1908-1910 low-salinity episode in the North Atlantic waters was not as extensive as the low salinity anomaly of the mid-1970s. The second example in the eastern Atlantic is also interesting, as it suggests there have been other fresh episodes with rapid changes, in addition to that of the 1970s. These encouraging results imply that it is possible to improve the data sets by elimination of dubious data along particular crossings and by the statistical correction of the systematic errors in some of the sets. This suggests that corrected surface data will help to sample the higher-frequency interannual variability (examples for subsets presented in KERNE, 1931, and REVERDIN, 1993) in order to avoid a serious aliasing problem when retrieving the low-frequency variability.

10. ACKNOWLEDGEMENT

The collection of information required spending many days in locating old sets of data. This effort was kindly supported by many people, colleagues, archivists or librarians who I besieged with obscure requests. It is a pleasure to acknowledge the support of librarians of numerous institutions (Lamont-Doherty Earth Observatory, Woods Hole Oceanographic Institution, Institute of Oceanographic Sciences at Wormley, the Bibliothèque de l'Institut Océanographique in Paris, the Bundesamt für Seeschiffahrt und Hydrographie, among others). Yves Gouriou spent many hours on this project locating various documents in Woods Hole. Stig Fonselius and the Archivist of the Museum of Natural History in Gothenburg were instrumental in locating late 19th century Swedish archives, and Else Nielsen located various data files and documents at ICES and other institutes in Copenhagen. We also benefitted from visits to the ORSTOM Laboratory in Brest (in particular to Jean-René Donguy) to the Dunstaffnage Marine Laboratory, to Bill Turrell at the SOAFD Marine Laboratory in Aberdeen, to Johann Blindheim at the Havforskningsinstituttet in Bergen, to the Geofysik Institutt in Bergen (Tor Gammelsrød, Mister Biskopshavn and Odd Saelen), to IFM Kiel (Thomas Müller and Birgit Klein), and to NMFS Narragansett (Robert Benway). We should emphasize that the data in institution reports, in particular for the early years, proved to be a very useful source of information. Also, the fact that the early station data were not 'cleaned' was instrumental in giving a sense of what bias can result from an iron bucket. Help from the ORSTOM Centre in Nouméa for testing the French bucket and from NODC, ICES and BODC (Roy Lowry) for additional data are gratefully acknowledged. The years 1904, 1950, 1955-1956 of the *Bulletin Hydrographique* lists were digitized by Pamela Tan, and 1925, 1936-1937 were digitized by Elizabeth Jackson. This work is supported by NODC under NOAA grant 40ANNE201417 and by the NOAA Atlantic Climate Change Program grant NA90AA-D-AC502.

11. REFERENCES

BRENNECKE, W. (1921) Die ozeanographischen Arbeiten der Deutschen Antarktischen Expedition 1911-1912. *Aus dem Archiv der Deutschen Seewarte*, **39**, n.1.

CLEVE, P.T., G. EKMAN and O. PETTERSSON (1901) Les variations annuelles de l'eau de surface de l'océan Atlantique. *Göteborg Künglija Svenska Vetenskaps Handlingar,* Ser.4, 3.

DELCROIX, T. and C. HÉNIN (1991) Seasonal and interannual variations of sea surface salinity in the tropical Pacific Ocean. *Journal of Geophysical Research*, **96**, 22135-22150.

DICKSON, H.N. (1901) The circulation of the surface waters of the North Atlantic Ocean. *Philosophical Transactions of the Royal Society, London*, Ser.A, **196**, 61-201.

DICKSON, R.R. (1971) A recurrent and persistent pressure-anomaly pattern as the principal cause of intermediate-scale hydrographic variation in the European shelf seas. *Deutsche Hydrographische Zeitschrift*, **24**, n.3, 97-119.

DICKSON, R.R., S-A. MALMBERG, S.R. JONES and A.J. LEE (1984) An investigation of the earlier Great Salinity Anomaly of 1910-1914 in waters west of the British Isles. *International Council for the Exploration of the Sea, Contributions to Council Meetings*, 1987/C:21.

DICKSON, R.R., J. MEINCKE, S-A. MALMBERG and A.J. LEE (1988) The 'Great salinity anomaly' in the northern North Atlantic 1968-1982. *Progress in Oceanography*, **20**, 103-151.

DONGUY, J.R. and A. DESSIER (1993) Sea surface salinity in the tropical Atlantic ocean (10°S-30°N). Seasonal and inter-annual variability (1977-1986). *Deep-Sea Research*, **41**, 81-100.

DOOLEY, H.D., J.H.A. MARTIN and D.J. ELLETT (1984) Abnormal hydrographic conditions in the northeast Atlantic during the 1970s. *Rapports et Procès-Verbaux du Conseil Permanent International pour l'Exploration de la Mer*, **185**, 179-187.

ELLETT, D.J. (1982) Long-term water-mass changes to the west of Britain. In: *World Climatic Research Programme Report*, **WCP21**, D.J. ELLETT, editor, WMO, Geneva, 245-254.

ELLETT, D.J. and J. BLINDHEIM (1992) Climate and hydrographic variability in the ICES area during the 1980s. In: *ICES Marine Science Symposium*, **195**, 11-31.

ELLETT, D.J. and N. MCDOUGALL (1983) Some monitoring results from west of Britain. *Technical Series, Intergovernmental Oceanographic Commission, UNESCO*, **24**, 21-25.

ELLETT, D.J. and W.R. TURRELL (1992) Increased salinity levels in the northeast Atlantic. *International Council for the Exploration of the Sea, Contributions to Council Meetings*, 1992/C:20, 12pp.

FOLLAND, C.K. (1991) Sea temperature bucket models used to correct historical SST data in the Meteorological Office. *Climate Research Technical Note*, **14**, Hadley Centre for Climate Prediction and Research, UK. 29pp.

FOLLAND, C.K. and D.E. PARKER (1990) Observed variations of sea surface temperature. In: *Climate-Ocean Interaction*, M.E. SCHLESINGER, editor, Kluwer, Dordrecht, 21-52.

GAMMELSRØD, T. and A. HOLM (1984) Variations of temperature and salinity at station M (66°N, 2°E) since 1948. *Rapports et Procès-Verbaux du Conseil Permanent International pour l'Exploration de la Mer*, **185**, 188-200.

HANSEN, B. and J. MEINCKE (1984) Long-term coastal surface temperature observations at the Faroe Islands. *Rapports et Procès-Verbaux du Conseil Permanent International pour l'Exploration de la Mer*, **185**, 162-169.

HELLAND-HANSEN, B. and F. JANSEN (1909) The Norwegian Sea. Its physical oceanography based upon the Norwegian researches 1900-1904. *Report on Norwegian Fishery and marine Investigations.* **Vol.II, n.2.**

ICES (1962) Mean monthly temperature and salinity of the surface layer of the North Sea and adjacent waters from 1905 to 1954. ICES Service Hydrographique, Charlottenlund. 26 charts and 293 tables.

JENKINS, W.J. (1982) On the climate of a subtropical ocean gyre: Decade timescale variations in water mass renewal in the Sargasso Sea. *Journal of Marine Research*, **40**, Suppl. 265-290.

KERNER, G. (1937) Die Salzgehaltschwankungen im Oberflächenwasser des Nordatlantischen Ozeans während der Jahre 1920-1931. *Aus dem Archiv der Deutschen Seewarte und des Marineobservatoriums*, **57, n.8.**

KNUDSEN, M. (1899) *Ingolf, 1895-96: Hydrographi.* Copenhagen, 1899.

KNUDSEN, M. (1905) Contribution to the hydrography of the North Atlantic Ocean. *Meddelelser fra Kommissionen for Havundersogelser. Serie Hydrografi*, **I, n.6.**

KIRKWOOD, D.S. and A.R. FOLKARD (1987) Results of the ICES salinity sample-bottle intercomparison, Hydrography Committee, 1986. *International Council for the Exploration of the Sea, Contributions to Council Meetings*, 1987/C:21.

LEVITUS, S. (1986) Annual cycle of salinity and salt storage in the world ocean. *Journal of Physical Oceanography*, **16**, 322-343.

LEVITUS, S. (1989a) Interpentadal variability of temperature and salinity at intermediate depths of the North Atlantic Ocean, 1970-1974 versus 1955-1959. *Journal Geophysical Research*, **94**, 6091-6131.

LEVITUS, S. (1989b) Interpentadal variability of salinity in the upper 150m of the North Atlantic Ocean, 1970-1974 versus 1955-1959. *Journal of Geophysical Research*, **94**, 9679-9685.

LEVITUS, S., J. ANTONOV, Z. ZENGSI, H. DOOLEY, V. TSERESCHENKOV, S. GULEV and T. MICHAELS (1992) Observational evidence of Decadal-scale variability of the North Atlantic Ocean. *In: . Natural climate variability on decade-to-century time scales*. D.G. MARTINSON, K. BRYAN, M. GHIL, M.M. HALL, T.R. KARL, E.S. SARACHIK, S. SOROOSHIAN and L.D. TALLEY, editors, National Academy Press, Washington D.C.

LUMBY, J.R. (1927) The surface sampler, an apparatus for the collection of samples from the sea surface from ships in motion. *Rapports et Procès-Verbaux du Conseil Permanent International pour l'Exploration de la Mer,* II, n.3, 332-342.

LUMBY, J.R. (1928) Modification of the surface sampler with a view to the improvement of temperature observation. *Rapports et Procès-Verbaux du Conseil Permanent International pour l'Exploration de la Mer,* **III**, 340-350.

LUMBY, J.R. (1935) Salinity and temperature of the English Channel. *Fishery Investigations, Series II,* **14**, n.3. 150pp and atlas.

MANNING, J. (1991) Middle Atlantic Bight salinity: Interannual variability. *Continental Shelf Research,* **11**, 123-137.

MARINI, L. (1912) Manuale per le osservazionne meteorologiche a mare. *R Comitato Tallassographico Italiano. Memoria IV,* C. FERRARI, editor, Venezia.

MATTHEWS, D.J. (1907) The surface waters of the north Atlantic Ocean south of 60°N latitutde, September 1904 to December 1905. *Marine Biological Association, Report Fishery and Hydrography Investigations North Sea and adjacent waters, 2nd Rep. (Southern area). Part I.* London.

PARKER, D.E. and C.K. FOLLAND (1991) Worldwide surface temperature trends since the mid-19th Century. In: *Greenhouse Gas induced climatic change: A critical appraisal of simulations and observations.* M.E. SCHLESINGER, editor, Elsevier Science Publications, Amsterdam, 173-193.

PETRIE, B., J.W. LODER, J. LAZIER and S. AKENHEAD (1992) Temperature and salinity variability on the eastern Newfoundland shelf: the residual field. *Atmosphere-Ocean,* **30**, 120-139.

PETTERSSON, O. (1894) A review of Swedish Hydrographic Research in the Baltic and the North Seas. *Scottish Geographical Magazine,* **10**, part I to VI.

PINGREE, R.D. and B. LE CANN (1990) Structure, strength and seasonality of the slope currents in the Bay of Biscay region. *Journal of the Marine Biological Association of the UK,* **70**, 857-885.

POLLARD, R.T. and S. PU (1985) Structure and circulation of the upper Atlantic Ocean northeast of the Azores. *Progress in Oceanography,* **14**, 443-462.

REVERDIN, G. (1993) Past surface salinity of the North Atlantic Ocean. LDEO-93-1. unpublished report.

ROBERTSON, A.J. (1907) Report on hydrographical investigations in the Faeroe-Shetland Channel and the northern part of the North Sea during the years 1904-1905. Second report (Northern area) on *Fishery and Hydrographical Investigations in the North Sea and adjacent waters, 1904-1905, Part I.* North Sea Fisheries Investigation Committee, London.

RODEWALD, M. (1972) Long-term variations of the sea temperature in the areas of the nine North Atlantic ocean weather stations during the period 1951-1968. *Rapports et Procès-Verbaux du Conseil Permanent International pour l'Exploration de la Mer,* **162**, 139-153.

SCHOTT, F. (1966) The surface salinity of the North Sea. *Deutsche Hydrographische Zeitschrift (Reihe A),* **8**, 1-58.

SCHUHMACHER, A. (1933) Zwei Oberflächenschnitte durch das nordwestafrikanische Auftriebwasser. *Annalen der Hydrographie und Marine Meteorologie,* **61**, 33-336 and Table 46.

SMED, J. (1943) Annual and seasonal variations in the salinity of the North Atlantic surface water. *Rapports et Procès-Verbaux du Conseil Permanent International pour l'Exploration de la Mer,* **112**, 77-94.

SUND, O. (1931) A new speed surface sampler. *Rapports et Procès-Verbaux du Conseil Permanent International pour l'Exploration de la Mer,* **VI**, 419-420.

SY, A. and H-H. HINRICHSEN (1986) The influence of long-term storage on the salinity of bottled seawater samples. *Deutsche Hydrographische Zeitschrift,* **39**, 35-40.

TALLEY, L.D. and M.E. RAYMER (1982) Eighteen degree water variability. *Journal of Marine Research,* **40**, Suppl. 757-775.

TAYLOR, A.H., P.C. REID, T.J. MARSH, J.A. STEPHENS and T.D. JONAS (1983) Year-to-year changes in the salinity of the southern North Sea, 1948-1973: A budget. In: *North Sea Dynamics,* SÜNDERMANN AND LENZ, editor, Springer-Verlag, Berlin, 200-219.

TAYLOR, A.H. and J.A. STEPHENS (1980) Seasonal and year-toi-year variations in surface salinity at the nine North Atlantic Ocean Weather Stations. *Oceanologica Acta,* **3**, 421-430.

TURRELL, W.R. and R.G. SHELTON (1993) Climatic change in the North-Eastern Atlantic and its impacts on salmon stocks. In: *Salmon at sea and new enhancement strategies.* D. MILLS, editor, Blackwell Scientific, Oxford, 40-78.

WISSER, S.W. (1938) *The Snellius Expedition in the eastern part of the Netherlands East Indies, 1929-1930. Vol.II Oceanographic Results. Part 4 Surface observations, temperature, salinity, density.* E.J. BRILL, editor, Leiden.

APPENDIX A: VALIDATION OF THE SURFACE DATA SETS

This section complements the investigation of statistical properties of deviations for a few sets of surface data relative to nearby hydrographic stations presented in section 7. Here we summarize the information available for each subset from similar comparison with either hydrological stations or other surface data. Unfortunately in many instances there are too few comparisons to establish comparable distributions to those in section 7 and we will only discuss the means and standard deviations of the differences (Table 3), although it is incomplete because the distributions are likely to be non-normal as were those in section 7.

A.1 UK and Dutch OWS samples

In the introduction, we reported that samples collected from the UK ocean weather ships (OWS) on its way to the Lima site at 57°N 20°30'W in 1988-1990 seem to show a slight bias and also contained positive outliers, which deviated from the conclusion for another period in section 7. More comparisons for UK and Dutch OWS are presented in Table 3. This could not be done for 1948-1949 and 1951-1954 when no country code is reported in our files. For the period investigated, UK OWS samples are drawn from the intake and Dutch OWS samples were drawn with a galvanized bucket. Comparison between the UK surface data and the station data suggests that there is no significant bias in salinity for different periods (within ±0.02psu at the 95% confidence interval). The Dutch surface data, on the other hand, usually show a small positive bias (of the order of 0.01psu), with some exceptions. Specifically, for the years 1965-1967, when compared with station data, the Dutch samples have an average positive bias of around 0.10psu in the southern north Sea, the English Channel and the near Atlantic south of 50°N (towards Ocean Station *Kilo*). This geographical distribution is not consistent with that obtained from comparisons with the UK surface samples for the same years, when the median differences show no large bias, suggesting inhomogeneities in the data set for those years.

There are also significant temperature differences between the sets. For instance, the Dutch temperatures are too low by 0.2°C compared to the station data before 1976. Between 1964 and 1976, they are also lower by 0.15 to 0.30°C than English OWS and other surface temperatures. This is compatible with the Dutch data for this period being measured from a bucket whereas the UK data were measured by a thermistor either hull-mounted or in an intake. Early OWS SSTs seem a little on the high side. More recently (after 1976) Dutch data no longer show significant differences with the station data.

A.2 French data sets

The French ORSTOM data set (close to 75,000 salinity data in the Atlantic) contains only samples drawn with the French bucket between 1977 and 1991 (we refer to this file as the ORSTOM file, in reference to the French agency ORSTOM which was responsible for collecting the data). Comparisons with station casts, calibrated thermo-salinometer data and SeaSoar data, and other surface data from English and Dutch Ocean wether ships (OWS) are presented in Table 3. Outliers larger than 0.5psu were initially removed because they usually are detectable as erroneous data in the eastern Atlantic. Throughout the ORSTOM salinities have a positive bias of around 0.08 to 0.10psu. It is not possible to distinguish whether this evolves in time or whether it presents a seasonal cycle. This average bias is close to that found for *in situ* measurements with the same bucket in the Pacific Ocean. The scatter in the comparison with station data has a *rms* deviation of 0.14psu; slightly larger than the *rms* deviation derived for UK or Dutch OWS surface data compared with station data in the same area of the North-east Atlantic. This suggests that ORSTOM measurements have a larger uncertainty, which is also our experience from direct use of the bucket. Assuming 0.10psu to be a typical estimate of the 'noise' associated with the measurements, more than 40 independent data are needed to reduce the effect of the random noise on the average to within 0.03psu at the 95% confidence limit (we have again assumed a normal distribution; if the distributions resemble the one of Fig.4, the median would converge a little faster than the mean). This is certainly a very stringent constraint for the study of the low-frequency signal, considering the fairly loose sampling. The average difference for temperature is compatible with a bias in the range ±0.1°C at the 95% confidence interval[1].

There is another French surface data set resulting from the compilation of data from different sources since 1957 (we refer to it as the FR set). It contains over 40,000 surface salinities, most probably collected with a bucket, usually with the French bucket (French OWS in particular). Comparison with station data shows that the set is not homogeneous. Some subsets have a slightly negative bias, for example in 1969-1974 off Brittany and at OWS site A (28°05'W, 62°N). At OWS site K (15°W, 46°N), on the other hand, there are many anomalously high salinities between 1957-1964. Undoubtedly, contamination was a serious problem during some of the oceanographic cruises (we removed some anomalous data before carrying out the comparison). The set includes some data which were not properly calibrated, for example in April 1974 or July and August 1977 to the west of France. The 'average' difference with station data is 0.04psu (within ± 0.01psu at the 95% confidence limit). It is interesting that the bias in this set is so different from that of the ORSTOM set, or from the bias we had determined during three cruises. One difference is that many vessels from which the FR data were collected, have a bridge closer to the water than the *ca* 30m bridge elevation common on vessels of the ORSTOM program. The frequency of sample collection in FR set is often higher than from the merchant vessels of the ORSTOM set (every 6 hours), so the buckets may have remained 'cleaner'. There may also have been undocumented changes in the buckets used for FR set, for which most data are earlier than the ORSTOM data (collected after 1976)[2].

A.3 National Marine Fisheries Service samples

Since 1971, the American National Marine Fisheries Service (NMFS) has been in charge of a collection of surface samples from merchant ships in the western Atlantic and Gulf of Mexico. Samples have been collected using a bucket without a rubber rim. Water is stored in small bottles (100cc) with a bakelite cap and plastic sub-cap. Surface samples collected during NMFS cruises on the American continental shelf are also included.

Large spatial gradients occur in the western Atlantic so the distribution of the deviations with nearby hydrographic stations has a large scatter. In the comparisons, a range of ±1psu retains 90% of the pairs (3% negative and 7% positive outliers in the differences NMFS-hydrographic station) within a total set of 555 pairs. The mean salinity difference is 0.08psu (±0.03psu 95% confidence interval assuming a normal distribution). The median is 0.06psu, but the large percentage of extreme positive salinity errors suggests that at least 2-3% of the samples have either not been

[1]The confidence interval estimate is computed from the standard deviation assuming that the data are independent and are normally distributed. The first assumption is not controversial, as each individual datum of the ORSTOM set is never considered more than once (i.e. if there is more than one station close to an ORSTOM sample, we average them before comparing). The second assumption is, however, not fully substantiated as is commented upon in Section 7. This is of little consequence for the uncertainty estimate.

[2]A few comparisons for FR data after 1976 exhibit a larger scatter, which could result from a different collection method.

reported correctly or originate from improperly sealed bottles. However, lower biases are to be expected from the southern mid-Atlantic Bight area when data are mostly collected during oceanographic cruises, and biases should be larger in particular in the earlier years of the sampling program when most samples from merchant ships were not analyzed quickly. Unfortunately, because of the large variability of the area, it is not possible to resolve these differences. The average SST deviation is 0 ± 0.1°C (95% confidence interval for a *rms* difference of 0.99°C).

A.4 ICES data from liners along shipping routes

Collection of data from merchant vessels was carried out within various national programs and coordinated by ICES until 1960. As mentioned earlier, there are many instances in those early years of surface sample at hydrographic stations also being taken by buckets, so the comparison between the surface samples from hydrographic stations and surface data do not provide a clear indication of the biases in the surface set (the station salinities are often too high, for instance for Scottish data in 1904, 1905 and 1909). Because the indication of the country is missing before 1955 (with the exception of 1950) and the ICES line number is not in our files, itis not easy to distinguish the different programmes. We tried to identify them by reconstructing individual crossings, but our efforts are prone to error in areas where lines intersect so our analysis does not include the North Sea, the English Channel, and the Irish Sea.

It is for the Danish data, primarily from the line to Iceland that the comparisons are most conclusive, and on which we will now comment. We find small average deviations from station data before 1914, and comparisons continue to be good until the late 1920s. Between 1931 and 1939, the line to Iceland was sampled by both Norway and Denmark. The Norwegian samples were collected every one to two hours and their salinity data exhibit no significant difference with station data; their temperature data, however, are much higher (more than 0.3°C). Interestingly, comparisons with published thermograms in *Bulletin Hydrographique* suggests that the SST reported with the salinity measurements are from the thermograph placed at the intake, and we are inclined to believe that the salinity samples were also drawn from the intake. The Danish surface salinities are too high both with respect to these Norwegian surface samples and to the station data. However, their SST is comparable to the station data but lower than the Norwegian SST. These differences in salinity are also found further west on the line to Greenland (also of the order of 0.05psu), but seem to have been less prior to 1931, although there are too few data to be conclusive.

After 1945, conditions in the Danish subset varied substantially from year to year, both for SST and salinity anomalies, suggesting the sampling techniques kept changing. For 1950-1952, for example, we find that temperature is too high by nearly 0.3°C and for 1949-1952 that salinity is too high by 0.1psu or more. The distribution is fairly narrow and fairly symmetrical (for instance, only 2 out of 70 differences in 1952 are negative and 2 positive exceed 0.25psu). One wonders how such an error distribution has arisen: it contrasts, for example, with the data collected with the French bucket which has a similar average error, but with a larger scatter. The positive temperature differences may have been the consequence of the SST having been measured by an intake for a couple of years, and mainly from a bucket for the others (there are suggestions that, even recently, canvas buckets were used on some Danish vessels to measure SST). It is also difficult to be sure whether or not there are differences between the two Danish lines monitored to Greenland and Iceland respectively, because there are far fewer comparisons with Greenland line although the comparisons appear to show similar biases.

Between 1955 and 1960 the set becomes more homogeneous. During this period comparison with station data suggests that the Danish surface salinities are too large by 0.05psu. Although the set is too small to be conclusive and a few outliers have a strong influence, the comparisons with the English and Dutch surface data imply similar mean (or median) bias. Interestingly, there are relatively few large outliers in the Danish data (4 out of 139 were too large by >0.4psu). There is a large negative SST bias for these Danish reports (about 0.2°C with respect to the Dutch data and 0.5°C with respect to the UK data, which suggests an error of about 0.4°C). This could originate from the temperature being measured by either a towed thermistor or in a canvas bucket.

The other area where samples can be compared with hydrographic stations is in the eastern Atlantic mainly between Ireland and Spain, monitored by UK liners in 1904-1909, 1913-1916 and 1922-1939, German liners in 1928-1939, and Dutch liners in 1912-1915. Based on a small set of comparisons with Danish, Irish or UK hydrographic stations, the UK data shows a small bias before 1914, but of a positive bias of at least 0.10psu in the 1920s, even after removal of the most outstanding error (crossings were removed if they presented at least 2 samples in the eastern Atlantic deviating from climatology and nearby crossings by more than 0.40psu). Salinity deviations diminished in the 1930s when there was the greater awareness of the factors effecting data quality (LUMBY, 1935). Salinity bias is less for the German data and in the late 1930s even occasional large deviations from climatology, which commonly happen in 1928-1934 disappeared. This may have resulted from the more frequent sampling in the late 1930s than before (6 times a day instead of twice a day). Temperature deviations of the UK sampling with respect to the stations were small, but the German SSTs are cooler, maybe because of the use of a small non-insulated iron bucket on board German liners (PARKER and FOLLAND, 1991). However, the comparisons are based on too few samples to be very reliable. Also, the hydrographic stations included in the comparisons may also have in certain years a large surface salinity bias (examples for France and Ireland are given in Table 2, and we suspect that many French salinities between 1924 and 1927 are too low), so that the estimates just given could be underestimates of the true bias.

The bias diminishes closer to the English channel, probably because the numbers of surface observations increases as a result of short crossings in the western Channel. Even there, the average deviation reaches 0.04psu (±0.02psu at the 95% level). Between 1905 and 1917, no large anomaly is found in this area (neither for temperature nor for salinity, although a positive salinity bias of up to 0.02psu is compatible with the data).