Canadian Technical Report of Hydrography and Ocean Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. The subject matter is related generally to programs and interests of the Ocean Science and Surveys (OSS) sector of the Department of Fisheries and Oceans.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in *Aquatic Sciences and Fisheries Abstracts* and indexed in the Department's annual index to scientific and technical publications.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page. Out of stock reports will be supplied for a fee by commercial agents.

Regional and headquarters establishments of Ocean Science and Surveys ceased publication of their various report series as of December 1981. A complete listing of these publications is published in the *Canadian Journal of Fisheries and Aquatic Sciences*, Volume 39: Index to Publications 1982. The current series, which begins with report number 1, was initiated in January 1982.

Rapport technique canadien sur l'hydrographie et les sciences océaniques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Le sujet est généralement lié aux programmes et intérêts du service des Sciences et levés océaniques (SLO) du ministère des Pêches et des Océans.

Les rapports techniques peuvent être cités comme des publications complètes. Le titre exact paraît au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la revue *Résumés des sciences aquatiques et halieutiques*, et ils sont classés dans l'index annuel des publications scientifiques et techniques du Ministère.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre. Les rapports épuisés seront fournis contre rétribution par des agents commerciaux.

Les établissements des Sciences et levés océaniques dans les régions et à l'administration centrale ont cessé de publier leurs diverses séries de rapports en décembre 1981. Une liste complète de ces publications figure dans le volume 39, Index des publications 1982 du *Journal canadien des sciences halieutiques et aquatiques*. La série actuelle a commencé avec la publication du rapport numéro 1 en janvier 1982.

Canadian Data Report of
Hydrography and Ocean Sciences 130

1994

OBJECTIVE ANALYSIS OF HYDROGRAPHIC DATA IN THE NORTHWEST ATLANTIC

by

B. deYoung, F. Perry and R. Greatbatch

Department of Physics
Memorial University of Newfoundland
St. John's, Newfoundland
A1B 3X7

INTRODUCTION

In this data report we will present the results of an objective analysis of hydrographic data for the Northwest Atlantic (Figure 1). Our purpose in carrying out this analysis is to generate gridded fields of temperature, salinity and density. We have used all the available data for this region in performing this analysis, roughly 124,000 stations spanning the period from 1910 to 1989. Recent data, collected after 1989 has not been incorporated into this present analysis.

The purpose of this data report is to give a brief summary of the techniques employed in the analysis and present a comprehensive series of data plots for the area. We shall present little analysis of the data, only that necessary to explain the data analysis techniques developed. In the next section we shall discuss the analysis techniques employed and indicate why we made the choices that we did.

It should be noted that there are a number of possible ways that data interpolation can be carried out, from linear interpolation to some form of data assimilation which includes the dynamics(Tziperman et al. 1992). We have chosen the middle ground and applied a straightforward numerical technique. In section 2 we shall present the results of the analysis indicating how we performed quality control and how the raw data were treated before analysis.

OBJECTIVE ANALYSIS

Objective analysis involves using the correlation function for a data set to provide the weights for the interpolation function. The principal advantage of this technique is that the weights are then determined by the characteristics of the data set. This approach was first developed by meteorologists (Gandin 1963) and only *found* by oceanographers in the 1970's (Bretherton et al.

1976). Different variations of the techniques have been developed in recent years (e.g. Clancy, 1983; Roemmich 1983; Smith et al. 1991). One principal aim of the recent work has been to develop better models of the error analysis, in particular the relative importance of random and systematic errors.

In this analysis, we have taken a simpler approach, one that does not lead us to much insight about the errors. The basic approach that we applied here follows that of Levitus (1982). In this approach, a fixed correlation analysis length scale is chosen based upon a preliminary analysis of the data. Rather than a single analysis of the data, as in the Gandin (1963) approach, which involves inverting an error matrix, iterative calculations are carried out at smaller and smaller length scale. This iterative difference-correction approach was first developed by Cressman (1959). We have modified the approach slightly in our application to data in the Northwest Atlantic.

The first step in the analysis procedure is to average the data in bins that are 1/12 by 1/12 degree. These data will form the input to the analysis routine. An influence of radius is chosen, based upon the data density, the desired resolution and the topography. To begin the analysis a first guess is generated that is corrected at each step in the iteration:

$$G_{i,j} = F_{i,j} + C_{i,j}$$

where $G_{i,j}$ is the analysed value, $F_{i,j}$ is the first guess and $C_{i,j}$ is the correction value. The analysis generates the values of $C_{i,j}$ using

$$C_{i,j} = \frac{\sum_{i=1}^{n} W_s Q_s}{\sum_{i=1}^{n} W_s}$$

where the weights in this equation W_s are determined from

$$W_s = e^{-4(r^2 R^{-2})}$$

for $r \leq R$. Here r is the distance between the observation and the gridpoint and R is the chosen influence radius. This is the analysis scheme as applied by Levitus (1982). The weight function applied in this analysis was developed by Sasaki (1960) and Barnes (1964).

In this analysis we will modify slightly the approach developed by Levitus. First, we shall carry out the analysis at standard depths that will ignore the presence of land. Thus, we shall not try to interpolate along the intersection of the bottom topography. This approach will greatly simplify the analysis. The approach outlined above is isotropic, i.e. it takes the influence radius to be the same in every direction. The distribution of the data do not justify this. Also, because we are interested in the cross-shelf structure of the data, we will make the radius of influence depend upon direction. Making this directional dependence of the radius will allow us much greater resolution in the cross-shelf direction, where the data support such an analysis. The modified weight function in this analysis will be

$$W_s = e^{-4(x^2 R_x^{-2} + y^2 R_y^{-2})}$$

where x, and y are the distances from the gridpoint to the observations in the east and west directions.

In this analysis we choose to make the cross-shelf direction for the Newfoundland Shelf region smaller than the alongshelf, half the length, down to a latitude of 47 ° N. Between 47 ° and 49 ° N the influence radii are linearly interpolated to minimise the effect of the transition (Figure 2). The results of the analysis did not appear to be sensitive to the nature of this transition zone, so long as it was not too abrupt. The analysis scales employed are shown in Table 1. In this report, we present the results from the fourth level of analysis down to scales of 150/75 km. The differences between the third and fourth levels are mainly in the more detailed structure apparent on in the shelf regions.

Table 1 Scales for the Objective Analysis

Level	North/South (km)	East/West (km)
1	1200	600
2	600	300
3	300	150
4	150	75

DATA ANALYSIS

The raw data for this analysis was obtained from the National Oceanic Data Centre (NODC) in Washington and the Marine Environmental Data Service (MEDS) in Ottawa. We obtained all the available data for the region shown in Figure 1, from 40 to 70 ° N and from 40 to 75 ° W. The data from these two centres were merged to form the single data set upon which most of the analysis took place. In discussing how much data were used, it is difficult to compare the MEDS and NODC data sets because many of the MEDS station profiles were broken up into multiple files. Thus a single station might exists as more than one file, broken into pieces. The NODC profiles did not tend to get broken up in this way. Table 2 shows the number of stations in the merged file by season. Although this is a large number of stations, it is clearly not enough. Indeed the most serious problem with the gridding of these data is the absence of data and the spatial patterns of the station locations. If all the stations were distributed through the domain, then there would be almost enough. However, they tend to be along standard lines and at standard locations (Figure 3). The weather observing stations which were occupied for many years (Bravo, Delta, Echo and Charlie) are all visible in Figure 3 as roughly circular areas where there are large numbers of stations. Standard oceanographic lines, such as the line along 47 ° N are also apparent on the shelf and in some of the deep-sea sections. This plot of the overall station numbers makes the distribution look even better than it truly is, if we consider the seasonal patterns. Figures 4-7 show the winter through fall station numbers. As one would expect, there are few stations in the north during the winter and spring period because of ice and poor winter weather. Most of the data for the northern regions comes from summer and fall cruises. This bias in the high-latitude observations towards summer data has been a long-term problem with oceanographic data.

In addition to the spatial bias, one must also recognise that temporal biases exist in the data set. Most of the data were collected after 1950 (Figure 8). A greater percentage of the data will come from the recent past as these newer data make their way into the archives. The falloff in data after 1985, apparent in Figure 8, is because of the delay in getting data into the archives after collection.

Table 2 Number of Stations per Season

Season	No. MEDS	No. NODC	Total No. Stations
All Seasons	89,502	34,529	124,031
Summer	31,174	14,713	45,887
Autumn	15,745	7,282	23,027
Winter	11,747	5,185	16,392
Spring	30,836	7,889	38,725

The first step in combining of the NODC and MEDS data was to put them in a common format. We choose the NODC format because these are the data with which we first started working. For the most part these data were already at standard depths however where there were missing standard depths we used the observations to linearly interpolate to obtain standard depths. Relatively little interpolation of this sort was necessary. It was usually done just to fill in missing depths. Duplicate stations were removed by checking the time and location of the observations. Because of reporting and transcription errors, we did not just look for exact matches. Instead, we considered a duplicate to be a station that occurred on the same day within 1 minute (1 nautical mile) of the other station. During this process we found about 65% of the stations to be duplicates.

After duplicate stations were removed we range checked the data, just to remove data that was beyond any physically realistic value. Thus we removed temperatures outside the range -1.9 to 25 °C, salinity outside 15 to 36 and sigma-t outside the range 15 to 30. Few data points were removed during this step, although it was a crucial step in removing bad data. After range checking we looked at the vertical structure to look for inversions in the density data, a good sign of problems in the profiles. If less than three inversions were found in a profile, they were linearly interpolated. If three or more were found, then the station was rejected. The rejection criteria was made based on the argument that if there were too many inversions then this cast doubt on the quality of the station itself and the data were rejected.

In order to perform any kind of statistical analysis it is necessary to bin the data. The binning of the data has also the advantage of reducing the size of the data set and the number of stations to handle, without limiting the real resolution of the analysis. We binned the data in 1/12 by 1/12 ° squares. Averages for temperature and salinity were calculated, and the means and standard deviations computed. Data beyond three standard deviations were rejected and then these T and S data were used in all other calculations, i.e. density was re-computed from the cleaned data set. About 2-3 % of the data were removed during this step.

In performing the binning, we worked on the seasonal and composite data fields. The composite field is just the average of all the data. For the seasonal analysis, we choose an oceanographic convention for the seasons, different from that chosen by Levitus: winter (Jan-Feb-Mar), spring (Apr-May-Jun), summer (Jul-Aug-Sep) and fall (Oct-Nov-Dec). For this high latitude area, where in the northern part of the domain ice does not leave until late June or early July, this breakdown of the seasons seems more reasonable than making the break about a month earlier, as is often done by meteorologists.

After gridding of the data, the resulting interpolated fields were smoothed with a 1/3 by 1/3 ° degree smoother. Even after performing this quality control it was still found that there were bad data. Some of this bad data would generate "bulls'eyes" in the resulting contour plots but for the most part they were not apparent in the contoured data fields. These bad data were found by applying a diagnostic circulation model (deYoung et al., 1993).

In areas where there were unreal density gradients the circulation field would clearly show the effects. Visual inspection of the data would usually reveal the bad data, which were then manually removed.

Some of the resulting fields were also used to investigate interannual variations in the Newfoundland Shelf region. During this analysis, we discovered that the summer temperature data from a station on the inner part of the Newfoundland Shelf (Station 27, Keeley 1981) had the sign of the temperature reversed over a period from July 1959 to July 1975. These sign reversals were found by analysing time series of temperature for this region. All the errors, in 285 stations, were found in the NODC data set. No such errors were found in the MEDS data. Subzero surface water temperatures were easy to pick out but the correction was aided by comparison with the original station data which is available here in St. John's. Although we found this one problem it has reduced our faith in the data set because we happened to find serious errors in the one part of the data set that we had extensively investigated. Without the original data to correct the erroneous temperatures, it would have been a difficult task.

After obtaining the binned temperature and salinity data from the merged data sets we computed derived quantities from the data. Potential temperature was computed using the formula of Bryden (1972). Potential density was computed using the potential temperature and the equation of state (see Gill 1982). In addition to σ_θ, we also computed the complete density from the equation of state (Gill, 1982).

CONTOURED FIELDS

A sample of a contoured fields are shown in Figures A1.1 and A1.2. The complete set of figures is included in the appendix of this report. We shall use this sample plot just to make a few comments upon the results. As the station position figures show (Figures 4-7) the results at high latitudes are the least reliable, since there are so few data there. The plot of station numbers versus time also shows that there is significant bias in the temporal representation of this data set. Most of the data collected here was obtained after 1950. Very few data (less than 3 %) were obtained before 1930.

There were some problems with nearshore data which tended to bias the results on the shelf. On the Labrador shelf in particular this problem is noticeable. Data collected in Lake Melville, a land-locked fjord located at about 55 ° N on the Labrador coast were removed from the analysis because of their excessive influence on the Labrador shelf. Even with their removal the effect of nearshore data, strongly influenced by runoff, is apparent in the final gridded results.

We suspected that similar problems might exists on near the Gulf of ST. Lawrence, where data inside the Gulf influence shelf grids on the Scotian Shelf or on the Newfoundland Shelf. To test this we ran a set of interpolations with the Gulf of St. Lawrence data excluded if the grid point was outside the Gulf. As one might expect this did have a large effect on the first gridding level, but there was no measurable effect at the final gridding level, which is at a much smaller scale (see Table 2). The final gridded results presented here did not have the Gulf exclusion algorithm applied.

We also generated a composite field, however, we have not presented the results here. The seasonal signal is so strong at these latitudes, that it was not felt useful to present the results here, though they may be of some use in certain applications.

In the Appendices, three types of contour plots are presented. The first shows potential temperature, salinity and σ_θ at selected standard depths by season. These contour plots have had the 1/3° smoother applied to them. The second set of contour plots are section plots at three latitudes, 40 °N, 47 °N and 54 °N, also presented for each season. The final set presents steric height by season relative to 75, 150, 300, 500 and 1000 m. We calculate the steric height from the seasonal density data

$$\rho(z) = \rho_o + \Delta\rho(z)$$

where ρ_o is the reference density and $\Delta\rho$ is the deviation of the density from the reference value. The steric height is then evaluated from

$$\zeta = -\frac{1}{\rho_o}\int \Delta\rho dz$$

where we choose $\Delta\rho_o$ to be the largest density at the reference depth.

SUMMARY

The analysis reported here was carried out as part of the Ocean Production Enhancement Network (OPEN) one of a Network of Centres of Excellence Programs in Canada. The purpose of this analysis was to generate gridded fields for the modelling of the circulation fields on the Newfoundland and Labrador shelves. A paper presenting that work is now in preparation (deYoung et al. 1994).

The data reported here are available in computer format on tape or via ftp over internet (bdeyoung@kean.ucs.mun.ca). Those interested in obtaining copies of the data should address their enquires to the first author.

ACKNOWLEDGEMENTS

We would like to thank Drs. W. Perrie and J. Helbig of the Department of Fisheries and Oceans for their support of this project.

REFERENCES

Barnes, S.L. (1964): A technique for maximizing details in numerical weather map analysis. *J. App. Meteor.*, **3**, 396-409.

Bretherton, F.P., R.E. Davis, and C.B. Fandry. (1976): A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep Sea Res.*, **23**, 559-582.

Bryden, H.L. (1973): New polynomials for thermal expansion, adiabatic temperature gradient and potential temperature gradient of sea water. *Deep Sea Res.*, **20**, 401-408.

Clancy, R.M. (1983): The effect of observational error correlations on objective analysis of ocean thermal structure. *Deep Sea Res.*, **30**, 985-1002.

Cressman, G.P. (1959): An operational objective analysis scheme. *Mon. Wea. Rev.*, **87**, 329-340.

deYoung, B., R.J. Greatbatch and K. Forward. (1993) A diagnostic coastal circulation model with application to Conception Bay, Newfoundland. *J. Phys. Oceangr.*, **12**, 2617-2635.

Gandin, L.S. (1963): *Objective analysis of Meteorological fields.*, Leningrad, Gidrometeorizdat, 266 pp.

Gill, A.E. (1982) Atmosphere-Ocean Dynamics. Academic Press, New York, 662 pp.

Keeley, J.R. (1981) Temperature, salinity and sigma-t at Station 27 (47°33′N, 52°35′ W). An analysis of historical data. Environment Canada, Marine Environmental Data Service Tech. Rep. No. 8, Ottawa, Canada, 56 pp.

Levitus, S. (1982): Climatological atlas of the World Ocean. NOAA Professional Paper 13.

Roemmich, D. (1983) Optimal estimation of hydrographic station data and derived fields. *J. Phys. Oceangr.*, **13**, 1544-1549.

Sasaki, Y. (1960): An objective analysis for determining initial conditions for the primitive equations. Ref. 60-16T, Atmospheric Research Lab., Univ. of Oklahoma Research Institute, Norman, 23 pp.

Smith, N.R., J.E. Blomley and G. Meyers (1991) A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans. *Prog. Oceanogr.*, **28**, 219-256.

Tziperman, E., W.C. Thacker, R.R. Long and S.-M. Hwang. (1992) Oceanic data analyses using a general circulation model. Part I: Simulations. *J. Phys. Oceangr.*,**23**, 1434-1457.

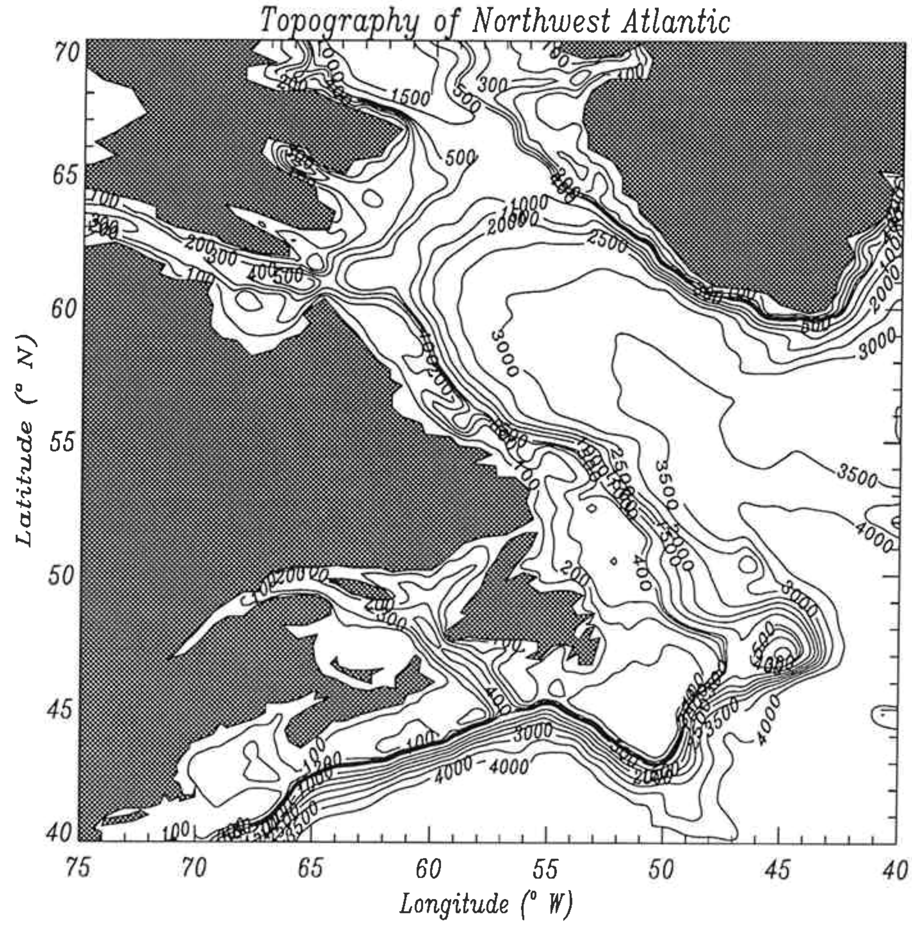

Figure 1: Bottom topography (m) for the Northwest Atlantic.

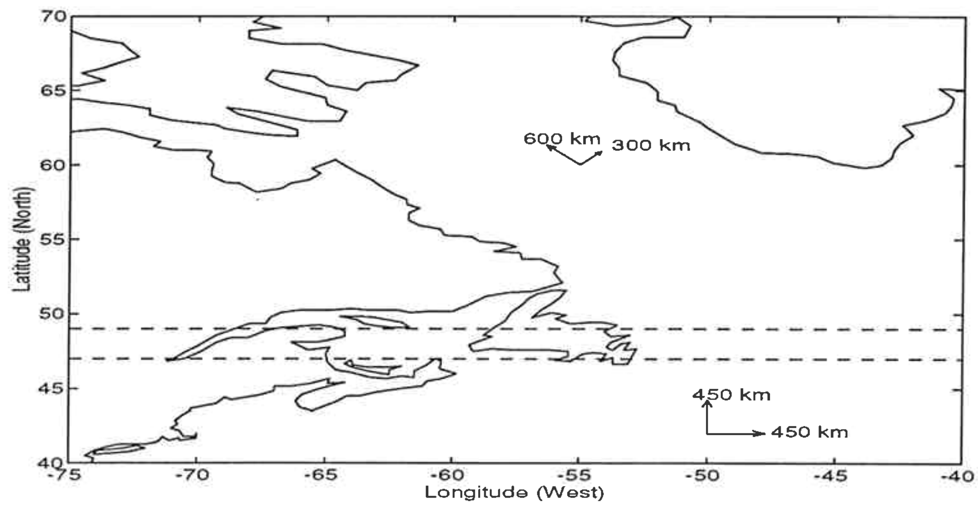

Figure 2: Orientation of the scale analysis for the objective analysis, shown for the second gridding level, at 600 and 300 kilometres. The dashed lines show the transition zone from the symmetric weights (south) to the alongshelf orientation (north of 49 °).

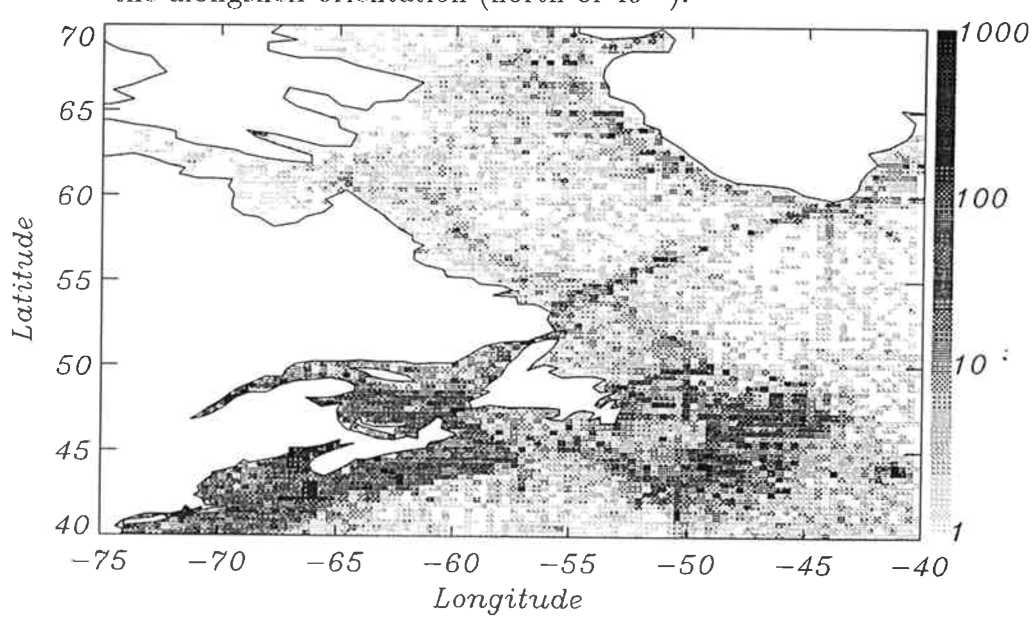

Figure 3: Overall station distribution in the merged NODC/MEDS data set for 1910-1989. Note the log scale indicated on the right.

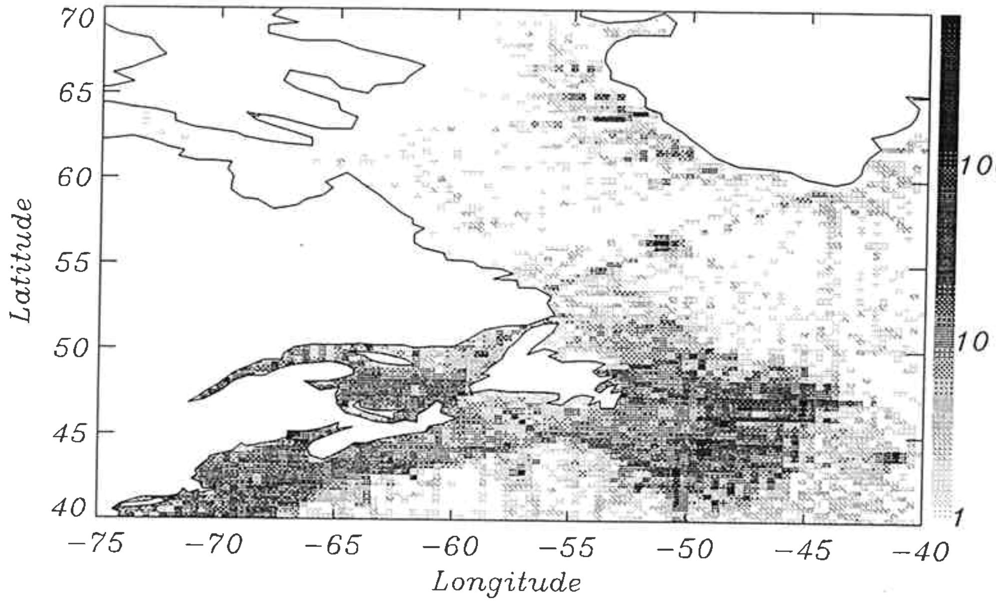

Figure 4: Spring station distribution in the merged NODC/MEDS data set for 1910-1989. Note the log scale indicated on the right.

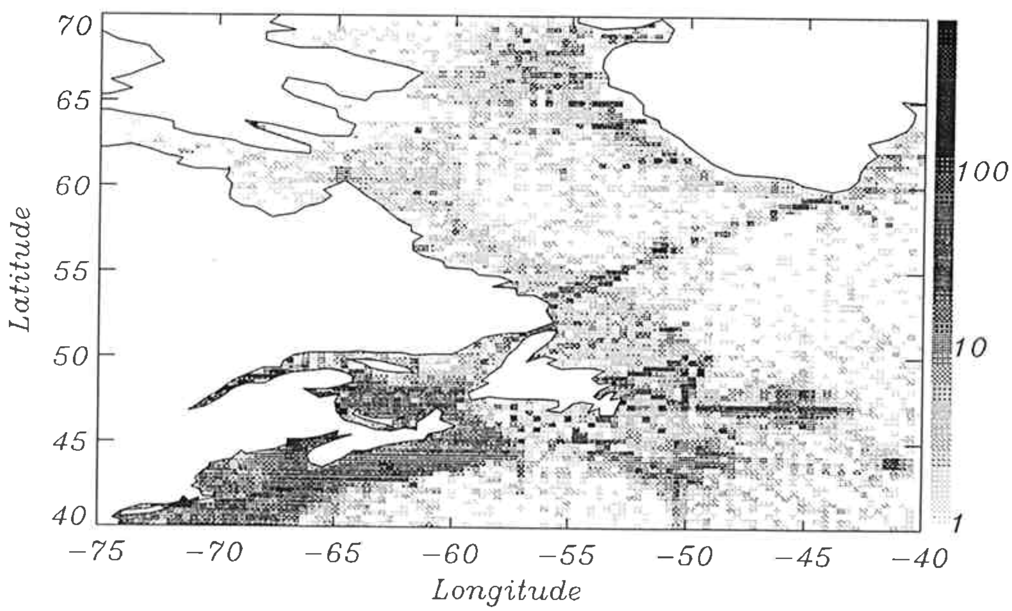

Figure 5: Summer station distribution in the merged NODC/MEDS data set for 1910-1989. Note the log scale indicated on the right.

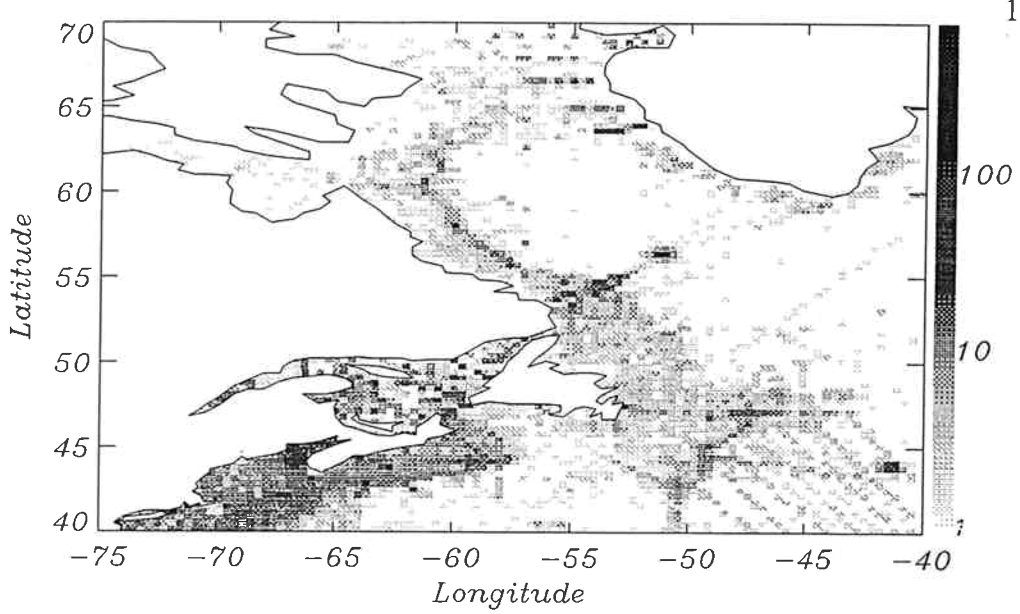

Figure 6: Autumn station distribution in the merged NODC/MEDS data set for 1910-1989. Note the log scale indicated on the right.

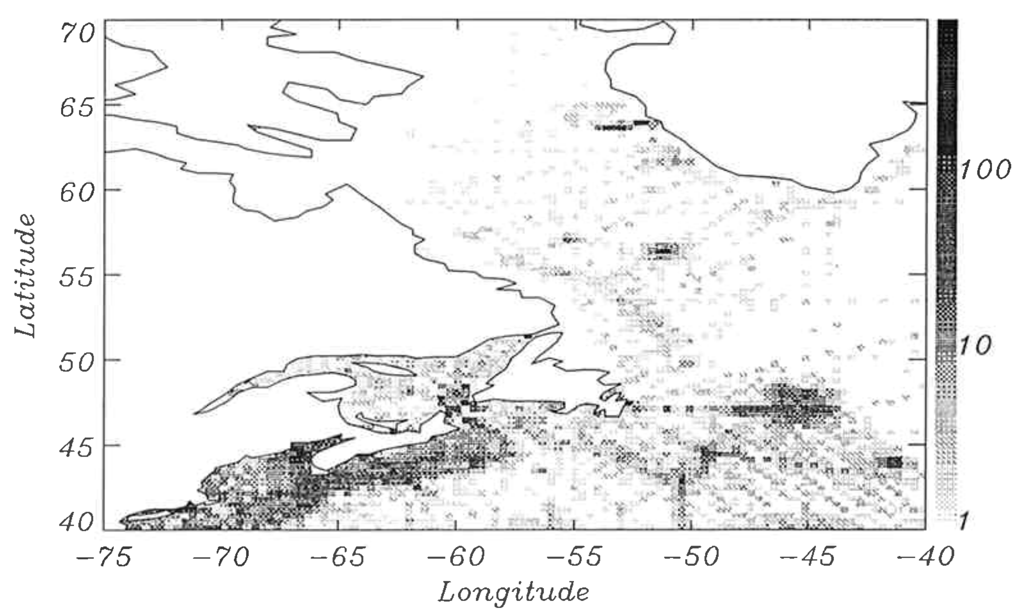

Figure 7: Winter station distribution in the merged NODC/MEDS data set for 1910-1989. Note the log scale indicated on the right.

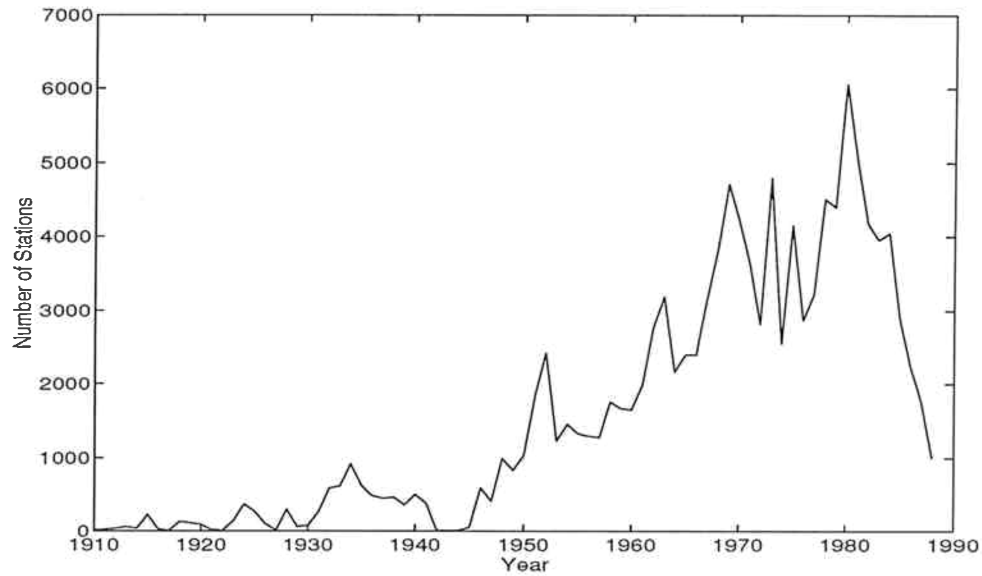

Figure 8: Number of stations in the analysis region (Figure 1) over the period 1910-1989.

APPENDIX 1

Contour plots of Horizontal Layers.

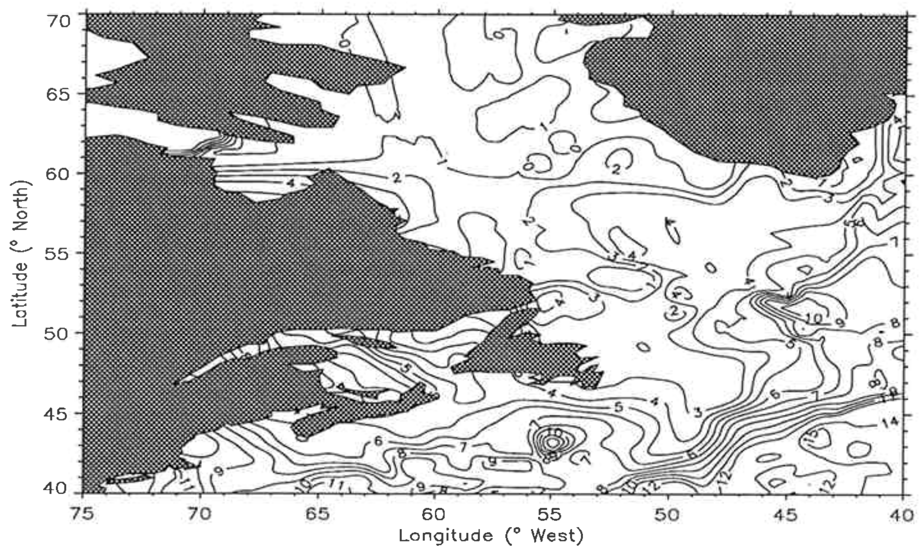

Figure A1.1: Potential temperature ($\theta(s,t,p)$) contours for spring (0 m) in °C.

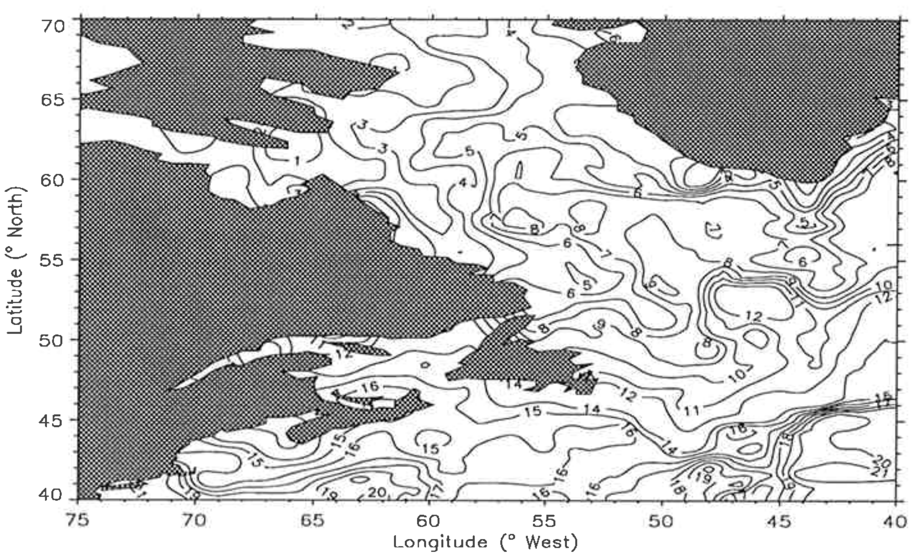

Figure A1.2: Potential temperature ($\theta(s,t,p)$) contours for summer (0 m) in °C.

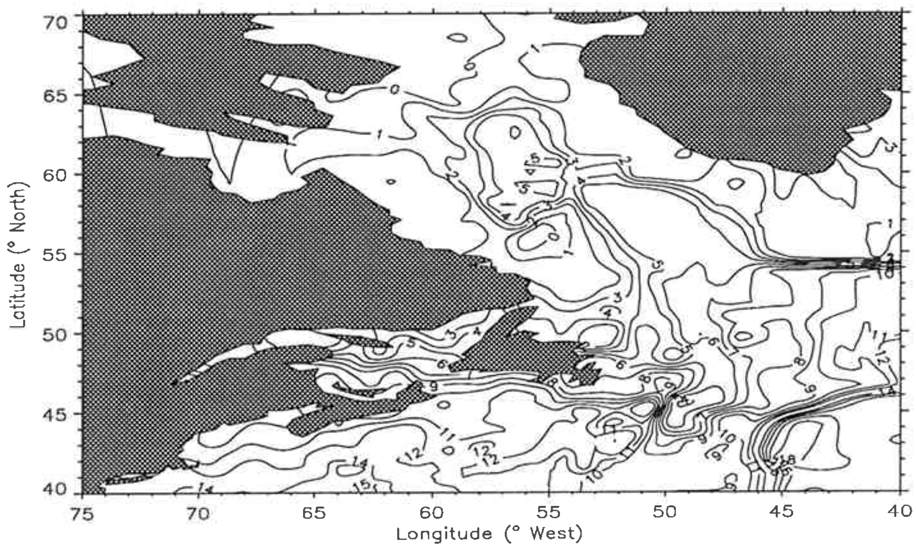

Figure A1.3: Potential temperature ($\theta(s,t,p)$) contours for autumn (0 m) in °C.

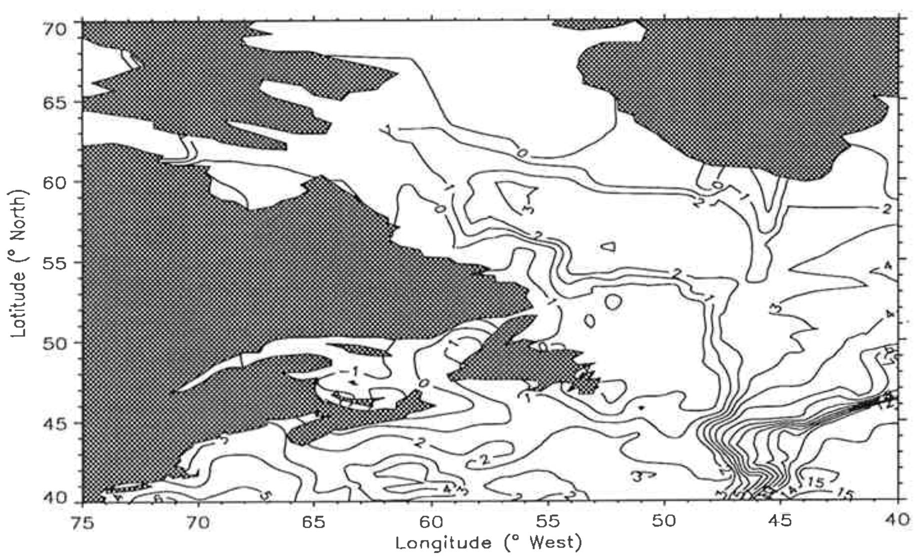

Figure A1.4: Potential temperature ($\theta(s,t,p)$) contours for winter (0 m) in °C.

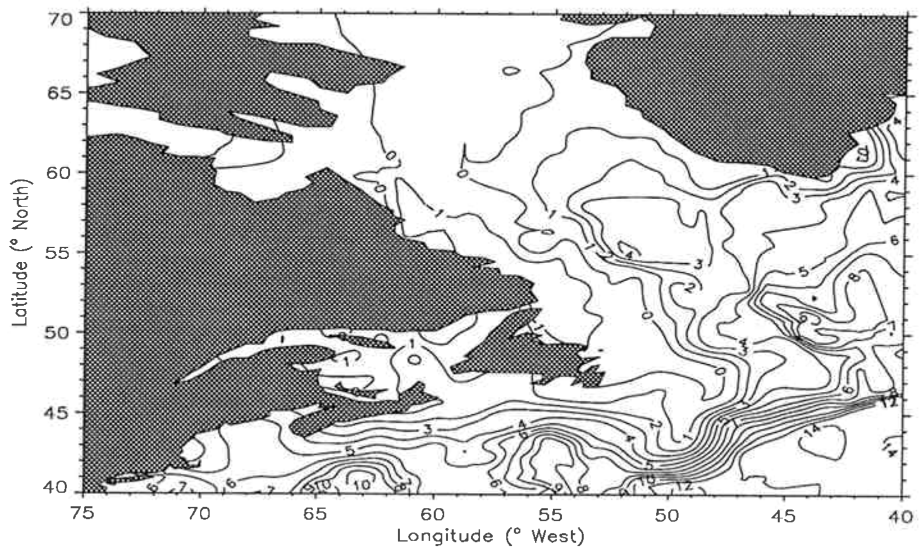

Figure A1.5: Potential temperature ($\theta(s,t,p)$) contours for spring (50 m) in °C.

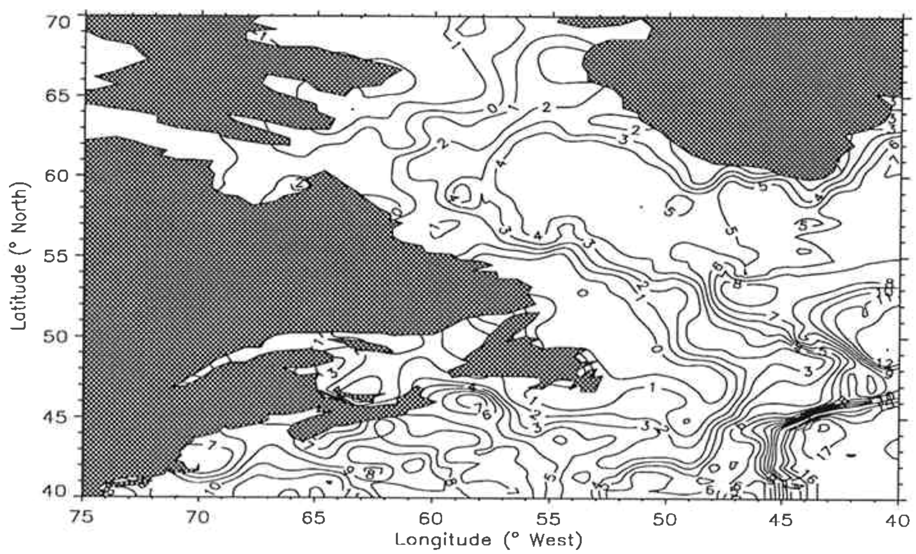

Figure A1.6: Potential temperature ($\theta(s,t,p)$) contours for summer (50 m) in °C.

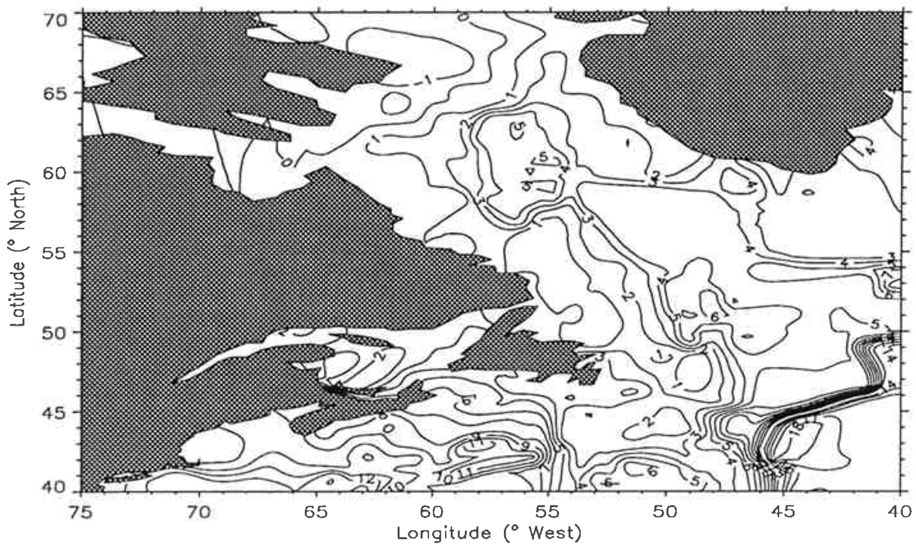

Figure A1.7: Potential temperature ($\theta(s,t,p)$) contours for autumn (50 m) in °C.

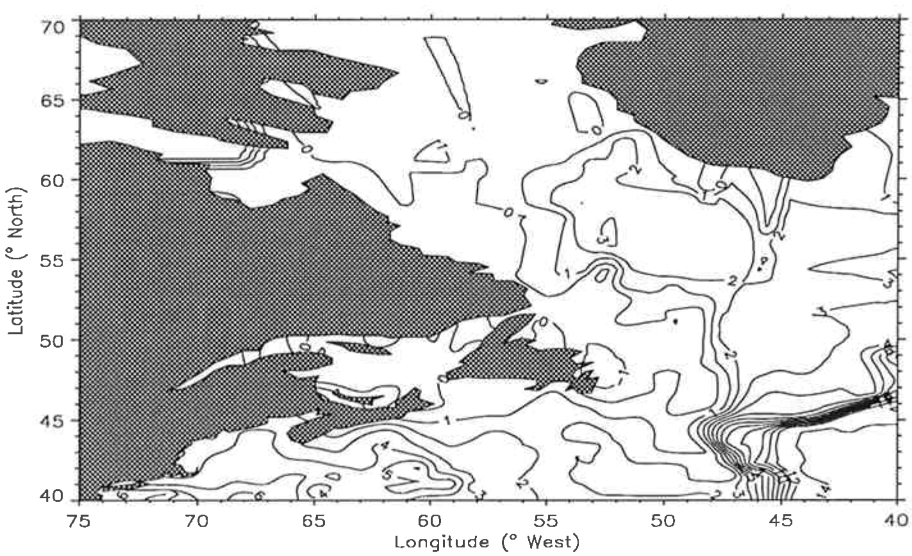

Figure A1.8: Potential temperature ($\theta(s,t,p)$) contours for winter (50 m) in °C.

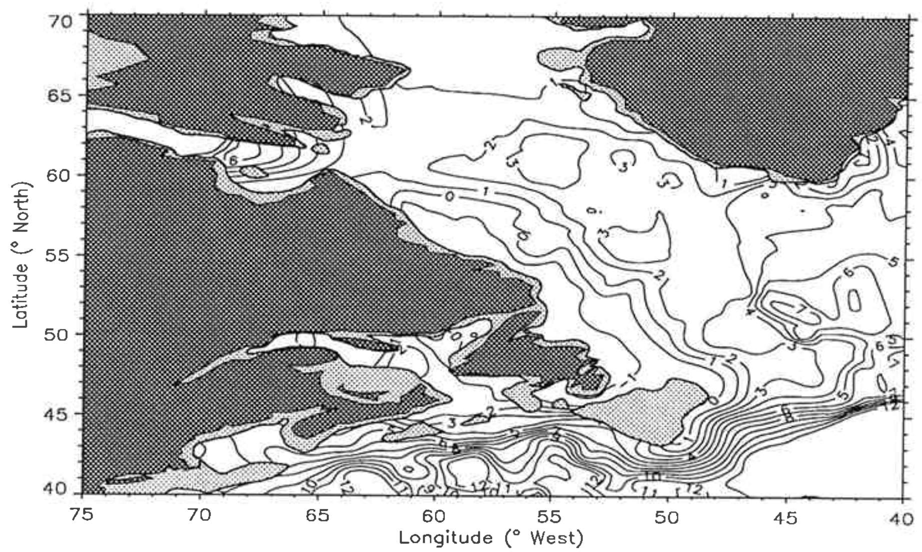

Figure A1.9: Potential temperature ($\theta(s,t,p)$) contours for spring (100 m) in °C.

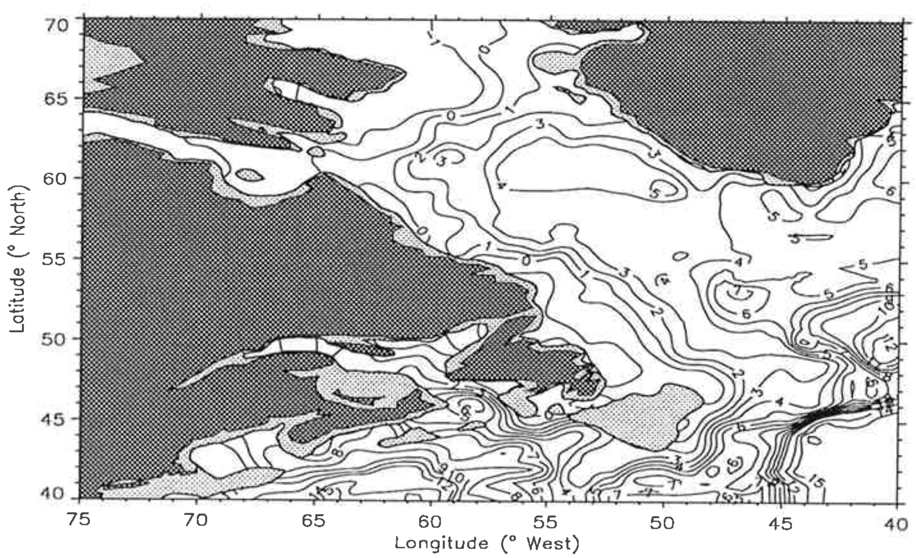

Figure A1.10: Potential temperature ($\theta(s,t,p)$) contours for summer (100 m) in °C.

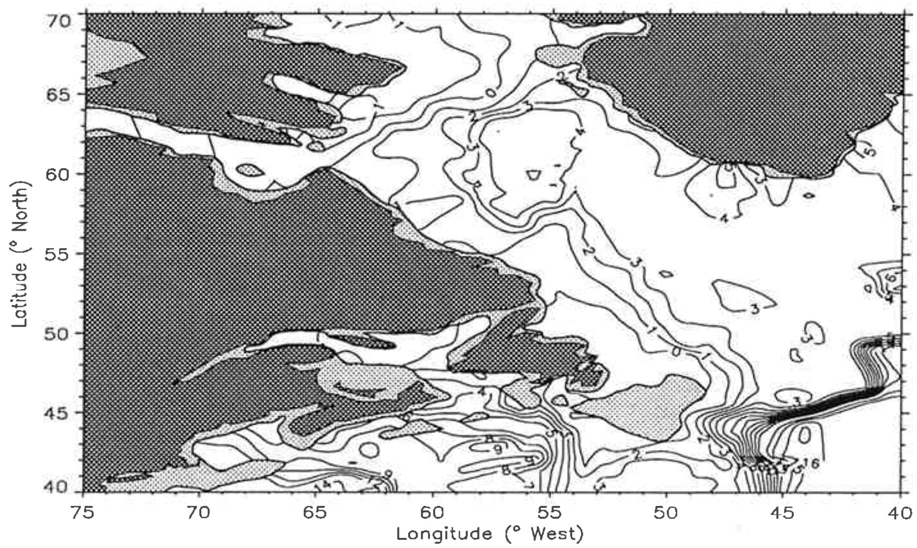

Figure A1.11: Potential temperature ($\theta(s,t,p)$) contours for autumn (100 m) in °C.

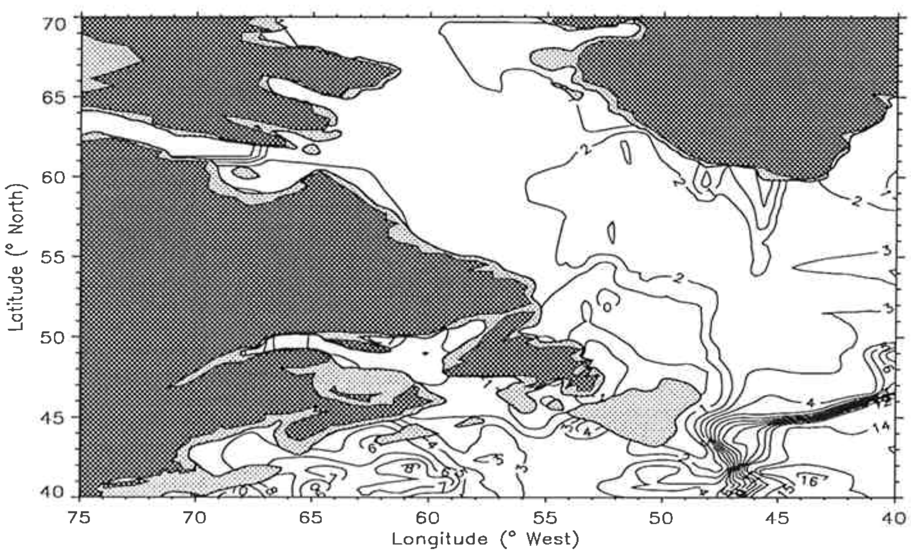

Figure A1.12: Potential temperature ($\theta(s,t,p)$) contours for winter (100 m) in °C.