Available online at www.sciencedirect.com
ScienceDirect
Journal of Marine Systems 65 (2007) 224–248
JOURNAL OF MARINE SYSTEMS
www.elsevier.com/locate/jmarsys

Estimating salinity to complement observed temperature: 1. Gulf of Mexico

W.C. Thacker

Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami FL 33149 USA

Received 23 September 2004; accepted 14 June 2005
Available online 13 November 2006

Abstract

This paper and its companion [Thacker, W.C., Sindlinger, L., 2007-this issue. Estimating salinity to complement observed temperature: 2. Northwestern Atlantic. Journal of Marine Systems. doi:10.1016/j.jmarsys.2005.06.007.] document initial efforts in a project with the goal of developing capability for estimating salinity on a region-by-region basis for the world oceans. The primary motivation for this project is to provide information for correcting salinity, and thus density, when assimilating expendable-bathythermograph (XBT) data into numerical simulations of oceanic circulation, while a secondary motivation is to provide information for calibrating salinity from autonomous profiling floats. Empirical relationships between salinity and temperature, which can be identified from archived conductivity–temperature–depth (*CTD*) data, provide the basis for the salinity estimates.

The Gulf of Mexico was chosen as the first region to explore for several reasons: (1) It's geographical separation from the Caribbean Sea and the North Atlantic Ocean makes it a "small ocean" characterised by a deep central basin surrounded by a substantial continental shelf. (2) The archives contain a relatively large number of *CTD* data that can be used to establish empirical relationships. (3) The sharp fronts associated with the Loop Current and its rings, which separate water with different thermal and haline characteristics, pose a challenge for estimating salinity. In spite of the shelf and the fronts, the relationship between salinity and temperature was found to be sufficiently regular that a single empirical model could be used to estimate salinity on each pressure surface for the entire Gulf for all seasons. In and below the thermocline, root-mean-square estimation errors are small — less than 0.02 psu for pressures greater than 400 dbar, corresponding to potential density errors of less than 0.015 kg/m^3. Errors for estimates nearer to the surface can be an order of magnitude larger.

Keywords: XBT; *CTD*; Regression; Data assimilation; HYCOM

1. Introduction

When assimilating expendable-bathythermograph (XBT) data into numerical models of the ocean's circulation, it is important to correct the model's salinity together with its temperature; otherwise, errors can be introduced into the density field that negatively impact the model's dynamics (Acero-Shertzer et al., 1997; Reynolds et al., 1998; Vossepoel and Behringer, 2000; Troccoli et al., 2002; Thacker et al., 2004). There is no dynamical relationship between salinity and temperature comparable to the geostrophic–hydrostatic relationship between momentum and density, but within the various water masses salinity and temperature can exhibit strong empirical relationships. By exploiting these correlations, XBT data can be used to estimate companion salinity

E-mail address: carlisle.thacker@noaa.gov.

0924-7963/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmarsys.2005.06.008

profiles, which can be assimilated along with the observed temperature profiles to guarantee that the model's density field preserves the properties of the water masses.

The need for salinity estimates is global. Unfortunately, salinity's relationship to temperature and to other observables varies from region to region. Thus, the task of developing capability for estimating salinity must be approached region by region. This paper focuses on the Gulf of Mexico as one such region. While it is a "small ocean", it is not so small, spanning roughly 10° in latitude and 20° in longitude. It has a deep central basin surrounded by broad shallow shelf, as indicated by the bathymetric[1] contours of Fig. 1. Its principal dynamical features are the Loop Current and associated anticyclonic rings, but there are also cyclonic rings and substantial river runoff. Given its size, shelf, and thermal fronts, can salinity be modelled for the entire Gulf, or are different models needed for different sub-regions or for different sides of the Loop Current front? An important result presented here is that salinity can be modelled for the entire Gulf of Mexico, in spite its size and diversity, without regard for the seasonal cycle.[2]

This assessment of the problems and possibilities of estimating salinity from observed temperature for the Gulf of Mexico is a first step toward implementing salinity estimating capability region-by-region for the global ocean. As the Gulf of Mexico will provide the context for a study comparing various techniques for assimilating data into a hybrid-coordinate ocean model (Bleck, 2002; Halliwell, 2004; Chassignet et al., 2003), examining this region first is particularly appropriate.

Over fifty years ago Stommel (1947) recognised that the co-variability of salinity with temperature can be exploited for estimating salinity. The basic idea is that much of salinity's variability is due to vertical displacements of water with relatively well-defined salinity and temperature: the salinity to expect for a given temperature is essentially what was observed previously at this same temperature. To implement Stommel's method properly, mean salinity on temperature surfaces should be extracted from previous measurements. However, there is an easy, approximate implementation that exploits the climatological mean values for temperature and salinity, which have been tabulated on a $1° \times 1°$ longitude × latitude grid at standard depths for the world's oceans (Conkright et al., 2002a): salinity is simply estimated by interpolating the climatological mean profiles to the observed temperature.

Many have built upon Stommel's idea, *e.g.*, Flierl (1978), Donguy et al. (1986), Kessler and Taft (1987), Vossepoel et al. (1999), and Troccoli and Haines (1999). In particular, Hansen and Thacker (1999) have expressed the temperature–salinity (*TS*) relationship via regression models: for any desired depth, salinity can be regressed on temperature and on any other appropriate variables such as latitude, longitude, seasonal index, surface salinity, etc., which might provide information about salinity.[3] A second result of this work is to show that Stommel's method does not perform as well as the regression approach of Hansen and Thacker. For example, the root-mean-square error when salinity is estimated with a parabolic function of temperature at 200 dbar is 0.05 psu, while it is 0.14 psu — almost three times larger — with the easy implementation of Stommel's method. This result suggests that the laborous examination of the world's *TS* data on a region-by-region basis is worth the effort.

Wong et al. (2003) have developed a method for calibrating the salinity sensors of autonomous profiling floats: salinity data from historical *CTD* profiles are statistically interpolated on potential-temperature surfaces to the locations of the float to be calibrated to give a climatological estimate for the float's salinity at that potential temperature. Because salinity generally varies less at greater depths, the calibration relies primarily on detecting drifts from climatology near the float's parking level, but the method does exploit the entire climatological profile. As their method is essentially the same as Stommel's, it is reasonable to expect that their calibration profile is less accurate than one based on the regression method used here. If that proves to be true under closer examination, then a second important application of this project's regional salinity-estimation models would be that of calibrating autonomous salinity sensors.

Neither Stommel's method nor regressing salinity on temperature were expected to perform well in the near-surface region where salinity is only weakly correlated with temperature. There, some other source of information is required. The regression approach allows inclusion

[1] Thanks are extended to Dong-Shan Ko of the Naval Research Laboratory for providing the DBDB2 data that are used for drawing the bathymetric contours.

[2] This result suggested that the second region to explore (Thacker and Sindlinger, 2007-this issue) should be a large, highly variable area in the North Atlantic characterised by the Gulf Stream and its recirculation. Excluding the shelf in the northwest and a small subarea in the southeast, this large region was also found to have sufficiently homogeneous temperature–salinity statistics that this entire region could be treated as a unit.

[3] An important distinction beyond the number of variables they can accommodate is that, while Stommel's method uses only means, the Hansen–Thacker method exploits variances and covariances.

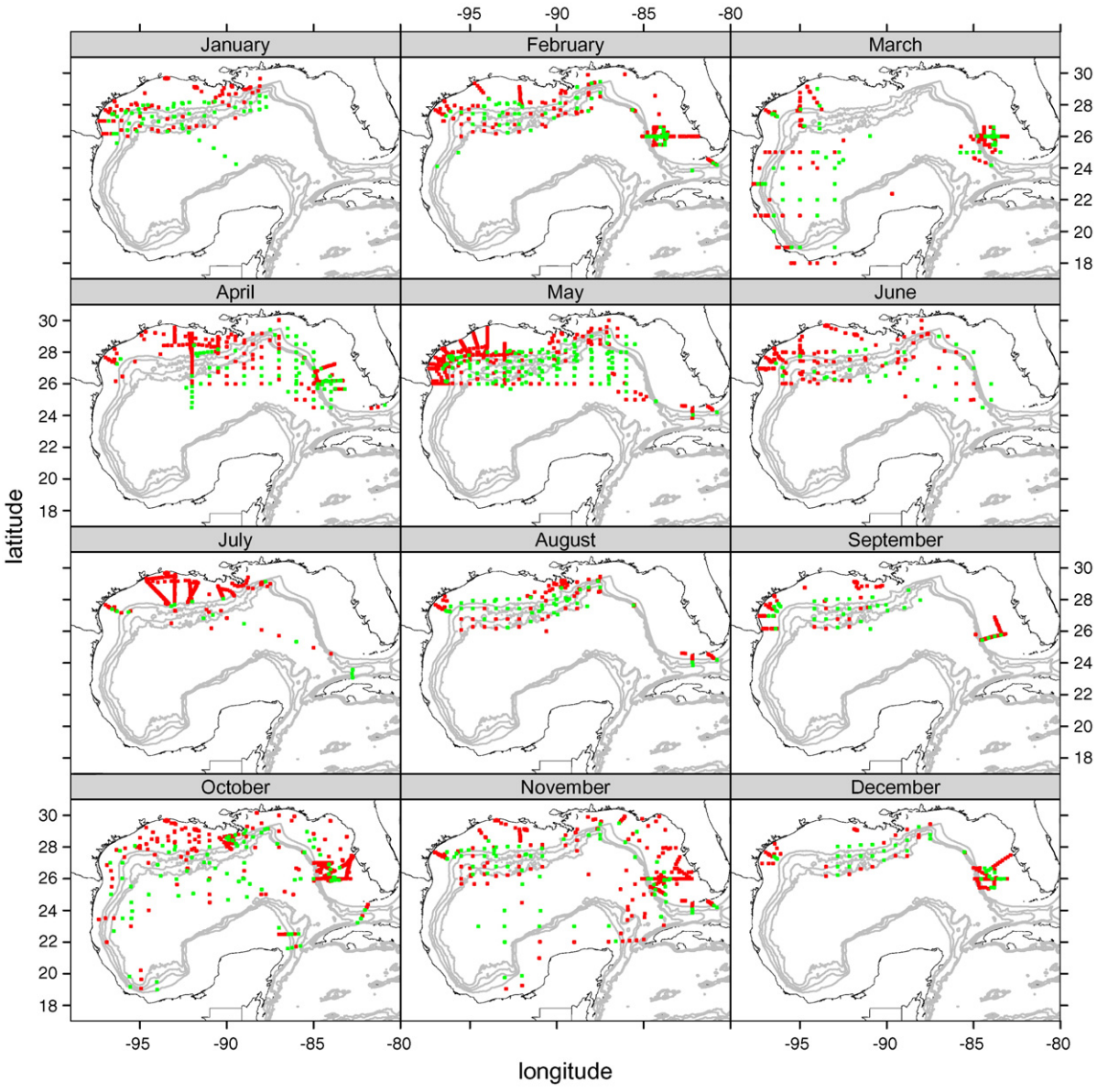

Fig. 1. Red and green dots, some obscuring others, indicate the locations of 3485 CTD stations from the World Ocean Database 2001 on maps of the Gulf of Mexico for each month. Only the 739 stations indicated by green dots were used in this study. Bathymetry is indicated by the gray contours, which are spaced at 500 m intervals.

of other correlates of salinity, such as satellite-based measurements of surface salinity or altimetric height, climatic indices, or even latitude, longitude, or day-of-year, to be included as predictors of near-surface salinity. While surface salinity proved to be quite useful in the upper 50 dbar in the eastern tropical Pacific (Hansen and Thacker, 1999), a surprising result of this work is that surface salinity in the Gulf of Mexico provides no useful information about sub-surface salinity. Because estimating near-surface salinity is expected to be problematic everywhere and requires extra care, our attention here is focused primarily on deeper water where salinity's co-variability with temperature can be exploited.

The nature of the *CTD* data for the Gulf of Mexico are described in Section 2, and Section 3 describes how a more homogeneous and error-free sample was chosen for this study. Section 4 describes how the data were interpolated to standard pressure levels and how density inversions were handled. The data were partitioned into a set for establishing the empirical models and another

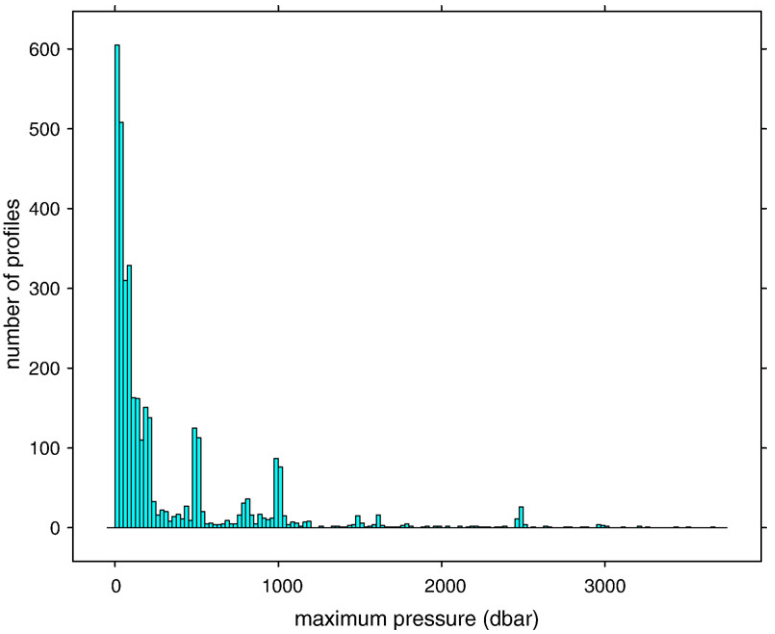

Fig. 2. Histogram of maximum pressure for 3489 archived *CTD* profiles. Bin width is 25 dbar.

for independent verification of model skill, as described in Section 5. Models describing salinity as polynomial functions of temperature are discussed in Section 6, and for comparison models based on Stommel's method are discussed in Section 7. Finally, Section 8 discusses how accurately potential density can be inferred from the estimated salinity and measured temperature profiles, and Section 9 offers a few concluding remarks.

2. *CTD* data

The National Oceanographic Data Center's World Ocean Database 2001 contains 3489 *CTD* profiles for the Gulf of Mexico during the period from 1973 to 1998 (Conkright et al., 2002b). The dots in Fig. 1 indicate how those *CTD* stations are distributed geographically by month, but not all months are sampled every year. While there are data for all calendar months, some months have more than others. Except for March, the northern half of the basin is much better sampled than the southern. There are many stations close to other stations with some dots obscuring others,[4] while there are regions with very few stations. Green dots indicate the stations that were selected for use in this study, while the red dots indicate those that were not used because of problems with the data or because of redundancy. The overlaid bathymetric contours indicate that relatively few stations are located in the deeper water, while many are on the continental shelf, so it isn't surprising that the histogram of maximum pressure (Fig. 2) shows that less than 25% of the profiles sample the water deeper than 500 dbar and 40% do not sample deeper than 100 dbar. A similar histogram of minimum pressures would show that most profiles start very near the surface, but 17 profiles provide no information above 50 dbar and 2 profiles provide no information above 1000 dbar. Clearly, the sampling is neither spatially nor temporally uniform. Nevertheless, there should be sufficient data to determine whether or not it is necessary to accommodate systematic horizontal or seasonal variability.

Not only is the sampling far from ideal, the data are not all reliable.[5] For example, in the March panel, several stations are indicated as being on land. When consistence between each profiles deepest measurement and the local bottom depth, one station in water shallower than 500 m had a cast going deeper than 1100 m; comparing with casts for stations with similar identification numbers suggests that the latitude for this station was incorrect. Furthermore, 30 stations were multiply occupied; i.e. each had multiple profiles at precisely the same location at precisely the same day and time of day, accounting for 259 profiles in all. A few of these were duplicates, but

[4] Closer examination reveals that many of the adjacent stations were occupied within hours of each other, so their data are redundant.

[5] While the data are flagged with a variety of codes to indicate possible problems, not all problems were flagged and some flagged data appeared to be usable. For this work the flags were not used.

most were not. As 14 of these 30 stations have 12 casts each, for some stations the problem might only be incorrect time of day, but large differences in the profiles for other stations suggest that something else might be wrong. As these 259 are among the numerous short profiles, they can easily be discarded. Nevertheless, this raises the spectre of problems with the other profiles.

Even with these ambiguous stations discarded, there are still other stations with almost the same location at almost the same time. They are so close that they cannot be distinguished in Fig. 1. For example, when the stations are ordered by date, time, then by latitude, and then by longitude, so that each can be compared to the next in sequence, there are 1070 stations within 0.1° latitude and 0.1° longitude of the subsequent station on exactly the same day. Furthermore, 1846, which amount to over half of the archived profiles for the Gulf of Mexico, are within 0.2° latitude and 0.2° longitude and 1 day of the subsequent station. Such stations contribute redundant information that can bias statistics derived from these data. To get a more uniform sample, most of these redundant profiles should be discarded. It is desirable to sample as much variety as possible while avoiding bad data. Ideally, the difference between true variability and bad data can be recognised by examining the distributions of values of observed salinity and temperature and of inferred density. But care is needed, as these distributions can be biased by the redundant sampling.

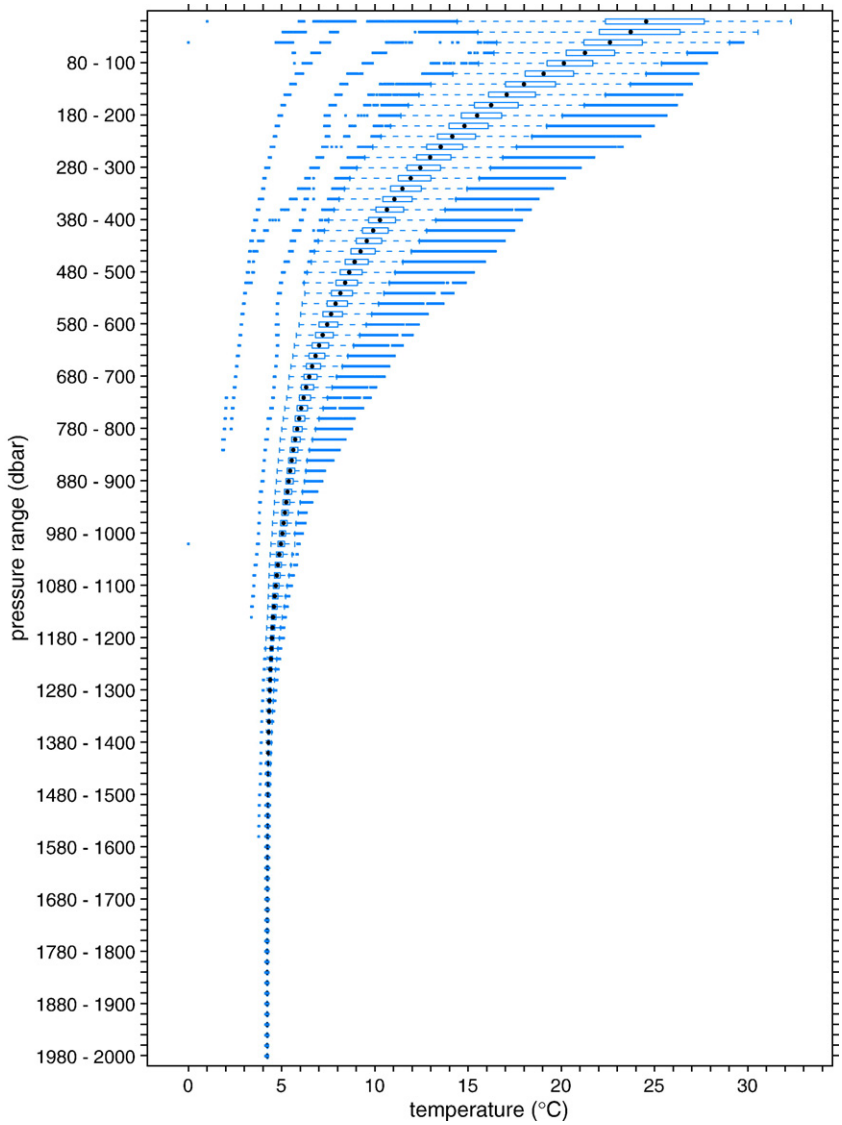

Fig. 3. Box-and-whisker plots of temperature within 20 dbar pressure intervals. Data are from 3489 *CTD* profiles for the Gulf of Mexico.

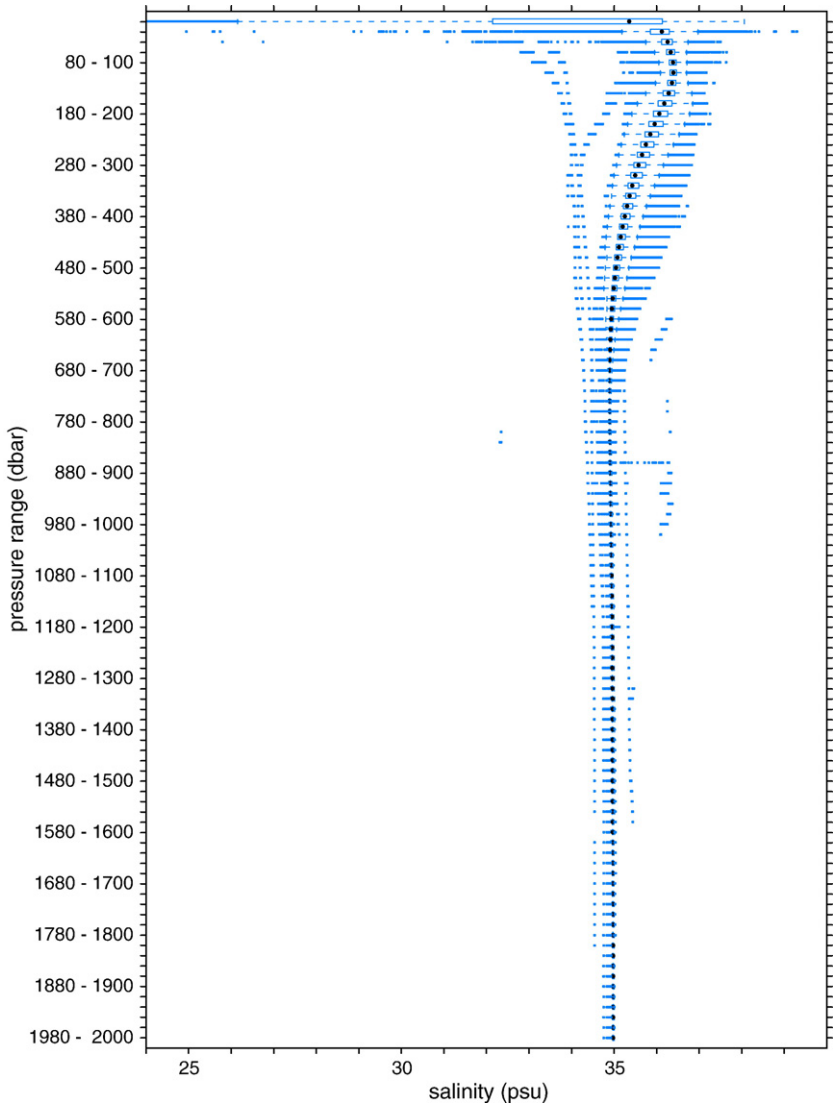

Fig. 4. Box-and-whisker plots of salinity within 20 dbar pressure intervals. Data are from 3489 *CTD* profiles for the Gulf of Mexico. Outliers with values smaller than 20 psu are not shown.

The distribution of values of observed temperature and salinity within 20 dbar pressure intervals from the surface to 2000 dbar are summarised with box and whisker plots[6] in Figs. 3 and 4. The range of the salinity axis has been restricted, omitting fresh outliers that extend to 0 psu even far below the surface, in order to focus on the bulk of the data. As the profiles vary in length, the number of data reflected in the plots decrease with increasing pressure. Also profiles contributing 20 or more samples within the 20 dbar interval have a larger impact on the distribution than those with only one or two measurements. Nevertheless, these plots give some idea of the way the data are distributed. If the sampling were uniform and if the data were distributed normally, the outliers could be considered to be highly unlikely and to be discarded; however, as emphasised earlier, the sampling is not uniform and the data are not normally distributed. Many of the warm outliers are likely to be associated with observations of the Loop Current and eddies that it spawned.

[6] The large dots indicate medians; boxes extend from 1st quartile to 3rd quartile; whiskers extend from quartiles to most extreme observation within 1.5 times the inter-quartile range; small dots indicate outliers. For example, in the interval from 180 dbar to 200 dbar, the middle half of the temperature data fall within an approximately 2 °C interval, and the whiskers extend an additional 3 °C in each direction, so everything outside an 8 °C interval is indicated as an outlier.

Fig. 5 shows the temperature profiles responsible for all of the warm outliers in the 180–200 dbar interval, which are indicated in Fig. 3 as a continuous row of dots extending to the right of the right whisker. Many appear to be reasonable Loop Current profiles, but some are bad. Similarly some of the cold outliers can correspond to cyclonic eddies. The Loop Current can also contribute what appear to be salty outliers to box-and-whisker plots, and the cyclonic eddies can contribute false fresh outliers. As the Mississippi River discharges a large volume of fresh water into the Gulf, the very-low-salinity outliers in the 0 to 20 dbar interval could be reasonable values, if the stations are near the mouth of the Mississippi. On the other hand, salinities less than 10 psu at depths of 20 m or more are not believable. Generally, the outliers that are widely separated from the others are likely to be bad data. However, there may also be bad data with values within the range of true variability.

Fig. 6 shows temperature vs. salinity for all data within the 180–200 dbar interval, including those associated with both warm and cold outliers. To focus on the bulk of the data, one point was excluded from the *TS* plot; it corresponded an outlier with S = 0 psu in the 180–200 dbar interval that was also excluded from Fig. 4. Note that the detached cold outliers indicated in Fig. 3 with temperatures less than 10 °C do not fall into the general pattern set by the bulk of the data; the same is true for the detached fresh outliers with salinity less

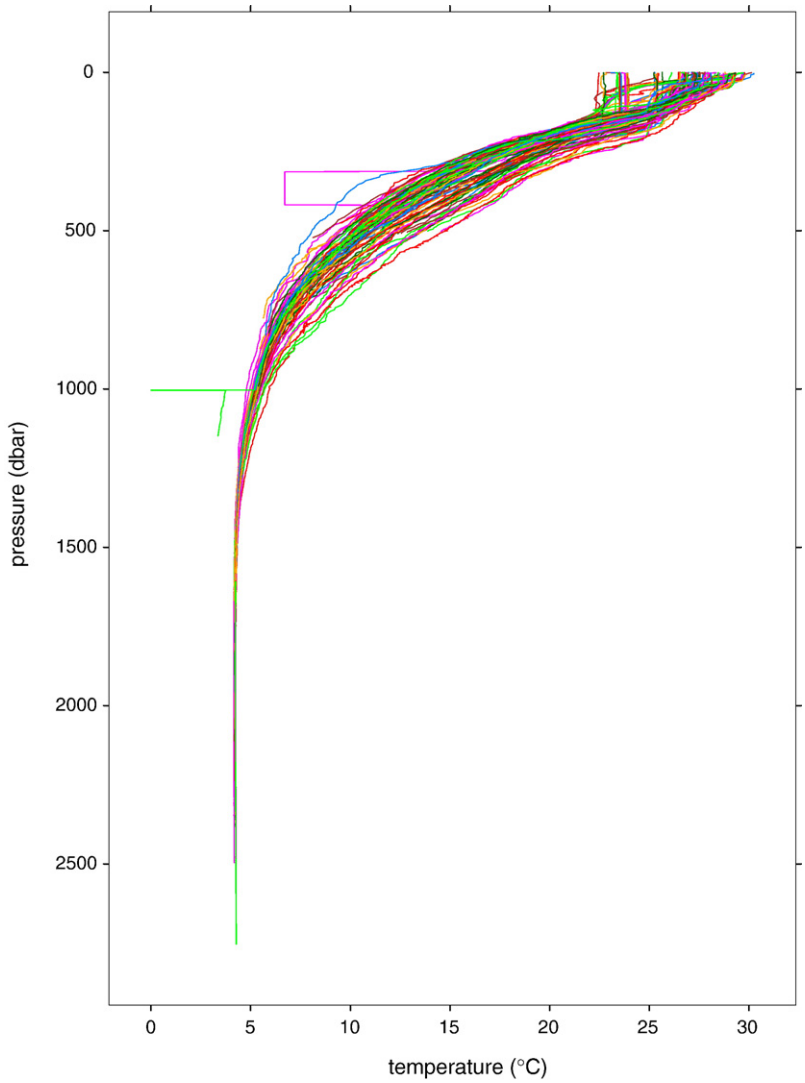

Fig. 5. One hundred and fifteen temperature profiles responsible for the warm outliers in the 180–200 dbar interval in Fig. 3. Most look like reasonable Loop Current profiles.

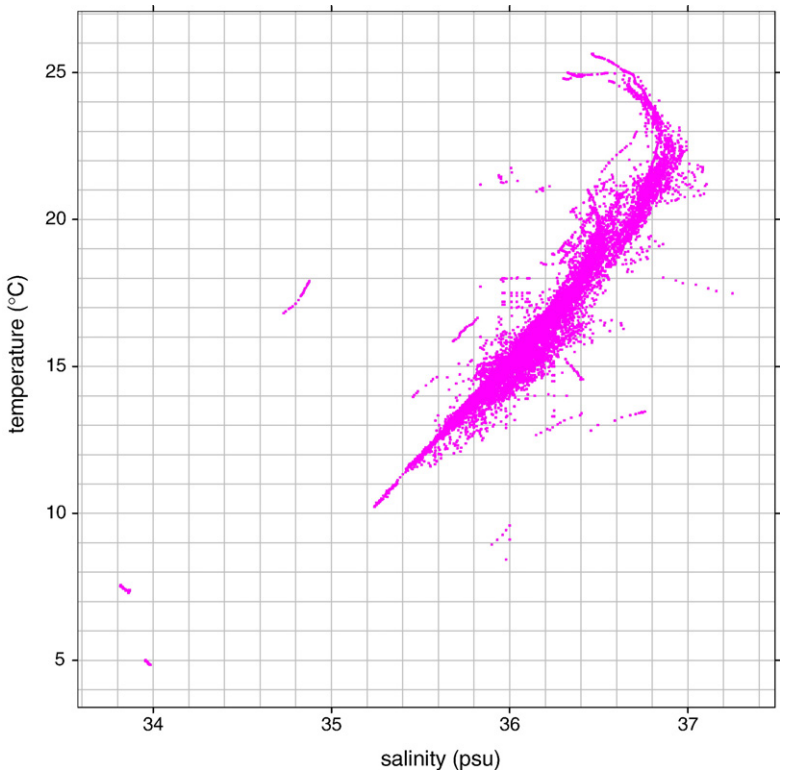

Fig. 6. Temperature vs. salinity for all data in the 180–200 dbar interval of Figs. 3 and 4 except for a single point with salinity value of 0 psu.

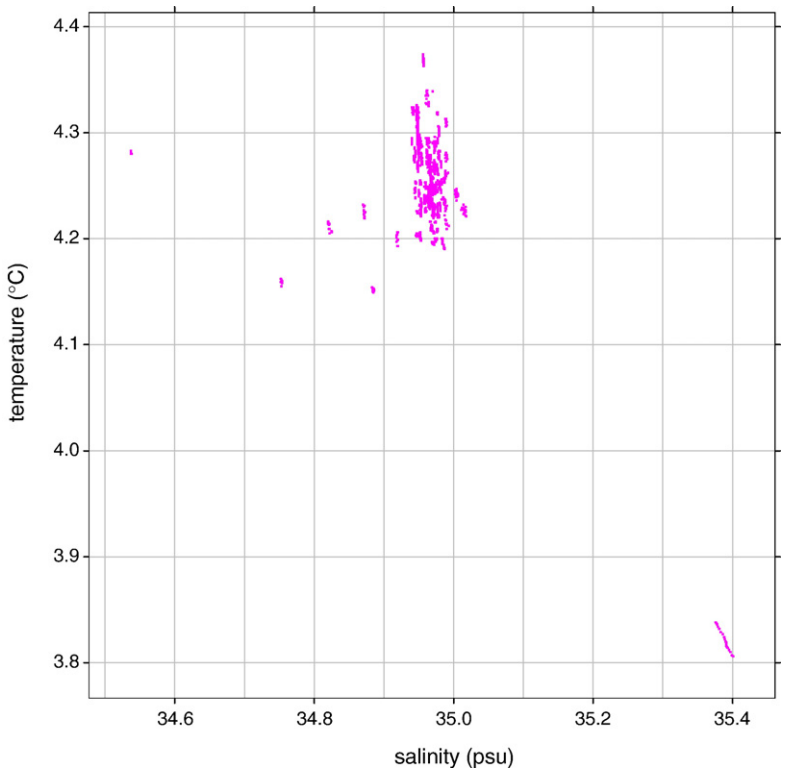

Fig. 7. Temperature vs. salinity for profiles in the 1480–1500 dbar interval of Figs. 3 and 4.

than 35 psu. Furthermore, there are quite a few points that appear to have erroneous salinity values in the *TS* plot, which do not correspond to outliers in Fig. 4. Most of the data show a well-defined relationship between salinity and temperature. In particular, the profiles of Fig. 5, which were indicated as non-detached outliers in Fig. 3, all having temperature above 20 °C in this pressure interval, extend the cluster formed by the other points in such a way that a smooth curve through the data in Fig. 6 should accommodate both sides of the Loop Current front. From the width of this cluster, you might expect such a curve to estimate salinity from temperature in this pressure range with expected root-mean-square error of no more than 0.1 psu and with the greatest error being for temperatures of approximately 15 °C.

It is also interesting to look at a *TS* plot for water well below the sills that separate the Gulf from the Caribbean and from the North Atlantic and far enough below the thermocline that surface influences are unlikely to have much effect. Fig. 7 shows the data for the interval 1480–1500 dbar. Note that the range of salinity is considerably less than for Fig. 6. Ignoring the obvious outliers, you can see a slight tendency for salinity to increase as temperature decreases. However, for any value of temperature there is a spread, which at this depth is more likely to be a reflection of the accuracy of the salinity measurements than an indication of true variability. This spread should set a lower bound for the accuracy of salinity that can be inferred from these profiles throughout the water column.

3. Selecting and rejecting data

The best way to proceed with the tasks of identifying and removing bad data and of thinning out redundant data is not at all clear. As these tasks must ultimately be done for the whole world, not just for the Gulf of Mexico, it is important to find a way to proceed that is not too time-consuming. Here, the task of thinning was addressed first, so that there would be fewer bad data to deal with after thinning.

To explore the extent to which thinning the data might impact the statistics of the data for the Gulf of Mexico, the choice of which data to discard was based entirely on convenience and without regard for optimality. First, the 259 profiles from the 30 multiply occupied stations were removed. Then all 1850 profiles that were within 0.2° latitude and 0.2° longitude and 1 day of the subsequent station in the ordered list were set aside. While some sequences of adjacent stations might be long enough for some of those stations to be well-separated from others, no effort was made to determine whether this was the case so that the sequence of stations could be sub-sampled. Also, no effort was made to select the most representative profile, the most error-free, or the one with the best vertical sampling. The decision to retain the first of each sequence and discard the others was made entirely for convenience with the thought that selection procedure could be improved if necessary, time permitting.

Box-and-whisker plots for this subset of profiles look remarkably like those in Figs. 3 and 4. The means and quartiles are much the same, many of the extreme outliers that correspond to bad data remain, as do the heavy tails that might be attributed to Loop Current water. At this stage we could continue with further thinning of the stations before addressing the issue of bad data. However, the fact that the box-and-whisker plots after removing over half the data is much the same as before indicates that the non-homogeneity of the sampling has not had a strong impact on the distributions. The data appear to be relatively homogeneous throughout the Gulf and details of their spatial and temporal coordinates seem not to be very important. In particular, the nature of the outliers is fairly insensitive to the sampling.

In light of this conclusion there are two strategies to consider. One is to attempt to use as many of the profiles as possible, discarding primarily those that are clearly bad and discarding far fewer on the grounds of redundancy. The other is to continue working with the smaller set of data, deeming it to be sufficient. For the sake of expediency, the second strategy was chosen. With fewer data, there are fewer bad data, so their identification should be less work.

The approach to handling bad data was motivated by Figs. 6 and 7 together with Figs. 3 and 4: outliers that are well-separated from the whiskers in the box-and-whisker plots are generally separated from the other data on the *TS* plots. While some of these detached outliers may indeed be good data, they would appear at the extremes of the *TS* plots and thus could have unwarranted influence on the regression curve used for estimating salinity; as there are generally sufficient data, these shouldn't be missed. Furthermore, as there were very few profiles with cold or fresh outliers, detached or not, they were all discarded, even though they might have corresponded to cyclonic rings. Retaining the non-detached warm, salty outliers guaranteed that most of the data for the Loop Current and its rings would be retained. Fig. 6 illustrates that there can be bad data that do not show up as outliers in box-and-whisker plots for temperature or salinity (Figs. 3 and 4 at 180–200 dbar)

and even lie in the inter-quartile interval. However, by discarding the entire profile when a temperature or salinity outlier at even a single level was suspicious, there was the chance of also eliminating unrecognised bad data at other levels. This in fact proved to be the case, as *TS* plots after discarding profiles with suspicious univatiate outliers exhibited far fewer points away from the principal clusters. As those remote points might be avoided by using robust regression techniques, they were not considered to pose a serious problem.

Thus, we eliminated from the 1380 less-redundant profiles the 50 profiles that contribute cold outliers and the additional 417 that contribute fresh outliers. While some of these profiles might have outliers in only one of the 20 dbar intervals, no effort was made to retain the "good" parts; the entire profile was eliminated if it contributes at least one cold or fresh outlier. No effort was made to check if any of these discarded profiles characterise cyclonic eddies and should be retained. Three more profiles, which contribute isolated salty outliers, were also discarded, leaving 910. This leaves us with a bit more than a quarter of the original profiles with which to work; 135 provide information below 1000 dbar and 630 provide information to at least

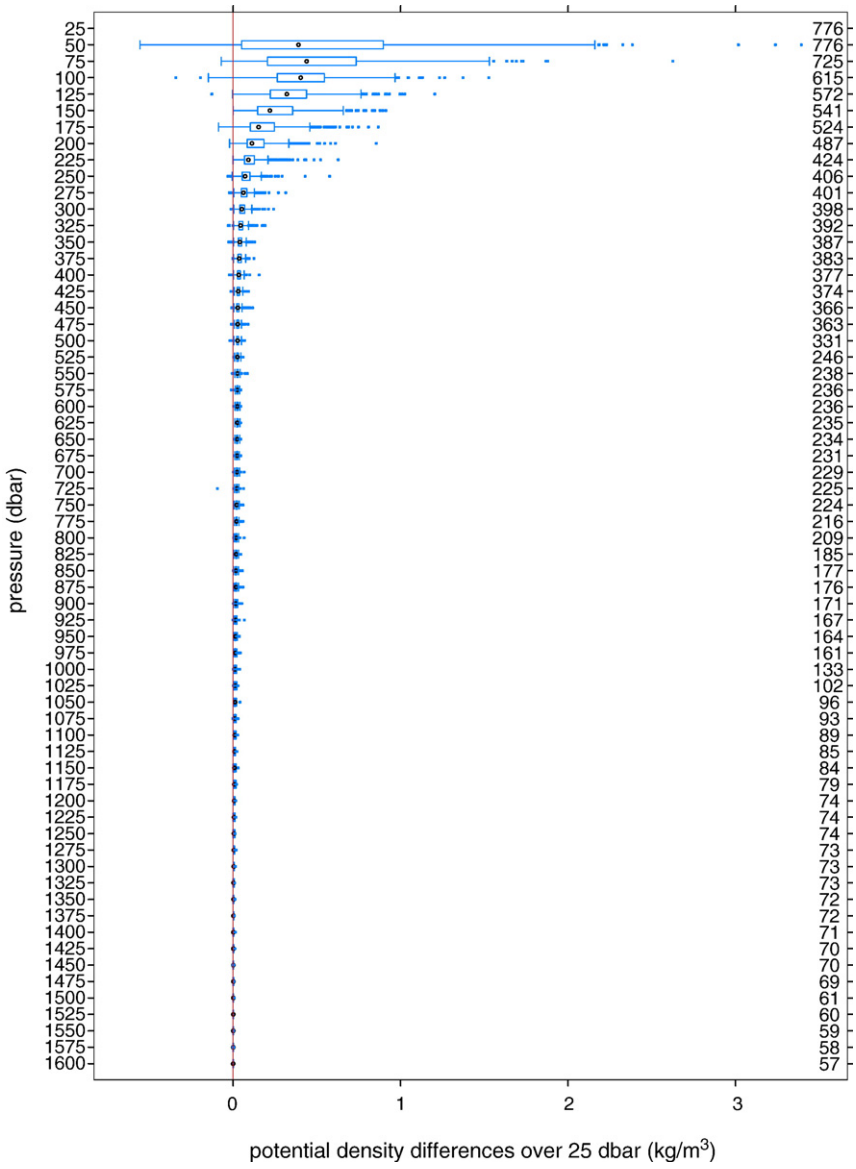

Fig. 8. Box-and-whisker plots of potential density differences between interpolated profiles at adjacent pressure levels. The numbers of profiles at each level are indicated to the right. Inversions appear to the left of the vertical red line.

100 dbar. The locations of the 910 stations were compared with Fig. 1 and seen to be a relatively uniform subsample with significantly less redundancy and fewer points near the mouth of the Mississippi River. One of the stations over land remained and was removed, bringing the count to 909 stations. Box-and-whisker plots for these profiles (not shown here) still have many warm and salty outliers, most of which are presumed to be due to the Loop Current and its rings, but some of which might be bad data that were not eliminated. Of the 909 remaining profiles, 16 having fewer than 5 measurements, all in shallow water, were discarded. In addition, 14 profiles having minimum pressure greater than 25 dbar and 103 having maximum pressure less than 50 dbar were removed, leaving 776 profiles for modelling salinity.

4. Vertical interpolation and density inversions

To avoid biases resulting from differences in vertical sampling among the profiles and to make the task of identifying density inversions easier, it is best to interpolate all profiles to standard levels. Examination of the *TS* plots of the data with the 20 dbar intervals show little

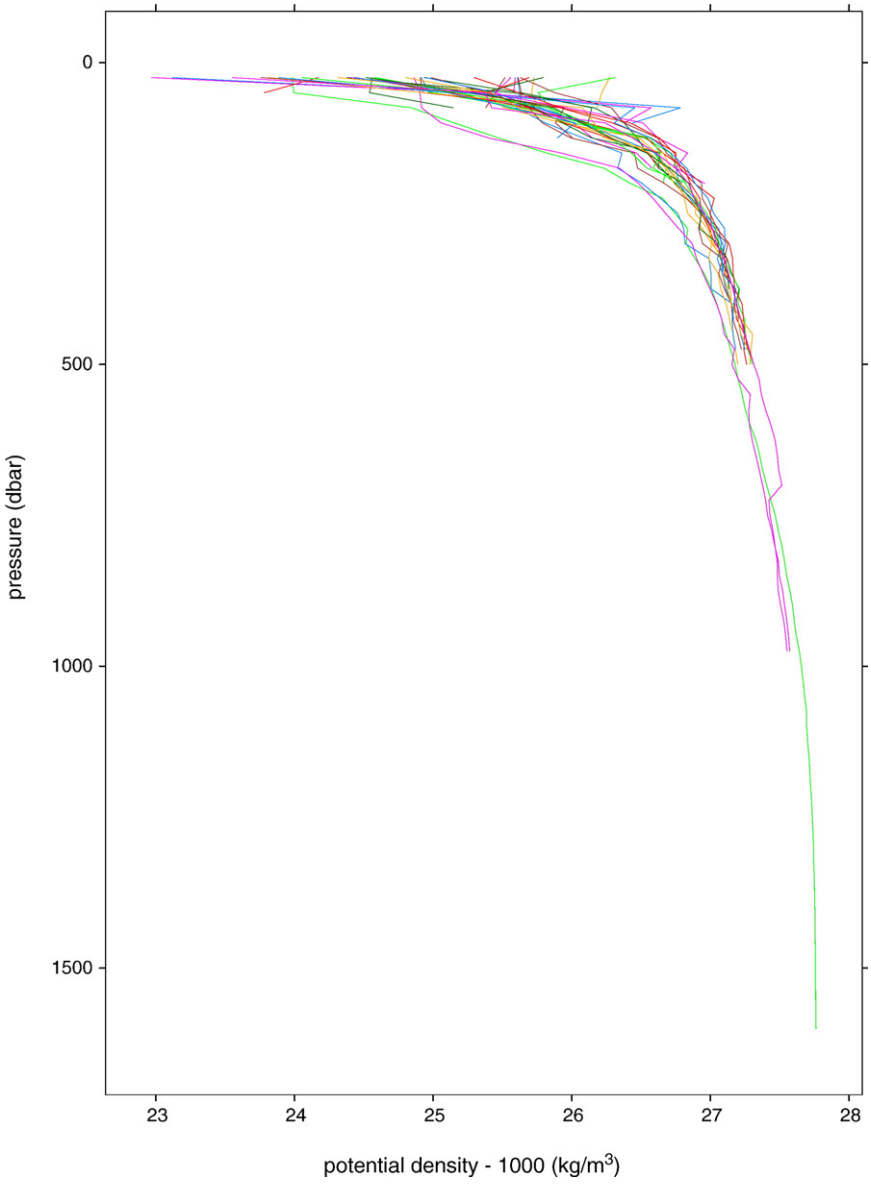

Fig. 9. Thirty-seven interpolated potential density profiles with inversions between adjacent 25 dbar pressure levels in excess of 0.01 kg/m^3.

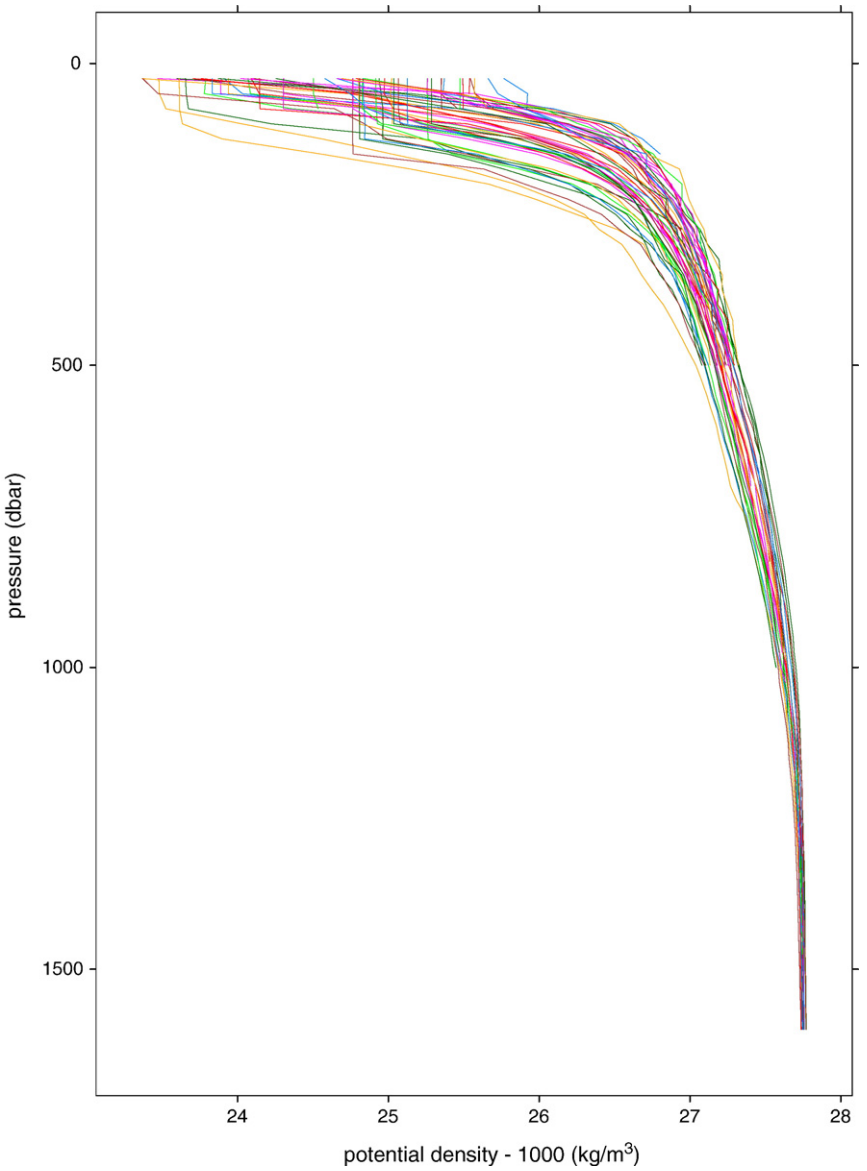

Fig. 10. Fifty-eight interpolated potential density profiles with negligible (less than 0.01 kg/m^3) inversions between adjacent 25 dbar pressure levels.

change from one interval to the next, so a slightly larger interval of 25 dbar was chosen for the standard pressure levels on which salinity is to be modelled.[7] As there is very little variability deeper than 1600 dbar, salinity there can be estimated by the mean salinity; in any case, XBT profiles are not expected to extend below 1600 dbar. A first try using smoothing splines[8] for interpolation, with the thought to remove unwanted small-scale spikes and offsets while preserving variability on scales greater than 5 dbar, proved to be problematic; for sparse vertical sampling in the vicinity of a sharp thermocline below a well-mixed layer, the smoothing splines occasionally gave unreasonable interpolated values. Consequently, simple linear interpolation was used.

The interpolated data are easier to examine for density inversions than are the original, high-density profiles, as the complications caused by cast-to-cast differences in vertical sampling have been eliminated. Fig. 8 shows the distributions of differences in values of potential density at adjacent 25 dbar levels. Ideally, they

[7] Between these standard levels estimates of salinity can be obtained by interpolation.

[8] Smoothing splines were computed using the R software function smooth.spline() with the degree of smoothing determined by generalised cross-validation (Venables and Ripley, 2002; R Development Core Team, 2004).

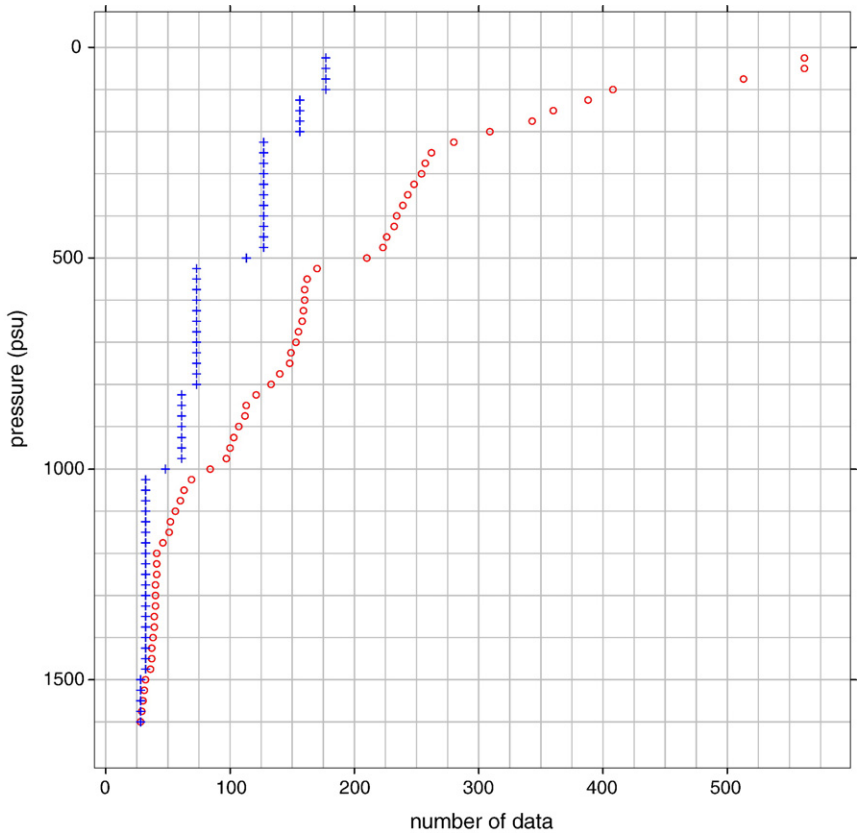

Fig. 11. Number of training data + and verification data ○ at each pressure level.

should all be positive, with no values lying to the left of the vertical red line. Most of the density inversions occur in the first 200 dbar, but a large inversion can be seen at 725 dbar. All inversions can be attributed to 95 of the 776 profiles. Given the accuracies with which salinity and temperature are recorded, inversions over 25 dbar of 0.01 kg/m^3 should not be considered significant.[9] The 37 potential density profiles with significant inversions are shown in Fig. 9. The small inversions of the remaining 58 unstable profiles, which are shown in Fig. 10, are imperceptible. Thus only the 37 profiles with larger inversions need be eliminated from the pool of data to be used for modelling salinity. Only 3 of these 39 extend to 800 dbar, reducing the number of data at that depth from 209 to 206, and only 1 extends to 1000 dbar, reducing the number of data there from 89 to 88, so there are ample data for modelling salinity over the range of most XBTs.

[9] A more tolerant value of 0.03 kg/m^3 was used for the Gulf Stream Recirculation region (Thacker and Sindlinger, 2007-this issue).

5. Partition of data into training and verification sets

The 739 profiles remaining after sub-sampling for uniformity and after elimination of outliers and inversions can be used for fitting statistical models of salinity. One strategy would be to use all of these data for determining the statistical models and to use the discarded data for verification. However, as there are a relatively large number of profiles to work with, a second strategy would be to divide these 739 profiles into two groups: one for fitting the models and the second to verify that the models perform well with independent data. The advantages of the second approach are (1) that both training and verification data would be relatively homogeneous and clean, so judging the performance of the models should be easier, and (2) that the verification data would be more independent of the training data, as redundant data have been eliminated. This advantage seems to outweigh the advantage of fitting to the larger number of data.

The fact that many more profiles provide data at 25 dbar than at 1600 dbar (as indicated in Fig. 8) is

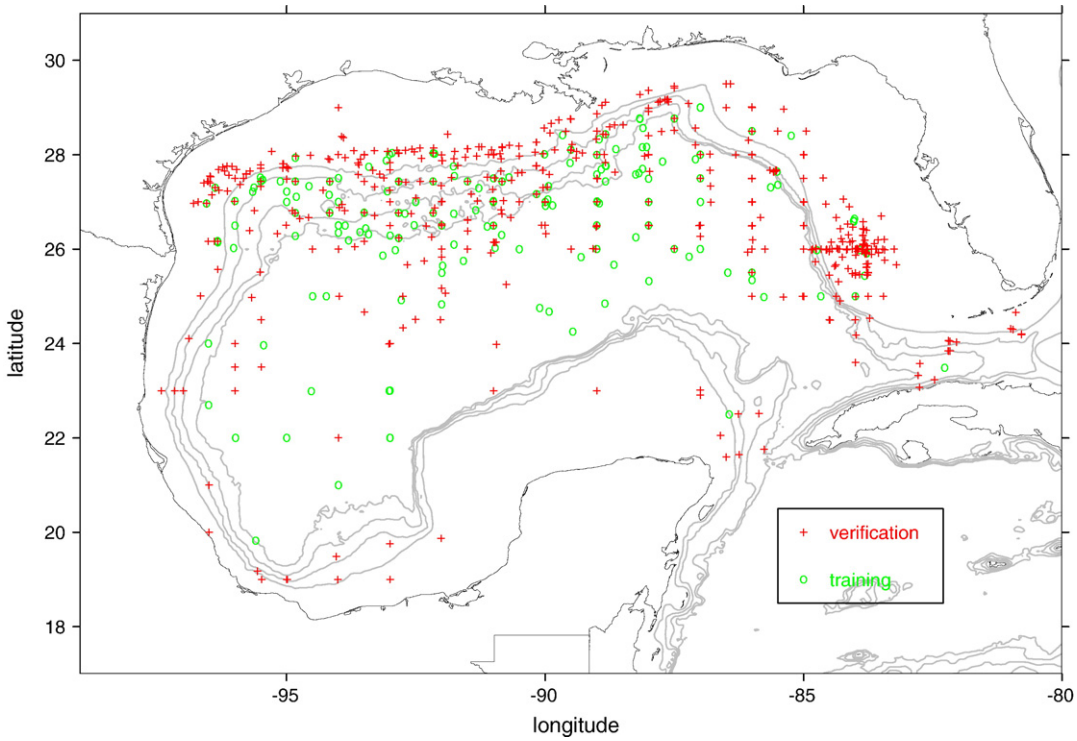

Fig. 12. Locations of stations used for training and for ○ and verification +.

central to the strategy for partitioning the data into training and verification sets. To maintain as much vertical continuity as possible, any profile to be used for training should be used at every level where it provides data. While using half of the profiles that reach 1600 dbar for training might be sufficient at that level, a larger sample is needed closer to the surface where variability is much larger. A random selection of the profiles providing the deepest data formed only a part of the training set. More profiles were selected from those that penetrate almost that deep, then more from the next shorter profiles, and so on for shorter and shorter profiles. To maximize vertical continuity, far less than half of the short profiles were selected for the training set.

Fig. 11 indicates the number of data used for training and for verification at each pressure level. To 100 dbar all 177 training profiles provide data, while only the 28 longest provide data all the way to 1600 dbar. As there were several levels that corresponded to the bottoms of many profiles, most of the other training profiles were chosen from those ending at these levels. Thus there are 28 verification profiles providing verification data throughout the water column to 1600 dbar, while there are 562 providing data at 25 and 50 dbar.

Fig. 12 shows the locations of the stations contributing data for training marked with a green circle and the locations of those contributing data for verification marked with a red plus. Except for more of the verification data being in shallow water, their spatial distribution can be seen to be quite similar. Both sets of stations reflect the distribution of the stations before sub-sampling (Fig. 1).

The *TS* plots for the training data are shown in Fig. 13. If all data were plotted on the same scale, details for the lower depths would be hidden, so three sets of temperature and salinity scales are used. In the first 100 dbar little relationship between salinity and temperature can be seen, so estimates based on fits to these scatter plots[10] are not expected to be much better than those based on the mean salinity at these levels. However, the *TS* relationship gets stronger with increasing pressure, indicating that reliable estimates can certainly be expected for pressures greater than 225 dbar. The *TS* plot for data at 750 dbar shows a minimum of salinity within the range of observed temperatures, suggesting that straight-line fits might not be the best

[10] When regression curves are plotted through scatter plots, it is more conventional to associate the independent variable with the abscissa and the dependent variable with the ordinate. However, oceanographic tradition for TS plots is to associate the temperature with the ordinate and salinity with the abscissa.

choice for this pressure level. Other pressure levels might also be better modelled by a curve than by a straight line. Note that the interval between ticks on the salinity scale for the lower set of panels is only 0.02 psu; considering the spread of the data, this gives an idea of what to expect for the size of estimation errors at these pressure levels.

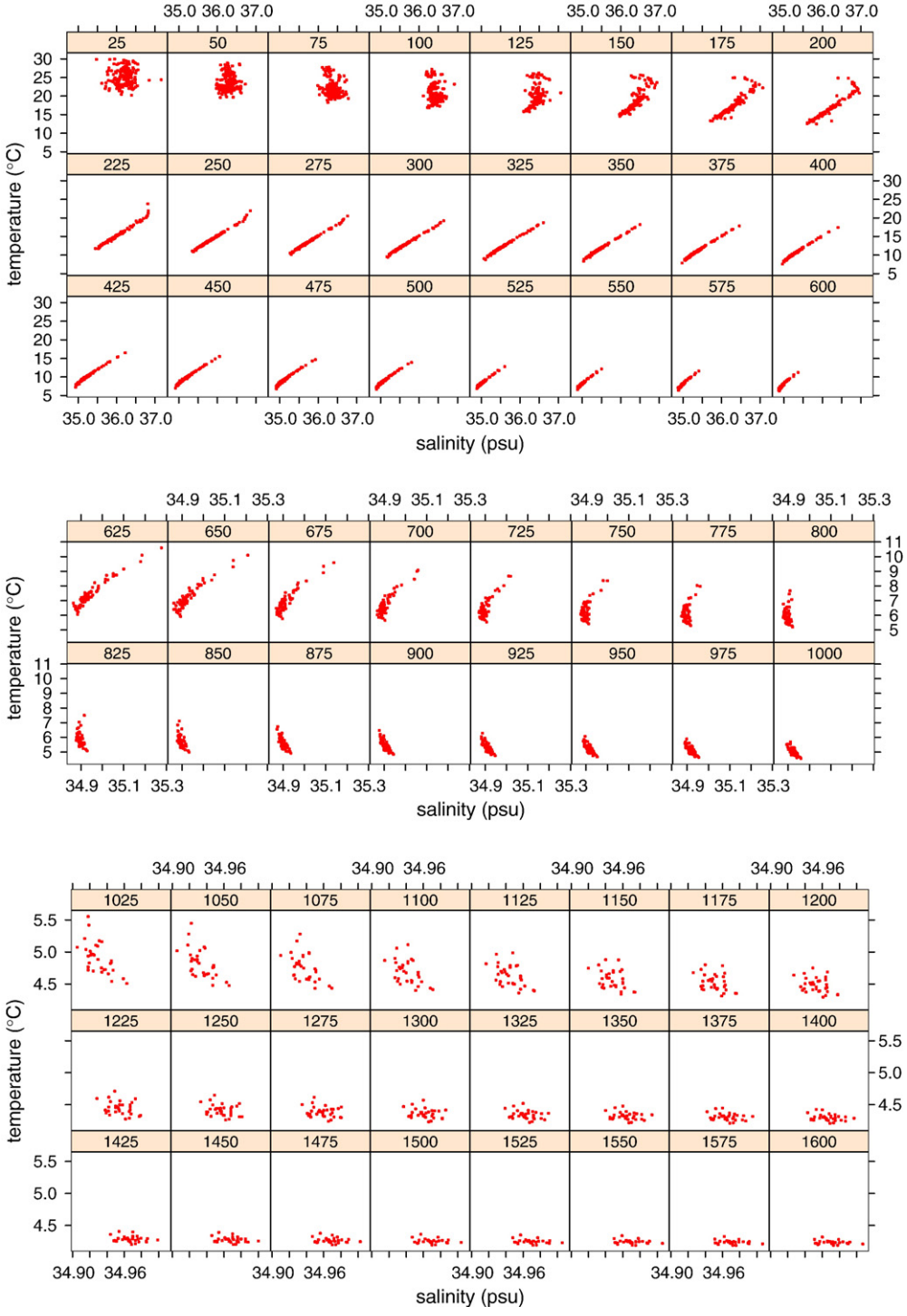

Fig. 13. *TS* plots for the training data at each pressure level as indicated in panel labels. Because variability diminishes with depth, three different sets of scales are used.

The range of observed temperatures for levels 1425 through 1600 dbar is about 0.5 °C, which is considerably smaller than that for levels 1025 dbar through 1200 dbar, whereas the ranges of salinity are comparable; this indicates that much of the variability of salinity for the deeper water can be attributed to measurement error. No effort has been made to explore whether the recent measurements are more reliable than those made earlier; instead, they are all treated equally, relying on the averaging effect of the relatively few data to provide usable estimates.

6. Regression models

The strategy for estimating salinity is to identify regression models for each pressure level that explain the data in the panels of Fig. 13. The skill of these models is to be assessed against the independent verification data for the corresponding levels. In order to see whether computing such fits is worth the effort, the performance of easier-to-implement models based on Stommel's method will also be evaluated.

The scatter plots shown in Fig. 13 suggest that straight lines or parabolas might approximate the data at all levels, although neither is expected to approximate the data well at the near-surface levels. The straight-line models,

$$\hat{S}_p = a_p + b_p T_p, \tag{1}$$

are identified by finding coefficients a_p and b_p that provide the best fit $\hat{S}_p$ to the *TS* plots at pressure-level p. Similarly, the parabolic models,

$$\hat{S}_p = c_p + d_p T_p + e_p T_p^2, \tag{2}$$

require finding the values of the coefficients c_p, d_p, and e_p. For each type of model, two methods for determining the coefficients were used.[11] The first was the usual linear regression, where the coefficients were chosen to minimise the sum of the squares of the differences between the estimated and observed values of salinity. Because a few unrepresentative points on the scatter plots can have a large influence on the sum of squares and thus on the resulting coefficients, robust regression, which discounts the influence of outliers, was also used. Thus, at each level four different regression models were examined.

While the root-mean-square residuals from fitting each type of model at each level could give an idea of how well the various models approximate the data, better indications of their expected accuracies for estimating salinity to accompany XBT data are their root-mean-square (rms) errors for approximating the independent *CTD* data that were set aside for verification. These rms verification errors are plotted in Fig. 14. While different coefficients were used at each pressure level, the rms errors at adjacent pressure levels are connected to show how errors are expected to vary between levels. For all four types of models, the rms errors generally decrease with depth. Near the surface, where variability is large and salinity does not reflect temperature, the models essentially estimate the salinity by the mean (or robust mean) salinity of the training data and the error reflects the variability of the data around this estimate of the climatological mean. At depth, where there is little variability, rms estimation errors are small and conform to expectations based on the scatter plots of Fig. 13. Between 100 dbar and 225 dbar and also between 300 dbar and 900 dbar parabolic models have smaller rms errors than those based on straight-line fits; everywhere else the rms errors for the two types are about the same. The robust fits show some advantage for the straight-line models[12] but hardly any for the parabolic models.

Displaying all 562 estimated profiles is impractical, but the 28 full-length profiles can be shown. This subset is a particularly good choice, as it provides examples where salinity is relatively poorly estimated. The blue curves in Fig. 15 are the salinity estimates for these profiles made with the robust parabolic models; the red curves are the measured salinity. While the blue curves generally hide the red curves for pressures greater than 300 dbar, nearer the surface differences are obvious. Salinity appears to be under estimated and over estimated with approximately the same frequency and by about the same amounts; it is most under estimated (by 0.41 psu) for profile number 3314012 at 50 dbar and most over estimated (by 0.37 psu) for profile number 3358719 at 75 dbar. These are not the only profiles with

[11] The computations were made using the R software. the function lm() was used for simple linear regression and rlm() for robust linear regression.

[12] The rms residuals for the fits to the training data are greater for the robust models, because their fits ignore problematic points that are not given special treatment by the scoring. Ignoring these points pays of with better estimates for independent data. Note that, even though bad training data are ignored, bad verification data are not ignored; when computing the rms errors shown in Fig. 14, no attempt has been made to identify and to exclude any bad data. Thus the bad data, such as those at 200 dbar shown in Fig. 19 below, cause the rms errors to be larger than they would be if evaluated using only good data.

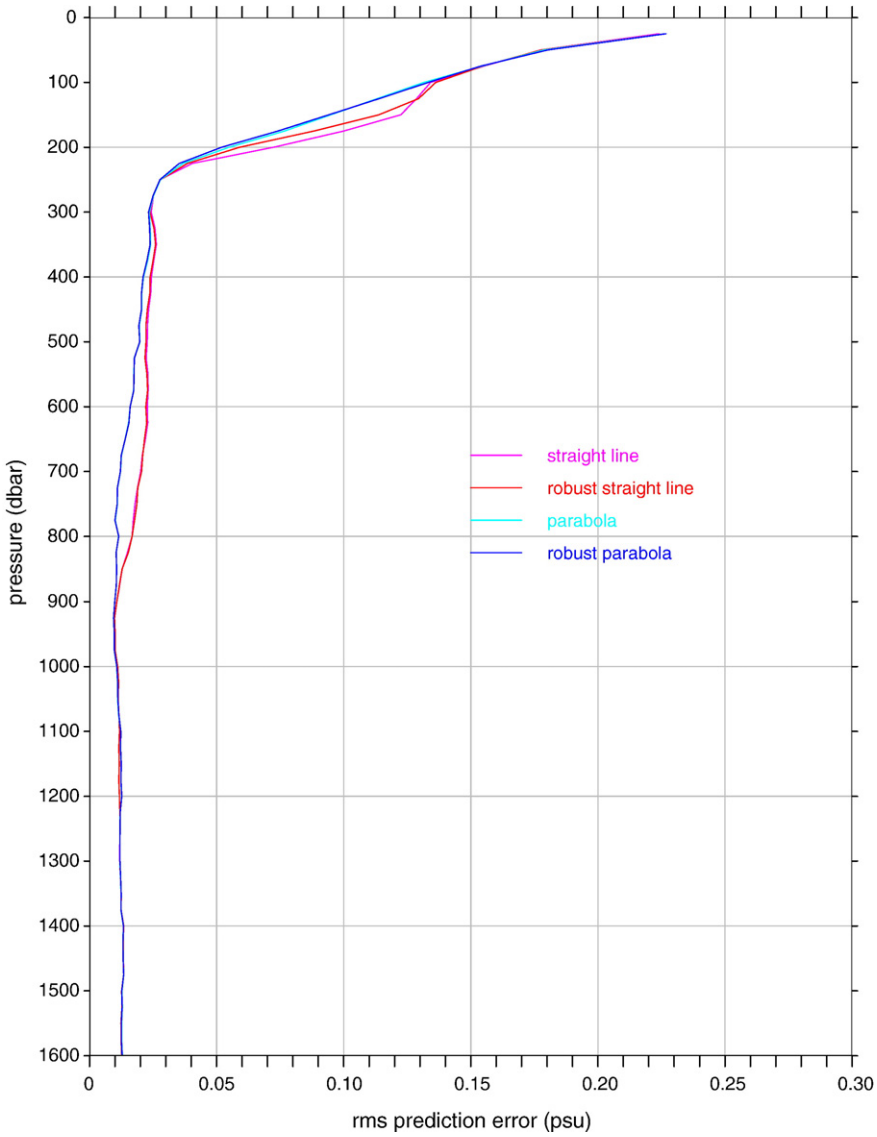

Fig. 14. Root-mean-square differences between verification data and their estimated counterparts for four regression models. Magenta and cyan curves, respectively, are for least-squares fits of straight lines and parabolas to the training data at each pressure level; red and blue curves are for the corresponding robust fits. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

large errors; for example, 3249485 and 3358538 also have errors greater than 0.4 psu and 0.3 psu, respectively. Considering how many profiles have relatively large errors, it is not surprising that, in the pressure ranges of 50–125 dbar and 200–225 dbar, the rms errors for this subset are greater than those for the entire verification set.

Fig. 16 shows the coefficients for the robust parabolic models that were determined by fitting to the training data at each standard pressure level. Also shown are standard-error intervals around the best-fit values indicating their uncertainties. Note that the coefficients vary smoothly between pressure levels, confirming the assumption that 25 dbar intervals between levels would be satisfactory. Note also that the uncertainties increase as pressure increases beyond 1000 dbar, reflecting the sparsity of the data for the deeper regions of the Gulf. Finally, neither this figure nor Fig. 14 should suggest that a second-degree polynomial function of temperature should be used at all levels. For example, the sharp bend at the salinity maximum seen in the *TS* plot for 180–200 dbar in Fig. 6 suggest something like a fourth-degree polynomial might be more suitable at 200 dbar, while Fig. 7 suggests that a more parsimonious straight

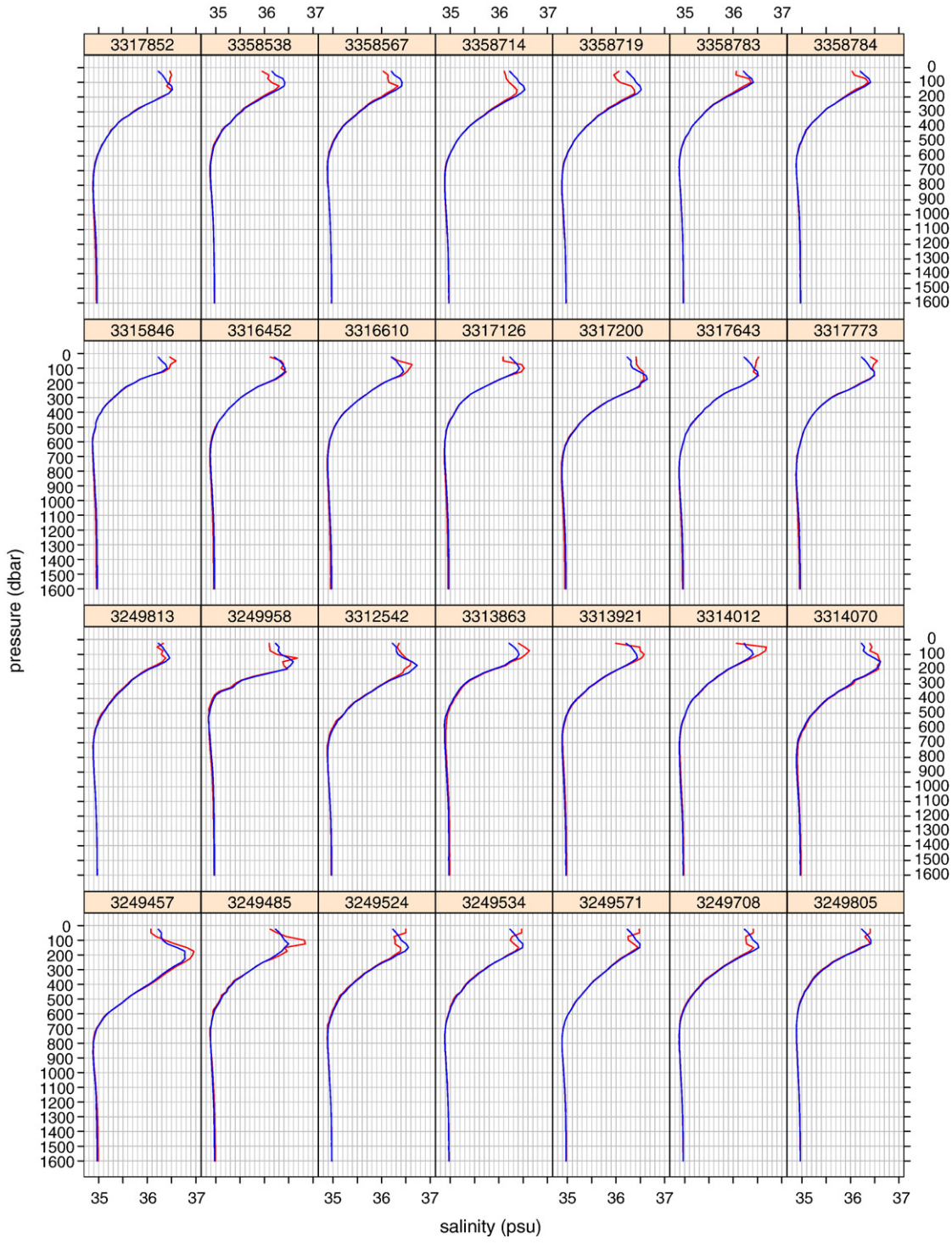

Fig. 15. Blue curves indicate salinity estimated with the robust parabolic model, red curves indicate observed salinity; and panel labels are identification numbers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

line might be more appropriate in the deeper regions, and models employing regressors other than powers of temperature might be best near the surface. Still, as the robust parabola does well over the entire range of pressures, it can be recommended for all levels until better models are available for estimating the near-surface

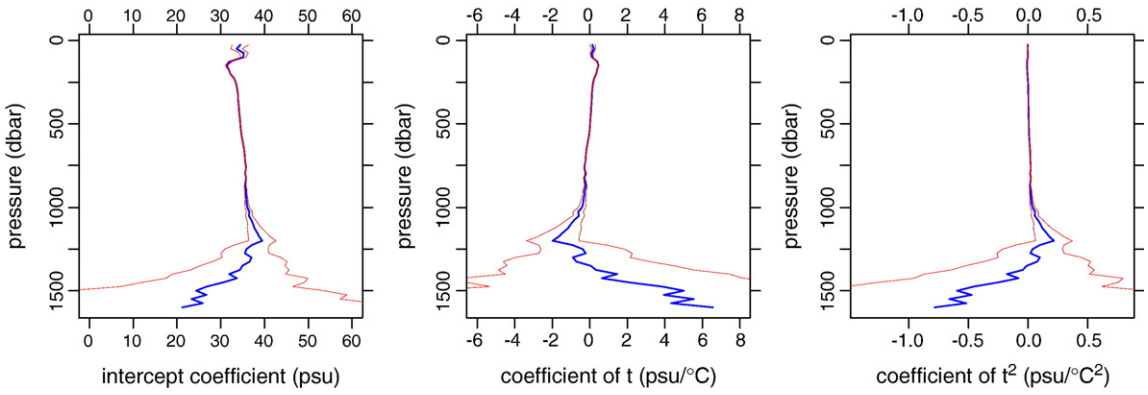

Fig. 16. Coefficients and their standard-error intervals for the robust parabolic models vary with pressure.

salinity. Computer codes based on the coefficients shown here can be easily implemented to give salinity counterparts for temperature profiles from XBTs for the Gulf of Mexico.

These models were all identified without regard for the spatial and/or temporal locations of the data. To check whether the models might be improved by using spatial and/or temporal coordinates as additional regressors, or by partitioning the data by location or by season, residuals can be plotted versus latitude, longitude, and day of year. As systematic spatial and temporal behaviour is most likely near the surface, such plots are presented in Fig. 17 for residuals of the robust parabolic fit at 25 dbar; no evidence of spatial or temporal dependence can be seen. Likewise, plots (not shown here) for all four models at this and other pressure levels give no indication that predictions might be improved by accounting for location or for time of year.

The estimates of salinity at 25 dbar are quite poor for all methods. As salinity correlates poorly with temperature this close to the surface, the regression models do no better than the mean salinity at this level, and the models based on Stommel's method are worse. With the prospect of satellite-based measurements of sea-surface salinity and with the positive results reported for the eastern tropical Pacific (Hansen and Thacker, 1999), it is useful to check whether sub-surface estimates of salinity might be improved when XBT profiles are supplemented with such data. Of the 739 *CTD* profiles in the combined training and verification sets, only 577 had values for salinity for pressures no greater than 2 dbar; the others were removed from the training and verification sets and the salinity at minimum pressure of the retained profiles were used to represent the surface salinity. When used as a regressor along with temperature, the improvement was negligible, indicating that for the Gulf of Mexico satellite observations of sea-surface salinity would not be very useful for improving estimates of salinity at or below 25 m.

Altimetric data from satellites may provide some additional information about near-surface salinity. As salinity estimates from XBT data are quite good below 400 m, it may be reasonable to expect that errors in dynamic height reflect errors in salinity above 400 m. If so, assuming that the altimetric data behave in the same way as the dynamic heights based on *CTD* data, it

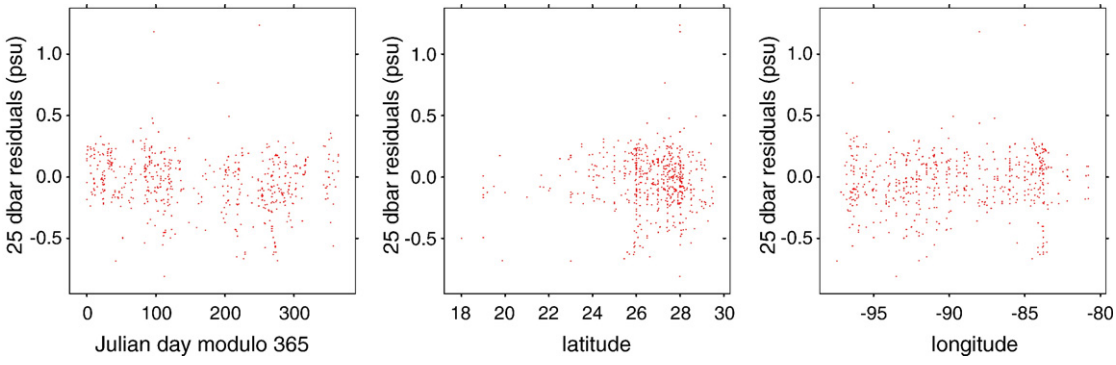

Fig. 17. Residuals from robust parabolic fit at 25 dbar vs. Julian day, latitude, and longitude.

should be possible to improve the estimated salinity profiles by supplementing the XBT data with altimetric data from satellites.

7. Stommel's method

In addition to regression models, Stommel's (1947) method should also be considered. This method estimates the salinity to accompany an observed temperature by the mean value of salinity observed at this temperature, regardless of pressure:

$$\hat{S} = \langle S \rangle_T, \tag{3}$$

where $\langle S \rangle_T$ is the mean salinity when temperature has the value T.

To implement this, the *CTD* profiles in the training set were linearly interpolated to standard temperatures at 0.1 °C intervals.[13] The mean salinity at the standard temperature levels were then linearly interpolated to the observed values of temperature for each of the verification profiles to get estimates that can be compared with the verification profiles' observed salinity.[14] The green curve in Fig. 18 shows the rms errors of these estimates. For comparison, the purple curve shows the rms estimates for the best regression method of Fig. 14. The two methods have essentially the same rms errors for pressures greater than 350 dbar. Stommel's method is a tiny bit better at 800 dbar and below 1550 dbar, probably because it doesn't require data on a single pressure surface and thus can benefit from more data. Generally, Stommel's method is substantially worse than the regression model. Thus, it seems to offer no advantage over regression.

Fig. 19 illustrates why the regression model outperforms Stommel's method at 200 dbar. The green curve shows the predictions for Stommel's method, and the purple curve, for the robust parabola (same colours as for Fig. 18). The predictions for Stommel's method fail for the warm, salty data characterising the Loop Current, while the robust parabolic model passes through those points. Even so, the robust parabola does a poor job at capturing the salinity maximum near 21 °C, which can also be seen for the training data in Fig. 13.

The red curve in Fig. 19 shows the predictions for a quartic function of temperature that was fit to the training data at this level; it appears to do a better job predicting the verification data than the quadratic function over the entire temperature range but especially in the region of the salinity maximum. Nevertheless, its rms error (0.050 psu)is only slightly better than that for the robust parabola (0.052 psu). Quartic fits could not be made for pressures greater than 1350 dbar, and their predictive skills were considerably worse than for parabolic models for pressures greater than 700 dbar. However, for pressures less than 700 dbar their rms errors would be essentially indistinguishable from those of the parabolic models in Fig. 18.

While determining mean salinity on temperature surfaces from archived *CTD* data involves essentially the same amount of work as fitting regression models, which give better results, there is an alternative implementation of Stommel's method that is attractive because it involves less work: $\langle S \rangle_T$ can be approximated using climatological mean temperatures and salinities, which are readily available at standard depths for the world's oceans. Plotting the mean temperature profile vs. the mean salinity profile gives a *TS* curve from which the estimated salinity can be read at the value of the observed temperature.

The World Ocean Atlas (Conkright et al., 2002a) provides mean salinity and temperature for $1° \times 1°$ longitude × latitude cells at selected depths for the whole world.[15] Means are provided for each calendar month, for the four seasons, and for all data without regard for time of year. This raises the question of whether to tailor the estimates to $1° \times 1°$ cells and to time of year or to consolidate the means into estimates with less resolution. The result of finding little need to account for spatiotemporal variability when using regression models was also established in this context.

[13] As some profiles have temperature inversions, giving multiple values for salinity for the same value of temperature, the interpolation process is not always well-defined. Because the objective is to compute mean salinity at the specified pressure over the ensemble of training profiles, it seemed reasonable to use the average of the profile's salinity values whenever this case was encountered. Also, the mean salinity for each standard temperature was computed from different numbers of values, as not all profiles observed temperatures in the same range.

[14] Estimates cannot be made for observed temperatures greater than the warmest standard level defined from the data or colder than the coldest. These limits are set jointly by the selection of standard temperatures and by the range of the temperatures encountered in the training data. Thus, the rms errors for this method can be computed from a different number of profiles, especially near the surface.

[15] Perhaps in the near future Hydrobase will provide means of salinity and temperature on potential density surfaces for all of the world's oceans (Lozier et al., 1995} (http://www.whoi.edu/science/PO/hydrobase). If so, such climatological data might provide a better basis for approximating $\langle S \rangle_T$ for use with Stommel's method for estimating salinity from observed temperature, as isopycnic surfaces are generally better approximations for isothermal surfaces than are isobaric surfaces.

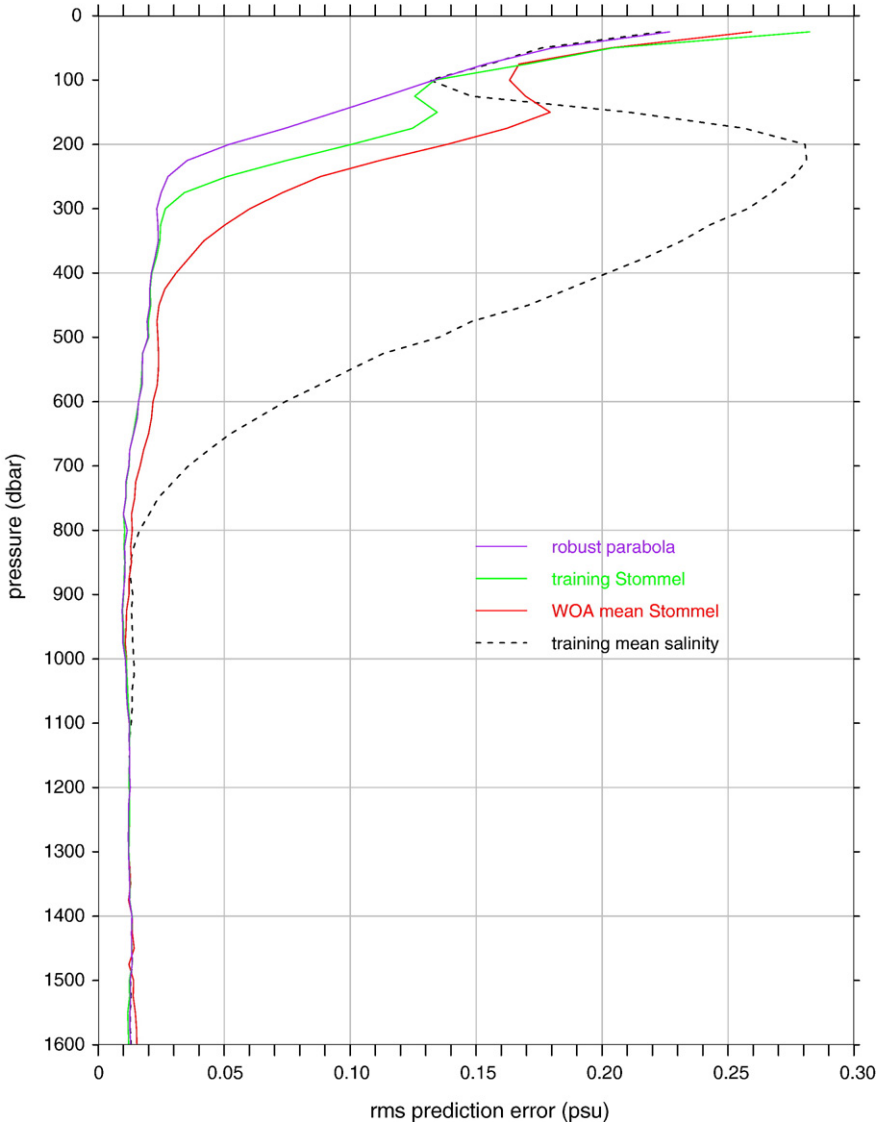

Fig. 18. Root-mean-square differences between verification data and their counterparts estimated using two different implementations of Stommel's method for estimating salinity. Green curve is for $\langle S\rangle_T$ computed from the training data; red, for $\langle S\rangle_T$ estimated using climatological mean profiles $\langle S(z)\rangle$ and $\langle T(z)\rangle$. Also shown are the rms errors for the robust parabolic model (purple) and for salinity estimated by the mean of the training data for specified pressure $\langle S\rangle_p$ (black, broken curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The red curve in Fig. 18 shows the root rms errors when $\langle S\rangle_T$ is approximated by the Gulf-wide average of the $1°\times 1°$ mean annual profiles. Separate estimates for $\langle S\rangle_T$ were also made of each cell, and the corresponding rms error estimates (not shown) were substantially larger, indicating that averaging over the Gulf provided better estimates. Nevertheless, the red curve indicates considerably larger errors than for the proper implementation of Stommel's method (green curve).

While discussing convenient estimates, it is important to consider the simplest, namely estimating the salinity to be the climatological salinity at the pressure level without regard for the temperature. The mean profiles of the World Ocean Atlas (Conkright et al., 2002a) can provide such estimates, but again with the issue of how much spatial resolution to use. The black dashed curve in Fig. 18 shows the rms errors when the verification data were estimated by the means of the training data at each pressure level. Errors for the upper 100 dbar are essentially the same as for the regression model (because temperature offers little help in this region), and also for the deep region where there is little

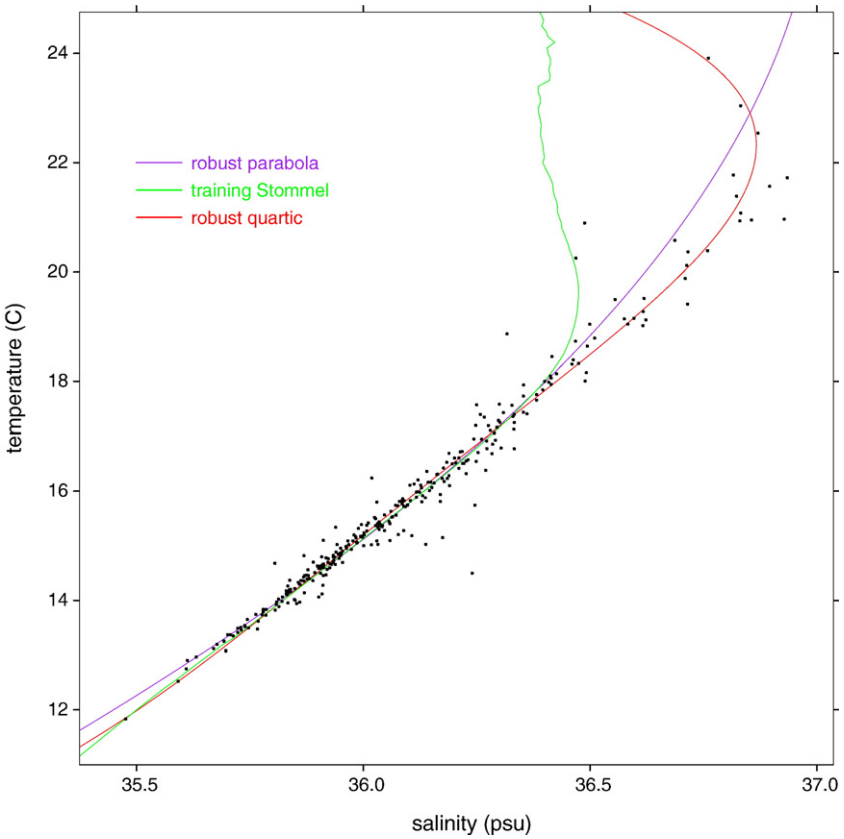

Fig. 19. Dots indicate verification data at 200 dbar with overlaid curves showing the predictions of the robust parabolic model (purple), Stommel's method (green), and a fourth-degree polynomial of temperature (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

variability, but in between this method, while convenient, is not competitive. Thus, there is little need to sort out how best to use the climatological mean profiles to estimate salinity. Using them together with Stommel's method is better, at least for the Gulf of Mexico, and not much more work.

8. Potential density

Using the equation of state for sea water (Fofonoff, 1977), estimates of salinity can be combined with observed temperature to give estimates of potential density, which are needed for inferring how the HYCOM model's layers should be configured (Thacker and Esenkov, 2002; Thacker et al., 2004) Fig. 20 shows the rms errors of such potential density[16] estimates for the verification data when the salinity is estimated using the robust parabolic models for each standard pressure level. The rms error is approximately 0.01 kg/m^3 for pressures greater than 700 dbar, which is generally an order of magnitude smaller than the smallest differences in target densities of adjacent layers of most HYCOM configurations. For pressures less than 200 dbar, where HYCOM's layers are generally isobaric and target potential densities vary in increments of approximately 0.5 kg/m^3, the rms potential density errors range from 0.04 kg/m^3 to 0.17 kg/m^3.

Potential density can be estimated directly using regression models[17] similar to those used for salinity, and salinity could be inferred from the measured temperature and pressure and the estimated potential density using the equation of state of sea water. Scatter plots of potential density vs. temperature (not shown) indicate that parabolic models are appropriate. Thus,

[16] Here, potential densities are referenced to the sea surface, *i.e.*, it is the density the water would have if it were adiabatically brought to the surface.

[17] Also, Stommel's method could be modified to estimate potential density (rather than salinity) by its mean on isothermal surfaces.

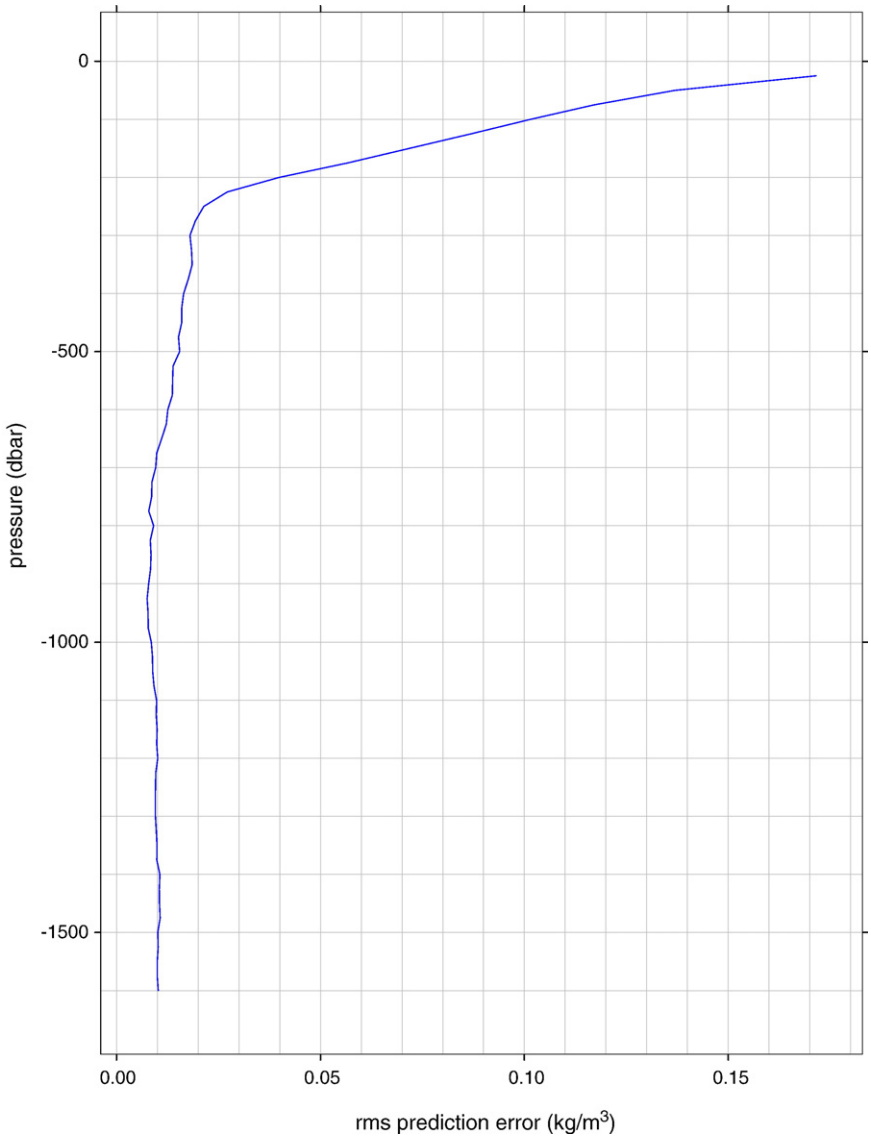

Fig. 20. Root-mean-square differences between verification data for potential density and their counterparts estimated using salinity values from the robust parabolic models together with the observed temperatures.

quadratic functions of temperature were robustly fit to the potential density data of the training set and used to approximate the verification data derived from observed salinity and temperature. The rms errors for the directly modelled potential density were almost exactly the same as for those based on the estimated salinity. The rms differences between the two estimates were less than 10^{-3} kg/m^3 at all pressure levels with the largest rms difference being at 25 dbar. The conclusion for the Gulf of Mexico is that it doesn't matter whether potential density is derived from an estimated salinity or whether it is estimated directly from observed temperature.

As estimates are made independently at each pressure level, it is interesting to ask whether the estimated potential density profiles are stable. For both types of estimated profiles — direct and via estimated salinity — only six inversions greater than 0.01 kg/m^3 over 25 dbar were encountered; for both types, the inversion involved precisely the same profiles at precisely the same levels, all in the hard-to-predict near-surface region. The largest was roughly 0.047 kg/m^3 between 25 dbar and 50 dbar levels, the others being less than half as strong. Except for the inversion extending over two levels from 100 dbar to 150 dbar with combined magnitude of

0.038 kg/m^3, the others were no deeper than 75 dbar. These inversions can easily be removed, and the salinity estimate can be corrected to be consistent with the equation of state.

9. Conclusion

For pressures greater than 200 dbar in the Gulf of Mexico salinity can be estimated from measurements of temperature with expected errors of no more than 0.05 psu. For pressures greater than 300 dbar they are only half this size, and for pressures greater than 800 dbar errors are halved again to the level set by the errors of the salinity data. Such accuracies are achieved using a robust fit of quadratic functions of temperature for each pressure level to data that are independent of those used for gauging the errors. Unfortunately, expected errors increase linearly with decreasing pressure to about 0.23 psu at 25 dbar with even greater errors expected at the surface.

Stommel's method gave almost identical root-mean-square errors for deep water, but was less accurate for pressures less than 350 dbar. As the work needed for preparing such a model is comparable to that required for a regression model and as that method lacks the flexibility needed for adding information beyond temperature that might improve its performance, it cannot be considered to be competitive. However, an approximate implementation of Stommel's method that can be constructed from available climatologies of temperature and salinity offers an expedient, although less accurate, alternative to regression as a stop-gap solution until regression models have been developed for other regions.

Salinity estimated with the robust quadratic function of temperature can be combined with observed temperatures to provide estimates of density. Near the sea surface where salinity estimates are poor, rms errors of potential density referenced to 0 dbar can be as large as 0.17 kg/m^3, but they rapidly decrease with depth. For pressures greater than 250 dbar, the rms errors are less than 0.02 kg/m^3; for pressures greater than 650 dbar, less than 0.01 kg/m^3. Estimating potential density directly as a robust quadratic function of temperature produced almost identical rms errors. Whichever approach is chosen, the results should be useful for correcting density when XBT data are assimilated into numerical ocean-circulation models.

Better estimates are still needed for both salinity and for potential density in the upper 200 m of the Gulf. No systematic variations with location or with season were exploitable. Even more surprising, given the opposite results of Hansen and Thacker (1999) for the eastern tropical Pacific, was the fact that surface salinity was uncorrelated with salinity as shallow as 25 m, ruling out the possibility that satellite-based measurements of surface salinity could help. Perhaps altimetric data carry information about the near-surface salinity for the Gulf of Mexico. As the near-surface estimation problem is much more difficult than that for the thermocline and below, for the sake of expediency, it has been deferred so that the project can advance as quickly as possible to provide deeper estimates for the entire world.

Two things have been learned here that might be exploited for other regions. First, it is possible to avoid spending too much time on separating the bad data from the good, a task that is complicated by the non-normal nature of the data distributions. For regions with as many data as the Gulf of Mexico, it is not so important if some good data are confounded with bad, so extreme outliers identified on univariate box-and-whisker plots can be assumed to indicate profiles to discard. Perhaps more important is to exploit multivariate relationships to identify bad data. In particular, bad data show up as outliers on *TS* scatter plots, so they might be identified as extreme outliers on box-and-whisker plots of residual of preliminary, smooth-curve fits to the data. Second, it is possible to work within relatively large geographical regions rather than being confined to $1° \times 1°$ or $5° \times 5°$ cells. More important is the fact that the Loop Current and its rings caused no problem; data from both sides of the thermal fronts could be handled with a single model. This result suggests that relatively large regions, even those with sharp fronts such as the region of the Gulf-Stream recirculation, can be modelled, thus allowing the project to proceed rapidly.

Acknowledgements

This work was supported by the National Oceanographic Partnership Program and by the Atlantic Oceanographic and Meteorological Laboratory.

References

Acero-Shertzer, C.E., Hansen, D.V., Swenson, M.S., 1997. Evaluation and diagnosis of surface currents in the National Centers for Environmental Prediction's ocean analyses. Journal of Geophysical Research 102, 21,037–21,048.

Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Modelling 37, 55–88.

Chassignet, E.P., Smith Jr., L., Halliwell, G.R., Bleck, R., 2003. North Atlantic simulations with the hybrid coordinate ocean model (HYCOM): impact of the vertical coordinate choice and resolution, reference density, and thermobaricity. Journal of Physical Oceanography 33, 2504–2526.

Conkright, M., Locarnini, R.A., Garcia, H.E., O'Brien, T., Boyer, T., Stephens, C., Antonov, J., 2002a. World Ocean Atlas 2001: Objective Analysis, Data Statistics, and Figures, CD-ROM Documentation. National Oceanographic Data Center Internal Report, vol. 17. NOAA.

Conkright, M., O'Brien, T.D., Boyer, T., Stephens, C., Locarnini, R.A., Garcia, H.E., Murphy, P.P., Johnson, D., Baranova, O., Antonov, J.I., Tatusko, R., Gelfeld, R., 2002b. World Ocean Database 2001. National Oceanographic Data Center Internal Report 16, NOAA, cD-ROM Data Set Documentation.

Donguy, J.R., Eldin, G., Wyrtki, K., 1986. Sealevel and dynamic topography in the western Pacific during 1982–1983 El Nino. Tropical Ocean-Atmosphere Newsletter 36, 1–3.

Flierl, G.R., 1978. Correcting expendable bathythermograph (XBT) data for salinity effects to compute dynamic heights in Gulf Stream rings. Deep-Sea Research 25, 129–134.

Fofonoff, N.P., 1977. Computation of potential temperature of seawater for an arbitrary reference pressure. Deep-Sea Research 24, 489–491.

Halliwell Jr., G.R., 2004. Evaluation of vertical coordinate and vertical mixing algorithms in the hybrid-coordinate ocean model HYCOM. Ocean Modelling 7, 285–322.

Hansen, D.V., Thacker, W.C., 1999. On estimation of salinity profiles in the upper ocean. Journal of Geophysical Research 104, 7921–7933.

Kessler, W.S., Taft, B.A., 1987. Dynamic heights and zonal geostrophic transports in the central tropical Pacific during 1979–1984. Journal of Physical Oceanography 17, 97–122.

Lozier, M.S., Owens, W.B., Curry, R.G., 1995. The climatology of the North Atlantic. Progress in Oceanography 36, 1–44.

R Development Core Team, 2004. R: A Language and Environment for Statistical Computing. R. Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-00-3. http://www.R-project.org.

Reynolds, R.W., Ji, M., Leetmaa, A., 1998. Use of salinity to improve ocean modeling. Physics and Chemistry of the Earth 23, 543–553.

Stommel, H., 1947. Note on the use of the T–S correlation for dynamic height anomaly calculations. Journal of Marine Research VI, 85–92.

Thacker, W.C., Esenkov, O.E., 2002. Assimilating XBT data into HYCOM. Journal of Atmospheric and Oceanic Technology 19 (5), 709–724.

Thacker, W.C., Sindlinger, L., 2007. Estimating salinity to complement observed temperature: 2. Northwestern Atlantic. Journal of Marine Systems 65, 249–267 (this issue). doi:10.1016/j.marsys.2005.06.007.

Thacker, W.C., Lee, S.-K., Halliwell Jr., G.R., 2004. Assimilating 20 years of Atlantic XBT data into HYCOM: a first look. Ocean Modelling 7, 183–210.

Troccoli, A., Haines, K., 1999. Use of the temperature–salinity relation in a data assimilation context. Journal of Atmospheric and Oceanic Technology 16, 2011–2025.

Troccoli, A., Balmaseda, M.A., Segschneider, J., Vialard, J., Anderson, D.L.T., Haines, K., Stockdale, T., Vitart, F., Fox, A.D., 2002. Salinity adjustments in the presence of temperature data assimilation. Monthly Weather Review 130, 89–102.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S-Plus. Springer–Verlag, New York.

Vossepoel, F.C., Behringer, D.W., 2000. Impact of sea level assimilation on salinity variability in the western equatorial Pacific. Journal of Physical Oceanography 30, 1706–1721.

Vossepoel, F.C., Reynolds, R.W., Miller, L., 1999. Use of sea level observations to estimate salinity variability in the tropical Pacific. Journal of Atmospheric and Oceanic Technology 16, 1401–1415.

Wong, A.P.S., Johnson, G.C., Owens, W.B., 2003. Delay-mode calibration of autonomous CTD profiling float salinity data by θ–S climatology. Journal of Atmospheric and Oceanic Technology 20, 308–318.