Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ocean Modelling 7 (2004) 183–210
Ocean Modelling
www.elsevier.com/locate/ocemod

Assimilating 20 years of Atlantic XBT data into HYCOM: a first look

W.C. Thacker [a,*], S.-K. Lee [b], G.R. Halliwell Jr. [c]

[a] *Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami, FL 33149, USA*
[b] *Cooperative Institute for Marine and Atmospheric Studies, 4600 Rickenbacker Causeway, Miami, FL 33149, USA*
[c] *Rosensteil School of Marine and Atmospheric Science, 4600 Rickenbacker Causeway, Miami, FL 33149, USA*

Abstract

Expendable bathythermographic (XBT) data for the years 1972–1991 have been assimilated into a Hybrid Coordinate Ocean Model (HYCOM) for the Atlantic Ocean. Climatological salinity profiles were combined with the observed temperature profiles to estimate companion potential-density profiles, which are used to determine the observation-based local structure of the model's hybrid layers. The model's density, temperature, and layer-interface-depth fields were corrected monthly via optimal interpolation. Preliminary results presented here show that the data have a major impact on the simulation, correcting model biases, and that the corrections persist between monthly assimilations.
Published by Elsevier Ltd.

Keywords: Oceanic data assimilation; Estimation theory; Salinity; HYCOM

1. Introduction

Thacker and Esenkov (2002) have described a method for assimilating XBT data into the Hybrid Coordinate Ocean Model (HYCOM). The model depicts the ocean as fluid layers, which are isopycnic when and where the ocean is stratified, terrain-following in shallow water, and isobaric otherwise (Bleck, 2002; Chassignet et al., 2003; Halliwell, 2003). When assimilating the XBT data the local nature of the layers must be determined from the observed temperature profiles so that both layer thicknesses and layer densities be corrected in addition to the correction of layer temperatures. Climatological estimates of salinity are combined with observed temperature to infer

* Corresponding author. Tel.: +1-305-361-4323; fax: +1-305-361-4392.
E-mail addresses: carlisle.thacker@noaa.gov (W.C. Thacker), sang-ki.lee@noaa.gov (S.-K. Lee), grh@rsmas.miami.edu (G.R. Halliwell Jr.).

1463-5003/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.ocemod.2003.08.002

companion profiles of potential density, which are analyzed into layers with the model's isopycnic targets; then layers for which no target water has been observed are assigned the minimum thicknesses allowed by the model. Layer averages of potential temperature and potential density and the depths of the layer interfaces provide the data for correcting the model state. After these data are assimilated via optimal interpolation, salinity is diagnosed from the equation of state and currents are adjusted so that velocity corrections geostrophically balance pressure corrections.

While Thacker and Esenkov (2002) have shown that the assimilation method can produce reasonable static multivariate analyses, they did not demonstrate that these analyses can be used as initial conditions for model integrations. Here we address that issue by assimilating XBT data for years 1972 through 1991 into a non-eddy-resolving configuration of HYCOM for the Atlantic Ocean to provide monthly corrections to a simulation forced by surface fluxes of heat, moisture, and momentum for this 20-year period. We have found that the largest corrections implied by the data can lead to numerical instability, but when these few data are excluded the model behaves well. Even so, there are still unreasonably large differences between the data that were assimilated and their model counterparts, which indicate problems with both the model and the data. As HYCOM has evolved substantially beyond the version used for this study, [1] it is not appropriate to spend too much effort characterizing and correcting the model-related errors encountered here. Understanding and eliminating errors in the data is an important but time-consuming task that will be the subject of a separate project. Here we only provide a preliminary examination of these errors in order to establish the need for that work; in particular, we find that better estimates of salinity are needed in order to estimate density from the XBT data. In spite of problems with both the model and the data, we are able to show that the assimilation method provides reasonable multivariate corrections and that there is only a weak tendency for the model to revert to its uncorrected state.

After discussing the details of the model configuration, the data, and the assimilation methodology in Sections 2–4, results of the 20-year simulation are presented in Section 5. A full examination of all of the model's evolving three-dimensional fields is neither practical nor appropriate, given the preliminary nature of the results. Distributions of innovations are presented to show that the nature of model-data differences. Impact of assimilation on the $\sigma_0 = 27$ kg/m^3 potential-density-anomaly surface is shown to provide a horizontal view, and impact along 25° W and on the zonally averaged meridional overturning circulation provide vertical views. The spatial scales of the corrections and the model's month-to-month response is presented in terms of potential density on the 500 dbar pressure surface. While our results demonstrate that assimilating XBT data can significantly improve HYCOM simulations, it is possible to do even better. Plans for improving the data-assimilation methodology are discussed in Section 6.

2. Model configuration

Bleck (2002) describes the architectural aspects of HYCOM, emphasizing the treatment of the model's hybrid vertical coordinate, which distinguish HYCOM from its isopycnic predecessor,

[1] Version 1.0 was used here. At present, version 2.1 is available.

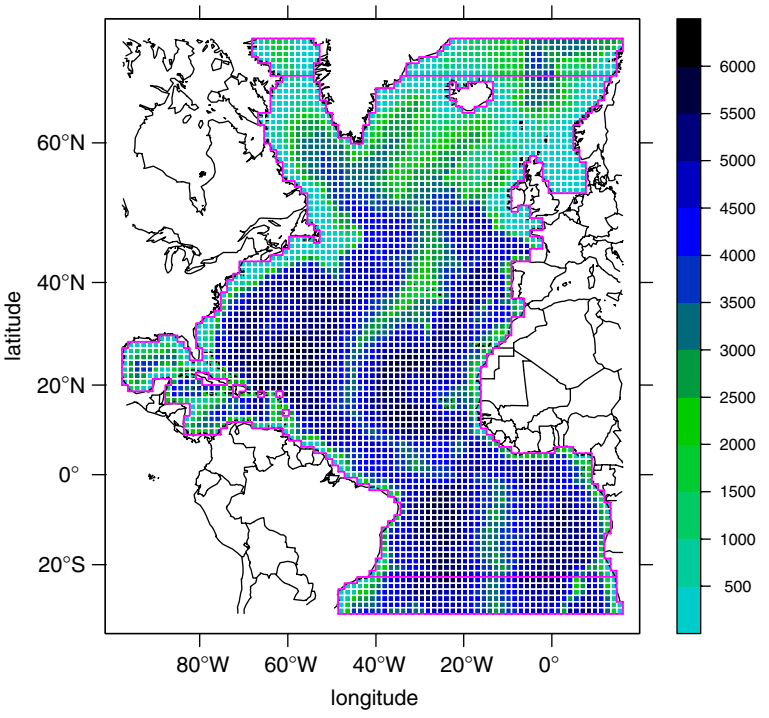

Fig. 1. Mercator projection of 1.4°-HYCOM grid with map overlay and with colors indicating cell depths in meters. The magenta lines at 22.6° S and 66.8° N indicate the boundaries of the computational "sponge" layers.

the Miami Isopycnic Coordinate Ocean Model (MICOM); more details are provided by Halliwell (2003) and by Chassignet et al. (2003). The configuration used in this study is exactly the same as used in two recent Atlantic climate studies (Halliwell et al., 2003). The model domain extends from 30° S to 70° N and from 98° W to 16° E as illustrated in Fig. 1. The spherical computational grid is designed to have approximately square cells, which appear to be uniform in size in a Mercator projection. Cell size decreases with latitude, reflecting the convergence of the meridians; at the equator their latitudinal extent is 1.4° and at 60° N it is 0.7°, while their longitudinal extent is 1.4° everywhere. The horizontal resolution of the model grid determines what features can be simulated. With this grid eddies cannot be handled and fronts will not be as sharp as they should be. Still, this resolution should be adequate for our goal of assimilating 20 years of XBT data to get the month-to-month evolution of the Atlantic circulation.

Flow through cuts connecting the Atlantic to the Caribbean Sea and the Gulf of Mexico is allowed but not well resolved; there is no flow through the English Channel. The model's northern and southern boundaries are treated as vertical walls; to compensate, temperature and salinity are relaxed to their climatological values (Levitus and Boyer, 1994; Levitus et al., 1994) within zones six cells wide adjacent to the walls. Mediterranean outflow is neglected, as is river runoff.

Sea ice is represented via an energy-loan model (Bleck, 2002) similar to that of Semtner (1976). Vertical mixing is modelled [2] by the so-called KPP scheme of Large et al. (1994). Horizontal mixing is parameterized as an average of Laplacian and biharmonic mixing.

The bathymetry (Fig. 1) can limit the number of active layers in a given grid cell. Because the deeper layers are isopycnic, they only exist where water of their target densities is found. [3] Lower layers collapse to zero thickness when there is no model water sufficiently dense locally to allow these layers to obtain their target values. Where there is no sufficiently light water to allow the upper layers to obtain their target values, e.g., at high latitudes, layer thickness is prescribed. [4] The transition to depth-proportionate coordinates in shallow water is achieved by re-specifying the minimum allowed layer thicknesses based on having at least 15 active layers: minimum allowed thickness is reduced to one-fifteenth the local depth. [5] Thus, the vertical resolution depends on the nature of the model state and varies both spatially and temporally. [6]

Starting from rest and from zonally averaged climatological distributions of temperature and salinity, the model was forced for 20 years with monthly [7] climatological mean wind stress, heat flux, [8] and moisture flux [8] to get an approximate initial state. Forcing fields were computed from the National Centers for Environmental Prediction's reanalyses (Kistler et al., 2001) of wind stress and wind speed, atmospheric temperature and specific humidity at 2 m above the surface, friction velocity, net radiative heat flux (long-wave and short-wave), and precipitation. Evaporation and turbulent heat flux are modelled using the bulk-formula representation with the Kara et al. (2000) parameterization of the turbulent drag coefficients. The model state at the end of this 20-year spin-up provided initial conditions for the simulations.

From January 1948 the model was forced more realistically by the actual monthly surface-forcing fields, rather than by their climatological means, to simulate conditions through 1991. This study starts with the model state at the beginning of 1972. The simulation continues for 20 more years without data assimilation to provide a basis of comparison. The simulation with XBT data for each month assimilated at mid-month starts with the same January 1972 state and is forced with the same surface fluxes.

[2] Halliwell (2003) evaluates alternative vertical mixing algorithms that are available to HYCOM.

[3] In units of kg/m^3, target values of potential-density anomaly σ_0 referenced to the air–sea surface for this 25-layer configuration of HYCOM are: 21.20, 21.60, 22.00, 22.40, 22.90, 23.40, 24.02, 24.70, 25.28, 25.77, 26.18, 26.52, 26.80, 27.03, 27.22, 27.38, 27.52, 27.63, 27.71, 27.78, 27.85, 27.92, 27.99, 28.06, 28.12. Potential density is potential-density anomaly plus 1000 kg/m^3.

[4] This configuration of HYCOM has minimum layer thicknesses that increase with depth. In units of 10^4 Pa (decibars), they are: 6.00, 6.84, 7.80, 8.89, 10.13, 11.55, 13.17 for layers 1–7 and 15.00 for layers 8–25. With the specified values of target densities, layers 1–4 always have their minimum thicknesses everywhere.

[5] The layer's thickness is taken to be the smaller of its deep-water minimum and the depth-proportionate value. Because the model's water depth is never less than 100 m, the minimum thickness of the surface layer is already less than the depth-proportionate value and never has to be respecified.

[6] Details of the algorithm for updating the structure of the model's layers are provided by Bleck (2002) and by Halliwell (2003).

[7] All model months have 30.5 24-h days. The surface-forcing fields, which are based on data from January 1948 through December 2000, are interpolated between months.

[8] Temperatures and salinity in the surface layer were relaxed to their climatological values (Levitus and Boyer, 1994) with a relaxation time of one month.

While HYCOM is at heart based on isopycnic coordinates but has been started from a z-coordinate climatological state, it is useful to compare its mean behavior with an isopycnic climatology. Fig. 2 shows a comparison of the model simulation with a climatology computed using Hydrobase [9] software and data (Lozier et al., 1995). Box and whisker plots for 10° bands of latitude indicate the distribution of the differences between simulated and climatological pressure (depth) of nine isopycnic surfaces, [10] which correspond to the target potential-density anomalies of model layers 11–19. Differences were computed at each point where there are data to be assimilated during the 20-year interval and where the isopycnic surface exists for both model and climatology. The tendency toward leftward displacement of the median indicators from the vertical zero-difference line shows the model's bias, preferring deeper than climatological isopycnic surfaces. Weaker stratification at depth results in larger uncertainties as where the model's layer interfaces should be; this is reflected in the generally larger spread of the majority of differences for the denser layers. However, the extremely wide range of differences, where the simulation and climatology might disagree by over 1000 m about how deep the water should be, reflects the model's dynamically varying layer structure.

3. Data

The XBT data used for this study are from the Atlantic Oceanographic and Meteorological Laboratory's Global Ocean Observing System (GOOS) Center where they were carefully examined and their problems were flagged. Independently, similar checking was done by Canada's Marine Environmental Data Service (MEDS). Those deemed acceptable by both were selected for use. A small fraction of the data had not been examined by both centers; of those, data acceptable to one or the other were used. In addition, we screened the data further to remove duplicate casts, those with temperatures that were unphysically hot or cold, and those that differed from the climatological mean by more than three standard deviations. [11] Data that were near to the boundaries of the model grid but not within any active grid cells were not used, even though they might provide some useful information, just as data within the grid boundaries are used to correct conditions in neighboring cells. Such restrictions were made for ease of implementation. When time allows, the selection and screening of the data can be improved. Nevertheless, we feel that we

[9] The complete set of "raw" CTD and bottle data for the North and South Atlantic from the Hydrobase web site (http://www.whoi.edu/science/PO/hydrobase) were used. The codes for averaging the data on isopycnic surfaces were also obtained from this web site.

[10] The potential densities for the model's layers were assigned to the layer's mid-depth, where the pressure was taken to be the average of the pressures for the layer's upper and lower interfaces, and the pressure on the isopycnic surfaces were found by interpolation.

[11] Some of the data that were discarded as being unlikely on the basis of climatology seemed to be quite reasonable when examined collectively as a single cruise. This indicates that assuming data to be normally distributed and thus unlikely to have fluctuations of more than three standard deviations might not provide an adequate basis for screening and that more elaborate procedures are needed. For example, Du and Tong (2001) and Kadanoff (2001) discuss exponentially distributed fluid-dynamic measurements exhibiting relatively frequent six-standard-deviation departures from the mean.

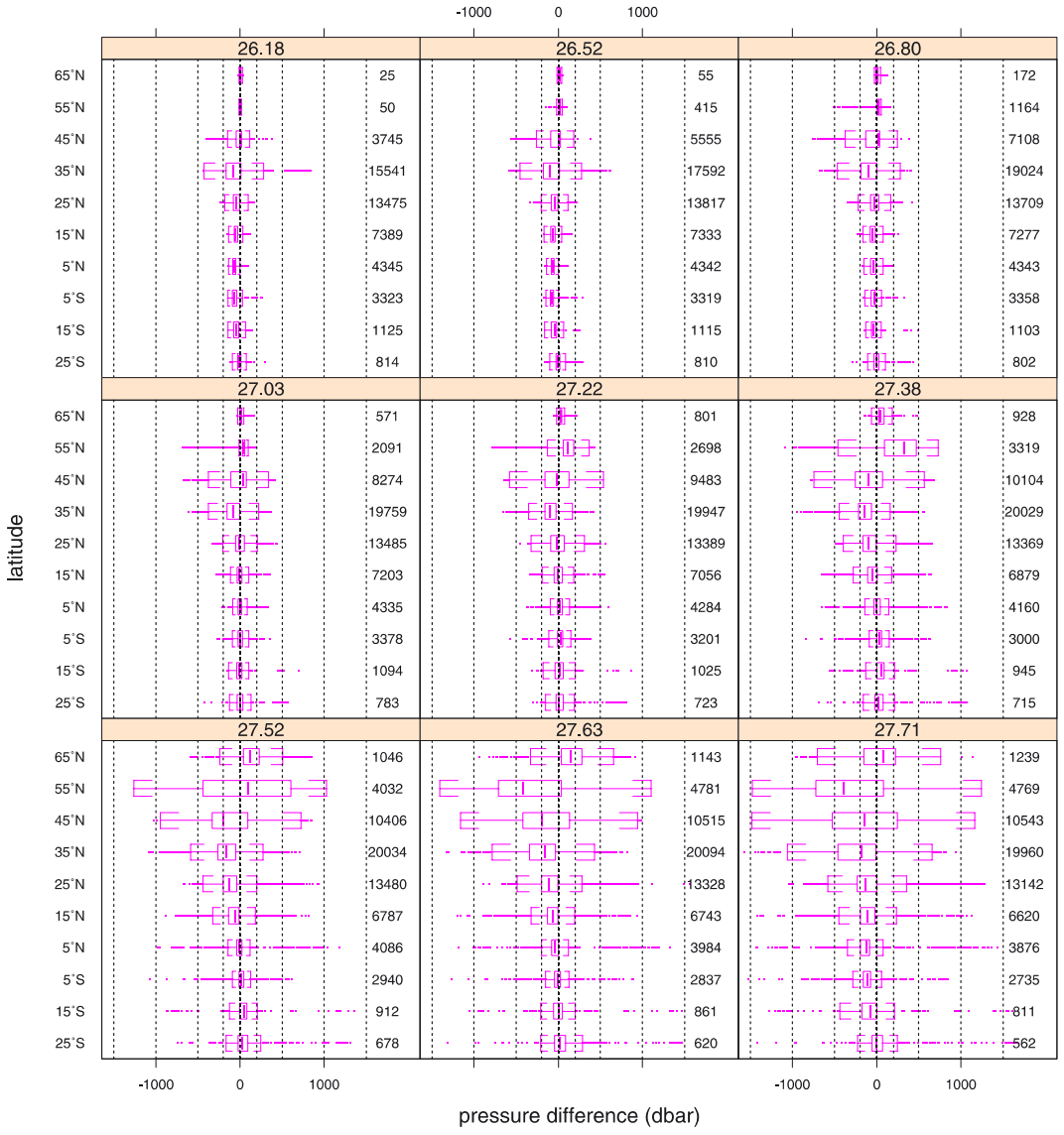

Fig. 2. Box and whisker plots showing distribution of Hydrobase–HYCOM differences in pressure of isopycnic surfaces for the HYCOM simulation without data assimilation. The central bars indicate the median differences, the boxes extend between the first quartile and the third quartile, the whiskers extend up to 1.5 times the interquartile range, and the dots beyond the whiskers indicate outliers. The number of samples for each plot is indicated to its right. Vertical guide lines indicate 0, 200, 500, 1000, and 1500 dbar differences. Panel labels indicate the potential density anomaly for the isopycnic surfaces in kg/m^3.

have assembled a large body of clean data that provide information for correcting the model's simulation of 20 years of Atlantic circulation.

Fig. 3 illustrates the distribution of the data from 1972 that are available to be assimilated. It is clear that the bulk of the data lie along shipping routes. Many of the model's 3995 grid cells are

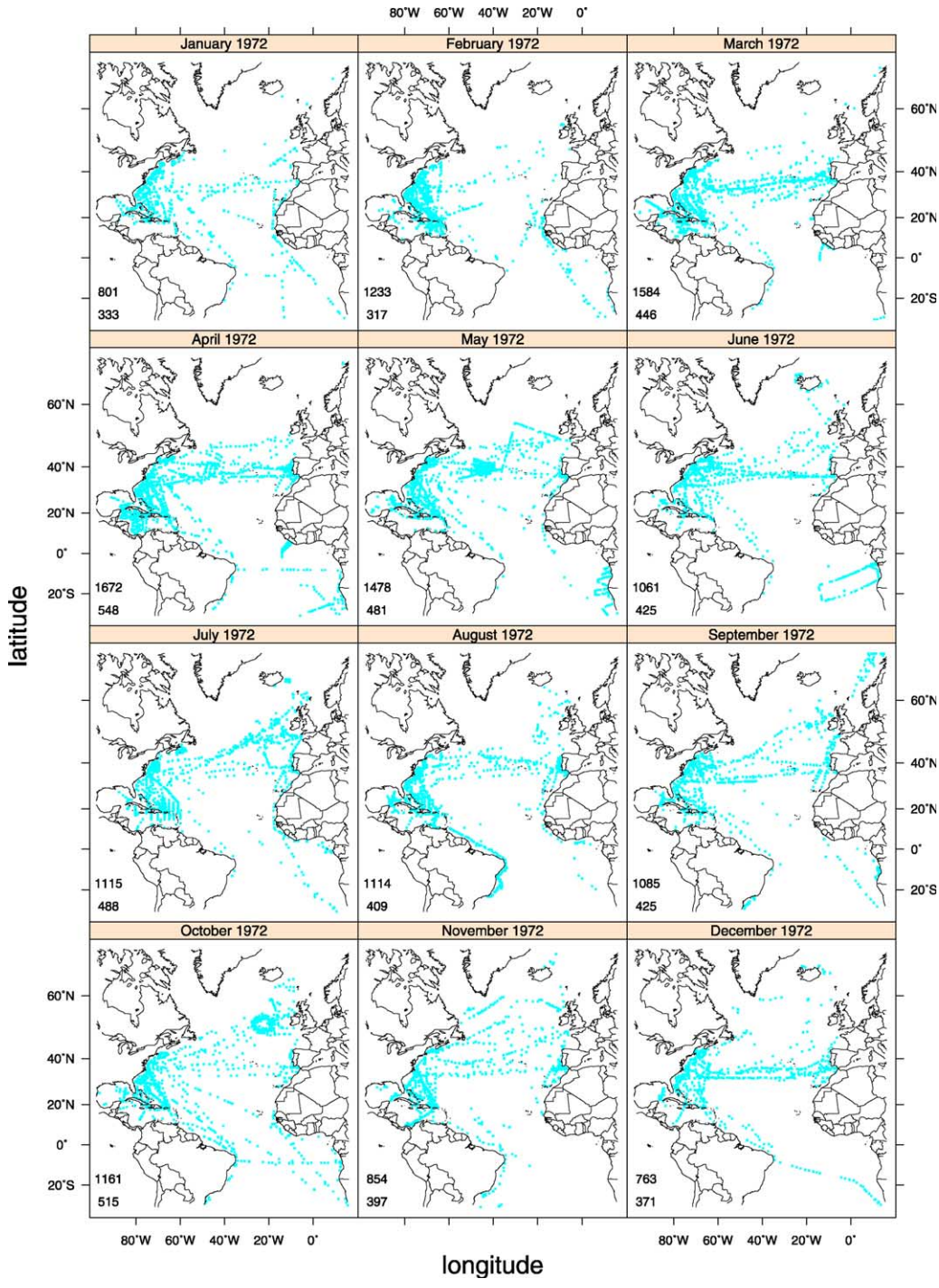

Fig. 3. Cyan dots mark locations of XBT casts within the domain of the 1.4-° HYCOM grid by month for the year 1972. The numbers in the southwest corner of each panel indicate how many casts (upper) were available for assimilation for that month and how many of the model's 3995 grid cells (lower) were sampled.

Table 1
Numbers of XBT casts available for assimilation in region of model grid by year and month

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1972	801	1233	1584	1672	1478	1061	1115	1114	1085	1161	854	763
1973	644	1333	893	873	1645	2075	1741	1402	1590	1270	1092	952
1974	671	1094	1358	931	1298	1272	1134	1372	2072	996	773	785
1975	657	674	1173	1190	1258	1231	1352	708	771	1300	1340	697
1976	654	930	1206	935	724	910	655	754	777	1285	956	892
1977	1033	865	719	1205	1366	877	1417	1229	1420	2447	1512	1233
1978	609	1169	1116	1180	2048	1052	1050	1179	960	786	734	679
1979	543	948	1020	764	1236	939	933	937	912	1765	1257	401
1980	431	500	922	914	877	667	873	552	1238	1015	799	563
1981	734	584	854	735	1211	692	1154	939	1467	1115	1126	623
1982	416	610	826	1172	1191	1346	862	1042	788	846	899	397
1983	528	766	1068	706	972	955	896	732	685	857	856	414
1984	502	841	760	826	1197	867	844	828	819	1000	925	456
1985	591	788	994	606	992	1028	580	895	751	1040	919	741
1986	1173	971	799	820	955	1253	708	868	584	386	592	209
1987	363	825	680	651	966	859	668	709	835	565	706	264
1988	349	407	495	397	660	624	554	706	473	500	445	271
1989	177	283	318	343	487	338	285	338	237	183	161	121
1990	437	924	834	846	1061	1280	1178	1314	1220	1245	1149	984
1991	1213	1175	1180	979	985	1304	1141	995	1114	891	895	748

not observed during a given month, while other cells have multiple observations. Because our objective is to simulate the monthly mean dynamical state of the ocean, observed differences within individual cells are ignored and the average [12] of the data are used to characterize the cells. Counts of both the XBT casts and the cells for which information is provided are shown on each panel. In general the sampling seen in 1972 can be said to be typical, as data for other years are also concentrated along shipping routes. Counts for all years of this study are shown in Table 1.

XBT data alone are insufficient to determine how the model's density structure should be corrected. Supplemental information about salinity is needed. Thacker and Esenkov (2002) discuss some options for estimating companion salinity profiles, which together with the XBT data provide profiles of potential density that are needed for determining the structure of the model's layers. The simplest option is chosen here, namely, to assume the salinity to have its local climatological mean value. For each XBT profile the climatological salinity for that location and that month was interpolated to the depths where temperatures were observed to estimate a companion salinity profile for the XBT data, and the equation of state was then used to estimate companion profiles of potential temperature and potential density. [13] Errors for the derived

[12] See footnote 15 below.

[13] By specifying appropriate covariances for background errors of temperature with density, it is possible in principle to prescribe corrections for the model's density field without using companion salinity profiles. However, cross-field error covariances are difficult to estimate for any model, and casting them in terms of HYCOM's hybrid layers makes it even more difficult. Furthermore, companion density profiles can be expected to be more effective for correcting problems associated with modelling bias.

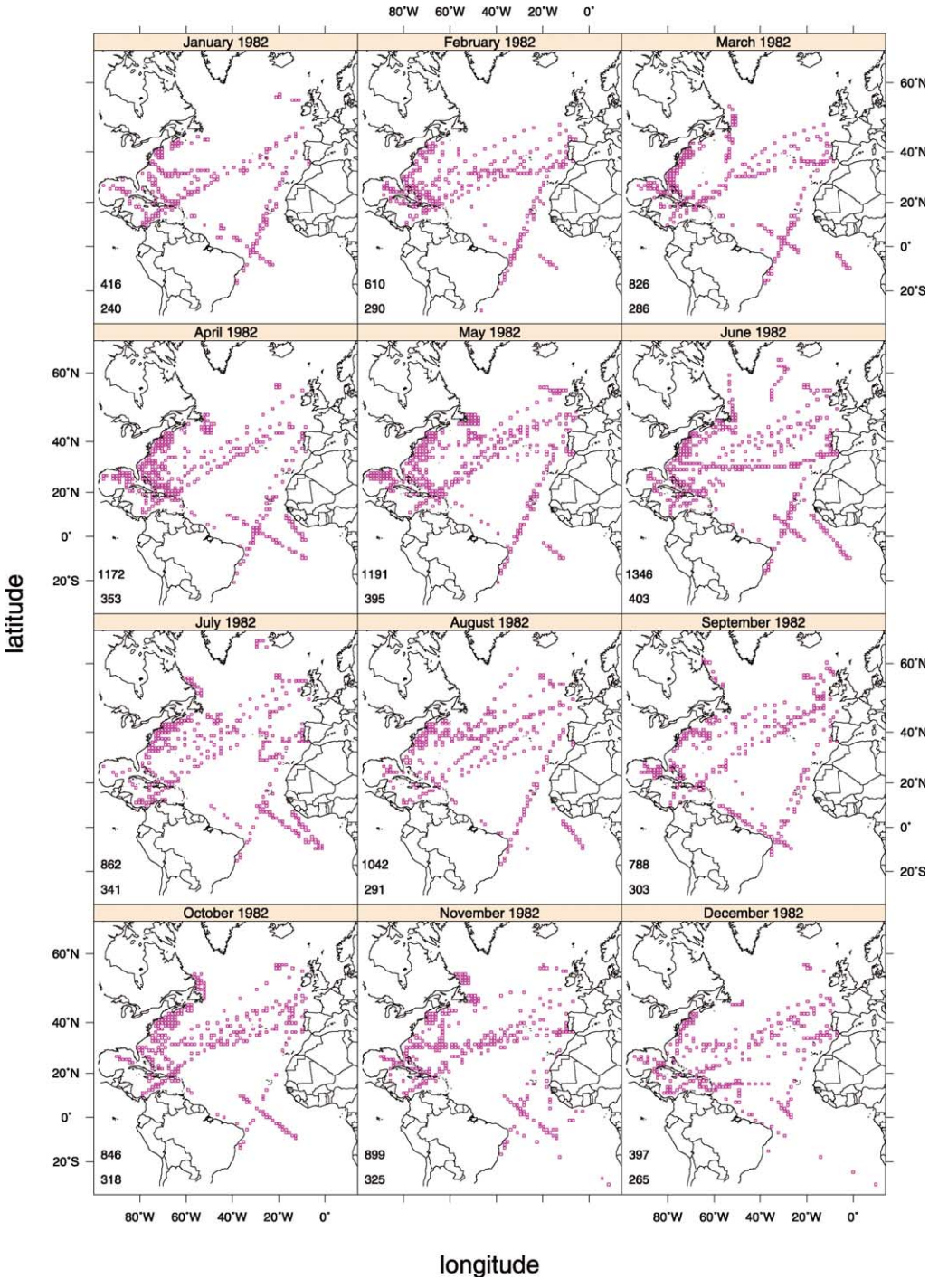

Fig. 4. Magenta squares mark locations of grid cells with data to be assimilated by month for the year 1982. The numbers in the southwest corner of each panel indicate the number of casts (upper) for the month and the number of superobs (lower) available for assimilation.

Table 2
Numbers of superobs by year and month

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1972	333	317	446	548	481	425	488	409	425	515	397	371
1973	284	414	431	410	511	681	557	515	570	476	458	309
1974	330	443	468	362	343	465	473	435	452	386	268	263
1975	234	235	389	343	392	386	454	310	333	448	402	329
1976	306	307	248	342	293	357	335	301	316	426	394	242
1977	372	358	301	274	403	366	466	254	347	489	418	375
1978	231	311	304	315	404	302	360	284	343	284	206	218
1979	250	233	292	252	324	273	293	219	306	393	276	202
1980	204	193	319	323	284	287	292	175	416	319	324	265
1981	220	247	272	264	318	291	363	403	421	406	391	310
1982	240	290	286	353	395	403	341	291	303	318	325	265
1983	237	339	394	279	330	324	354	253	273	307	293	201
1984	254	305	296	323	358	296	321	338	342	393	411	243
1985	216	261	364	281	288	335	239	339	380	470	329	283
1986	275	293	261	254	227	324	275	344	273	230	249	116
1987	163	215	235	218	251	340	287	269	318	220	218	126
1988	166	167	245	167	228	194	202	339	175	187	206	124
1989	142	157	142	182	250	205	162	191	118	76	129	99
1990	282	432	442	401	450	485	537	566	559	579	545	428
1991	511	471	537	454	454	511	467	368	416	440	465	377

potential density and potential temperature profiles are estimated from the climatological variability of salinity and temperature as described by Thacker and Esenkov (2002).

The profile from an XBT cast might characterize the interior of an eddy or conditions on one side of a front, and without further information [14] there is no way of knowing. At present no attempt has been made to determine the extent to which each cast might characterize the monthly mean conditions for its grid cell. All data are treated as being equally representative of what the model is simulating. Because observed within-cell differences cannot be represented by the model and because of the sensitivity of the assimilation computations to such differences, the data [15] within each cell are averaged to form *superobs*. No effort has been made to prevent several casts that are very close spatially and temporally from dominating a superob at the expense of less redundant data. Such improvements are not fundamentally difficult, but they do take time.

Fig. 4 illustrates the spatial distribution of superobs by month for 1982. The fraction of cells that have been observed can be obtained by dividing the number of observed cells, indicated in the lower-left corner of each panel, by 3995, which is the total number of active (water) cells. The number of superobs available for assimilation by year and month is indicated in Table 2.

[14] It may be possible to infer which side of a front a profile characterizes by comparing the data to climatologies of the two sides of fronts, but such climatologies have not yet been compiled. Useful information about local features might also be obtained from satellite data.

[15] The superobs are averages of the companion potential density and potential temperature profiles before they have been further preprocessed to determine the local structure of the model's layers. The profiles within each grid cell are first interpolated to a common set of depths and are then averaged.

Before the profiles can be assimilated, they must be reduced to information about the thickness of the model's layers and their local potential density and potential temperature. Details of this preprocessing are given by Thacker and Esenkov (2002). First, positions of the layer interfaces are determined by fitting to the potential-density profile a piecewise-constant function having the prescribed target values. The resulting surface layers have zero thickness when there is no water with their target densities. Then, starting from the top, these layers are inflated to their minimal allowed thickness. After the interfaces have been determined, the layer averages of potential density and potential temperature are computed from the profile data. Finally, estimates of the errors of the layer-specific data are estimated from the climatological uncertainties assigned to the superob profiles. These derived data are what are to be assimilated into HYCOM.

While our intent is to pre-determine which data are meaningful, at this stage of development, because of large model-data differences, we found it necessary to screen the data further by comparing them to the model state during the assimilation process. This is not ideal, as the data should be used to correct the model state, not the other way around. For example, in the vicinity of the Gulf Stream, one part of the problem is the model's tendency to have the front too far north, while another is our failure to account for covariability of temperature and salinity during preprocessing. Fig. 5 shows an example from August 1972 of the sort of data that were skipped. The XBT cast indicates unusually cold water for this location (39.9° N, 46.4° W), so the density could be assumed to be higher than average. Being exceptionally cold, it can also be expected to be

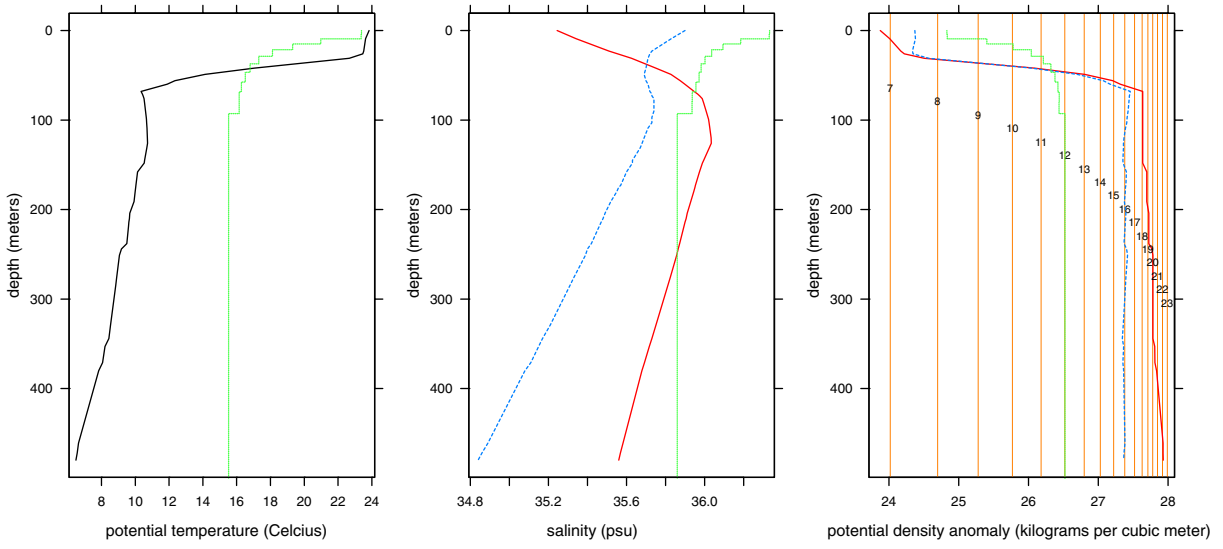

Fig. 5. Left panel: solid black curve is observed potential temperature; dotted green curve is potential temperature from HYCOM simulation without data assimilation. Center panel: solid red curve is climatological salinity; dashed blue curve is salinity estimated from observed potential temperature by MODAS; dotted green curve is salinity from HYCOM simulation without data assimilation. Right panel: solid red curve is potential density based on observed temperature and climatological salinity; dashed blue curve is potential density based on observed temperature and MODAS salinity; dotted green curve is potential density from HYCOM simulation without data assimilation; vertical orange lines indicate target values for HYCOM's layers; locations of numbers labeling vertical lines indicate minimum depth at which layers can be isopycnic.

fresher than usual, so the use of climatological mean salinity *over estimates* the density. The HYCOM simulation without data assimilation (control run) has warmer water than usual in this cell. [16] The salinity profile estimated [17] using TS relationships from the Navy's Modular Ocean Data Assimilation System (Fox et al., 2002) gives a more realistic potential-density profile, which has layer 16 at 450 m rather than layer 22. This suggests that, by using better estimates of salinity during the preprocessing, more density profiles would not have such large model-data differences and fewer XBT data would have to be skipped.

4. Assimilation methodology

Once superob profiles have been preprocessed to reflect the local nature of the model's layers, they provide model-appropriate data for correcting the simulation. For each observed cell, there is information about the pressure (depth) of layer interfaces, as well as the potential density and potential temperature of the layers from the air–sea surface to the bottoms of the profiles. [18] The model-appropriate data are assimilated layer-by-layer and field-by-field using the method of optimal interpolation as described by Thacker and Esenkov (2002). Only two of potential temperature θ_0, potential density σ_0, and salinity S are independent, as the third can be diagnosed using the equation of state. During the model simulation, potential temperature and salinity evolve independently and potential density is diagnosed. On the other hand, during assimilation, θ_0 and σ_0 are corrected directly and the corrected salinity field is diagnosed. After assimilating layer-interface pressures, adjustments are made to ensure that the minimum allowed thicknesses of the model's layer are not violated and that no interfaces lie below the sea floor. [19] Corrections to the velocity field are computed geostrophically [20] from the corrections to the baroclinic Montgomery potential implied by the corrections to pressure and potential density. [21]

One departure from the description by Thacker and Esenkov (2002) is the treatment of pressure. Instead of assimilating layer thickness, we have chosen to assimilate depths of layer interfaces. Assimilating layer thicknesses was most easily accomplished by preserving thicknesses of all unobserved layers except for the deepest, while assimilating interface depths had the most dramatic impact on the shallowest unobserved layer. Difficulties associated with large model-data differences, such as illustrated in Fig. 5 which cause large vertical displacements of water masses were less dramatic when interface depths were assimilated. The question of what to do below the

[16] When data were assimilated for January through July, 1972, the model water in this cell was not quite this warm and light, as corrections had moved the Gulf Stream toward the south.

[17] Thanks to Dan Fox for the MODAS salinity profile.

[18] The depth to which XBT probes provide data varies; for most probes, it is between 450 and 750 m. Which model layers an XBT samples is determined during the preprocessing.

[19] The corrections are made independently in each grid cell by increasing the interface depth, starting from the top, to guarantee minimum thickness and raising the lower interfaces until they coincide with the sea floor.

[20] These corrections are not applied within 20° of the equator.

[21] When pressure, density and temperature corrections were spread over the entire month rather than at mid-month (Bloom et al., 1996) and no velocity corrections were made, relying instead on the model's ability to dynamically adjust to the minor changes in the mass field, results were quite similar. Because of its computational economy, we chose to have the corrections lumped at a single time.

profiles needs further attention. In some cases it is reasonable to expect that the unobserved layer structure should be preserved just below the data; in other cases it can be equally reasonable to think that conditions seen at the bottom of the profile require adjustments to the next deeper layer. If this problem were not further complicated by the fact that the profiles are of differing lengths, it would be analogous to that of assimilating satellite altimetric data (Cooper and Haines, 1996).

The change to assimilating interface pressures rather than layer thicknesses required estimates of uncertainty of the interface pressures, which lead to a modification of the Thacker and Esenkov (2002) estimates of uncertainty of the layer densities. As before, the idea is that the uncertainty in interface pressure should be small when the layer is isobaric and uncertainty in potential density should be small when the layer is isopycnic. Both are related to the climatological standard deviation of potential density SD (σ), which is estimated as before from the climatological standard deviations for temperature and salinity:

$$\begin{aligned} \sigma_\sigma &= \sqrt{(\min \sigma_\sigma)^2 \sin^2 \psi + (\max \sigma_\sigma)^2 \cos^2 \psi} \\ \sigma_p &= \sqrt{(\min \sigma_\mathrm{p})^2 \cos^2 \psi + (\max \sigma_\mathrm{p})^2 \sin^2 \psi} \\ \tan \psi &= \frac{|p - p_{\min}|/p_{\min}}{|\sigma - \sigma_\mathrm{T}|/\sigma_\mathrm{T}} \end{aligned} \tag{1}$$

where σ_σ denotes the standard deviation of the errors of a layer's potential density and σ_p denotes the standard deviation of the errors of the pressure of the layer's lower interface. The maximum value for the standard deviation of potential density is the climatological standard deviation SD (σ) interpolated to the middle of the layer; the minimum value is taken as 0.001 kg/m^2. The minimum value of the standard deviation of the interface pressure is taken to be the pressure equivalent of 1 cm; the maximum is estimated by multiplying SD (σ) by the slope [22] of the density profile $\mathrm{d}p/\mathrm{d}\sigma_0$, but is limited so that it cannot exceed the pressure at the bottom of the water column. The angle ψ is determined by the relative departure of the interface pressure p from $p_{\min}$, the minimum value allowed by the model configuration for this layer, and that of the potential density σ_0 from its target value σ_T.

Background error covariances are much as described by Thacker and Esenkov (2002). Covariances between errors of the model state in different grid cells are exponential functions of their separation with the width chosen to reflect the model's resolution, so observations have their influence limited to the span of a few grid cells. The amplitude of the exponential, which varies with layer, is chosen [23] so that the background error variance will be less than the observational error variance assigned to the relatively few observations that are far from the climatological mean but greater than the observational error variance for the bulk of the data.

[22] When computing the slope, the denominator was taken as the larger of the difference in potential density at the upper and lower interfaces or the value 0.01 kg/m^2.

[23] The background error covariance, which is constant for the entire domain, is set to the third quartile value of the observational error covariances, except for interface-pressure errors; in that case it is set to the third quartile of the observational error covariances for pressures greater than the minimum allowed value for the layer.

The question of estimating the error covariances of the interface pressures should be revisited in the future. In particular, the covariability of temperature and salinity should be taken into account. In this study, the covariances for interface pressures were computed exactly as described for layer thicknesses by (Thacker and Esenkov, 2002). More generally, future parameterizations of error covariances for all variables should be based on a careful examination of model-data difference.

5. Results

Innovations, i.e., the differences between the data that are assimilated and their model counterparts, are the basis for the monthly corrections. Examining them collectively can reveal both systematic and random differences between the model state and the observations. However, when examining innovations for HYCOM, it is important to keep in mind that model-data disagreements might to some extent be attributable to the way the XBT observations have been transformed into data specifying pressures of layer interfaces and layer-averaged potential temperatures and potential densities.

The nine panels of Fig. 6 indicate how innovations for potential temperature are distributed for nine HYCOM layers. [24] When and where these layers are isopycnic, potential density should be largely unchanged, with the innovations indicating differences in temperature that are compensated by differencess in salinity. However, in higher latitudes where the layers are generally isobaric, the potential density can change and the temperature innovations are not balanced by corresponding salinity innovations. The 20 years of θ_0 innovations for each surface are grouped into 10° bands of latitude with box and whisker plots showing their medians, quartiles, expected range, and outliers. Generally the medians of the θ_0 innovations are close to zero, indicating model values are colder than the data as often as they are warmer. However this is not the case in the deeper layers at lower latitudes, where there is a systematic tendency for the model's water before assimilation to be colder than the data being assimilated. It is interesting to note that, while based on only 16 samples, the opposite bias is found for layer 18 at 25° N, where the median temperature of the data is about 1 °C colder than that of the model, while the situation has reversed in the layer below with the median innovation based on five samples being approximately 3 °C. While such systematic differences might indicate problems with the model, because they occur where HYCOM's layers are expected to be isopycnic, they might also indicate problems with the manner in which the model's layers are inferred from the XBT data. The inter-quartile boxes are generally symmetric about the median, and their widths indicate that half of the innovations differ from zero by less than 1 °C in either direction. The character of the boxes changes as the medians depart from zero, reinforcing the need for further examination of these innovations. The outliers occur mostly in the upper layers at high latitudes, and are generally symmetrically distributed. The outliers suggest heavier-than-Gaussian tails, which reflects the non-normal distribution of temperatures.

[24] Recall that the nine isopycnic surfaces of Fig. 2 were chosen to correspond to the target densities of these nine layers.

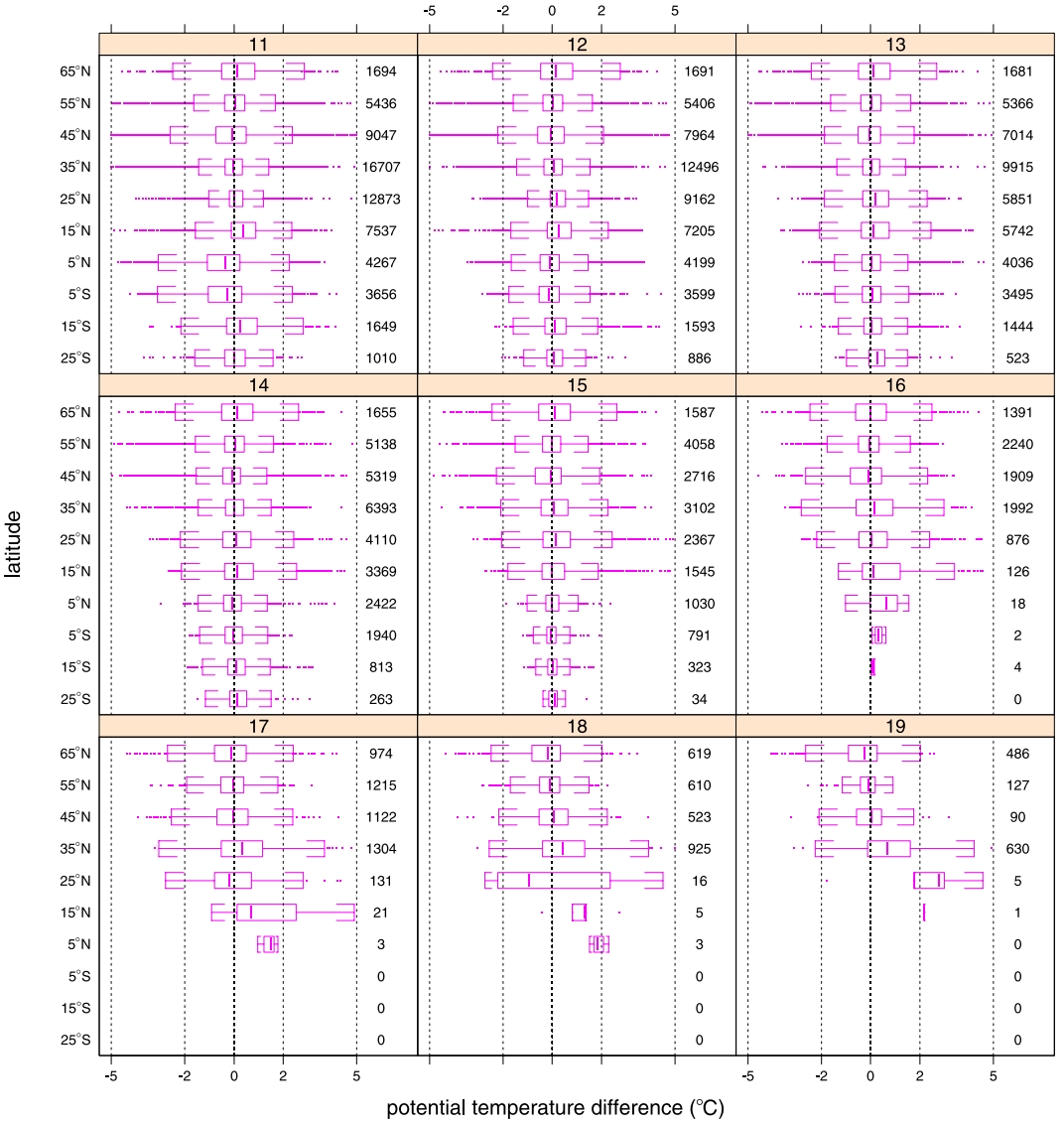

Fig. 6. Distributions of innovations (data-model) for the potential temperature of layers 11–19 for the 20-year HYCOM simulation with data assimilated. Conventions for the box and whisker plots are the same as for Fig. 2.

Fig. 7 shows corresponding box and whisker plots for the innovations of the pressure at the upper surfaces of the same nine HYCOM layers. The absence of variability for the upper layers at higher latitudes is because these layers are almost always isobaric; those few non-zero innovations appear as outliers with the inter-quartile box shrinking to zero width. More generally, as we saw for the potential-temperature innovations, the mean pressure innovations are close to zero throughout most of the basin, indicating corrections with upward displacements occur as often as those with downward displacements. The systematic departure of the means from zero occur in

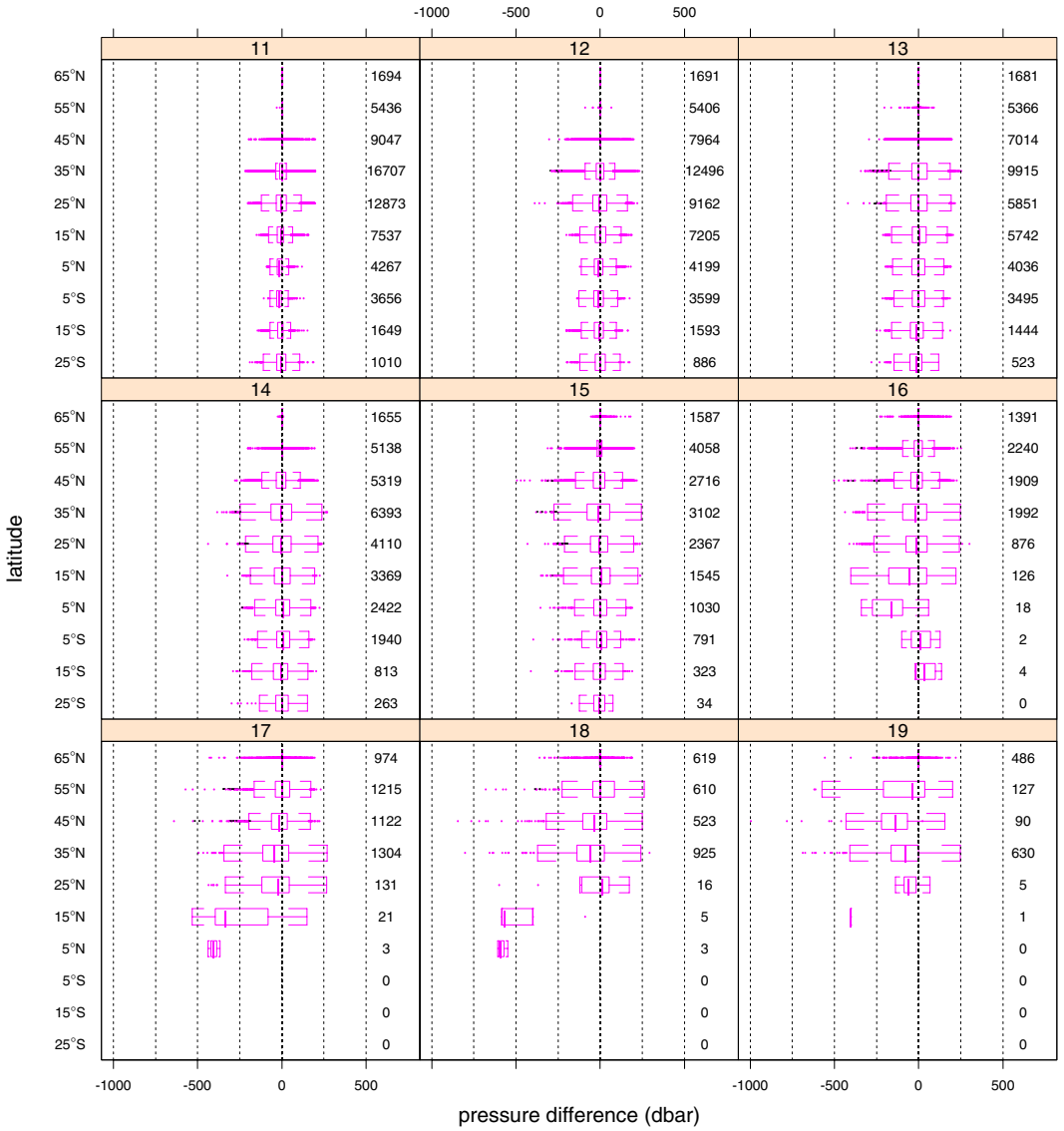

Fig. 7. Distributions of innovations (data-model) for upper-interface pressure of layers 11–19 for the 20-year HYCOM simulation with data assimilated. Conventions for the box and whisker plots are the same as for Figs. 2 and 6.

the lower layers where mean temperature innovations also differed from zero, reinforcing the view that the systematic problems might be associated with the preprocessing of the temperature profiles to get layer-specific data. The inter-quartile ranges are generally consistent with layer displacements of less than 100 m, but outliers indicate corrections with larger displacements, e.g. a 1000 m upward displacement of layer 19 in the 45° N band. Nevertheless, the innovations show a considerably smaller spread than the differences between the uncorrected run and the Hydrobase climatology of Fig. 2, indicating that assimilation is keeping the model closer to the data.

Comparing Figs. 2 and 7 does give a feel for the degree to which the model improves the assimilation, but other figures can make the case mode directly. Plots like those of Figs. 6 and 7 but for differences between the superobs and their counterparts in the 20-year run without data assimilation show much more directly that data assimilation brings the simulation into much closer agreement with the data. But that is to be expected. Similar plots showing differences between the observations and the Hydrobase climatology show a range of variability that is comparable with that seen in Figs. 6 and 7. In other words, when data are assimilated, the results are about as close to the data as is climatology. A goal for the future is a model-based analysis that is closer to the data than climatology. To achieve this goal will require a close examination of model-data differences.

Available space limits our showing extensive details of the evolution and correction of the model's three-dimensional fields over the 20-year data-assimilation period. Thus, we primarily discuss temporal averages, computed as composites of fields at mid-month just before and just after data are assimilated. As HYCOM is at heart isopycnic, for a horizontal view of the impact of assimilation, we examine the potential temperature, salinity, and pressure on a surface of constant potential density. Because of the past focus on the meridional overturning circulation, we also examine the impact that the data have on the zonally averaged volume and heat flux. Going beyond 20-year averages to address month-to-month response, we show that the model's responses to the data-based corrections is quite similar to the model's usual behavior when no data are assimilated. For this, rather than examining an isopycnic surface, for variety we discuss potential density on an isobaric surface.

Fig. 8 shows temporally averaged [25] potential temperature on the surface where potential-density anomaly has the constant value of $\sigma_0 = 27$ kg/m^3, which is a very close approximation to the target value for layer 14. This surface was chosen, because it spans the entire basin, at least for the summer months, and because it is generally within the data-gathering range of the XBT probes. Three of the panels show averages computed from HYCOM simulations for the 20-year data-assimilation period. The lower-left panel shows the 240-month-mean potential-temperature field from a simulation without data assimilation, which provides a baseline for judging the impact of the data. The upper two panels show the mean potential temperature for a simulation with exactly the same initial conditions, forcing, mixing, etc., but with the XBT data assimilated; the upper-left panel shows the average conditions at mid-month just before data were assimilated, while the upper-right shows the average immediately following assimilation; together they show that there is very little tendency to revert to the baseline average during the intervals between assimilation. The fourth, lower-right, panel shows the climatological mean potential temperature on the $\sigma_0 = 27$ kg/m^3 surface computed using Hydrobase software and data. As the World Ocean Atlas (Levitus and Boyer, 1994; Levitus et al., 1994) provided climatological values for initializing the model state and for estimating companion salinity profiles for the XBT data, the isopycnic

[25] The $\sigma_0 = 27$ kg/m^3 surface does not exist everywhere at all times of the year. The temporal averages for the HYCOM simulation was computed for each cell using mid-month potential-temperature values during the 20-year interval only when there was water of this density for this cell. The Hydrobase averages were computed by summing all values, regardless of the time of year, within $1^\circ \times 1^\circ$ latitude–longitude cells and dividing by the number of values.

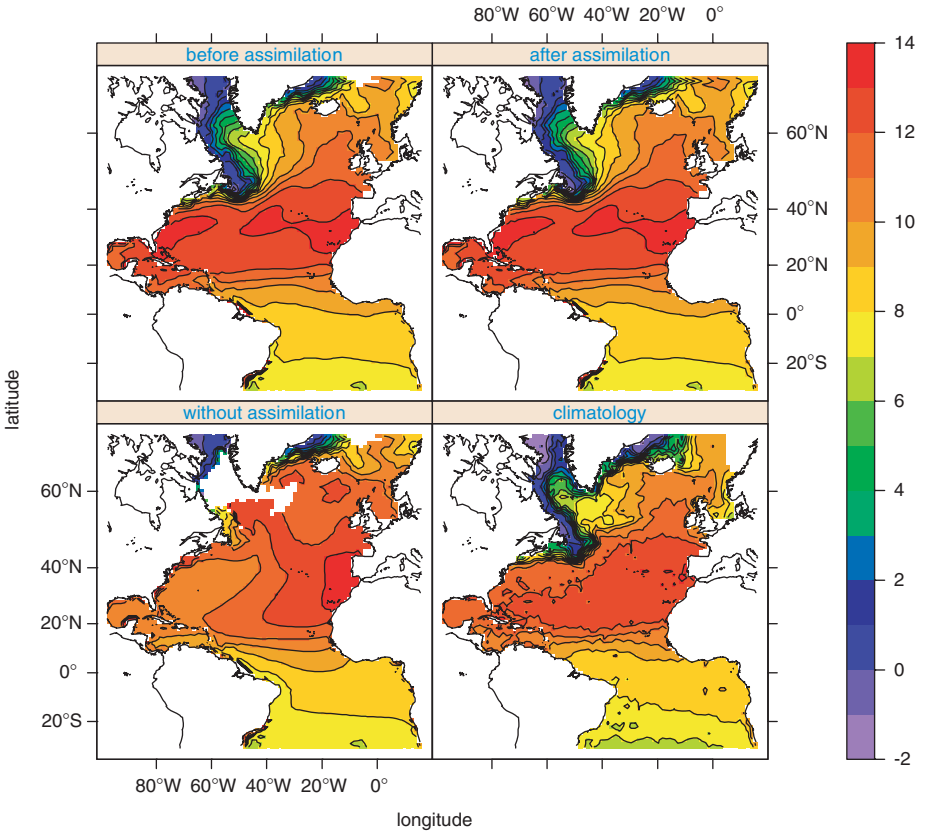

Fig. 8. Climatological potential temperature (°C) for $\sigma_0 = 27$ kg/m^3: from HYCOM simulation without data assimilation (lower left); average conditions immediately before assimilating data (upper left); immediately after assimilating data (upper right); from Hydrobase climatology (lower right). The color bar at the right indicates the color-coded values of potential temperature on the maps; white indicates absence of the $\sigma_0 = 27$ kg/m^3 surface for the entire 20-year simulation.

climatology of Hydrobase provides a somewhat independent view of the potential temperature on this potential-density surface.

The most striking feature for the case without assimilation is the data void near Greenland, where HYCOM never has water with $\sigma_0 = 27$ kg/m^3. Assimilation is seen to correct this problem and to bring the model state into rough agreement with the Hydrobase climatology in this region. Furthermore, the near-coastal temperatures from Labrador to Massachusetts, which were much too warm without assimilation, have been brought into better agreement with climatological expectations. Comparing the two lower panels of Fig. 8 shows that, without data assimilation, the HYCOM simulation is generally colder than the climatological mean in middle latitudes, while the upper panels indicate that the assimilated XBT data cause the estimated temperature to be warmer than climatology. This difference must be due to differences in the temperature–salinity (TS) make-up of the $\sigma_0 = 27$ kg/m^3 water (Fig. 9), which could stem from the initialization or could be a result of failure to preserve the TS relationship during assimilation; this will be a focus

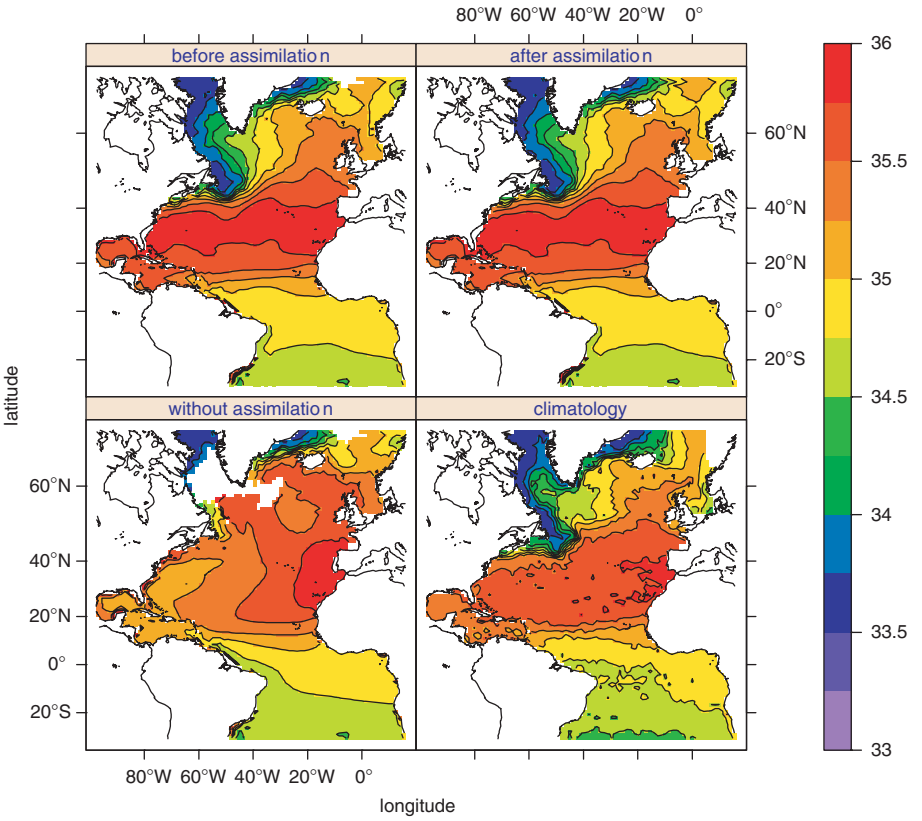

Fig. 9. Climatological salinity (psu) for $\sigma_0 = 27$ kg/m^3: from simulation without data assimilation (lower left); average conditions immediately before assimilating data (upper left); immediately after assimilating data (upper right); from Hydrobase climatology (lower right).

of future work. The climatological temperatures are a bit colder near the equator than the model, both with and without assimilation, which again must be attributed to differences in TS make-up.

Fig. 9 shows salinity on the $\sigma_0 = 27$ kg/m^3 surface for the same four cases. As salinity is determined by potential temperature for fixed potential density, these figures provide no information beyond what is implied by the temperature plots. Nevertheless, they have been included, because they might be difficult to imagine. Fig. 10 shows pressure for the $\sigma_0 = 27$ kg/m^3 surface. While the model without assimilation has this surface too deep in the mid-ocean gyre (and so shallow that it is missing in the Labrador Sea), data assimilation causes it to be a bit too shallow in comparison with Hydrobase climatology. This suggests that the companion salinity profiles should be somewhat fresher, as this would make the density less, putting the $\sigma_0 = 27$ kg/m^3 surface deeper, reinforcing our suspicion that the TS make-up of the water is in error and suggesting that better salinity estimates are needed.

One problem with the model simulation without data assimilation, which stands out in Figs. 8–10, is the warm, salty water that extends north from 40° N into the Labrador Sea. This feature is not seen in the climatology, and it does not appear when data are assimilated. When the 20-year averages shown in the upper-left panels were replaced by one-year averages for years 1972, 1973,

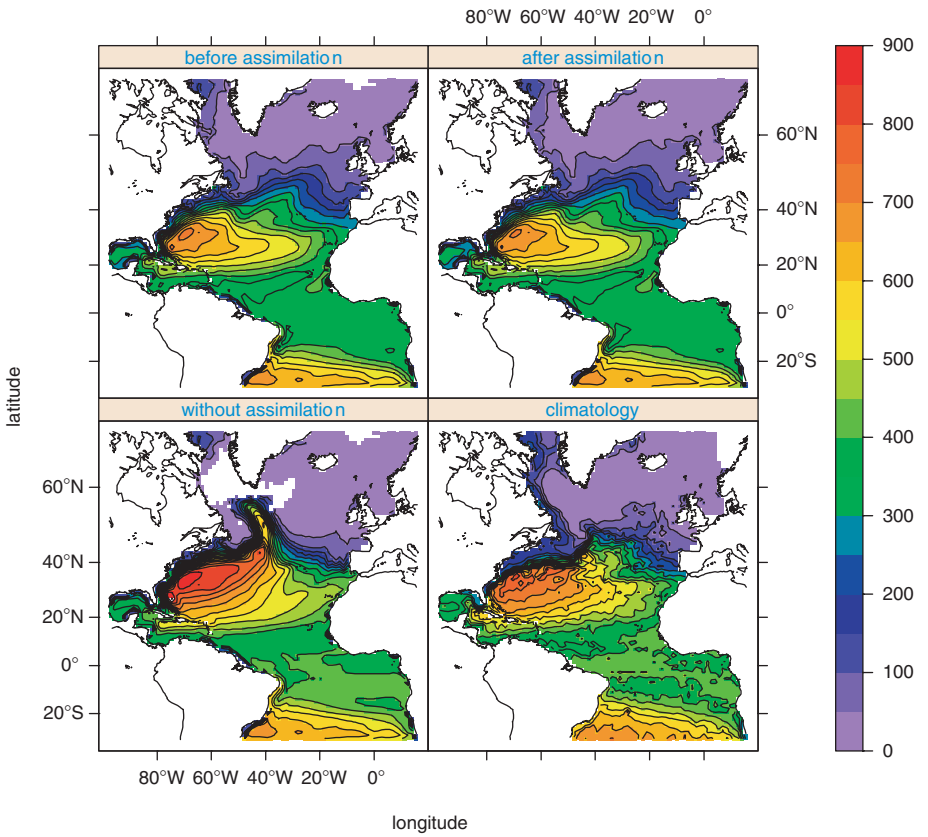

Fig. 10. Climatological pressure (dbar) for $\sigma_0 = 27$ kg/m^3: from simulation without data assimilation (lower left); average conditions immediately before assimilating data (upper left); immediately after assimilating data (upper right); from Hydrobase climatology (lower right).

and 1974, we found that this feature was greatly reduced after one year of data assimilation, and that only a hint of the feature remained after two years. Below, we show that the corrections are sufficiently well-adjusted to allow the model to evolve smoothly.

It is also useful to check the impact of assimilating XBT data by looking at a vertical section. This is best done by comparing differences between the model's fields and their climatological counterparts. Fig. 11 shows differences between the model's 20-year mean potential temperature and potential density fields and their Hydrobase counterparts along the 25° W mid-ocean meridian. As before, the simulations are averaged at mid-month, and the results with data assimilation are computed from the model states just before the monthly corrections. Without assimilation HYCOM's temperatures are too warm near 50° N below 200 m and also too warm in the tropics; [26] assimilating the XBT data brings them much closer to climatological expectations.

[26] Halliwell (2003) has shown that this large warm bias can be greatly reduced by using a modification to the algorithm for reconfiguring the model's layers, which is substantially less diffusive.

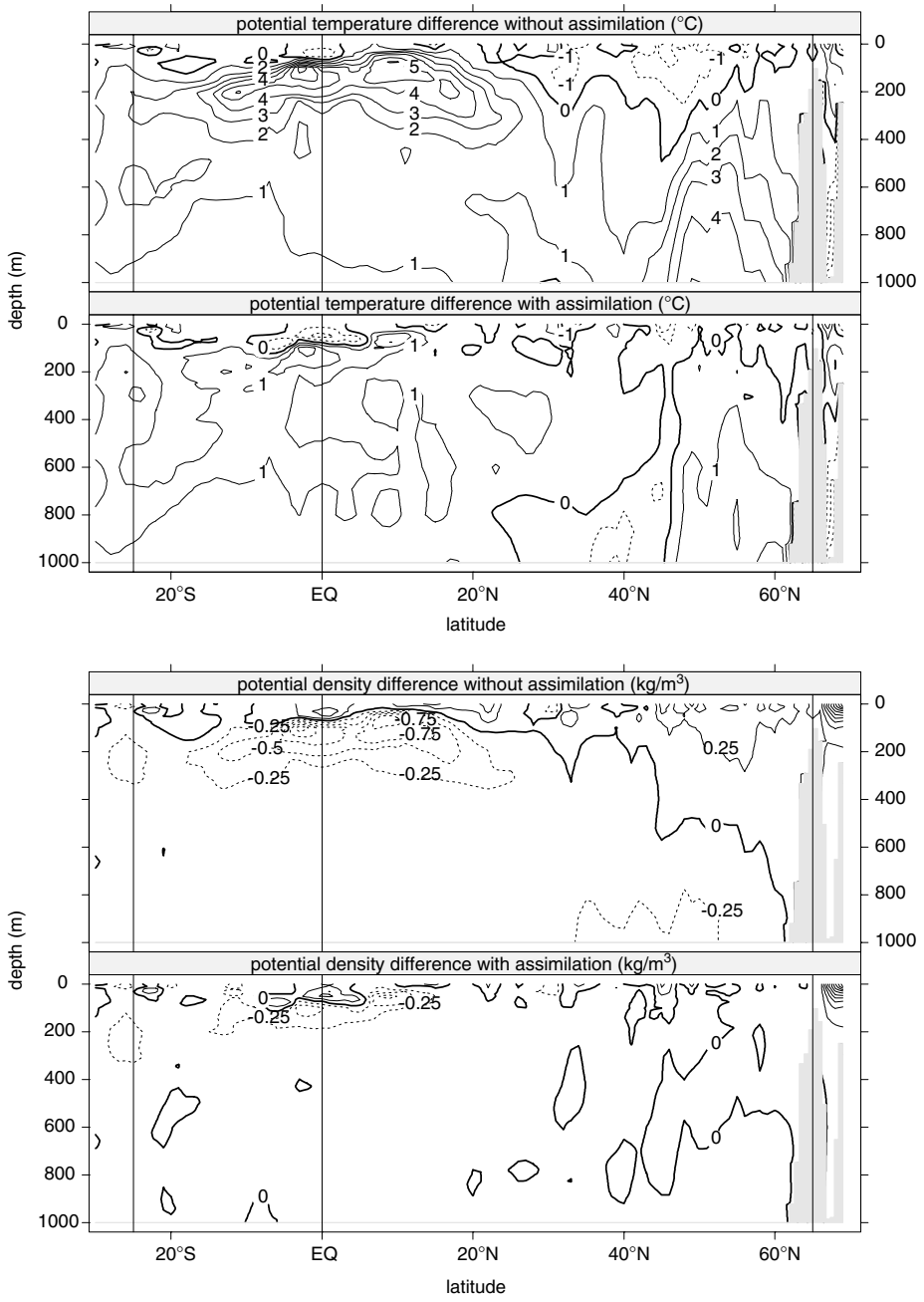

Fig. 11. Contours of HYCOM–Hydrobase differences along 25° W for 20-year simulation with and without data assimilation. Vertical lines indicate northern and southern boundaries of the computational "sponge" layers and location of the equator.

While the simulated potential density without benefit of data assimilation shows sizable departures from climatology in the same regions, the different patterns of the contours indicate that the

differences involve salinity as well as density. Nevertheless, in spite of our questionable correction of salinity, the XBT data bring the simulated density field into much closer agreement with climatology.

Fig. 12 shows the impact of data assimilation on the meridional overturning circulation. Without data assimilation, the average volume flux [27] associated with the overturning circulation exceeds 21×10^6 m^3/s at all latitudes, and reaches 27×10^6 m^3/s between 25° and 40° N. This magnitude exceeds estimates obtained from observations (Bryden and Imawaki, 2001) and other HYCOM simulations (Halliwell et al., 2003; Chassignet et al., 2003) by about 30–40%. Our southern boundary conditions and our use of the air–sea interface as the reference level for potential density [28] both contribute to the excessive overturning. [29] When data are assimilated, we find an enhanced sinking between 25° and 35° N. Comparison of the stream functions for the 20-year composites just before and just after assimilation is performed shows the downward motion is not primarily induced during the assimilative corrections. Instead it can be understood as a subsequent slow downward settling of the water, which has been relocated upward. Data assimilation has increased the strength of the overturning, not reduced it: the volume flux exceeds 27×10^6 m^3/s at all latitudes and reaches about 36×10^6 m^3/s near the equator. Furthermore, with assimilation, roughly half of the North Atlantic Deep Water (NADW) is formed between 25° and 40° N, much farther south than is observed in nature. The induced downward motion might indicate the model is attempting to correct for the assimilations displacement of isopycnal surfaces to shallower than expected depths, or it might reflect a slow reversion to the uncorrected simulation's preference for deeper isopycnals. A full understanding of these problems will be the focus of future work, which will involve a careful cast-by-cast examination of model-data differences and how they respond to changes in the specifications of the model's southern and northern boundaries.

Fig. 13 illustrates the impact of the data on the spatial and temporal scales of variability. The upper panel shows the corrections for potential density on the 500 dbar surface [30] for three months in the final [31] year of the 20-year simulation. These as well as corrections at other depth and for other fields are added at mid-month to a simulation that has had similar corrections in previous months. The spatial scales of the corrections reflect, first, that the influence of each observation is limited to a few grid cells and, second, that the data are generally located along shipping tracks. The month-to-month variability in the correction field is due in part to month-to-month differences in the distributions of observations.

[27] The vertical component of the meridionally integrated volume flux is computed from the northward component to satisfy the continuity equation.

[28] Use of σ_2, corresponding to a reference level of 20 MPa (2000 dbar), together with thermobaricity (Sun et al., 1999) results in a smaller volume flux.

[29] More recent versions of HYCOM, which reference potential density to 20 MPa and account for thermobaric compressibility , typically produces maximum volume fluxes in the range of 16×10^6–18×10^6 m^3/s.

[30] Results are presented on a pressure surface as an alternative view of horizontal variability. Values for potential density and its correction were assigned to the middle of each layer, where the pressure was taken to be the average of the layer's upper and lower interface pressure, and then interpolated to 500 dbar.

[31] The final year was chosen so that we can show the magnitude of the model-data differences after the model has benefited from 19 years of data assimilation. Still, these correction fields are typical of those for most of the 20-year period. After the first year, the size of the corrections have leveled off.

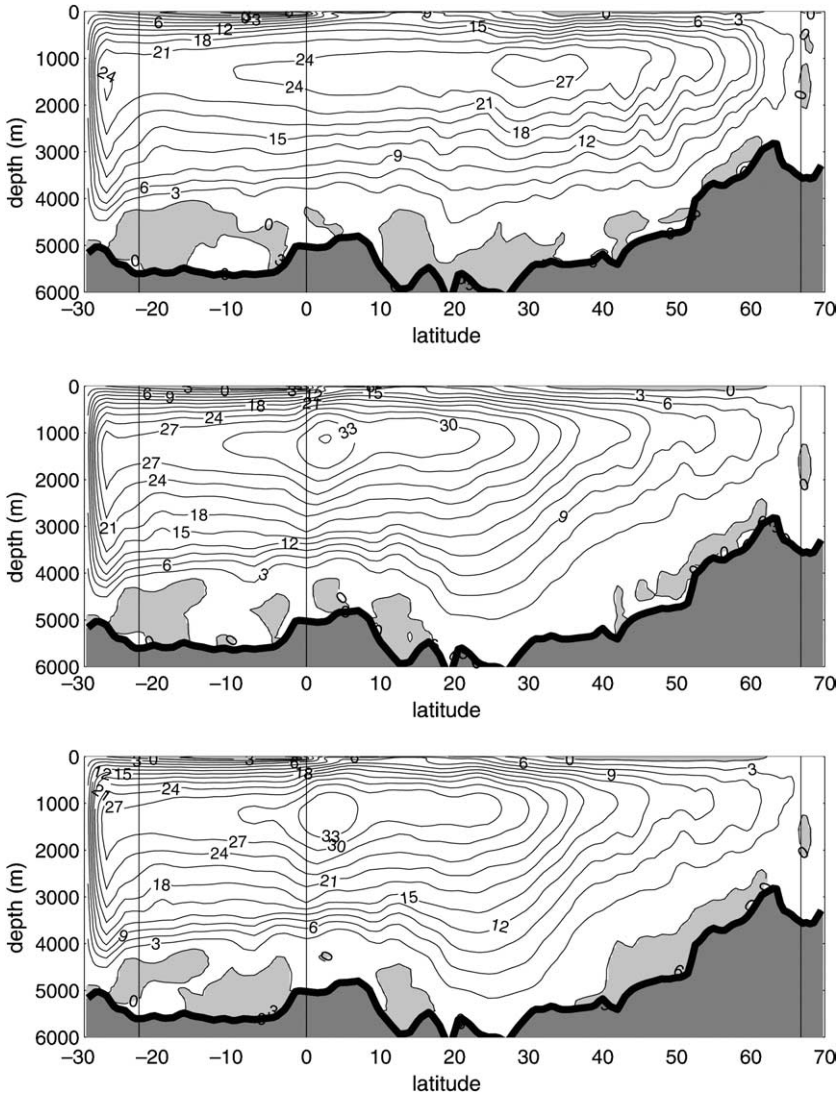

Fig. 12. Contours of the stream function for the zonally and temporally averaged volume transport (10^6 m^3/s) of the meridional overturning circulation: from 20-year simulation without data assimilation (upper panel); averages computed from monthly conditions immediately before data are assimilated (middle panel); averages computed from conditions immediately after data are assimilated (lower panel).

The middle panel of Fig. 13 shows the response over the subsequent month until just before the next mid-month correction. The response is computed as the change in the potential density at 500 dbar over this interval. This can be compared with the change over the corresponding interval for a simulation without data assimilation, which is shown in the lower panel. The qualitative similarity of the monthly differences indicates that the model can evolve from the corrected state without difficulty. For both sets of difference fields small-scale spatial contrasts can be seen; these are characteristic of the model at this resolution and can be diminished by increasing viscosity and

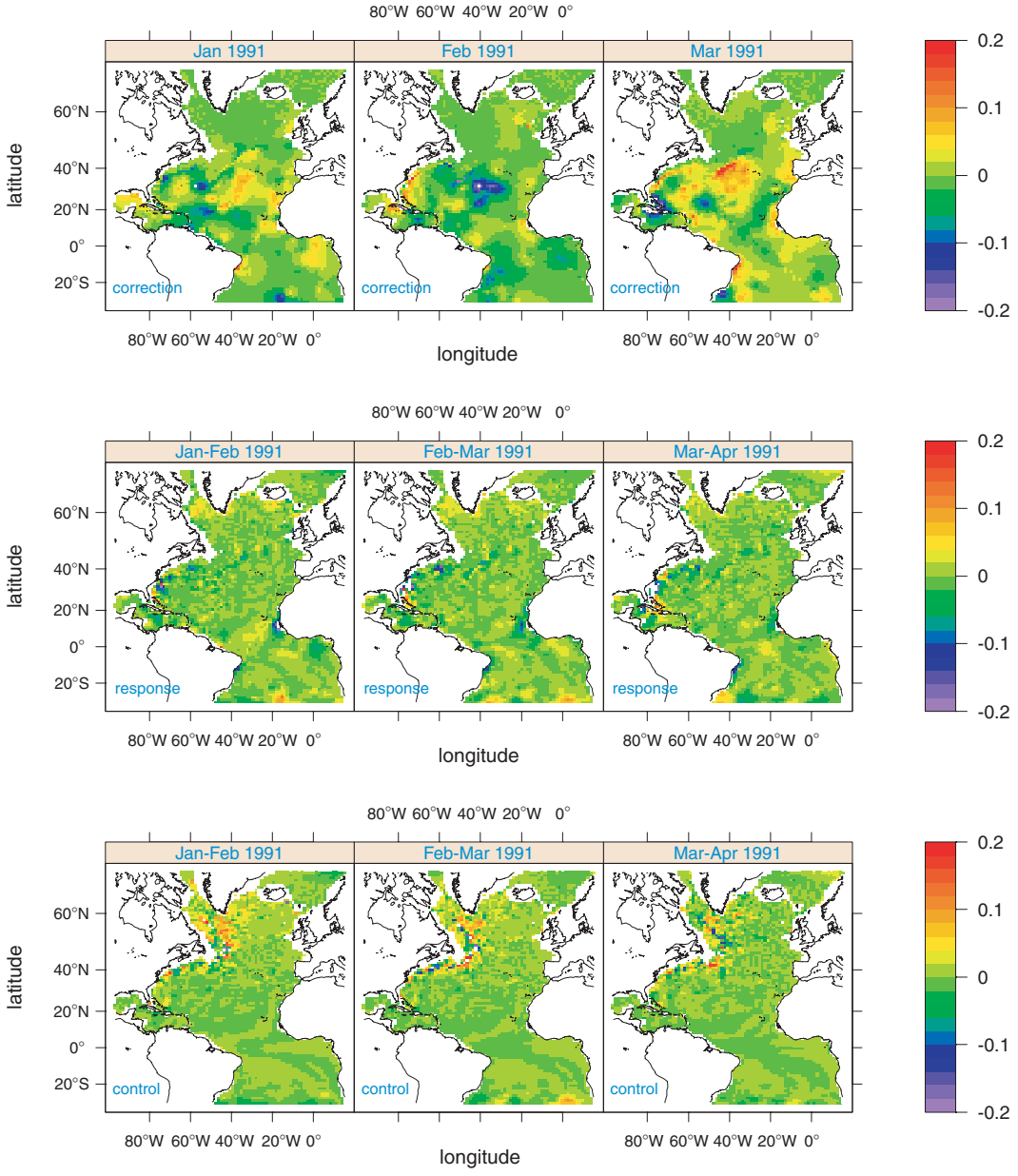

Fig. 13. Potential-density correction (upper panel) response (middle panel) and uncorrected evolution (lower panel) at 500 dbar for three successive months in 1991. The response and the uncorrected evolution are month-to-month differences.

diffusivity. The response is generally much smaller than the correction, indicating once again that there is no appreciable tendency to revert to the uncorrected state over periods of a few months. Furthermore, there is little evidence of propagation of the corrections. They have been adequately assimilated into the model state so that it evolves normally without excessive adjustment.

6. Conclusion

We have shown that the method of Thacker and Esenkov (2002) provides a framework that might be developed into a system for assimilating XBT data into HYCOM to produce a model-based re-analysis of the ocean's evolving state. The results presented here demonstrate that the data have a substantial impact and bring the simulation into closer agreement with reality. In particular, when the XBT data are assimilated, the model's mean fields computed as 20-year averages are much closer to climatological estimates based on data alone than are the mean fields when no data have been assimilated. Furthermore, the monthly corrections are well-adjusted and do not provoke unusual large-amplitude responses.

Because of the preliminary nature of this work, many important issues have not been addressed. In particular, the question of the impact of the XBT data on the simulation at various spatial and temporal scales deserves considerably more attention. And, while the temporally averaged re-analysis might be regarded as a new climatology, with all the issues of irregular sampling and influence scales that plague all such efforts, there is the question of exactly what the model adds to the process. Such discussion will be deferred to a future paper discussing the merits of the final product.

The greatest obstacle to the assimilation of XBT data is the underlying problem of estimating density from measurements of temperature when salinity is unknown. Because HYCOM is at heart formulated in terms of density, this issue immediately comes to the front and can be easily recognized. While it may not be so obvious when dealing with other models and may be overlooked, this problem is intrinsic to the assimilation of XBT data, as it concerns their limited information regarding oceanic dynamics. The temperature-to-density issue must be confronted regardless of the model into which the data are to be assimilated and regardless of the assimilation method. Here we have approached the density estimation problem by using climatological salinity together with the equation of state for sea water. While this often provides a reasonable estimate for the density, at times the estimates are extremely bad. This is often the case in the vicinity of the Gulf Stream front, where both thermal and haline variability are large and correlated. The results shown here indicate that the salinity errors can manifest as erroneous temperature–salinity characteristics for water masses.

The most important improvement would be to account for the covariability of salinity with temperature. For example, the method of Stommel (1947) could be used to estimate TS covariability associated with vertical motions: salinity is estimated by its climatological mean value at the depth where the climatological mean temperature matches the observed temperature. This approach has been been exploited by Vossepoel and Behringer (2000) when assimilating XBT data into an oceanic circulation model; however, rather than estimating companion salinity profiles and correcting salinity during assimilation, they incorporate a term into the prognostic equation for salinity to keep the model's TS relationships close to the climatological TS relationship. Recognizing that such TS relationships have little utility above the thermocline, Vossepoel et al. (1999) explore how it can be modified when observations of sea-surface salinity are available. Troccoli and Haines (1999) recognize that TS relationships are time dependent, so the function derived from climatology might not describe the situation encountered with any particular profile. They suggest using relationships derived from data that are close in time to the XBT data, whenever such data are available, and using TS relationships derived from the model, otherwise.

Troccoli et al. (2002) extend this approach to preserving the model's TS relationship below the XBT profiles where there are no observations of temperature. While this approach has the virtue of not creating new water types, it has the disadvantage of not being able to recognize when the model water properties are incorrect. To recognize incorrect water properties, observed temperatures must be supplemented by good estimates of salinity.

A problem with estimates based on Stommel's method is that they presume that all variability is associated with vertical displacements of otherwise unchanging water properties. A more general approach to covariability requires using information beyond that provided by the mean profiles; second-order statistics are needed. Recognizing this, Maes and Behringer (2000) estimate salinity profiles by fitting temperature profiles to joint temperature–salinity empirical–orthogonal functions. Hansen and Thacker (1999) suggest linear regression as a more straightforward way to estimate salinity from temperature; their method has the advantage of being easily expanded to incorporate other predictors and was found to perform well for difficult barrier-layer situations. Exploring the relative merits of alternative approaches to estimating salinity and ensuring correct water-mass properties is beyond the scope of this paper. A project is underway to develop the capability for estimating companion salinity profiles using second-order statistical methods; once that has been completed, such a comparison of the alternative approaches is planned.

There are several other issues that are closely related to that of salinity/density estimation. One concerns recognizing how the data are related to features such as fronts and eddies. For example, knowing that an XBT from the archive characterizes a cold-core ring could allow for a better treatment when it is assimilated: for a low-resolution model, the impact of the data could be discounted, and for a high-resolution model, short-range corrections might be needed. While compiling the salinity-estimation statistics, it will be useful to develop water-type identifiers similar to those discussed by Sauzé and Cummings (1994). Not only would this provide a step toward recognizing the relationship of the data to mesoscale features, it would also improve the statistics underlying salinity estimation by allowing for stronger TS correlations, and it could be useful for confining the influence of the data to the regions they characterize.

Another important issue that needs attention is how to handle the corrections below the XBT data. As this should reflect the expected local behavior, it should also be addressed statistically. Closely related is the issue of how to assimilate satellite altimetric data. A compilation of vertical correlations for various regions and water masses would be most valuable, as would statistics characterizing the information from the altimeter beyond what is provided by an observed temperature profile. Perhaps such a study could be combined with that directed toward the salinity-estimation problem.

There are several technical issues that need work. One concerns how data are averaged to create cell-based superobs: care should be taken to guarantee that a cluster of redundant data do not dominate the superob. Another concerns the treatment of data near the boundary of the model grid: data which fall outside the stair-step boundary need not be discarded. Conversely, if the model really does not handle coastal processes adequately, near-coastal data might be omitted. These are things that should be considered, whatever the model. For HYCOM, an issue that requires more study concerns the influence of data near a sloping bottom: how should the model's zero-thickness isopycnic layers be treated when nearby observations indicate the presence of heavier water?

There are also several issues related to the estimation of errors for the data and for the background state, which govern how the data correct the model when they are assimilated. One is a better estimate of the uncertainties associated with HYCOM's layerized representation of the data. In particular, better estimates of the errors of the depths of the interfaces between the isopycnic layers could be estimated using an ensemble of temperature profiles consistent with the XBT data and their errors. At present the background-error covariances are postulated to have an exponential shape with an influence scale set by the grid resolution. Better would be to derive them from a careful study of model data differences. Some model-independent technical issues associated with the error covariances should also be addressed. For example, care should be taken so that influence does not cross land barriers: observations in the Gulf of Mexico off the west coast of Florida should have little or no impact on corrections in the Atlantic across the intervening peninsula.

In spite of all these ways in which the data-assimilation system can be improved, it is still useful. We have shown that it provides an improved characterization of the ocean beyond what the model can provide without data-based corrections. More important, it provides the basis on which to build. As each of the above issues is addressed, the system can be improved. In particular, we expect that the better salinity estimates will have a dramatic impact and that, once they have been incorporated, progress will be sufficient for a detailed study of model data differences. Of course, that study will reveal the next most important weakness and provide the direction for subsequent improvements.

Acknowledgements

This work has been supported in part by the National Oceanographic Partnership Program, research project number N00014-03-IP20040.

References

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

Bloom, S.C., Takacs, L.L., da Silva, A.M., Ledvina, D., 1996. Data assimilation using incremental analysis updates. Monthly Weather Review 124, 1256–1271.

Bryden, H.L., Imawaki, S., 2001. Ocean heat transport. In: Seidler, G., Church, J., Gould, J. (Eds.), Ocean Circulation and Climate: Observing and Modelling the Global Ocean. Academic Press, pp. 455–474.

Chassignet, E.P., Smith, L., Halliwell Jr., 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, in press.

Cooper, N.S., Haines, K., 1996. Altimetric assimilation with water property conservation. Journal of Geophysical Research 24, 1059–1077.

Du, Y.-B., Tong, P., 2001. Temperature fluctuations in a convection cell with rough upper and lower surfaces. Physical Review E 63, 46303.

Fox, D.N., Barron, C.N., Carnes, M.R., Booda, M., Peggion, G., Gurley, J.V., 2002. The modular ocean data assimilation system. Oceanography 15, 22–28.

Halliwell Jr., G.R., 2003. Evaluation of vertical coordinate and vertical mixing algorithms in the hybrid-coordinate ocean model HYCOM. Ocean Modelling, in press.

Halliwell, Jr., G.R., Weisberg, R.H., Mayer, D.A., 2003. A synthetic float analysis of the upper-limb meridional overturning circulation interior ocean pathways in the tropical/subtropical Atlantic. In: Goñi, G., Malanotte-Rizzoli, P. (Eds.), Interhemispheric Water Exchange in the Atlantic Ocean. Elsevier, accepted for publication.

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

Kadanoff, L.P., 2001. Turbulent heat flow: structures and scaling. Physics Today 54.

Kara, A.B., Rochford, P.A., Hurlburt, H.E., 2000. Efficient and accurate bulk parameterizations of air–sea fluxes for use in general circulation models. Journal of Atmospheric and Oceanic Technology 17, 1421–1438.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., Fiorino, M., 2001. The NCEP/NCAR 50-year reanalysis: monthly means, CD-ROM and documentation. Bulletin American Meteorological Society 82, 347–367.

Large, W.G., McWilliams, J.C., Doney, S.C., 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Reviews of Geophysics 32, 363–403.

Levitus, S., Boyer, T., 1994. In: World Ocean Atlas, 1994, Volume 4: Temperature. NOAA Atlas NESDIS. U.S. Department of Commerce, Washington, DC.

Levitus, S., Burgett, R., Boyer, T., 1994. In: World Ocean Atlas, 1994, Volume 3: Salinity. NOAA Atlas NESDIS. U.S. Department of Commerce, Washington, DC.

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

Maes, C., Behringer, D., 2000. Using satellite-derived sea level and temperature profiles for determining the salinity variability: a new approach. Journal of Geophysical Research 105, 8537–8547.

Sauzé, M.J., Cummings, J.J., 1994. Real-time water mass analysis of bathythermograph data at Fleet Numerical Meteorology and Oceanography Center. In: MTS94. Marine Technology Society, pp. 440–452.

Semtner Jr., A.J., 1976. A model for the thermodynamic growth of sea ice in numerical simulations of climate. Journal of Physical Oceanography 63, 379–389.

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

Sun, S., Bleck, R., Rooth, C.G., Dukowitz, J., Chassignet, E., Killworth, P., 1999. Inclusion of thermobaricity in isopycniccoordinate ocean models. Journal of Physical Oceanography 29, 2719–2729.

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

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

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.

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, 1415–4101.