

## Salinity Assimilation Using $S(T)$ : Covariance Relationships

K. HAINES, J. D. BLOWER, J.-P. DRECOURT, AND C. LIU

*Environmental Systems Science Centre, University of Reading, Reading, United Kingdom*

A. VIDARD

*European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom*

I. ASTIN AND X. ZHOU

*Environmental Systems Science Centre, University of Reading, Reading, United Kingdom*

(Manuscript received 3 December 2004, in final form 12 July 2005)

### ABSTRACT

Assimilation of salinity into ocean and climate general circulation models is a very important problem. Argo data now provide far more salinity observations than ever before. In addition, a good analysis of salinity over time in ocean reanalyses can give important results for understanding climate change. Here it is shown from the historical ocean database that over large regions of the globe (mainly midlatitudes and lower latitudes) variance of salinity on an isotherm  $S(T)$  is often less than variance measured at a particular depth  $S(z)$ . It is also shown that the dominant temporal variations in  $S(T)$  occur more slowly than variations in  $S(z)$ , based on power spectra from the Bermuda time series. From ocean models it is shown that the horizontal spatial covariance of  $S(T)$  often has larger scales than  $S(z)$ . These observations suggest an assimilation method based on analyzing  $S(T)$ . An algorithm for assimilating salinity data on isotherms is then presented, and it is shown how this algorithm produces orthogonal salinity increments to those produced during the assimilation of temperature profiles. It is argued that the larger space and time scales can be used for the  $S(T)$  assimilation, leading to better use of scarce salinity observations. Results of applying the salinity assimilation algorithm to a single analysis time within the ECMWF seasonal forecasting ocean model are also shown. The separate salinity increments coming from temperature and salinity data are identified, and the independence of these increments is demonstrated. Results of an ocean reanalysis with this method will appear in a future paper.

### 1. Introduction

As more salinity observations become available from Argo floats (Roemmich et al. 2001), it is becoming more important to develop methods to assimilate salinity profile data into ocean circulation models. Representing the salinity field correctly in ocean models is important in a number of contexts. Salinity has an impact on the density field and hence on ocean currents and transports (e.g., Cooper 1988; Roemmich et al. 1994; Vialard and Delecluse 1998a,b). Salinity is also important in certain places in the mixed layer where it

controls the stability of the water column and hence to a degree, mixing and air–sea interaction, for example, in the barrier layer around the western equatorial Pacific Ocean. In addition, the relationship between temperature and salinity contains important information about the nature of the thermocline and subduction rates and areas (Iselin 1939). Temperature–salinity scatterplots are a standard tool in the armory of physical oceanographers. They are used to track water masses in the deep ocean and to infer information about mixing rates using end-member analyses (Tomczack 1981) and other inverse methods. Temperature and salinity data have also been compared in repeat section work in an attempt to identify climatic changes over decadal time scales (e.g., Bindoff and McDougall 1994; Wong et al. 1999; Bryden et al. 2003). Indeed, Dickson et al. (2003) and Curry et al. (2003) discuss

---

*Corresponding author address:* Keith Haines, Environmental Systems Science Centre, Reading University, 3 Earley Gate, Reading RG6 6AL, United Kingdom.  
E-mail: kh@mail.nerc-essc.ac.uk

high-latitude and global salinity changes as providing the dominant signal of climate change in the oceans over the past few decades. It is important therefore to recognize that the correct treatment of salinity data in the context of ocean data assimilation will allow analyzed ocean fields to be used for more detailed studies of all of the above phenomena.

Several methods have been proposed to update salinity in a multivariate fashion during the assimilation of other data (Troccoli and Haines 1999; Vossepoel and Behringer 2000; Maes and Behringer 2000), but there has been very little work so far attempting to assimilate salinity observations themselves. This paper takes as its starting point the conceptual methods advocated in Haines (1994, 2003), Cooper and Haines (1996), and Troccoli and Haines (1999). The aim is to use the salinity data in the most effective way by seeking to identify independent information from that available from other more abundant assimilated datasets. For example, Cooper and Haines (1996) argued that during altimeter data assimilation water columns should be changed in such a way as not to alter the volume, or the temperature  $T$  and salinity  $S$  characteristics, of water masses in the model water column. The justification is 1) that these quantities are not directly observed by the altimeter and 2) that much of the sea level variability is due to dynamical advection associated with wave motions, while budgets of water mass volumes and properties are controlled largely by separate thermodynamic processes.

Similarly Troccoli and Haines (1999, hereinafter TH99) argued that when temperature profile data are assimilated into models, the volume of water in each temperature range is observed but the  $T$ - $S$  relation is not. Therefore it is useful to introduce multivariate salinity changes with the aim of keeping the salinity on an isotherm [ $S(T)$ ] relations in model water columns unchanged. This is a better solution than leaving salinity measured at a particular depth [ $S(z)$ ] unchanged when the salinity is not observed. Troccoli et al. (2002) and Ricci et al. (2005) have shown that this constraint provides many benefits in assimilation of temperature profiles, in particular leading to improvements of both the salinity and temperature fields. Fox and Haines (2003) describe the application of both these methods in a high-resolution global ocean model and discussed the contributions of each dataset to the success of the final assimilation. However the above studies, and others, have also shown that, even with appropriate constraints, the salinity fields of ocean models tend to drift away from realistic values because of poor knowledge of the surface freshwater fluxes as well as poor repre-

sentation of other internal processes such as mixing and thermohaline circulation strength. Hence the urgent need for an appropriate salinity assimilation scheme exists.

This paper aims to demonstrate the value of analyzing salinity on surfaces of constant temperature  $S(T)$ , in comparison with analyzing on level surfaces, or  $S(z)$ . It is argued that this allows the salinity data to provide independent information to the analysis from that inferred from either altimeter or temperature profile data. In section 2, data from the *World Ocean Database 2001 (WOD01)*, from the Argo array, and from two different ocean models are used to study the variance and spatial covariance of  $S(T)$  as compared with  $S(z)$ . In section 3, we develop a data assimilation scheme for  $S(T)$  data. In section 4, we show preliminary results of salinity analyses that result from  $S(T)$  assimilation as applied in the European Centre for Medium-Range Weather Forecasts (ECMWF) ocean model. Section 5 provides conclusions and discussion. A future paper will describe the results from applying this  $S(T)$  assimilation scheme in the ECMWF seasonal forecasting model.

## 2. Variance and covariance of salinity from observations and models

### a. Observed salinity variability

The *WOD01* dataset has been recently analyzed and quality controlled by the Met Office as part of the European Union (EU) enhanced ocean data assimilation and climate prediction (ENACT) project (Ingleby and Huddleston 2005). These data cover a period of over 40 yr from 1958 to 2001 and were supplemented by more recent data, for example, from the World Ocean Circulation Experiment. To demonstrate the different results obtained by analyzing the salinity variability on depth levels and on temperature surfaces, we calculated the salinity variance. The ENACT data were first classified into  $1^\circ \times 1^\circ$  bins, where bins with less than five samples, or where the variance was in the lowest 5% of values, were omitted. The remaining populated bins typically contained 10–100 salinity profiles in the upper part of the water column. The  $S(z)$  variance was calculated at two different depth levels, 300 and 700 m. Then the mean temperature was identified at each of these two depth levels (this mean temperature varies strongly according to location), and the  $S(T)$  variance was calculated for these two isotherms at each location. This idea was used by TH99 (their Figs. 2 and 3) to demonstrate the reduced variations of  $S(T)$  at a couple of locations in the tropical Pacific.

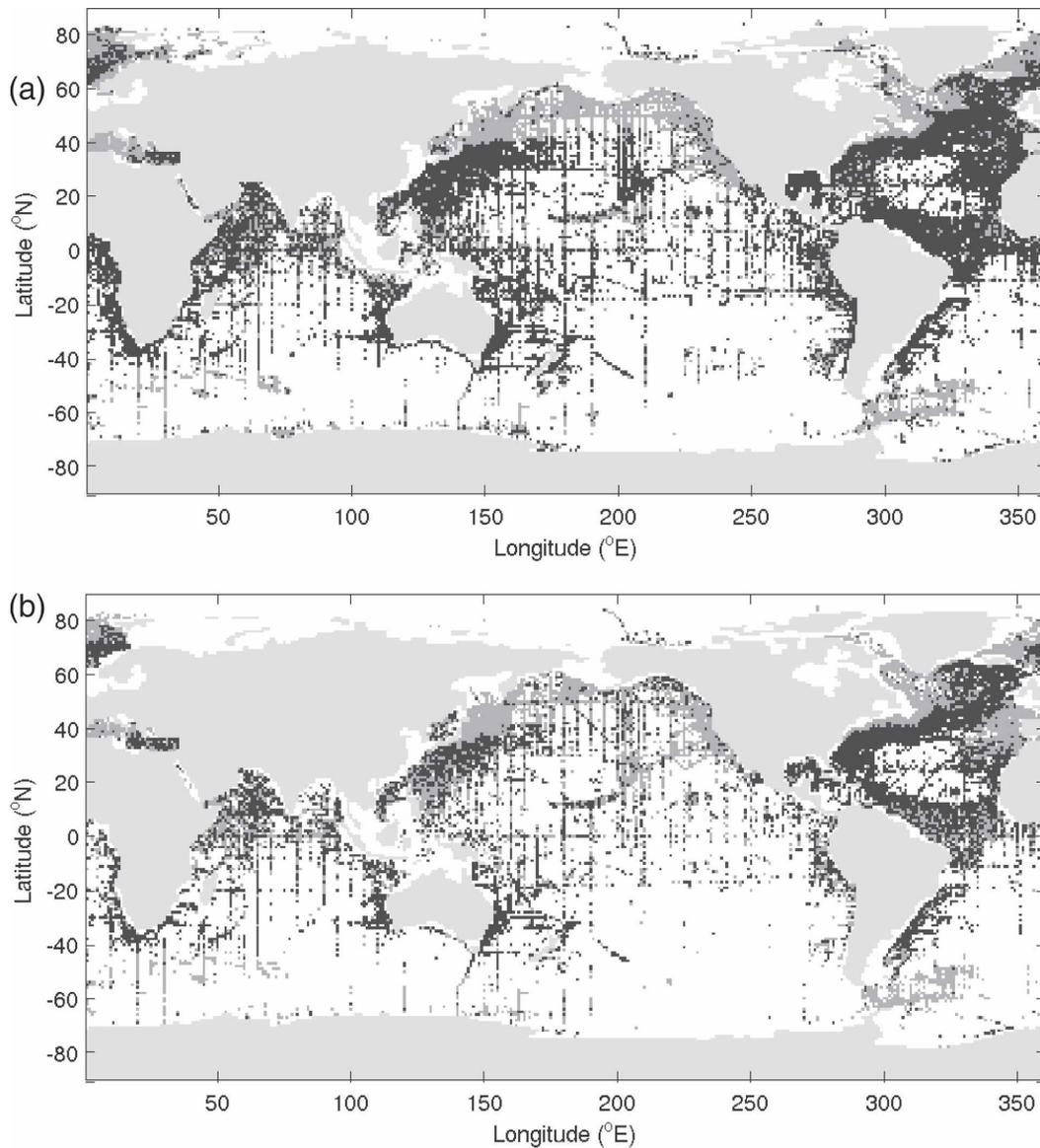


FIG. 1. Ratio of  $S(z)$  variance over  $S(T)$  variance in  $1^\circ \times 1^\circ$  bins for the 40 yr of ENACT data: (a) 300-m depth, and the mean isotherm at that depth, and (b) 700-m depth, and its mean isotherm. Points at which the ratio is greater than 1 are black, and points at which the ratio is less than 1 are dark gray.

Figures 1a and 1b show the ratio of the  $z$ -level salinity variance over the appropriate isotherm variance for the 300- and 700-m depth levels, respectively,  $\text{var}[S(z)]/\text{var}[S(T)]$ . For display purposes, for the black bins this ratio is greater than 1; that is, the  $z$ -level variance is larger. The shaded bins have a ratio of less than 1; that is, the isotherm salinity variance is larger.

The salinity variance on both mean isotherms is reduced relative to that measured on depth levels (ratio  $> 1$ ) for most of the data bins south of the subpolar gyres in both the North Atlantic and North Pacific Oceans. In addition, at 300 m the variance reduction for

$S(T)$  is valid right up into the northeast Atlantic, in the Indian Ocean, and down as far south as the Antarctic Circumpolar Current fronts. Notable exceptions are in the western Mediterranean Sea and a band in the Pacific down the west coast of the United States. At 700 m, the results are less clear cut but still have a predominance of lower  $S(T)$  variance in the subtropical North Pacific and Indian Oceans. However, in the eastern Atlantic near the Mediterranean Outflow the  $S(z)$  variance is clearly lower at 700 m. By inspecting particular bins, for example, in the Indian Ocean, it was determined that there were still some residual bad salinity

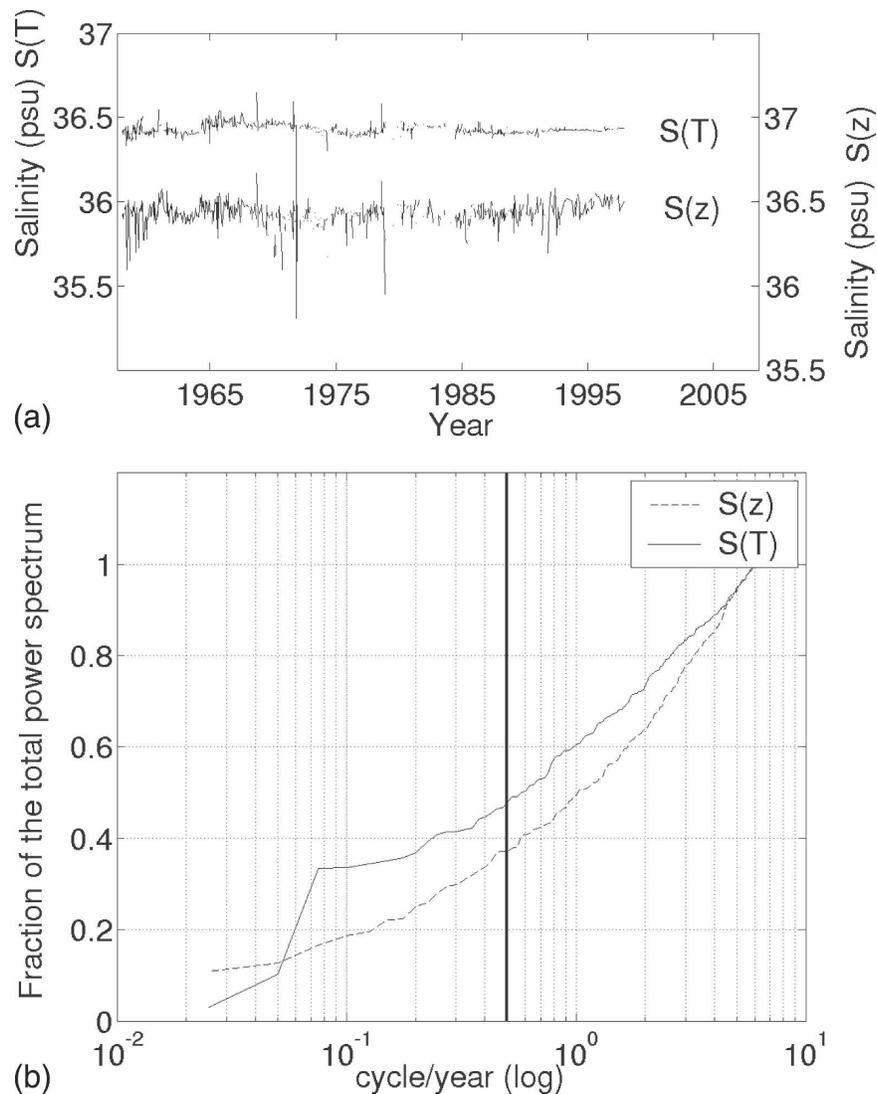


FIG. 2. (a) Salinity time series at Bermuda. The lower line (right-hand scale) shows salinity at 400-m depth, and the upper line (left-hand scale; displaced 0.5 psu) shows the salinity on the 17.4°C isotherm, which is the mean temperature at 400 m. (b) Cumulative normalized power spectra for the two time series, showing that a considerably larger fraction of the power in  $S(T)$  lies at lower frequencies, with periods longer than the marked line at 2 yr [in (b), the time series are linearly interpolated to fill gaps].

profiles in this ENACT dataset, and we would expect that removing these may reduce the  $S(T)$  variance still further and extend the black areas in Fig. 1.

Overall there is a clear indication of the tendency for  $S$  and  $T$  to vary together and hence for a reduced  $S(T)$  variability in most areas where the water column has a good thermal stratification. It is in the regions of colder water masses, in the subpolar gyres and down the west coast of the United States, and in regions where salinity is known to contribute strongly to stratification, as in the western Mediterranean, where this does not hold. If we consider the different kinds of wave motions in the

thermocline, including internal waves and Rossby waves, as well as mesoscale eddy activity, and seasonal variability due to Ekman pumping, all of these phenomena will contribute to increased salinity variance on a depth surface but will have very little impact on salinity variance on a  $T$  surface.

It is also relevant to look at the temporal variability of  $S(z)$  relative to  $S(T)$ . There are very few actual time series of salinity data available, but one place for which a reasonably populated series is available is Bermuda (Joyce and Robbins 1996). Figure 2a shows the time series of salinity at 400-m depth at Bermuda, along with

the time series of the salinity on the 17.4°C isotherm, which is the mean temperature at 400 m. It is clear from the two time series that the variability of  $S(z)$  is greater than the variability of  $S(T)$ , as expected from Fig. 1. However we also see that the dominant temporal variability in  $S(T)$  occurs on a longer time scale than the dominant variability in  $S(z)$ . Some of this  $S(T)$  variability takes place on very long time scales; for example, salinity on the 17.4°C isotherm during 1968–73 is a little higher than during 1995–2000. This dominance of longer time scales in the variability of  $S(T)$  is emphasized in Fig. 2b, which shows the cumulative normalized power spectra of the two time series. Clearly a considerably larger portion of the total power occurs at periods longer than 2 yr (marked 0.5 cpy) for the  $S(T)$  time series.

The processes that will change  $S(T)$  are processes of horizontal advection along isopycnals bringing in water of a different water type to a region. This ultimately occurs because of changing ventilation patterns at the sea surface where water masses are formed and subsided or because of large-scale changes in mean circulation. Ventilation changes may be larger or smaller in different regions, on different isotherms, and over different time periods. It has been suggested that the  $S(T)$  relationships in the North Atlantic have not changed much on interdecadal time scales (Levitus 1989); however in the Indian and South Pacific Oceans there is evidence of significant interdecadal change (e.g., Bindoff and McDougall 1994; Bryden et al. 2003). Ideally one would like to analyze any changes that do occur in  $S(T)$  by deriving the spatial covariance structure, and, given the different and considerably fewer processes involved in producing  $S(T)$  variability, we would expect different scales of spatial variability than for  $S(z)$ . However the sparseness of available observational data makes it very difficult to calculate such spatial scales in  $S(T)$  variability. Therefore we turn to model data to study spatial covariance.

#### *b. Modeled salinity variance and covariance*

As suggested above much of the variability of salinity on temperature surfaces within the thermocline occurs on long time scales and comes about because of variability in surface flux conditions where the isotherms (and isopycnals) outcrop. Through slow processes of ventilation, the changes in  $S(T)$  then penetrate into the subsurface ocean through advection. We therefore expect that models exhibiting such variability would have to be run for long periods. The third Hadley Centre Coupled Ocean–Atmosphere General Circulation Model (HadCM3) has been run for more than 1000

model years without flux correction (Gordon et al. 2000). The model climate is reasonably realistic and the drifts are very small after the preliminary adjustment, especially within the thermocline, making this model suitable for studying natural variability of water properties on long time scales. The ocean component is  $1.25^\circ \times 1.25^\circ$  with 20 vertical levels, and it is therefore not eddy permitting. Monthly mean temperature and salinity data from this model were available for a 100-yr period from the National Environment Research Council (NERC) Coupled Ocean–Atmosphere Processes and European Climate (COAPEC) program. A second Hadley Centre coupled model (HadCEM; Roberts et al. 2004), is an ocean eddy-permitting version of HadCM3 with an ocean resolution of  $1/3^\circ \times 1/3^\circ$  and 40 vertical levels but with the same atmospheric model as HadCM3. HadCEM has been run for 150 model years as a free coupled model, and the last 10 yr of these data were also available to us as monthly mean fields.

Figure 3 shows a set of four one-point correlation maps for salinity variability  $S(z)$  at a depth of 400 m from two locations in the Pacific and Indian Oceans, based on these HadCM3 and HadCEM data. Figure 4 shows the equivalent one-point correlations for salinity variability on an isotherm. In each case, the isotherm chosen (shown in the legend) corresponds to the mean temperature at 400 m for the same locations as in Fig. 3.

First of all it can be seen in Fig. 3 that the spatial covariance scales of the HadCM3 model and the HadCEM model for  $S(z)$  are very different, with the HadCEM model showing much smaller spatial covariances. This is the result of the HadCEM model being eddy permitting, and therefore the spatial scale is dominated by the mesoscale variability at the Rossby deformation radius. There is therefore every reason to believe that the  $S(z)$  covariances in the real world wherever mesoscale variability is strong would be dominated by these same short spatial scales.

However, when we look at the correlations for  $S(T)$  variability in Fig. 4 the spatial scales from HadCEM and HadCM3 are much more similar to each other, and these scales are much larger than for the  $S(z)$  correlations, especially those in the HadCEM model. Even when the  $S(z)$  and  $S(T)$  correlation scales are compared within the HadCM3 model, the spatial scales for the  $S(T)$  covariances are a little larger than for  $S(z)$ . There are some spuriously large remote covariances detected in these plots, particularly for the HadCEM dataset. This is due to only using 10 yr of monthly data, which is a short period of time to quantify  $S(T)$  variability accurately, for some of the reasons discussed previously. However, the data are sufficient to demonstrate that there are scale differences between the  $S(z)$

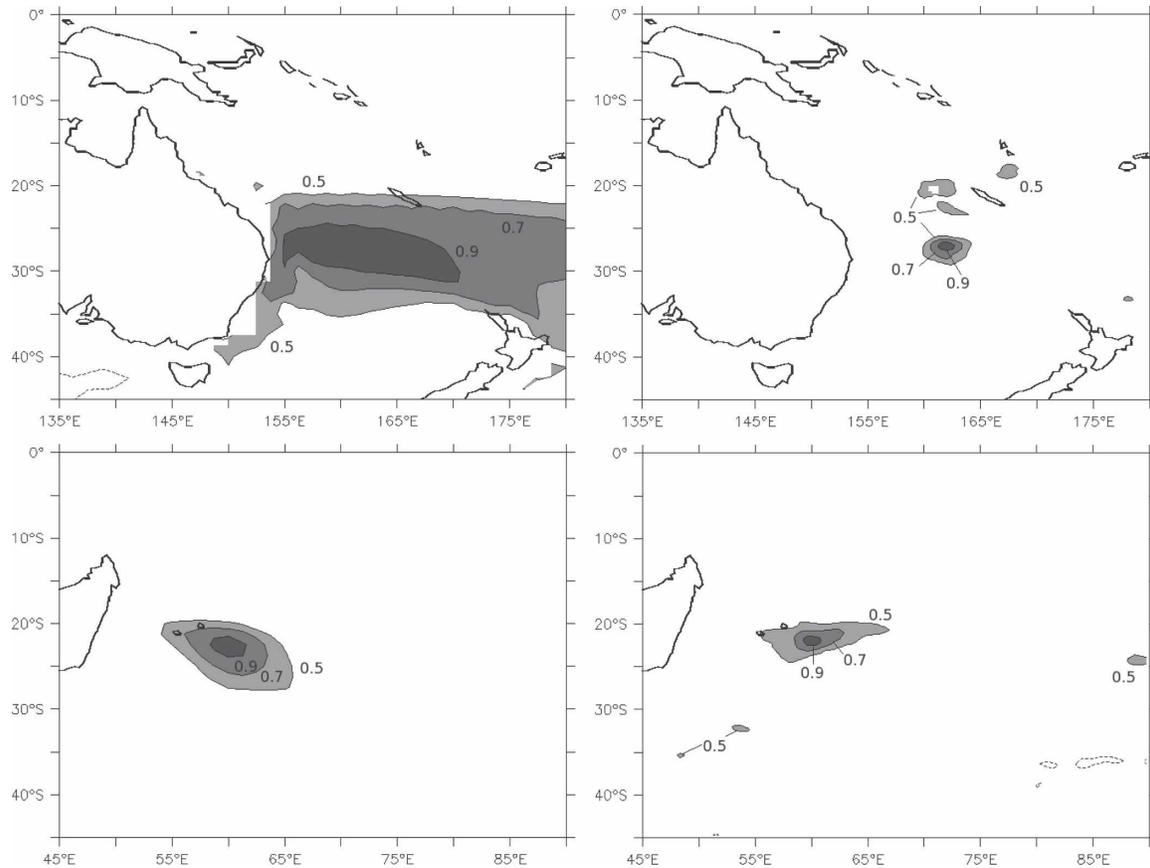


FIG. 3. One-point correlation maps for  $S(z)$  at 400 m at (top)  $27^{\circ}\text{S}$ ,  $162^{\circ}\text{E}$  in the South Pacific off Australia and (bottom)  $22^{\circ}\text{S}$ ,  $60^{\circ}\text{E}$  in the Indian Ocean east of Madagascar. (left) Results from 100 yr of the low-resolution climate model HadCM3; (right) results from 10 yr of the higher-resolution HadCEM model at the same locations. Contour intervals  $\pm 0.5$ ,  $\pm 0.7$ , and  $\pm 0.9$  are shown.

and  $S(T)$  variations. The 400-m depth level of this comparison is within the main thermocline at both of the locations, and temperature will dominate the stratification.

If we simplify the assimilation problem to one of capturing variability not present in the climatology then it is the covariance scales detected in Figs. 3 and 4 that would be relevant for the assimilation of observations. Even if the model a priori has some representation of the variations, the error covariances are still likely to show similar scale differences. Typically if the spatial covariance scales for  $S(T)$  are 3 times those for  $S(z)$ , then each observation of  $S(T)$  can be used to influence an area of ocean which is 9 times the area influenced by an observation of  $S(z)$  at the same location. Effectively this is achieved because the error covariance for  $S(T)$  will be naturally flow dependent, since the  $T$  field is flow dependent, whereas the  $S(z)$  covariances are not naturally flow dependent.

Therefore, for assimilation, the results of this section demonstrate two important facts:

- 1) At midlatitudes and lower latitudes a significant fraction of the salinity variability on a depth level can be modeled by re-referencing the salinity properties to an isotherm. This is achieved by the TH99 assimilation method in which salinity is modified to remain unchanged on an isotherm during temperature profile assimilation.
- 2) When attempting to assimilate salinity observations, the key additional information is the  $S(T)$ . This  $S(T)$  information exhibits larger spatial covariance scales than  $S(z)$  [or  $T(z)$ ], which are both dominated by the mesoscale. Thus  $S(T)$  data can be given a wider influence radius during the salinity data assimilation step. In the following section we show how an assimilation method for  $S(T)$  can be constructed.

### 3. Salinity assimilation methods

All of the assimilation algorithms described here are presented in the context of simple optimum interpola-

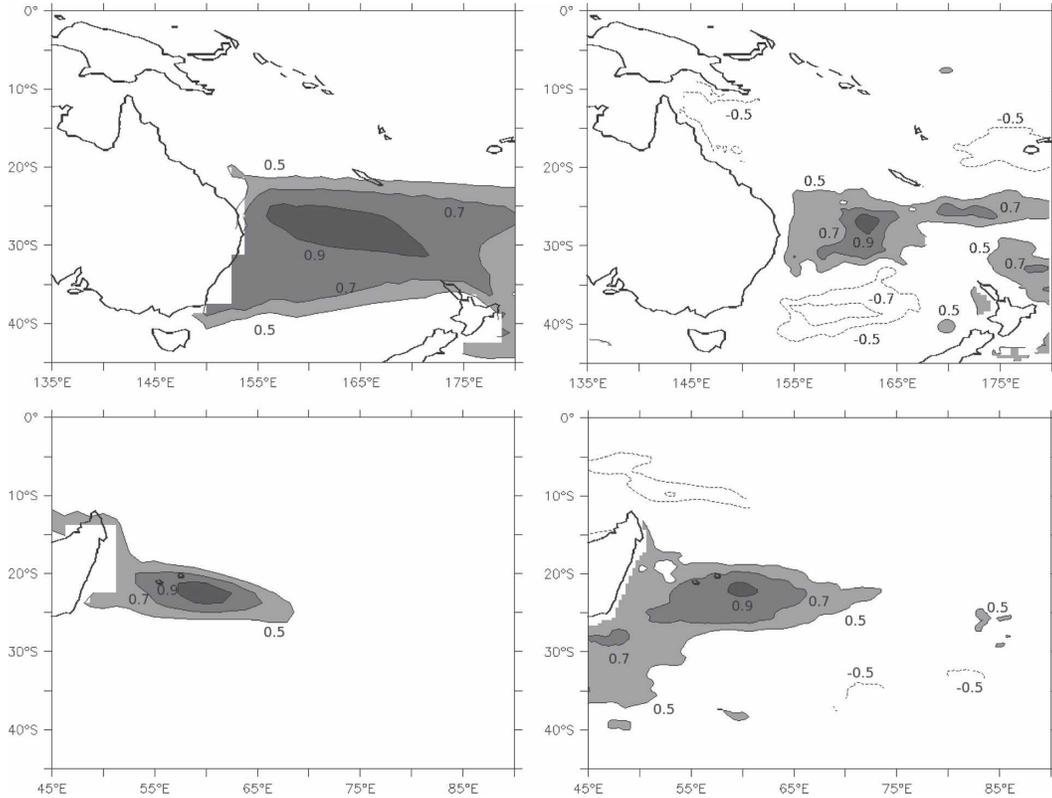


FIG. 4. As in Fig. 3 but now for  $S(T)$  variations. The  $12^{\circ}\text{C}$  temperature surface is used in the upper panels, being the mean temperature at 400 m in the Pacific. The  $15^{\circ}\text{C}$  isotherm is used in the lower panels, being the mean temperature at 400 m in the Indian Ocean.

tion (OI) methods. Nonetheless this does not detract from the physical content and indeed makes the consequences easier to explore. It is possible that these ideas could be extended to more advanced three-dimensional and four-dimensional variational methods of assimilation with the correct developments of the background error covariance matrix, and there is some further discussion in section 5.

*a. Salinity OI on  $z$  levels;  $S(z)$*

We begin by assuming a univariate assimilation scheme for observed  $S_O(z)$  profiles, which will mirror the OI assimilation normally performed on  $T$  profiles. We can write the salinity assimilation analysis  $S_a$  as

$$S_a = S_b + \mathbf{K}_1(S_O - HS_b),$$

where  $H$  is the observation operator to map the predicted model background salinity  $S_b$  to the observation space location of  $S_O$ . The gain matrix  $\mathbf{K}_1$  reflects several important effects: 1) it accounts for relative errors in the background and observations, 2) it reflects a representivity error or filter to remove elements of the data that cannot be modeled, and 3) it reflects the mapping

from the observation location onto the model grid, which includes spreading the influence of each observation profile quasi horizontally over a considerable area. The spatial weighting of each column in the gain will typically reflect the horizontal distance  $r$  between the analysis and the observation locations; for example,

$$\mathbf{K}_1 \sim \exp(-r^2/2R^2),$$

where  $R$  represents a correlation scale on which the influence of the observations decay (e.g., see Daley 1991). Other structure functions, which may be anisotropic, are also possible. In another simplification (used in the ECMWF seasonal forecast system 2) this analysis is only performed level by level so that  $\mathbf{K}_1$  would not reflect vertical correlations, and only salinity observations at the same level would be used in the analysis:

$$S_a(z) = S_b(z) + \mathbf{K}_1[S_O(z) - HS_b(z)]. \quad (1)$$

This salinity analysis can be carried out entirely independently of the simultaneous measurement and analysis of temperature; however, in doing so it misses the

opportunity of taking full advantage of the relationships between them.

Trocchi et al. (2002) have already shown that temperature assimilation can considerably improve the salinity field of an ocean model by taking advantage of the large fraction of salinity variance that is strongly correlated with temperature variance, as demonstrated in Fig. 1. Let us consider that a  $T$  analysis has been carried out prior to assimilating salinity data and that salinity has been updated as part of this, using TH99, in order to keep the  $S(T)$  relationship of the background unchanged. If the temperature analysis is denoted  $T_a$  then we already have a preliminary analysis of salinity:

$$S'_a(z) = S_b(z) + \Delta S_T(z), \quad (2)$$

where  $\Delta S_T$  ensures that

$$S'_a(T_a) = S_b(T_a),$$

within any model profile. This preliminary salinity analysis now becomes the background for the assimilation of the observed salinity data:

$$S_a(z) = S'_a(z) + \mathbf{K}_2[S_O(z) - HS'_a(z)]. \quad (3)$$

Given that one salinity increment has already been applied, the second increment in Eq. (3) is likely to be smaller, and it is not immediately clear what gain  $\mathbf{K}_2$  to apply. However, it is very enlightening to rewrite Eq. (3) as

$$S_a(T_a) = S'_a(T_a) + \mathbf{K}_2[S_O(T_O) - HS'_a(T_a)], \quad (4)$$

which is, term for term, exactly equivalent to Eq. (3). Where model fields appear in Eq. (3) the temperature value at level  $z$  is the analyzed temperature  $T_a$ , whereas in the observation profile the temperature at level  $z$  is, of course,  $T_O$ . Only in exceptional circumstances will these two temperatures be equal. The reformulation in Eq. (4) is highly suggestive of an increment in  $S(T)$ , which would be elegantly complementary to the TH99 assimilation, Eq. (2) above, which deliberately does not change  $S(T)$ . However, unless the two salinities in brackets in Eq. (4) are on the same isotherm, we will not have a true representation of an  $S(T)$  increment. The problem is that by doing this second analysis in  $z$  coordinates we are not taking full advantage of the fact that the first salinity increment from the TH99 scheme leaves the salinity unchanged on  $T$  surfaces.

#### *b. Salinity OI on temperature surfaces; $S(T)$*

The temperature and salinity provide two separate pieces of information about the hydrographic structure of the ocean, but, since temperature is always available whenever salinity is measured, it is possible to consider

that the two separate pieces of information are  $T(z)$  and  $S(T)$ . Provided we do not have temperature inversions, then this information will always allow  $S(z)$  to be reconstructed. Other considerations suggest that  $S(T)$  and  $T(z)$  are much more independent than are  $S(z)$  and  $T(z)$  since  $T(z)$  and  $S(z)$  profiles are both strongly influenced by fast dynamical processes, whereas  $S(T)$  is only affected by slow thermodynamic processes. The separation here is therefore a physically based choice and provides an alternative way of representing the relationship between  $S$  and  $T$  (see section 5 for more on this).

Assume that at a particular location at level  $z$  the temperature analysis has already been performed and has yielded the temperature  $T_a$ . What we would really like to be able to do when assimilating the salinity data is

$$S_a(T_a) = S_b(T_a) + \mathbf{K}_3[S_O(T_a) - HS_b(T_a)]. \quad (5)$$

Notice that this is almost the same as Eqs. (3) and (4) because  $S_b(T_a)$  and  $S'_a(T_a)$  are the same; however, now we are calculating salinity innovations on the same isotherm. Written in this way, fully in temperature coordinates, the orthogonality with the TH99 scheme is clear, with Eq. (5) correcting an entirely different aspect of the salinity field error. This elegant property means that the appropriate gain  $\mathbf{K}_3$  can be set without reference to the  $\mathbf{K}_1$  used in the temperature assimilation. As we argued in section 2, it is entirely appropriate that we take into account the different decorrelation scales and representivity errors associated with  $S(T)$  data in setting  $\mathbf{K}_3$ . The absence of internal gravity wave impacts on  $S(T)$  in particular will give this quantity a lower representivity error than  $S(z)$  when assimilating into ocean models.

This temperature coordinate formulation seems to be a highly desirable approach to the assimilation of salinity data; however, in practice we may still need to express the result of Eq. (5) in  $z$ -level coordinates when applied to a  $z$ -level model. In  $z$ -level coordinates Eq. (5) implies a second salinity increment that is additive to the first of Eq. (2). This can be defined as

$$\Delta S_S[z(T_a)] = \mathbf{K}_3[S_O(T_a) - HS_b(T_a)] = \mathbf{K}_3 \delta S(T_a), \quad (6)$$

in addition to the first salinity increment  $\Delta S_T(z)$  associated with temperature assimilation. Notice that this is different to the increment from Eq. (4). The novel step is that the salinity increments should be calculated on the same  $T$  surface and not the same  $z$  level. Although the final salinity increments will be based on analysis isotherms  $T_a$ , in Eq. (6), there will be a very wide range of analyzed model temperatures covering the entire spatial area affected by each observation. Therefore, as

is usual in assimilation, the first step is to calculate innovations in observation space, and that means on the observation isotherms  $T_O$  in the observed profile,

$$\delta S(T_O) = [S_O(T_O) - HS_b(T_O)]. \quad (7)$$

This then compares the observed salinity not with the model background at the same  $z$  level but with the model background salinity projected to each observed isotherm at the location of the profile. This is the key advance.

### c. Determining the gain for $S(T)$ assimilation

The following discussion gives some practical ideas of how to go from the observed innovations in Eq. (7) to the applied increments in Eq. (6). First we address the choice of isotherm. Even with a high-resolution observation profile it makes sense to store a finite number of salinity innovations  $\delta S(T_O)$  for a finite number of isotherms  $T_O$ , probably similar to the number of model levels from which independent background data are available. This set can then be used to infer salinity increments on all intermediate isotherms. This could be done by interpolation:

$$\begin{aligned} \delta S(T_a) = & \delta S(T_O^-) + (T_a - T_O^-)[\delta S(T_O^+) \\ & - \delta S(T_O^-)]/(T_O^+ - T_O^-), \end{aligned} \quad (8)$$

where  $T_O^+$  and  $T_O^-$  are the temperatures above and below  $T_a$ , on which the salinity increments are known. Alternatively, if only a single  $\delta S(T_O)$  value is available to infer  $\delta S(T_a)$  then one could modify the correlation function in the gain matrix by writing, for example,

$$\mathbf{K}_3 \sim \exp(-r^2/2R^2) \exp[-(T_a - T_O)^2/2T_R^2], \quad (9)$$

where the  $r^2/2R^2$  factor is the normal decay of influence with horizontal distance, but now  $T_R$  is an additional “temperature scale” that determines how salinity measurements on one temperature surface should influence salinity on another temperature surface. This dependency of the salinity covariances on another state variable, temperature, contains a similar idea to that suggested by Riishojgaard (1998) in the meteorological assimilation context. The implementation in ECMWF seasonal forecasting system 2, where assimilation has been coded level by level, has required this approach, and Eq. (9) is used for the results in section 4, with  $R = 400$  km in midlatitudes (somewhat larger than for  $T$  assimilation) and  $T_R = 2^\circ\text{C}$ .

We now discuss other reasons for differences between  $\mathbf{K}_3$  in Eq. (6) and  $\mathbf{K}_1$  in Eq. (1). The first is that  $S(T)$  measurements should be usable with a smaller representivity error than measurements of  $S(z)$ . This is

because  $S(T)$  is not altered by passing gravity, inertial, and Rossby wave oscillations that cause variability on short time scales. This is reflected in the time variability and the power distribution in the two time series in Fig. 2. The second reason lies in the larger decorrelation scales for  $S(T)$  measurements relative to  $S(z)$  measurements, as discussed in section 2b. The  $S(T)$  gain  $\mathbf{K}_3$  falls off more slowly as we move away from an observation profile because the  $R$  in Eq. (9) can be larger than the equivalent term in the definition of  $\mathbf{K}_1$ . These two effects are related because it is clear that the expected a priori spatial scale for  $S(z)$  and  $T(z)$  data in midlatitudes is the Rossby deformation radius, reflecting the rapidly varying ocean mesoscale. However, for  $S(T)$  data the mesoscale eddies have only a small signature, just as the gravity and inertial oscillations have a very small signature on  $S(T)$ . The formulation of  $S(T)$  assimilation is therefore very appropriate for assimilating scarce historical salinity data (pre Argo) and may also point the way to better methods for assimilating other Lagrangian tracers whose variability could reasonably be determined to first order from the temperature field. We leave to the discussion at the end of the paper the relative merits of using isotherms rather than isopycnals as the basis for the assimilation.

In the next section we go on to look at the results of salinity analyses using the scheme proposed in section 3b.

## 4. Implementation of $S(T)$ assimilation at ECMWF

To demonstrate the application of the  $S(T)$  assimilation method, it has been implemented within the ECMWF seasonal forecasting system. In this section, we focus on the results of the two separate salinity assimilation increments, associated with temperature and salinity data, and show the differences and also demonstrate that both are required in order to achieve a good analysis. The results of applying this scheme repeatedly in a model integration are left to a future publication.

The ocean model used at ECMWF is the Hamburg Ocean Primitive Equation model (HOPE). The model uses an Arakawa E grid with a horizontal resolution of  $1^\circ \times 1^\circ$  (latitude–longitude) plus a refinement to  $0.3^\circ$  meridionally within  $10^\circ$  of the equator. In the vertical direction, there are 29 levels, 21 of which are in the upper 425 m. Vertical mixing uses a Richardson number-dependent diffusivity based on Pacanowski and Philander (1981). The model is forced by daily average momentum, heat, and freshwater fluxes taken from the ECMWF atmospheric analysis system. More details of the model setup can be found in Anderson et al. (2003).

The model was initialized from climatological  $T$  and

$S$  from the *World Ocean Atlas 1998* (WOA98; Levitus et al. 1998) data, was run for a period of 5 years forced by climatological forcing, and then was run for 20 more years until 4 August 2002, forced by the ECMWF 40-yr reanalysis (ERA-40) and ECMWF operational fluxes with a weak relaxation (3 yr) to WOA98 subsurface  $T$  and  $S$ . At this point a first assimilation of both temperature and salinity data was carried out. This period was chosen because by this time the data from the Argo float network had become extensive, giving considerably more salinity data than were available in previous periods. The implementation of salinity adjustments,  $\Delta S_T(z)$ , as part of the temperature assimilation process was already available as part of the operational seasonal forecasting suite (Troccoli et al. 2002). The new salinity assimilation scheme was implemented as described in section 3b above, with the covariance function formulation in Eq. (9), with  $T_R = 1^\circ\text{C}$ .

Figure 5 shows for a single water column just south of the equator in the Indian Ocean how the salinity is altered by data assimilation. For convenience this observation profile is fairly isolated, and we analyze the model changes at the location of the profile so that the assimilation impact is not reduced by distance. In addition, the error variance associated with the observations is made small so that the assimilation is made to reproduce the observation profile. The black line shows the model background, and the dashed line shows the observed salinity profile. These two profiles have some level of agreement at the surface, at 150–450-m depth, and at 1000 m but largely disagree elsewhere. The light gray line shows the profile resulting from temperature data assimilation using the  $S(T)$ -preserving salinity increment  $\Delta S_T(z)$ . It can be seen that this gives a clear improvement in the salinity profile over the 150–450-m depth range but does not have a significant impact outside this depth range. The dark gray line shows the profile after also applying the  $\Delta S_S(z)$  salinity increment, and this now agrees very well with the observed profile at all depths. Note that both assimilation increments are needed for different parts of the water column. Over most of the water column  $\Delta S_S(z)$  is dominant, but within the 150–450-m depth range the  $\Delta S_T(z)$  increment is doing most of the work. The real advantage of separating the increments would come if we were to apply different treatments of the error covariances and hence different gains for the two components, both at the observation profile and as one moves off into the surrounding water.

Figure 6 shows the impact of the two salinity increments as spatial fields over the top 300 m of the water column over the whole globe. In this case the errors associated with the observations are chosen to give the

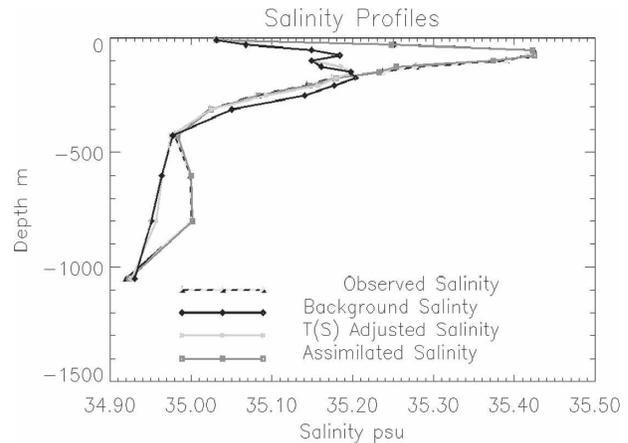


FIG. 5. Salinity profiles at a salinity assimilation point near the equator in the Indian Ocean. The  $T(S)$  adjusted salinity is the profile that can be deduced from temperature assimilation alone; the dark gray line, marked assimilated salinity, shows the result of both the  $T(S)$  adjustment and the salinity assimilation itself.

more typical 50% weighting to the new data. Notice that the increments associated with salinity assimilation are considerably larger than those associated with temperature assimilation outside the tropical band, and they also tend to influence a larger spatial area. We suggest that this is indicative of the poor salinity of the background model state and the poor  $S(T)$  relationships within that state. The surface forcing of salinity is probably poorly represented in the model, and also the WOA98 initial conditions are a poor representation of actual salinity conditions in August 2002. Notice also that there is no obvious spatial correspondence between the two salinity increments shown in Fig. 6. If the correlation coefficient between the two salinity increments is calculated, the value is very low:  $<0.1$ . This emphasizes the fact that the two salinity increments are correcting for very different processes causing variation in the error field of the model background salinity state. This is further justification of the suggestion of this paper that the salinity assimilation be separated in this way.

## 5. Conclusions and discussion

In this paper we have first explored some of the variability present in the observational record of ocean salinity and also in salinity within ocean circulation models. We have demonstrated that the different processes involved in controlling salinity measured on a depth surface  $z$  and salinity measured on a temperature surface  $T$  are reflected in the different amplitude, space, and time scales that dominate the salinity variability on  $z$  and  $T$  surfaces. In particular, salinity variability on  $T$

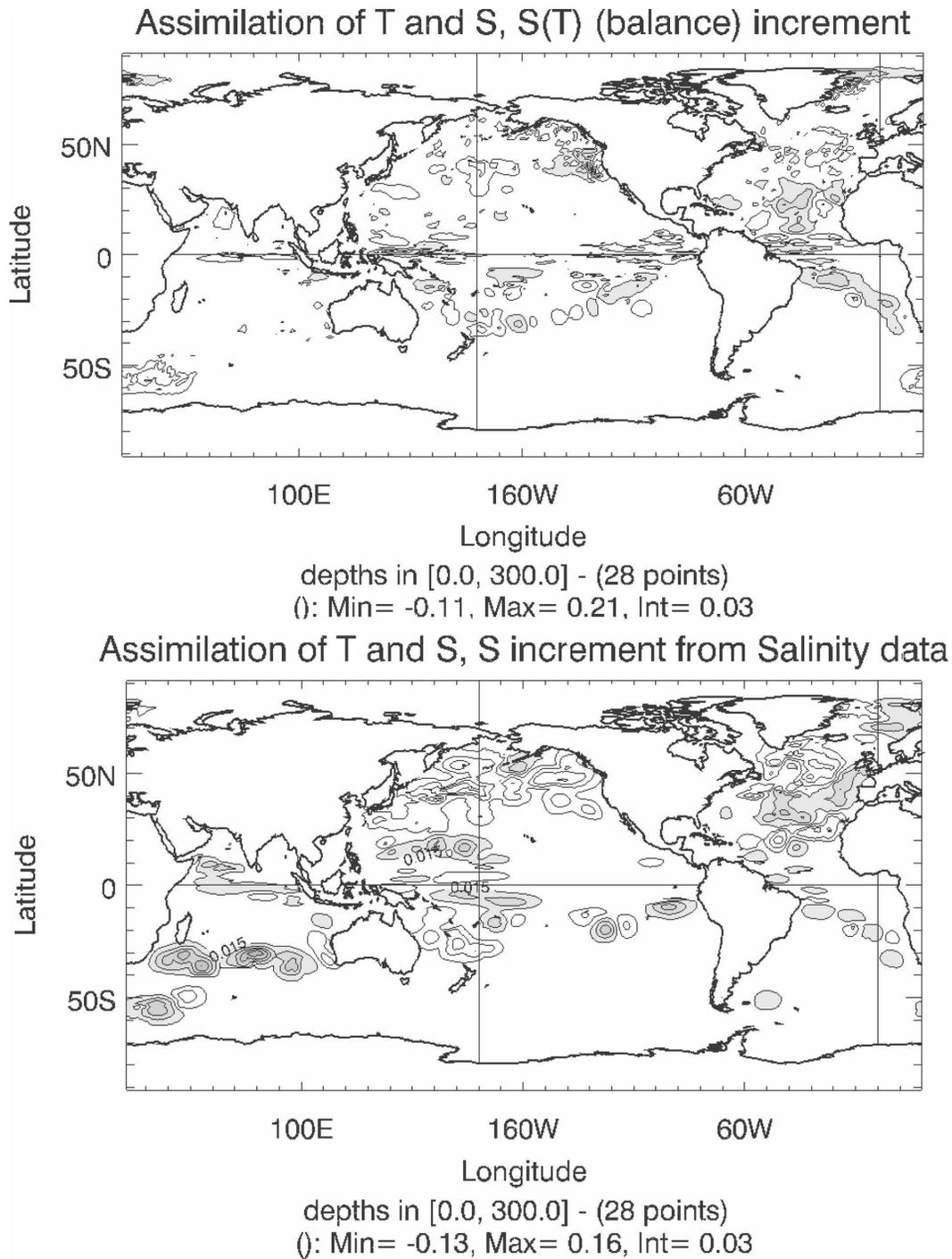


FIG. 6. (top) The salinity increments  $\Delta S_T$  over the top 300 m of the water column on the first assimilation step, deduced from the temperature assimilation using preservation of the  $S(T)$  relationship within the model background. (bottom) The salinity increments  $\Delta S_S$  over the top 300 m of the water column deduced from the salinity observations calculated to update the background  $S(T)$  relationship. Contour interval is 0.03 psu, starting at 0.015 psu; shaded contours are positive.

surfaces shows smaller amplitudes and larger spatial and temporal correlations than does salinity variability on a  $z$  surface, at least for midlatitudes and lower latitudes and at depths with a well-stratified water column.

The paper then uses these results to develop a new data assimilation algorithm for application to salinity data. A two-stage salinity assimilation algorithm is advocated in which the first salinity increment  $\Delta S_T$  allows the  $S(T)$  functional relationship to remain fixed when temperature data are assimilated. This first increment was already advocated in TH99. A second assimilation increment  $\Delta S_S$  is derived here using the observational salinity itself, and this allows the  $S(T)$  of the model to be updated. The earlier results on variability suggest that an improved gain can be used for deriving the  $\Delta S_S$  increments, thus extracting more information from the measurements and increasing the influence of the scarce salinity observations on the final analyzed model fields. Some preliminary test results from implementation in the ECMWF seasonal forecasting model are also shown.

It might be argued that a better physical separation to impose upon the two hydrographic variables  $T$  and  $S$  would be to assimilate potential density  $\rho(z)$  [or even  $z(\rho)$ ] and then to assimilate  $S(\rho)$  or equivalently  $T(\rho)$ , ignoring cabbeling effects. This probably would be a better separation, but it ignores the fact that at least historically we have far more  $T$  data than  $S$  data and therefore we simply do not have  $\rho$  observations in many cases. In the future, with Argo coming to dominate the hydrographic record, we may seek to implement a scheme focused on the potential density variable.

A more common approach to accounting for the co-variability of  $S$  and  $T$  during assimilation would be to use an a priori (i.e., state independent but perhaps spatially and temporally variable) background error covariance matrix reflecting the appropriate level of correlation. However this approach has its own weaknesses. It is a large job to calculate, store, and retrieve spatially and temporally varying error covariances. It is also questionable whether such a priori covariances can be realistic if they are fixed to an Eulerian reference frame and are not flow dependent. Derber and Bouttier (1999) have presented a generalized framework for transforming background error covariances, which would allow for a level of flow dependence; however, it is not clear whether their framework could be extended to consider the salinity assimilation method presented here. In this paper the salinity innovation itself must be calculated from the beginning in the temperature coordinate reference frame, as in Eq. (7), and this information cannot be recovered by a linear transformation of the Eulerian salinity innovation. That said, it would cer-

tainly be worthwhile exploring whether the insights obtained here with a relatively simple OI-type assimilation method and, in particular, the advantages of the improved gain matrices could be transferred to be used in a more advanced data assimilation scheme.

*Acknowledgments.* This paper benefited from the close scrutiny and advice from one of the reviewers. All members of the seasonal forecasting group at ECMWF have provided helpful comments in discussion. We acknowledge the EU under project ENACT and the NERC e-Science program for supporting this work. We also acknowledge the help of BADC in getting access to the Hadley Centre model datasets.

#### REFERENCES

- Anderson, D., and Coauthors, 2003: Comparison of the ECMWF seasonal forecast systems 1 and 2, including the relative performance for the 1997/98 El Niño. ECMWF Tech. Memo. 404, 93 pp.
- Bindoff, N. L., and T. J. McDougall, 1994: Diagnosing climate change and ocean ventilation using hydrographic data. *J. Phys. Oceanogr.*, **24**, 1137–1152.
- Bryden, H. L., E. L. McDonagh, and B. A. King, 2003: Changes in ocean water mass properties: Oscillations or trends? *Science*, **300**, 2086–2088.
- Cooper, M. C., and K. Haines, 1996: Data assimilation with water property conservation. *J. Geophys. Res.*, **101** (C1), 1059–1077.
- Cooper, N. S., 1988: The effect of salinity in tropical ocean models. *J. Phys. Oceanogr.*, **18**, 1333–1347.
- Curry, R., B. Dickson, and I. Yashayaev, 2003: Ocean evidence of a change in the freshwater balance of the Atlantic over the past four decades. *Nature*, **426**, 826–829.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Derber, J., and F. Bouttier, 1999: A reformulation of the background error covariance in ECMWF global data assimilation system. *Tellus*, **51A**, 195–221.
- Dickson, B., R. Curry, and I. Yashayaev, 2003: Recent changes in the North Atlantic. *Philos. Trans. Roy. Soc. London*, **A361**, 1917–1934.
- Fox, A. D., and K. Haines, 2003: Interpretation of water transformations diagnosed from data assimilation. *J. Phys. Oceanogr.*, **33**, 485–498.
- Gordon, C., C. Cooper, A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST sea ice extent and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **16**, 147–168.
- Haines, K., 1994: Dynamics and data assimilation in oceanography. *Data Assimilation: Tools for Modelling of the Ocean in a Global Change Perspective*, P. P. Brasseur, and J. C. J. Nihoul, Eds. NATO Series 1, Vol. 19, Springer-Verlag, 1–32.
- , 2003: Ocean data assimilation. *Data Assimilation for the Earth System*, R. Swinbank, V. Shutyaev, and W. Lahoz, Eds., NATO ASI Series, Vol. 26, Kluwer Academic, 289–320.
- Ingleby, B., and M. Huddleston, 2005: Quality control of ocean temperature and salinity profiles—Historical and real-time data. *J. Mar. Syst.*, in press.

- Iselin, C. O., 1939: The influence of vertical and lateral turbulence on the characteristics of the waters at mid-depths. *Trans. Amer. Geophys. Union*, **20**, 414–417.
- Joyce, T. M., and P. Robbins, 1996: The long-term hydrographic record at Bermuda. *J. Climate*, **9**, 3121–3131.
- Levitus, S., 1989: Interpentadal variability of temperature and salinity at intermediate depths in the North Atlantic Ocean, 1970–1974 versus 1955–1959. *J. Geophys. Res.*, **94**, 6091–6131.
- , and Coauthors, 1998: *Introduction*. Vol. 1, *World Ocean Database 1998*, NOAA Atlas NESDIS 18, 346 pp.
- Maes, C., and D. Behringer, 2000: Using satellite-derived sea level and temperature profiles for determining the salinity variability: A new approach. *J. Geophys. Res.*, **105**, 8537–8548.
- Pacanowski, R. C., and S. G. H. Philander, 1981: Parameterization of vertical mixing in numerical models of tropical oceans. *J. Phys. Oceanogr.*, **11**, 1443–1451.
- Ricci, S., A. T. Weaver, J. Vialard, and P. Rogel, 2005: Incorporating state-dependent temperature–salinity constraints in the background error covariance of variational ocean data assimilation. *Mon. Wea. Rev.*, **133**, 317–338.
- Riishojgaard, L. P., 1998: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus*, **50A**, 42–57.
- Roberts, M., and Coauthors, 2004: Impact of an eddy-permitting ocean resolution on control and climate change simulations with a global coupled GCM. *J. Climate*, **17**, 3–20.
- Roemmich, D., M. Morris, W. R. Young, and J. R. Donguy, 1994: Fresh equatorial jets. *J. Phys. Oceanogr.*, **24**, 540–558.
- , and Coauthors, 2001: Argo: The global array of profiling floats. *Observing the Oceans in the 21st Century*, C. J. Kobalinsky and N. R. Smith, Eds., Bureau of Meteorology, 248–257.
- Tomczack, J. R., 1981: A multi-parameter extension of temperature/salinity diagram for the analysis of non-isopycnal mixing. *Progress in Oceanography*, Vol. 10, Pergamon, 147–171.
- Troccoli, A., and K. Haines, 1999: Use of the temperature–salinity relation in a data assimilation context. *J. Atmos. Oceanic Technol.*, **16**, 2011–2025.
- , and Coauthors, 2002: Salinity adjustments in the presence of temperature data assimilation. *Mon. Wea. Rev.*, **130**, 89–102.
- Vialard, J., and P. Delecluse, 1998a: An OGCM study for the TOGA decade. Part I: Role of the salinity in the physics of the western Pacific fresh pool. *J. Phys. Oceanogr.*, **28**, 1071–1088.
- , and —, 1998b: An OGCM study for the TOGA decade. Part II: Barrier layer formation and variability. *J. Phys. Oceanogr.*, **28**, 1089–1106.
- Vossepoel, F. C., and D. W. Behringer, 2000: Impact of sea level assimilation on salinity variability in the western equatorial Pacific. *J. Phys. Oceanogr.*, **30**, 1706–1721.
- Wong, A. P. S., N. L. Bindoff, and J. C. Church, 1999: Large scale freshening of intermediate waters in the Pacific and Indian Oceans. *Nature*, **400**, 440–443.