



The ECMWF Ocean Analysis System: ORA-S3

Magdalena Balmaseda, Arthur Vidard, David Anderson

► To cite this version:

Magdalena Balmaseda, Arthur Vidard, David Anderson. The ECMWF Ocean Analysis System: ORA-S3. Monthly Weather Review, American Meteorological Society, 2008, 136 (8), pp.3018-3034. 10.1175/2008MWR2433.1 . inria-00182186

HAL Id: inria-00182186

<https://hal.inria.fr/inria-00182186>

Submitted on 25 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The ECMWF Ocean Analysis System: ORA-S3

MAGDALENA A. BALMASEDA *

ECMWF, READING, UK

ARTHUR VIDARD

INRIA, GRENOBLE, FRANCE

DAVID L.T. ANDERSON

ECMWF, READING, UK

* *Corresponding author address:* Magalena Alonso Balmaseda, European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK.
E-mail: neh@ecmwf.int

ABSTRACT

A new operational ocean analysis/reanalysis system (ORA-S3) has been implemented at ECMWF. The reanalysis started from 01/01/1959: it is continuously maintained up to 11 days behind real time and is used to initialize seasonal forecasts as well as to provide a historical representation of the ocean for climate studies. It has several innovative features, including an on-line bias-correction algorithm, the assimilation of salinity data on temperature surfaces and assimilation of altimeter-derived sea level anomalies and global sea level trends. It is designed to reduce spurious climate variability in the resulting ocean reanalysis due to the non-stationary nature of the observing system, while still taking advantage of the observation information. The new analysis system is compared with the previous operational version; the equatorial temperature biases are reduced and equatorial currents are improved. The impact of assimilation in the ocean state is discussed by diagnosis of the assimilation increment and bias correction terms. The resulting analysis not only improves the fit to the data, but also improves the representation of the interannual variability.

In addition to the basic analysis, a real-time analysis is produced (RT-S3). This is needed for monthly forecasts and in future may be needed for shorter-range forecasts. It is initialised from the near-real-time ORA-S3 and run each day from it.

1. Introduction

A new operational ocean analysis system has been implemented at ECMWF to provide initial conditions for seasonal forecasts. In what follows we will refer to this system as ORA-S3. The ORA-S3 ocean analysis extends back to 1959, and can be regarded as a historical ocean re-analysis which is continuously updated on a daily basis. This ocean re-analysis is used to initialize the calibrating hindcasts of the ORA-S3 seasonal forecasting system (Anderson *et al.*, 2007) for the period 1981-2005. The earlier period of ORA-S3 ocean analysis will be used to initialize seasonal and decadal predictions within the ENSEMBLES project. As well as providing initial conditions for coupled model forecasts, the ORA-S3 ocean re-analysis, based on the synthesis of surface and subsurface ocean observations, surface fluxes from atmospheric analyses and reanalyses, and a general circulation ocean model, constitutes an important resource for climate variability studies (Balmaseda *et al.* 2007a). The ocean analysis system also has a real time component (RT-S3), used to initialize the monthly forecasts. It is envisaged that in the near future it will also be used to initialize coupled medium range weather forecasts.

Since 1997, ECMWF has produced operationally daily global ocean analyses to provide initial conditions for the seasonal forecasting system. There have been two versions of the ocean analysis, linked to the operational seasonal forecasting system. System 1 (ORA-S1) started in 1997 (Alves *et al.*, 2003) and provided the initial conditions for the first ECMWF operational seasonal forecasting system (Stockdale *et al.*, 1998). System 2 (ORA-S2) was introduced in 2001 (Balmaseda 2004), and has provided initial conditions for the ECMWF operational seasonal forecasts since 2002 (Anderson *et al.* 2003, Oldenborgh *et al.* 2005a,b, Vialard *et al.* 2005). A comparison between ORA-S2 and ORA-S1 ocean analyses is given in Balmaseda 2004. In 2004 an extension of ORA-S2 was introduced in order to initialize the monthly forecasting system (Balmaseda 2005, Vitart 2005). In summer 2006 the new ocean analysis system, consisting of the ORA-S3 ocean re-analysis and the real-time counterpart (RT-S3) was implemented operationally, and used to initialise the seasonal forecasts from

March 2007, and the monthly forecasts from June 2007.

This paper describes the different components of the ORA-S3 ocean analysis system, and compares its performance with the previous operational system (ORA-S2). The impact of data assimilation on the mean state and internannual variability is also discussed, with emphasis on the tropical oceans. A more detailed description of the climate variability in ORA-S3 will be presented in a separate paper. This paper is organized as follows: section 2 offers a summary of the new features in ORA-S3 compared to ORA-S2; section 3 describes the observational data sets and the quality control procedure; section 4 provides a detailed description of the assimilation algorithm and analysis step; in section 5 the impact of assimilation is assessed by considering the assimilation increment, the bias correction term and the fit to the data; the RT-S3 real-time extension is presented briefly in section 6.

2. Summary of new features: ORA-S3 versus ORA-S2

The ocean data assimilation system for ORA-S3 is based on the HOPE-OI scheme: The first guess is given by forcing the ocean model with daily fluxes of momentum, heat, and fresh water, while the observations are assimilated using an Optimal Interpolation (OI) scheme.

Several modifications to the original HOPE ocean model (Wolff *et al.* 1997) have taken place over the years at ECMWF (Balmaseda 2004, Anderson and Balmaseda 2005): the horizontal resolution was increased to $1^\circ \times 1^\circ$ with equatorial refinement, i.e., the meridional resolution increases gradually towards the equator, where it is 0.3° in the meridional direction; there are 29 levels in the vertical, with a typical vertical thickness of 10 meters in the upper ocean; the vertical mixing is based on Peters *et al.*, 1998; the barotropic solver, originally implicit, was made explicit as described in Anderson and Balmaseda (2005).

In ORA-S3, major upgrades have been introduced in the assimilation system. In addition to subsurface temperature, altimeter derived sea-level anomalies and salinity data are now assimilated. All the observations in the upper 2000m are assimilated whereas in ORA-S2 only

the observations in the upper 400m were used. In ORA-S3, the observations come from the quality controlled dataset prepared for the ENACT and ENSEMBLES projects until 2004 (Ingleby and Huddleston 2006), and from the GTS thereafter (ENACT/GTS), whereas in ORA-S2 the data were from GTSP (Global Temperature and Salinity Pilot Project). More details about the data used in ORA-S3 are given in the following section. The OI scheme is now 3-dimensional, the analysis being performed at all levels simultaneously down to 2000m, whereas in ORA-S2, the analysis was carried out on each model level independently and only to 400m. In addition, the decorrelation scales depend on the density gradient, which favours the propagation of information along isopycnals.

The first-guess is obtained by integrating the ocean model from one analysis time to the next, forced by ERA40 fluxes for the period from January 1959 to June 2002 and NWP operational analyses thereafter. In ORA-S2, the fluxes were from ERA15/OPS, but the wind stresses were not directly used: instead, the wind stress was derived from the analyzed winds using an off-line bulk formula. The representation of the upper ocean interannual variability is improved when using the ERA40 wind stress (Uppala *et al.*, 2006), although the stresses are biased weak in the equatorial Pacific. The fresh water flux from ERA-40 (Precipitation - Evaporation, denoted P-E) is known to be inaccurate. ORA-S3 uses a better estimate, obtained by 'correcting' the ERA-40 precipitation values (Troccoli and Kallberg 2004). A pictorial view of the various data sets used in ORA-S3 is given in fig 1.

When designing a data assimilation system for seasonal forecasts several considerations need to be taken into account. It is important to represent the interannual/decadal variability in the ocean initial conditions, and therefore strong relaxation to climatology is not desirable. On the other hand, spurious trends and signals due to the non-stationary nature of the observing system should be avoided as much as possible. It is also important to avoid large initialization shocks in the coupled model, which may damage the forecast skill. In ORA-S3 we have tried to strike a balance between the above requirements: the weight to observations has been reduced and the relaxation to climatology is considerably weaker than in ORA-S2.

This has been possible because an additive bias correction has been included (Balmaseda et al 2007b). The bias-correction consists of a prescribed a-priori correction to temperature, salinity and pressure gradient, as well as a time-dependent bias term estimated on-line, which acts mainly on the pressure gradient. The on-line bias correction is adaptive and allows for flow-dependent errors.

An important feature of the ECMWF ocean analysis system is that not just a single analysis but five simultaneous analyses are performed. The purpose of the multiple analyses is to sample uncertainty in the ocean initial conditions and to contribute to the creation of the ensemble of forecasts for the probabilistic predictions at monthly and seasonal ranges. The five simultaneous ocean analyses are created by adding perturbations to the wind stress, commensurate with the estimated uncertainty in the wind stress analysis, while the model is being integrated forward from one analysis time to the next. The wind perturbations have been revised in ORA-S3 relative to those used in ORA-S2 to better represent the perceived uncertainty in ERA40/OPS wind stress. Further information is available at http://www.ecmwf.int/research/EU_projects/ENSEMBLES/documents/docu_perturbations.pdf.

3. The data sets used, Observation coverage and Automatic Quality Control

As shown in fig 1, the temperature and salinity profiles prior to 1st of January 2005 come from the ENACT/ENSEMBLES quality-controlled data set (Ingleby and Huddleston 2006). From January 2005, the observations come from the GTS (Global Telecommunication System). An automatic quality control procedure is used with several stages: first, daily averages are created if applicable as is the case for the TRITON moorings which report hourly. Data close to the coast are rejected. A level-by-level quality control is then performed as in Smith *et al.*, 1991, which checks both the distance between model values and observations in

relation to the error statistics as well as the consistency between observations by means of a buddy-check. Profiles which are close in space and time are superobbed, following the same criteria as in Smith *et al.*, 1991. In ORA-S3 there is an additional check for completeness of the profiles: a profile is considered incomplete, and therefore rejected, if the sparsity of the remaining observations in the vertical is judged insufficient to resolve the vertical temperature gradients. An observation profile will be rejected if the temperature difference between consecutive levels is larger than $5K$ or if it contains a vertical temperature gradient larger than $0.1K/m$.

Figure 2 shows a typical recent distribution of profiles in a 10-day window, in this case for the window centered around the 9th of January, 2007. Also shown are the results of the automatic quality control decisions. Maps of observation coverage throughout the reanalysis period can be seen at <http://www.ecmwf.int/products/forecasts/d/charts/ocean/reanalysis/obsmap>. Figure 2 shows that as a result of the Argo system, the observation coverage is now almost global. The figure also shows a few XBT tracks (only reporting temperature), and the TRITON/PIRATA/TAO moorings. Only the TRITON moorings in the West Pacific and Indian ocean provide salinity in real time. The PIRATA and TAO moorings report only temperature in real time, even though the sensors measure salinity. Most of the data from XBTs and Mooring are superobbed. Figure 2 shows a considerable number of Argo profiles partially rejected by our QC system in the Atlantic ocean. Many of these partially rejected floats correspond to SOLO/FSI instruments, later reported to have faulty sensors (<http://www.arago.ucsd.edu>).

The analysis of SST is not produced using the OI-Scheme. Instead, the model SSTs are strongly relaxed to analyzed SST maps which are daily interpolated values derived from the OIv2 SST product (Smith and Reynolds 1998, Reynolds et al 2002) from 1982 onwards. Global SST maps from NCEP are received every Monday at midday, representing the average of the previous week SST values. Daily SST maps are obtained by interpolation of the weekly products. Prior to 1982, the same SST product as in the ERA40 reanalysis was used.

The altimeter data consists of twice-weekly maps from AVISO. Two consecutive maps are interpolated in time to produce daily maps. Only the daily map corresponding to the center of the assimilation window is assimilated.

4. ORA-S3 Data Assimilation System

In ORA-S3, the different data streams are assimilated sequentially, as illustrated by the following scheme:

$$\begin{array}{c} \begin{pmatrix} \tilde{T}^b \\ \tilde{S}^b \\ \tilde{\eta}^b \\ \tilde{u}^b \end{pmatrix} \end{array} \xrightarrow[\text{bias}]{\text{remove}} \begin{array}{c} \begin{pmatrix} T^b \\ S^b \\ \eta^b \\ \vec{u}^b \end{pmatrix} \end{array} \xrightarrow[\eta^{t^o}]{\text{alti}} \begin{array}{c} \begin{pmatrix} T_1^a \\ S_1^a \\ - \\ - \end{pmatrix} \end{array} \xrightarrow[T^o]{\text{Temp}} \begin{array}{c} \begin{pmatrix} T_2^a \\ S_2^a \\ - \\ - \end{pmatrix} \end{array} \xrightarrow[S^o]{\text{Sal}} \begin{array}{c} \begin{pmatrix} - \\ S_3^a \\ - \\ - \end{pmatrix} \end{array} \xrightarrow[\text{SL-trend}]{\text{Geostr}} \begin{array}{c} \begin{pmatrix} T^a \\ S^a \\ \eta^a \\ \vec{u}^a \end{pmatrix} \end{array}$$

First, the model background $(\tilde{T}^b, \tilde{S}^b, \tilde{\eta}^b, \tilde{u}^b)$ is bias-corrected according to the scheme described in section a. Then, the detrended altimeter-derived sea level anomalies (η^{t^o}) are combined with the bias-corrected model first-guess $(T^b, S^b, \eta^b, \vec{u}^b)$ using the Cooper and Haines 1996 scheme (CH96 hereafter) to produce a first analysis $(T_1^a, S_1^a, \eta_1^b, -)$. This analysis is then used as a first guess for a second assimilation step, where only subsurface temperature data T^o are assimilated, and salinity is updated by imposing conservation of the model temperature/salinity (T/S) relationship, while the sea level and velocity fields remain unchanged. This second analysis is denoted as $(T_2^a, S_2^a, -, -)$. In a third assimilation step, the information provided by the salinity observations S^o is used to modify the model T/S relationship. In this step, the T/S information is spread along isotherms following the scheme of Haines *et al.*, 2006. Only salinity is modified in this step which results in the analysis $(-, S_3^a, -, -)$. After this 3rd assimilation step, velocity updates are derived from the temperature and salinity increments imposing geostrophic balance (Burgers *et al.*, 2002). Finally, the trend in global (area averaged) sea level is assimilated. By combining the altimeter-derived trend

in global sea level with the model trend in global dynamic height, it is possible to make the partition between changes in the global volume and changes in the total mass. By doing so, the global fresh water budget is closed and the global surface salinity and sea level adjusted accordingly. Each of the steps is described further below.

The analysis is performed every 10 days. All the observations within a centered 10-day window are gathered and quality controlled. Analysis increments in temperature, salinity and velocity are calculated using the procedure outlined above. To avoid exciting gravity waves, and to allow the model dynamics to adjust gradually to the changes in the density field, an Incremental Analysis Update (IAU) method (Bloom *et al.*, 1996) is used: the increment is added slowly over the subsequent 10 days (IAU-10), after which a new background field is available, and the cycle repeated.

a. Bias-correction algorithm

The presence of bias in an ocean data assimilation scheme is a serious obstacle to the reliable representation of climate by historical ocean reanalysis. The standard procedure to deal with systematic error in a data assimilation system is to augment the model state with a set of systematic error or bias variables. In sequential data assimilation, this approach requires two analysis steps: one for the bias estimation and a second for the state vector. Assuming that the bias is nearly constant in time, and that the bias error covariance matrix is proportional to the forecast error covariance matrix, with the proportionality constant small, the algorithm can be approximated so it only requires one analysis step, and thus the bias term can be updated at little extra cost (Dee 2005). However the requirement of proportionality between the bias and forecast covariance matrices is not generally appropriate since the bias and the model state vector can have different control variables and/or multivariate balance relationships.

An alternative approximation for the one-step bias correction algorithm for the general case has been implemented in ORA-S3, following Balmaseda *et al.*(2007b), which uses differ-

ent control variables for the bias and for the state vectors. The control variable of the bias vector is pressure gradient, while the state vector consists of the 3D temperature, salinity, velocity and sea level fields. The pressure gradient correction is derived from the analysis increments in temperature. The final temperature bias is the effect of applying the pressure gradient correction to the momentum equations. Balmaseda *et al.* (2007b) show that applying the correction in the pressure gradient improves the equatorial circulation, whereas applying the bias correction directly to the temperature field degrades the circulation.

Modifications have also been introduced in the equation for the time evolution of the bias, which is now described by a simple parametric model as in Balmaseda *et al.*, 2007b. The introduction of a memory term limits the influence in time of isolated or sporadic observations. A side effect is that the magnitude of the bias can be underestimated but to compensate for that, a constant term is also introduced. The constant term is not affected by the on-line estimation and has to be estimated *a priori*, preferably with independent information. The *a priori* term has the potential to provide a smoother analysis by preventing abrupt changes in the analysis associated with the introduction of new observing systems. In ORA-S3 the bias correction algorithm consists of an *a priori* correction to temperature, salinity and pressure gradient, as well as a time dependent bias estimated on-line. The on-line bias correction is only applied to the pressure gradient. The inclusion of an a-priori bias-correction term enabled the subsurface relaxation to climatology to be weakened, from a time scale of 18 months in ORA-S2 to 10-years in ORA-S3. Because of large uncertainties in the fresh water flux, the relaxation to climatology is stronger for surface salinity (approx 3-year time scale), but still considerably weaker than in the ORA-S2 analysis system, where it is approx. 6 months.

b. Assimilation of altimeter-derived sea level anomalies

In ORA-S3, altimeter data are assimilated for the first time in the ECMWF operational ocean analysis. The altimeter information is given by maps of merged satellite product,

provided by SSALTO/DUACS and distributed by AVISO, with support from CNES. Twice a week (on Wednesday and on Saturday mornings) $\frac{1}{3}^\circ \times \frac{1}{3}^\circ$ Maps of Sea Level Anomaly (MSLA) for a merged product combining all satellites (Envisat, Jason, Topex/Poseidon, ERS2, GFO) using optimal interpolation and accounting for long wavelength errors are produced (Le Traon *et al.*, 1998, Ducet *et al.*, 2000). Prior to assimilation, these maps are smoothed to remove features that are not resolved on the model grid. These are then interpolated in time to produce daily maps. Only the map corresponding to the centre of the assimilation window is used.

The anomaly maps (η^o) distributed by AVISO are relative to a 7 year mean (1st Jan 1993 to 31st Dec 1999). To enable comparison with the model background field (η^b), a reference mean sea level ($\bar{\eta}$) is required to compute the absolute mean sea level value:

$$\eta^o = (\eta^{lo} + \bar{\eta}). \quad (1)$$

In ORA-S3, $\bar{\eta}$ is the 7-year mean sea level from an ocean analysis spanning the period 1993-1999. The possibility of using a reference $\bar{\eta}$ derived from the GRACE gravity mission has also been explored, but it was found that sensitivities to the reference mean sea level are large, and could potentially introduce abrupt jumps in the analysis (Vidard *et al.*, 2007).

As the ocean model uses the Boussinesq approximation, the first guess can not represent steric changes in global volume. Therefore, prior to the assimilation, the global mean ($\langle \eta^o \rangle, \langle \eta^b \rangle$) is removed from both the model first-guess and observations. The difference in surface elevation $\delta\eta$ is then calculated as:

$$\delta\eta = (\eta^{lo} + \bar{\eta} - \langle \eta^o \rangle) - (\eta^b - \langle \eta^b \rangle). \quad (2)$$

The scheme used to project sea level increment information onto subsurface temperature and salinity consists of calculating the equivalent vertical displacement of the model water

column to a difference in surface elevation $\delta\eta$ between background and observations, subject to the constraint that the bottom pressure is not changed. Effectively, the CH96 vertical displacement of the water column translates into increments $(\delta T_1, \delta S_1)$ to the temperature and salinity fields.

$$\begin{pmatrix} \delta T_1 \\ \delta S_1 \end{pmatrix} = CH96(T^b, S^b, \eta^b, \eta'^o, \bar{\eta}) \quad (3)$$

from which a temporary analysis (T_1^a, S_1^a) is obtained:

$$\begin{pmatrix} T_1^a \\ S_1^a \end{pmatrix} = \begin{pmatrix} \tilde{T}^b \\ \tilde{S}^b \end{pmatrix} + \omega \begin{pmatrix} \delta T_1 \\ \delta S_1 \end{pmatrix}, \quad (4)$$

with ω a weighting factor depending on latitude:

$$\omega = \min(\alpha, \beta [\cos(lat)]^\gamma). \quad (5)$$

In ORA-S3, the values for α , β and γ are set to 0.3, 1 and 8 respectively.

c. OI assimilation of Temperature

ORA-S3 uses an Optimal Interpolation (OI) scheme for the assimilation for subsurface temperature. Originally derived from the scheme described by Smith *et al.*, 1991, it has evolved substantially: from the original univariate 2-dimensional OI scheme, where the analysis was performed on each model level separately, to a 3-dimensional scheme, where the analysis is performed at all levels simultaneously, with isopycnal formulation for the covariance matrices and a-posteriori multivariate updates of velocity and salinity. The OI interpolation is carried out on overlapping sub-domains of the model horizontal grid (in

order to reduce the cost of the matrix inversions). Where domains overlap, the analyses are blended together. The subdivisions of the globe into sub-domains depends on the observation distribution and is done so that the maximum number of observations within the domain is less than 500 (for ORA-S2, the maximum number of observations was 200).

In ORA-S3, the assimilation of subsurface temperature is performed with 3D domains down to 2000m. In ORA-S3 there are 2 novel aspects: the introduction of vertical scales and the introduction of a density-dependent term which favours isopycnal spreading of information. The background errors \mathbf{B}_{TT} are represented by Gaussian functions which are anisotropic and inhomogeneous:

$$\mathbf{B}_{TT} \sim e^{-\frac{1}{2} \left(\frac{|\Delta x|}{W_x} + \frac{|\Delta y|}{W_y} + \frac{|\Delta z|}{W_z} + \frac{|\Delta \rho|}{W_\rho} \right)} \quad (6)$$

The terms W_x, W_y in eq 6 represent the weighting given to the observations depending on the x and y distance respectively; within 4 degrees of the equator the correlation length scale in the E/W direction is 1000 km while in the N/S direction it is 150 km. The correlation scales change with latitude, such that polewards of 15 degrees, it is 300 km in all directions. Between the equatorial strip and the sub-tropics there is a smooth transition in correlation scales. Eqn 6 shows explicitly the introduction of vertical scales (W_z), which weight the influence of the observations depending on the vertical distance to the model grid point ($|\Delta z|$), as well as the introduction of a density dependent term W_ρ , which weights the observations depending on the background density difference between the observation location and the model grid point ($|\Delta \rho|$), thus preventing the spread of information across water masses with very different characteristics. In ORA-S3, W_ρ is 0.5 Kg/m^3 and W_z is depth-dependent, with the values proportional to the thickness of the model levels. Observation errors are assumed to be correlated in space and time, with a spatial correlation function with length scale of 2 degrees and a time correlation scale of 3 days. In general, the observations are given half the weight of the background, although the relative weight

given to the data compared to the background field varies with depth to account for an increase in uncertainty associated with the large gradients near the thermocline (see Balmaseda 2004). This function has been modified in ORA-S3 compared ORA-S2 so that near the thermocline, the observations in ORA-S3 are given less weight than in ORA-S2 (about four times less). This reduction of the weight given to the data near the model thermocline is especially important in the Equatorial Atlantic, where the model has a diffuse thermocline and the vertical stratification is weak: experimental results show that giving less weight to the observations improves the analysis in this region.

Fig.3 shows the assimilation increment coming from 3 single observations of temperature along the equator for an experiment using covariances as in ORA-S2 (top) and as in ORA-S3 (bottom). Contours are background temperature isotherms. One can see the effect of the two differences described above: the information is spread vertically in ORA-S3¹ and the spread is not uniform but depends on density. At 200E/210m where the water-masses are well stratified the increment is almost an ellipsoid whereas at 120W/100m it has a much more complicated shape.

The assimilation of subsurface temperature follows the equation:

$$\begin{pmatrix} T_2^a \\ S_2^a \end{pmatrix} = \begin{pmatrix} T_1^a \\ S_1^a \end{pmatrix} + \begin{pmatrix} \mathbf{K}_{TT} \\ \mathbf{K}_{ST} \end{pmatrix} (T^o - \mathbf{H}T_1^a), \quad (7)$$

where the gain matrix \mathbf{K}_{TT} , used to calculate the increments in T is :

$$\mathbf{K}_{TT} = \mathbf{B}_{TT} \mathbf{H}^T [\mathbf{H} \mathbf{B}_{TT} \mathbf{H}^T + \mathbf{R}_{TT}]^{-1}. \quad (8)$$

Following the assimilation of temperature data, the salinity field is updated through \mathbf{K}_{ST} . Strictly speaking \mathbf{K}_{ST} is not a matrix, but to simplify the notation we represent it as a gain matrix in eq (7). The salinity corrections are derived from the temperature increments

¹It looks like there is a vertical spread in the top panel but it is an artifact of the plotting. It is in fact confined to a model level.

by conserving the background water mass properties. The basic idea is to use locally the model background T/S relationship to reconstruct the salinity profile from temperature information only (Troccoli and Haines (1999), Troccoli *et al.*(2002)), hereafter TH99). It has been shown that temperature assimilation can 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. Once the T_1^a analysis has been carried out, salinity is updated following:

$$\begin{aligned} S_2^a(z) &= S_1^a(z) + \delta S_2(z) \\ \delta S_2^a(z) &= \mathbf{K}_{ST}(T_1^a(z), S_1^a(z), T_2^a(z)) \end{aligned} \quad (9)$$

where $\delta S_2(z)$ ensures that the T/S relationship is the same before and after the analysis of temperature, i.e.

$$S_2^a(T_2^a) = S_1^a(T_2^a). \quad (10)$$

The correction to S is not applied in the mixed layer or at higher latitudes where the hypothesis behind this scheme is less applicable. This is achieved by applying a latitude filter that reduces the effect of the scheme linearly to zero between 30° and 60° latitude.

d. Assimilation of salinity data on temperature surfaces

Another new feature of ORA-S3 is the assimilation of salinity data: with the recent development of the Argo network we now have an unprecedentedly good spatial coverage of salinity observations.

The conventional approach to assimilating salinity is to use covariance relationships formulated in (x, y, z) coordinates. However, by doing the analysis in (x, y, z) coordinates we are not taking advantage of the fact that the salinity increments in the TH99 scheme leave the salinity unchanged on T surfaces. Haines *et al.*, 2006 proposed an assimilation scheme

by which the temperature and salinity provide two separate pieces of information about the hydrographic structure of the ocean: the temperature information is used to correct the temperature and salinity field by preserving the T/S relationship, and the salinity information can be used to correct the model T/S relationship. They also propose a change of variable when assimilating salinity information; instead of using geographical coordinates, the salinity assimilation is done in temperature space. We refer to this scheme as S(T) in what follows, and the conventional scheme as S(z).

Let us assume that at a particular location at level z the temperature analysis has already been performed and has yielded the temperature T_2^a , and salinity analysis S_2^a by preserving the first-guess S/T relationship (CH99). Following the S(T) scheme, the salinity analysis is carried out as follows:

$$S_3^a(T_2^a) = S_2^a(T_2^a) + \mathbf{K}_{SS}[S^o(T_2^a) - \mathbf{H}S_2^a(T_2^a)] \quad (11)$$

Written in temperature coordinates, it is clear that the correction from the S(T) scheme is correcting an entirely different aspect of the salinity field error from the correction by TH99 (Haines et al 2006). This elegant property means that the appropriate gain \mathbf{K}_{SS} in the salinity assimilation and \mathbf{K}_{ST} in the temperature assimilation are additive, which is a nice feature.

In practice we need to express the result of eq (11) in z coordinates. In z level coordinates eq (11) implies a second salinity increment which is additive to the first salinity increment of eq (9), defined as

$$\begin{aligned} S_3^a(z) &= S^o(T_2^a(z)) + \delta S_3(T_2^a(z)) \\ \delta S_3(T_2^a(z)) &= \mathbf{K}_{SS}[S^o(T_2^a(z)) - \mathbf{H}S_2^a(T_2^a(z))]. \end{aligned} \quad (12)$$

The $\delta S_3(T_2^a(z))$ salinity increment is applied in addition to the first salinity increment $\delta S_2(T_2^a(z))$ associated with temperature assimilation. The spreading of the information

along isotherms is determined by the gain matrix \mathbf{K}_{SS} :

$$\mathbf{K}_{SS} = \mathbf{B}_{SS}\mathbf{H}^T [\mathbf{H}\mathbf{B}_{SS}\mathbf{H}^T + \mathbf{R}_{SS}]^{-1} \quad (13)$$

$$\mathbf{B}_{SS} \sim W_R(\Delta r) \exp(-(T_2^a - T^o)/W_T).$$

where $W_R(\Delta r) = e^{-\frac{1}{2}(\frac{|\Delta x|}{W_x} + \frac{|\Delta y|}{W_y} + \frac{|\Delta z|}{W_z})}$ as in eq (8) and W_T is an additional *temperature scale* that determines how salinity measurements on one temperature surface should influence salinity on another temperature surface. In ORA-S3, $W_T = 1K$. The assumption that a large fraction of salinity variance is strongly correlated with temperature variance is not valid everywhere, especially at high latitudes. For that reason the assimilation switches gradually between 30° and 60° from S(T) to S(z), using the same latitude filter as the salinity adjustment. Additionally S(T) is not used in the mixed layer where the search for the right temperature class may fail. Sensitivity experiments show that S(T) produces a better fit to the data than S(z), not only in salinity but also in temperature. As for the temperature, the observation error \mathbf{R}_{SS} is diagonal, the relative ratio of the diagonal elements of \mathbf{B}_{SS} and \mathbf{R}_{SS} is the same as the ratio used for temperature, and the vertical dependence is also parameterized in the same way.

e. Assimilation of Sea Level Trends

The assimilation of sea level data requires careful considerations of the treatment of global sea level trends. If this trend is produced by thermal expansion due to global warming, it can not be represented by the ocean model as, in common with most ocean models used for climate activities, the Boussinesq approximation is made, which means that ocean models preserve volume and will not see a global sea level rise in response to a global warming. Therefore, if not treated correctly, the trend in sea level can be a problem when assimilating altimeter observations. In ORA-S3, the global sea level trend is removed from the altimeter sea level anomalies before they are assimilated. The information on global the sea level trends

can be used, however, in combination with the trends in ocean subsurface to determine how much of the sea level rise is due to thermal expansion and how much is due to mass change of the ocean (Church and White 2006). In principle, ocean reanalyses provide an estimation of steric height η_s by combining model first guess with subsurface data. By comparing the total trend in sea level given by the altimeter data with the trend in steric height given by the ocean analysis, it is possible to estimate the component of the trend due to mass variations. In ORA-S3 the information given by the altimeter data about changes in the global mean sea level between consecutive analysis cycles ($\Delta\bar{\eta}$) is compared every assimilation cycle with the changes in ocean analysis steric height ($\Delta\bar{\eta}_s$). The residual

$$\Delta\bar{\eta}_m = \Delta\bar{\eta} - \Delta\bar{\eta}_s \quad (14)$$

is applied as a fresh water flux, uniform in space. The partition between volume change and mass change is valuable information, since it can help to close the fresh water budget over the oceans, although only in a global sense.

5. Results

This section focuses on the impact of data assimilation on the mean state and on the representation of ocean currents, with emphasis on the tropics. The impact on climate variability and initialization of seasonal forecasts is shown in (Balmaseda *et al.* 2007a). A comparison with the previous operational system ORA-S2 is provided, together with a discussion of the assimilation increment and bias correction. The impact of data assimilation is then assessed by comparing the ORA-S3 analyses with equivalent ocean analyses in which no data have been assimilated. Everything else (spin up, forcing fields, subsurface relaxation, the prescribed additive component of the bias correction and the relaxation to SST) remains the same. In what follows we refer to this latter experiment as ORA-nobs.

a. Comparison with ORA-S2

Figure 4 shows the equatorial section of the mean temperature increment (a) and mean vertical velocity (b) for ORA-S3. For comparison, the corresponding figures for ORA-S2 are shown in panels c and d. The resulting mean increment in ORA-S3 (fig 4a) is smaller than in ORA-S2 (fig. 4c), and does not show the pronounced dipolar structure. Although there are several differences between ORA-S2 and ORA-S3, Balmaseda *et al.* (2007c) show that the differences in the assimilation increment along the equator is mainly due to the bias correction. The spurious vertical circulation present in ORA-S2 (fig. 4d) is also notably reduced in ORA-S3 (fig 4b), a direct consequence of applying the bias correction to the pressure gradient as shown in Balmaseda *et al.* (2007b). In addition to reducing the mean assimilation increment, the bias correction also leads to a better analysis: in ORA-S3 the fit to the data is better. Figure 5 shows the vertical profiles of the mean difference between the analyses and observations in the Western (EQ3: 5N-5S, 150E-170W) and Eastern Pacific (Niño3: 5N-5S, 150W-90W). Positive/negative differences are indicative of warm/cold bias. Both the warm bias in the Eastern Pacific and the cold bias in the Western Pacific in ORA-S2 have been significantly reduced in ORA-S3, resulting in a stronger east-west slope of the thermocline.

The velocity measurements from the TAO moorings provide an independent data set for the validation of ocean analyses, since the currents have not been assimilated in either ORA-S2 or ORA-S3. Figure 6 shows the vertical profiles of the time correlation of the currents in ORA-S2 (grey) and ORA-S3 (black) with the TAO currents at different mooring locations. The better correlation shown by ORA-S3 indicates that not only is the density field better constrained by observations in ORA-S3, but it is also more dynamically consistent.

b. Impact of data assimilation

An equatorial cross-section of the salinity assimilation increment in ORA-S3 is shown in figures 7a and b, separated into balanced and unbalanced components respectively. In the Eastern Pacific, the salinity increment is mainly due to the balanced operator that preserves the T/S relationship (CH96): the temperature assimilation, by applying a negative temperature increment, lifts the position of the thermocline; the salinity increment from the balance operator is consistent with lifting the density profiles. The impact of the salinity observations is more noticeable in the Western Pacific, where the model water mass properties are corrected by reducing the salinity.

Figure 8 shows the equatorial cross-section of the differences ORA-S3 minus ORA-nobs in both temperature (a) and salinity (b). The vertical structure of the differences suggests that in the Pacific, the main effect of the assimilation is a net increase in the heat content by deepening the thermocline. Only in the far eastern Pacific does the data assimilation have a cooling effect below the thermocline. The pattern of differences in the Equatorial Indian and Atlantic Oceans suggests that the assimilation is correcting for too much vertical mixing at the base of the thermocline, consistent with the pattern of differences in salinity. The pattern of the difference between the ORA-S3 and ORA-nobs is very different to that of the assimilation increment (shown in figure 4).

The non-local effect of the assimilation increments is also visible in the figure 9. The left column of figure 9 shows the impact of data assimilation on the upper ocean heat and salinity content, as measured by the average temperature and salinity over the upper 300m (T300 and S300 respectively). The figures show the difference between ORA-S3 and ORA-nobs. The right column shows the average assimilation increment in the upper 300m in temperature and salinity. By comparing the impact of data in the left column with the applied increment in the right it is obvious that the impact of data is far from local. For instance, the largest temperature increment is applied in the eastern Equatorial Pacific (negative), where the differences in T300 between ORA-S3 and ORA-nobs are not so obvious.

Equally, the differences ORA-S3 minus ORA-nobs in T300 in the Equatorial Atlantic are quite large but the temperature increment is small. There are two reasons for the 'non-locality' of the increment: the ocean circulation, which can move the increment around, and bias correction in the pressure gradient.

In the Pacific, the data assimilation increases the equatorial ocean heat content (fig 9a). The heat gain has a banded latitudinal structure, with maxima both sides of the Equator (at around 2-4N/2-4S) and a larger maximum in the area of the NEC (12/15 N). There is also a slight increase in the east-west slope of the thermal gradient. It appears as if the data assimilation is correcting for two kinds of errors in the equatorial Pacific: firstly it corrects the deficit of equatorial heat content in the ORA-nobs, due either to an excessive export of heat via the Indonesian Throughflow, too large meridional heat transport or an excessive vertical mixing; secondly, it corrects the slope of the thermocline, which is too flat in the ORA-nobs experiment. In the Equatorial Atlantic, data assimilation also steepens the east-west zonal gradient, especially by cooling the eastern part of the basin. In the tropical Indian ocean, data assimilation decreases the heat and salinity content: the excess of heat/salt in the Indian ocean in the ORA-nobs experiment is consistent with too large a transport via the Indonesian Throughflow. In both the tropical Pacific and Atlantic, it appears as if the whole structure associated with the subtropical gyres (both north and south of the equator) is displaced equatorward. In the northern extratropics, the differences in T300 and S300 suggest that data assimilation affects the path of the Kuroshio and Gulf stream currents, which are not as zonal in ORA-S3 as in ORA-nobs. In the North Pacific, data assimilation has a large-scale impact on S300, which is much reduced in ORA-S3 compared to ORA-nobs. In the South Pacific, the salinity in ORA-S3 is higher than in ORA-nobs.

c. The bias correction

Figures 10a,b show the average correction in zonal and meridional pressure gradient (reverse sign) resulting from the on-line bias estimation in ORA-S3 integrated over the top

60m. For comparison, the zonal and meridional components of the wind stress are shown in the bottom panels. To make possible a detailed comparison, only the region within 10N/10S degrees of latitude is shown, which is one region where the bias correction is large. The level of 60m has been chosen because below that the equatorial zonal pressure gradient is small. However, the meridional component only starts to decay below approximately 200m. The bias correction in the pressure gradient is also large in the areas of the western boundary currents, where the correction remains significant at all depths (not shown).

In the Pacific, the equivalent correction to the zonal component is negative over most of the equator, and amounts to about 5-10% of the mean zonal wind stress. This value is consistent with the perceived bias of the ERA40 winds from validation of wave products (Sterl and Caires 2005). The vertical structure of the bias correction is also consistent with a wind stress bias.

In contrast, the correction to the meridional component has deeper vertical structure, larger zonal spatial correlation scale and larger amplitude than the meridional component of the wind stress. The meridional bias term is probably related to errors with the mixing. Near the equator, east of the dateline, the equivalent wind stress corresponding to the meridional pressure gradient correction shows an increased surface meridional divergence. West of the date line the meridional correction is equivalent to an increased northward flow. North of the Equator, there is an increased surface meridional convergence in the area of the NECC, which is probably responsible for the increased heat content in that area. The structure of the pressure gradient correction in the Atlantic also shows an increased equatorial divergence as well as corrections for the western boundary current.

From figures 4, 9 and 10 it can be inferred that the increased equatorial heat content in ORA-S3 relative to ORA-nobs is likely due to the positive temperature increments in the Western Pacific; the warmer water is then advected eastward by the equatorial undercurrent. The equatorial circulation in both the Pacific and Atlantic is more vigorous with data assimilation: i.e. stronger equatorial undercurrents and increased divergence in the eastern

part of the basins.

d. Comparison with observations

Data assimilation improves the fit to the observations, mainly by correcting the ocean mean state. Figure 11 shows the mean difference between analyses and observations from ORA-S3 and ORA-nobs in different equatorial regions (EQ3: 5N-5S, 150E-170W; Niño3: 5N-5S, 150W-90W; EQATL:5N-5S, 70W-30E; EQIND: 5N-5S, 40E-120E). The top two rows show the mean difference in temperature and the lower two rows show the mean difference in salinity. As expected, assimilating data improves the fit to the observations. Data assimilation reduces the cold bias of ORA-nobs in the Western Pacific, and corrects the diffuse thermocline of the Atlantic and Indian oceans. In general, the improvements in the mean state, in both temperature and salinity extend to all geographical areas and most depths. Exceptions are the eastern Pacific (Niño3) at thermocline depth and the North Atlantic between 50m and 150m (not shown). In these areas the data assimilation changes the sign of the bias, with ORA-S3 being warmer than observations while ORA-nobs is colder than observations. For the eastern Pacific, this is a well known problem, which was also present in ORA-S2, and which in ORA-S3 has been reduced as a result of the bias correction to the pressure gradient, but has not been completely eliminated. In general, data assimilation reduces the bias in salinity and temperature in the extratropics (not shown).

Comparison with the OSCAR currents (Bonjean and Lagerloef, 2002) shows that data assimilation improves the representation of surface currents in the system, mainly in the tropical band. The Ocean Surface Current Analysis - Real-time (OSCAR) project provides analyses of oceanic surface currents derived from satellite altimeter and scatterometer data from the end of 1992 up to near real-time and now covers the whole ocean from 60°S to 60°N. Comparisons between OSCAR and data from the World Wide Drifter Buoy Deployment, and between OSCAR and TAO/TRITON/PIRATA mooring data shows that OSCAR products are of good quality especially in the tropical areas (<http://www.oscar.noaa.gov/>). Fig 12

shows the comparison with OSCAR zonal surface velocities: these maps show the correlation between the zonal component of the surface velocities from OSCAR monthly means and ORA-nobs and ORA-S3 for the period 1993-2006. The contour interval is non linear and correlations below 0.5 are not plotted. OSCAR data are not available along the coasts. Fig 12 shows that data assimilation improves the representation of the surface currents consistently: within 10 degrees of the equator the improvement is mainly a result of assimilating in situ temperature observations. The assimilation of SLA improves the currents almost everywhere, though there are some exceptions such as the Gulf-Stream and Kuroshio regions, and the southern oceans, where the correlation with Oscar remains below 0.5.

OSCAR currents are not completely independent from sea level data since altimetry data are used in their production. Although SLA data are used in ORA-S3 they are not assimilated directly but anomalies in T and S derived from them and therefore there is no guarantee that the use of altimetry would lead to improved velocity analyses, and so using OSCAR currents is a useful metric to assess the quality of our analysis.

6. The real-time ocean analysis RT-S3

In order to make monthly forecasts, or indeed medium range forecasts, a delay of 11 days in producing the ocean analysis is not acceptable; an ocean analysis is necessary with less than 1 day delay. To this end an additional ocean analysis is run each day, starting from the delayed mode analysis described above. The model is forced with the operationally-analysed surface fluxes and assimilates any available in situ data. Special treatment is required for the SST and sea-level data as they are received with some delay. For the delayed ocean analysis, daily SST maps are obtained by interpolation of the weekly products, which requires the existence of two consecutive weekly values. This can introduce a delay of up to 12 days. The RT-S3 analysis does not wait for the second map of SST to be available. Instead, a daily SST product is created by adding the latest SST anomaly to the daily climatology, i.e.,

persisting the anomaly.

This procedure of persisting the latest anomaly is also used in the preparation of the altimeter product that goes into the real-time analysis. In the RT-S3 analysis, the daily map at the center of the assimilation window is produced either by time interpolation between two consecutive weekly maps where these are available, or by persisting the sea level anomaly from the latest map available. The anomaly is then added to the daily 1993-2001 climatology.

In the production of the RT-S3 there are two assimilation cycles. The first takes place in the first time step, using all the observations in a centered 10-days window. The second is done at day 7 (5 days behind real-time), also using observations in a centered 10-day window, with the 3 first days overlapping the previous cycle, and the last day (current day) with few observations. To avoid giving extra weight to the observations, only a fraction (7/10) of the assimilation increment of the first analysis cycle is applied.

7. Summary and Conclusions

A new ocean analysis system has been implemented operationally at ECMWF. It consists of a historical reanalysis (ORA-S3) and a real-time component (RT-S3). The ORA-S3 reanalysis period starts in 1959 and is continuously updated with a delay of 11 days. Considerable modifications have been made to the assimilation system relative to the previous operational system (ORA-S2). An innovative bias correction algorithm has been implemented and an ocean analysis of salinity allows corrections to the water mass properties to be made. These corrections are complimentary to changes to salinity made to preserve the water mass properties in conjunction with an analysis using in situ temperature and altimetry data. Altimetric data are now used operationally but it is not yet possible to use the mean dynamic topography; only the anomalies relative to a longterm (7 year) mean are used. Estimates of bias correction to pressure rather than temperature reduces the spurious vertical circulations that can occur in equatorial regions. The assimilation of temperature

data is done in such a way as to minimise spreading across water masses with very different properties.

Compared to ORA-S2, the assimilation increment in ORA-S3 is considerably smaller, the equatorial currents are improved, and the fit to the data in the tropical oceans is better. The bias correction scheme is effective in correcting the slope of the thermocline in the Equatorial Pacific. At the Equator, the corrections to the zonal pressure gradient are equivalent to an increase of 5-10% of the easterly stress in line with estimates of wind error derived from wave data. Corrections to the meridional component of the pressure gradient are much larger, probably related to errors in the mixing formulation in the model.

Data assimilation has a significant impact on the mean state and variability of the upper ocean heat content. In the Equatorial Pacific, it steepens the thermocline. In the Indian ocean it sharpens the thermocline, making it shallower. The assimilation also corrects for a too diffuse thermocline in the Atlantic. The assimilation has a large impact on the salinity field. As expected, the fit to the data improves with data assimilation almost everywhere.

REFERENCES

- Alves O., M. Balmaseda, D Anderson, T Stockdale, 2003, Sensitivity of dynamical seasonal forecasts to ocean initial conditions. *Q. J. R. Meteorol. Soc.*,130, Jan 2004, 647-668. Also ECMWF Technical Memorandum 369, available at www.ecmwf.int
- Anderson, D. L. T., and M. Balmaseda, 2005: Overview of ocean models at ECMWF. ECMWF Seminar Proceedings. Seminar on Recent developments in numerical methods for atmospheric and ocean modelling, 6-10 September 2004, 103-111.
- Anderson, D. L. T., T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitart, P. Doblaser-Reyes, R. Hagedorn, T. Jung, A. Vidard, A. Troccoli and T. Palmer, 2003: Comparison of the ECMWF seasonal forecast Systems 1 and 2, including the relative performance for the 1997/8 El Nino. ECMWF Technical Memorandum 404.
- Anderson, D. L. T., T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitart, F. Molteni, F. Doblaser-Reyes, K. Mogensen and A. Vidard , 2006. Development of the ECMWF seasonal forecast System 3. ECMWF Technical Memorandum 503.
- Balmaseda, M.A., 2004: Ocean data assimilation for seasonal forecasts. ECMWF Seminar Proceedings. Seminar on Recent developments in data assimilation for atmosphere and ocean, 8-12 September 2003, 301-326.
- Balmaseda, M.A., 2005: Ocean analysis at ECMWF: from real-time ocean initial conditions to historical ocean reanalysis. ECMWF Newsletter No. 105. Autumn 2005. <http://www.ecmwf.int/publications/newsletters/pdf/105.pdf>
- Balmaseda, M.A., D.L.T. Anderson, and F. Molteni 2007a: Climate variability from the ECMWF System 3 ocean reanalysis. ECMWF Newsletter No. 113. Autumn 2007.
- Balmaseda, M.A., D. Dee, A. Vidard and D.L.T. Anderson, 2007b: A Multivariate Treatment of Bias for Sequential Data Assimilation: Application to the Tropical Oceans. *Q. J. R. Meteorol. Soc.*, **133**, 167-179.
- Balmaseda, M.A., D.L.T. Anderson, and A. Vidard 2007c: The ECMWF System 3 ocean analysis. ECMWF Technical Memorandum 508.

http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/501-600/tm508.pdf

Balmaseda, M.A., D.L.T. Anderson, and A. Vidard 2007d: Impact of Argo on analyses of the global ocean. *Geophys. Res. Lett.*, **34**, L16605, doi:10.1029/2007GL030452.

Bloom, S. C., Takacs, L. L., Da Silva, A. M. and Ledvina, D., 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, 124, 1256-1271.

Burgers G., M.Balmaseda, F.Vossepoel, G.J.van Oldenborgh, P.J.van Leeuwen, 2002: Balanced ocean-data assimilation near the equator. *J Phys Oceanogr*, **32**, 2509-2519.

Bonjean F. and G.S.E. Lagerloef, 2002: Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean, *Journal of Physical Oceanography*, Vol. 32, No. 10, pages 2938-2954.

Church, J., and N.J. White, 2006: A 20th century acceleration in global sea-level rise. *Geophys. Res. Letters*, VOL. 33, L01602, doi:10.1029/2005GL024826.

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 on tropical ocean models. *J Phys Ocean.*, 18, 697-707.

Dee, D. P., 2005. Bias and data assimilation. *Q. J. R. Meteorol. Soc.* ,**131**, 3323-3343.

Ducet, N., P.-Y. Le Traon, and G. Reverdin, 2000: Global high resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *J. Geophys. Res.*, 105, 19477-19498.

Haines K., J. Blower, J-P. Drecourt, C. Liu, A. Vidard, I. Astin, X. Zhou, 2006. Salinity Assimilation using S(T) relationships. *Mon. Wea. Rev.* Vol. 134, No. 3, pages 759–771.

Ingleby, B and M. Huddleston, 2006. Quality control of ocean temperature and salinity profiles - historical and real-time data. *J. Mar. Sys.*, in press.

Le Traon, P.-Y., F. Nadal, and N. Ducet, 1998: An improved mapping method of multi-satellite altimeter data. *J. Atmos. Oceanic Technol.*, 15, 522-534.

van Oldenborgh G. J., M.A. Balmaseda, L. Ferranti, T.N. Stockdale, D.L.T. Anderson:

2005a: Did the ECMWF Seasonal Forecast Model Outperform Statistical ENSO Forecast Models over the Last 15 Years? *Journal of Climate*, 18, No. 16, 3240-3249.

van Oldenborgh G. J., M.A. Balmaseda, L. Ferranti, T.N. Stockdale, D.L.T. Anderson: 2005b: Evaluation of atmospheric fields from the ECMWF Seasonal Forecasts over a 15-year period. *J. Clim.*, 18, No. 16, 3250-3269. See also corrections Vol 18,

Peters, H, Gregg, M C and Toole, J M, 1988: On the parameterization of equatorial turbulence. *J. Geophys. Res.*, **93**, 1199-1218.

Reynolds R., N Rayner, T Smith, D Stokes, W Wang 2002: An improved in situ and satellite SST analysis for climate. *J Clim* , 15, 1609-1625.

Smith, Thomas M., Richard W. Reynolds, 1998: A High-Resolution Global Sea Surface Temperature Climatology for the 1961-90 Base Period. *Journal of Climate* **11**, 3320-3323.

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

Sterl, A. and S. Caires, 2005: Climatology, Variability and Extrema of Ocean Waves - The Web-based KNMI/ERA-40 Wave Atlas. *Int. J. Climatology*, 25(7), 963-997, doi:10.1029/joc.1175.

Stockdale T. N., D. L. T. Anderson, J. O. S Alves, and M. A. Balmaseda, 1998. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392**, 370-373.

Troccoli, A., and K. Haines, 1999: Use of the Temperature-Salinity relation in a data assimilation context, *J. Atmos. Ocean Tech.*, 16, 2011-2025.

Troccoli A., M. Balmaseda, J. Schneider, J. Vialard and D. Anderson 2002, Salinity adjustments in the presence of temperature data assimilation. *Mon. Wea. Rev.*, **130**, 89-102.

Troccoli, A. and Kallberg, P., 2004. Precipitation correction in the ERA-40 reanalysis, ERA-40 Project Report Series, 13.

Uppala, S., and coauthors, 2005. The ERA-40 Reanalysis. *Q. J. R. Meteorol. Soc.* **131**, Part **B**, 2961-3012.

Vialard, J., Vitart, F., Balmaseda, M. A., Stockdale, T. N., and Anderson, D. L. T., 2005.

An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. *Mon. Wea. Rev.*, **131**, 1379–1395. See also ECMWF Technical Memorandum No 417.

Vidard, A., D.L.T. Anderson, M. Balmaseda, 2006: Impact of ocean observation systems on ocean analysis and seasonal forecasts. *Mon. Wea. Rev.*, **135**, pp. 409429 DOI: 10.1175/MWR3310.1.

Vidard, A., M. Balmaseda, 2007: Assimilation of altimeter data in the ECMWF ocean analysis system. In preparation

Vidard, A., M. Balmaseda, 2007: Assimilation of in situ observations in the ECMWF ocean analysis system. In preparation.

Vitart, F., 2005: Monthly Forecast and the summer 2003 heat wave over Europe: a case study. *Atmospheric Science Letter*, 6(2), 112-117.

Wolff, J., E. Maier-Reimer and S. Legutke, 1997. The Hamburg Ocean Primitive Equation Model. *Deutsches Klimarechenzentrum, Hamburg*, Technical Report No. 13.

List of Figures

1	Upper panel shows the surface forcing used in the ocean analysis and the initial conditions for the calibration hindcasts for ORA-S3. Lower panel shows the origin of the subsurface data and surface temperature fields used.	32
2	Typical 10-day distribution of temperature profiles in the ORA-S3 ocean analysis. The upper left panel of each block shows the coverage; the other two panels show the quality control and super-obbing decisions. Source: http://www.ecmwf.int/products/forecasts/d/charts/ocean/reanalysis/obsmap	33
3	Temperature increment coming from the assimilation of three observation at different locations in the equatorial Pacific using the ORA-S2 (top) and ORA-S3 (bottom) covariances. The background temperature field is also plotted.	34
4	Equatorial longitude-depth section of mean temperature increment (a,c) and vertical velocity (b,d) for ORA-S3 and ORA-S2 respectively. Contours are $2 \cdot 10^{-4} C/hr$ for the temperature increment and $0.2 m/day$ for the vertical velocity. The zero contour is not plotted. The mean corresponds to the time average over the period 1987-2001.	35
5	Vertical profiles of the mean difference analysis minus observations in the western (left) and eastern (right) Pacific, for ORA-S3 (black) and ORA-S2 (grey). Units are $^{\circ}C$. In both regions the bias in ORA-S3 is smaller than in ORA-S2.	36
6	Correlation with TAO currents: ORA-S2 (grey) and ORA-S3 (black)	37
7	Equatorial longitude-depth section of mean salinity increment in ORA-S3 for the balanced (a) and unbalanced (b) components. Contour interval is 0.210^{-5} psu/hr. The zero contour is not plotted. Values above (below) the reference level are in light (dark) shading. The mean corresponds to the time average over the period 1987-2001.	38
8	Equatorial zonal section of the mean difference ORA-S3 minus ORA-nobs in temperature (left), and salinity (right) over the period 1987-2001. Contour interval is $0.2^{\circ}C$ for temperature and 0.05 psu for salinity.	39
9	Impact of data assimilation: the left column shows the mean ORA-S3 minus ORA-nobs differences for T300 (a) and S300 (c). The right column shows the average assimilation increment in the upper 300 meters for temperature (b) and salinity (d). The statistics are for the period 1987-2001. Contour interval is $0.2^{\circ}C$ for T300 and 0.04 psu for T300, $5 \cdot 10^{-5} C/hr$ for the temperature increment and $5 \cdot 10^{-6} psu/hr$ for the salinity increment.	40
10	The upper row shows the equivalent wind stress corrections in the zonal (a) and meridional (b) directions resulting from vertically averaging the bias correction in the upper 60m. For reference, the lower row shows the applied zonal and meridional component of the wind stress. Note the different colour scale for the zonal bias correction. The units are $0.01 N/m^2$. The statistics are for the period 1987-2001.	41
11	Mean difference between analyses and observations in the equatorial oceans. The top two rows show the differences for temperature (in $^{\circ}C$) and the lower rows for salinity (in psu). ORA-S3 is shown in black and ORA-nobs in grey.	42

- 12 Correlations with Oscar zonal component of the surface currents for ORA-S3 (top), and ORA-nobs (bottom). The contour interval is irregular (0.5,0.6,0.7,0.8,0.85,0.9,0.95,0.98,0.99).

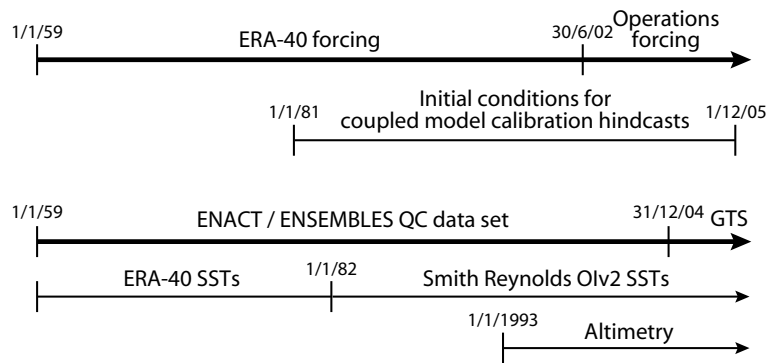


FIG. 1. Upper panel shows the surface forcing used in the ocean analysis and the initial conditions for the calibration hindcasts for ORA-S3. Lower panel shows the origin of the subsurface data and surface temperature fields used.

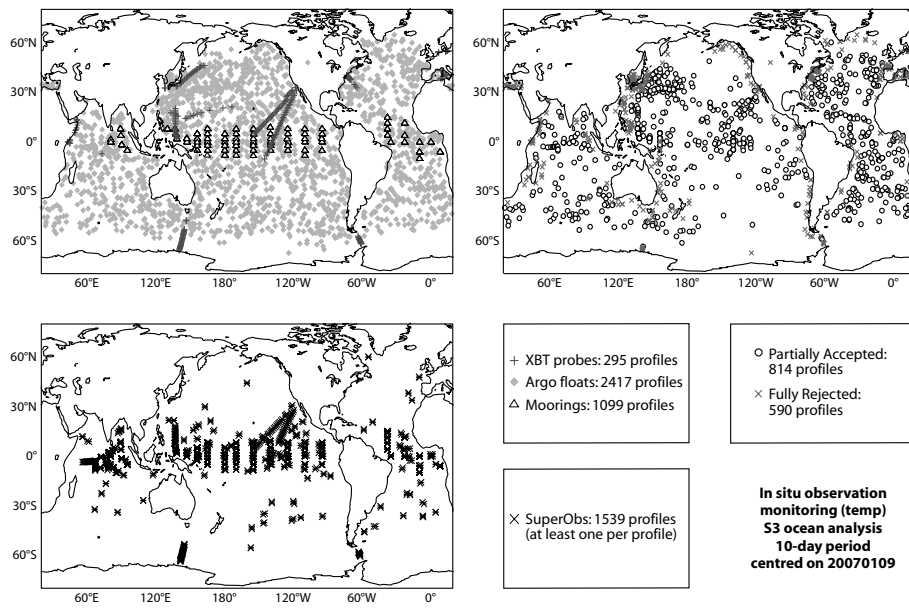


FIG. 2. Typical 10-day distribution of temperature profiles in the ORA-S3 ocean analysis. The upper left panel of each block shows the coverage; the other two panels show the quality control and super-obbing decisions. Source: <http://www.ecmwf.int/products/forecasts/d/charts/ocean/reanalysis/obsmap>.

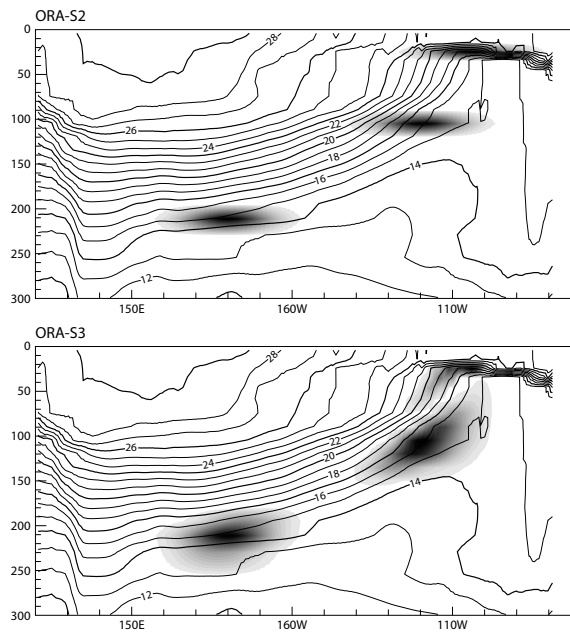


FIG. 3. Temperature increment coming from the assimilation of three observation at different locations in the equatorial Pacific using the ORA-S2 (top) and ORA-S3 (bottom) covariances. The background temperature field is also plotted.

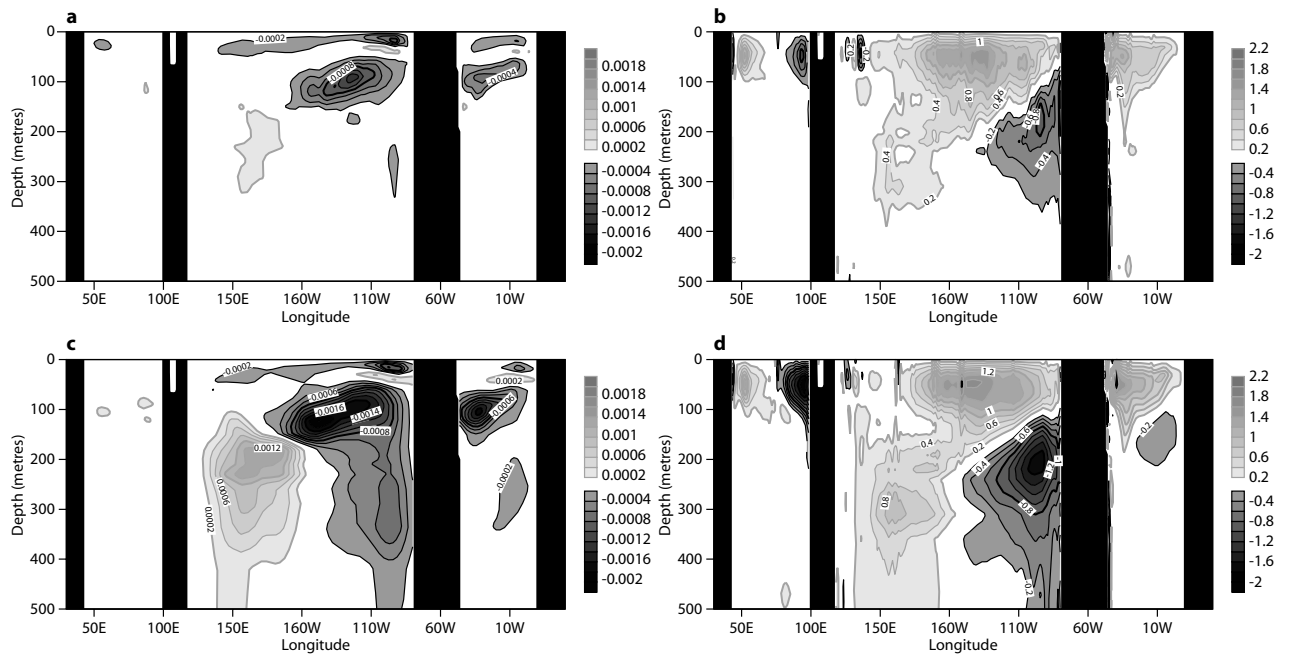


FIG. 4. Equatorial longitude-depth section of mean temperature increment (a,c) and vertical velocity (b,d) for ORA-S3 and ORA-S2 respectively. Contours are $2 \cdot 10^{-4} \text{C/hr}$ for the temperature increment and 0.2m/day for the vertical velocity. The zero contour is not plotted. The mean corresponds to the time average over the period 1987-2001.

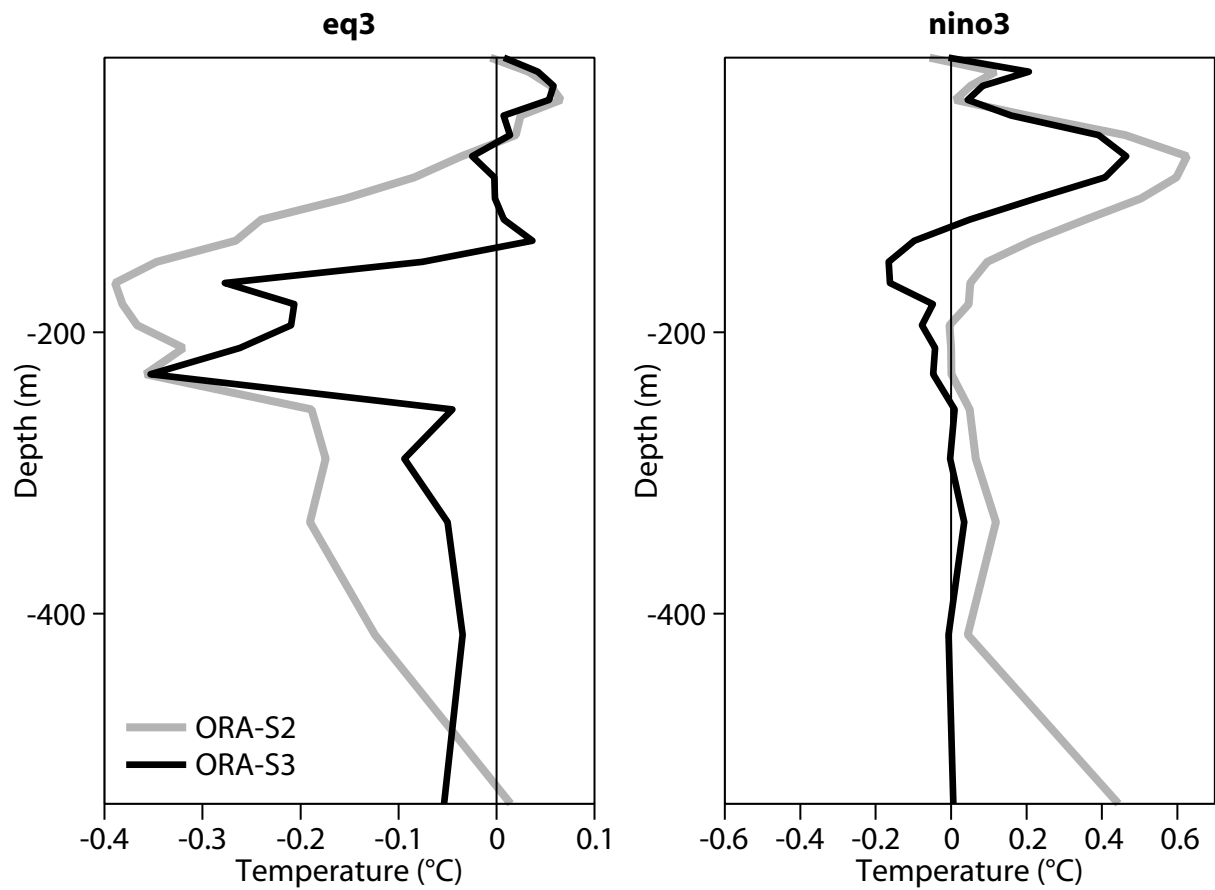


FIG. 5. Vertical profiles of the mean difference analysis minus observations in the western (left) and eastern (right) Pacific, for ORA-S3 (black) and ORA-S2 (grey). Units are $^{\circ}C$. In both regions the bias in ORA-S3 is smaller than in ORA-S2.

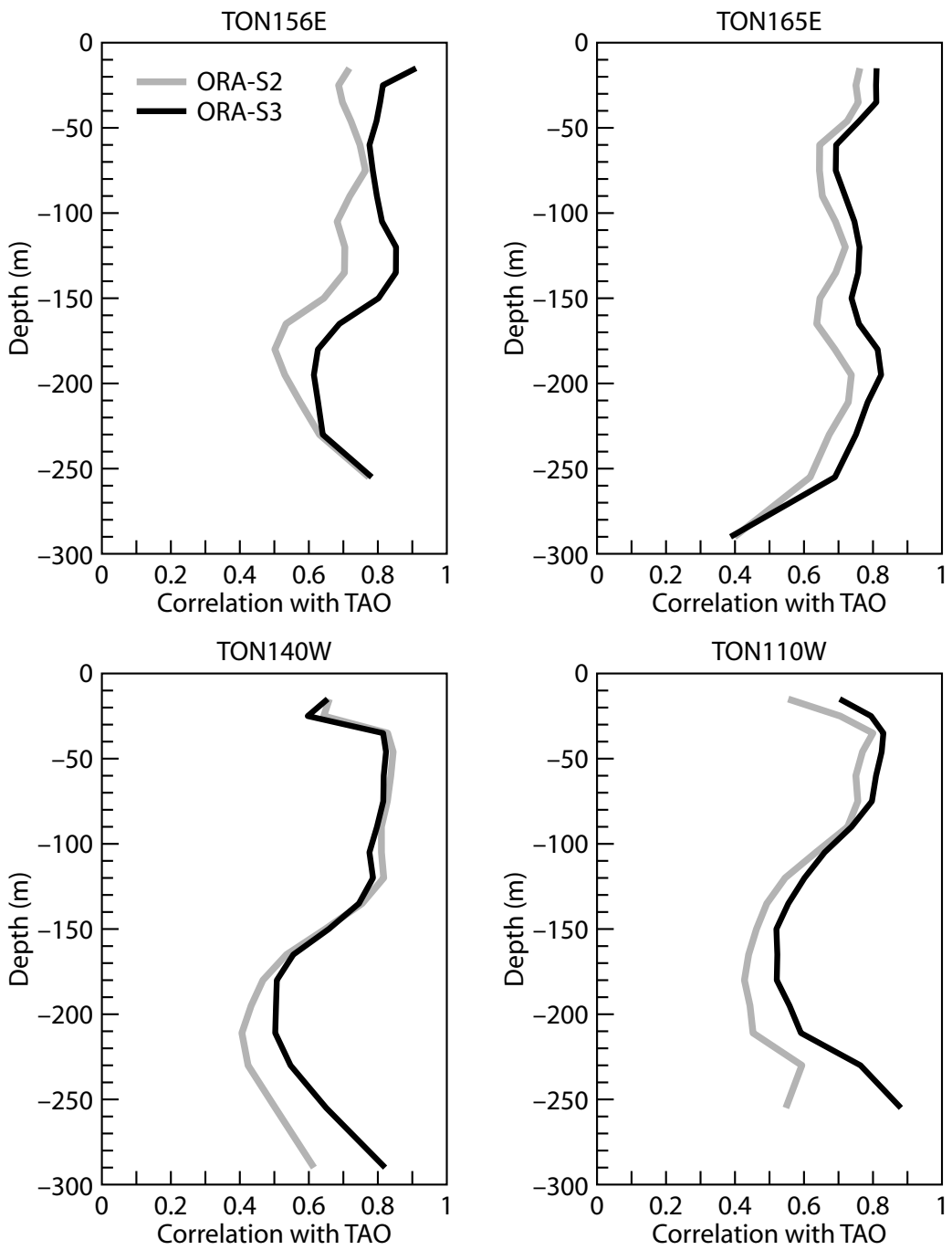


FIG. 6. Correlation with TAO currents: ORA-S2 (grey) and ORA-S3 (black)

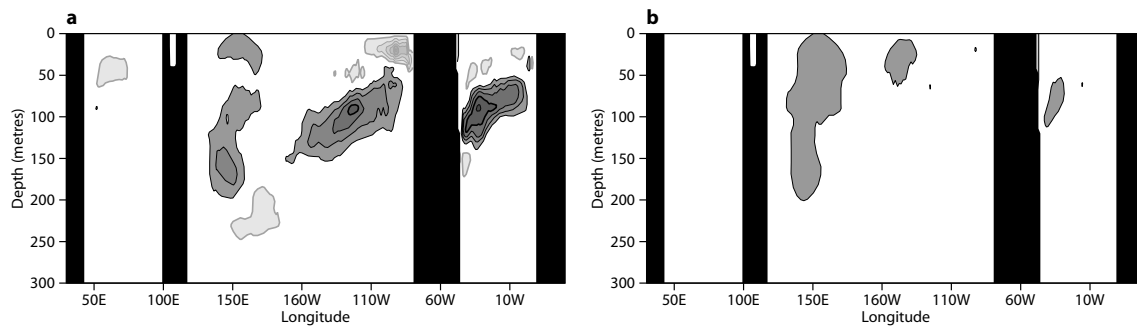


FIG. 7. Equatorial longitude-depth section of mean salinity increment in ORA-S3 for the balanced (a) and unbalanced (b) components. Contour interval is 0.210^{-5} psu/hr. The zero contour is not plotted. Values above (below) the reference level are in light (dark) shading. The mean corresponds to the time average over the period 1987-2001.

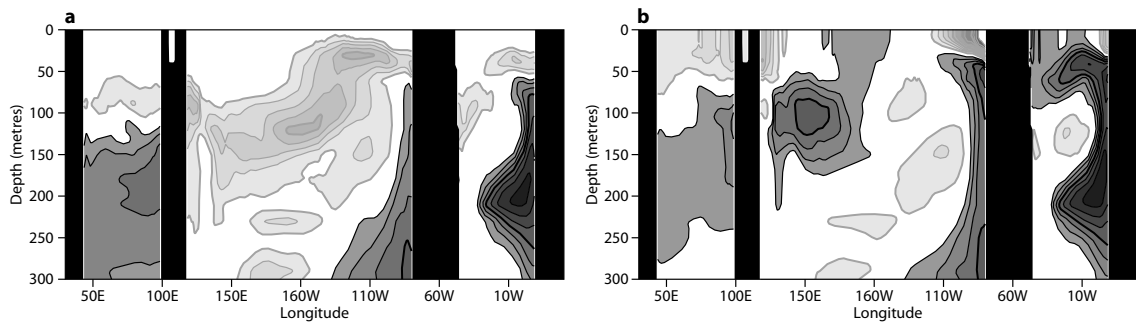


FIG. 8. Equatorial zonal section of the mean difference ORA-S3 minus ORA-nobs in temperature (left), and salinity (right) over the period 1987-2001. Contour interval is 0.2°C for temperature and 0.05 psu for salinity.

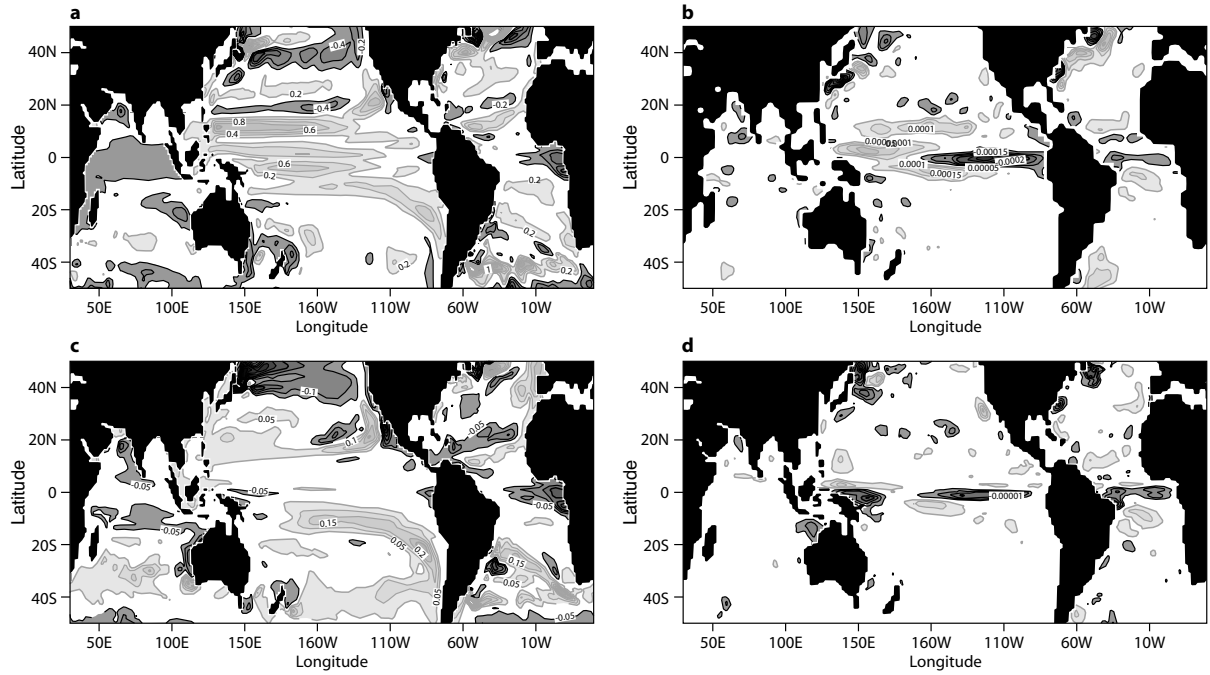


FIG. 9. Impact of data assimilation: the left column shows the mean ORA-S3 minus ORA-nobs differences for T300 (a) and S300 (c). The right column shows the average assimilation increment in the upper 300 meters for temperature (b) and salinity (d). The statistics are for the period 1987-2001. Contour interval is $0.2^{\circ}C$ for T300 and 0.04 psu for T300, $5.10^{-5}C/hr$ for the temperature increment and $5.10^{-6}psu/hr$ for the salinity increment.

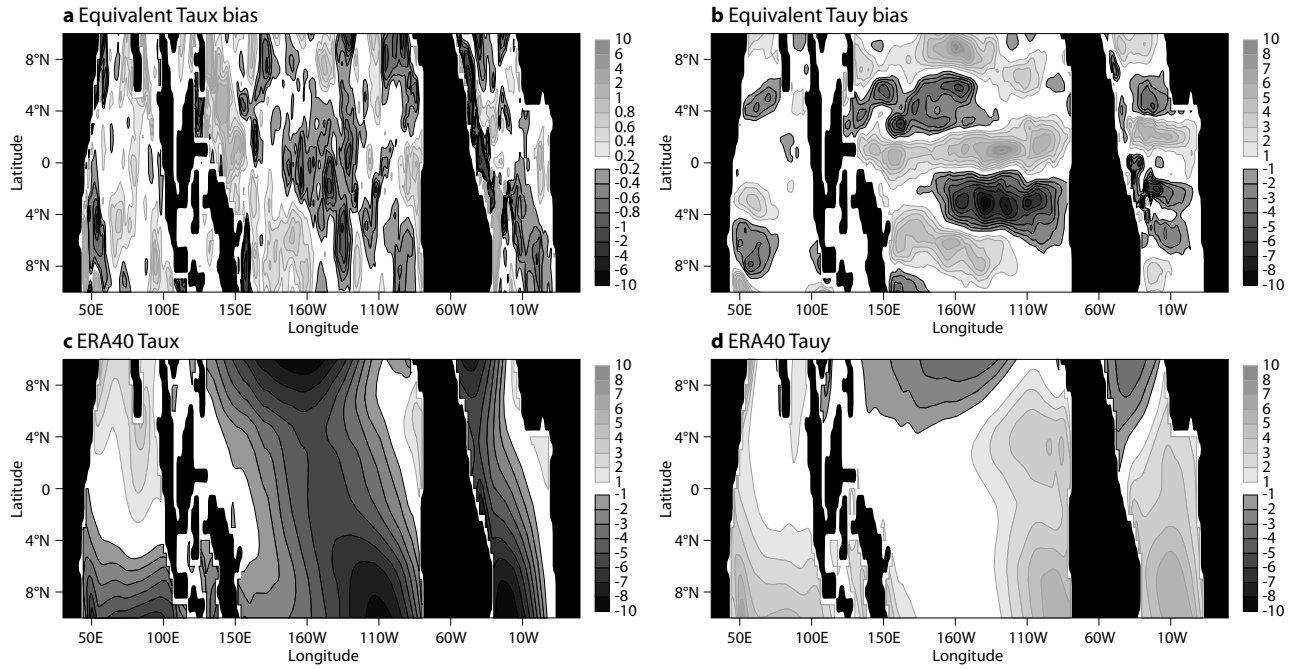


FIG. 10. The upper row shows the equivalent wind stress corrections in the zonal (a) and meridional (b) directions resulting from vertically averaging the bias correction in the upper 60m. For reference, the lower row shows the applied zonal and meridional component of the wind stress. Note the different colour scale for the zonal bias correction. The units are $0.01 N/m^2$. The statistics are for the period 1987-2001.

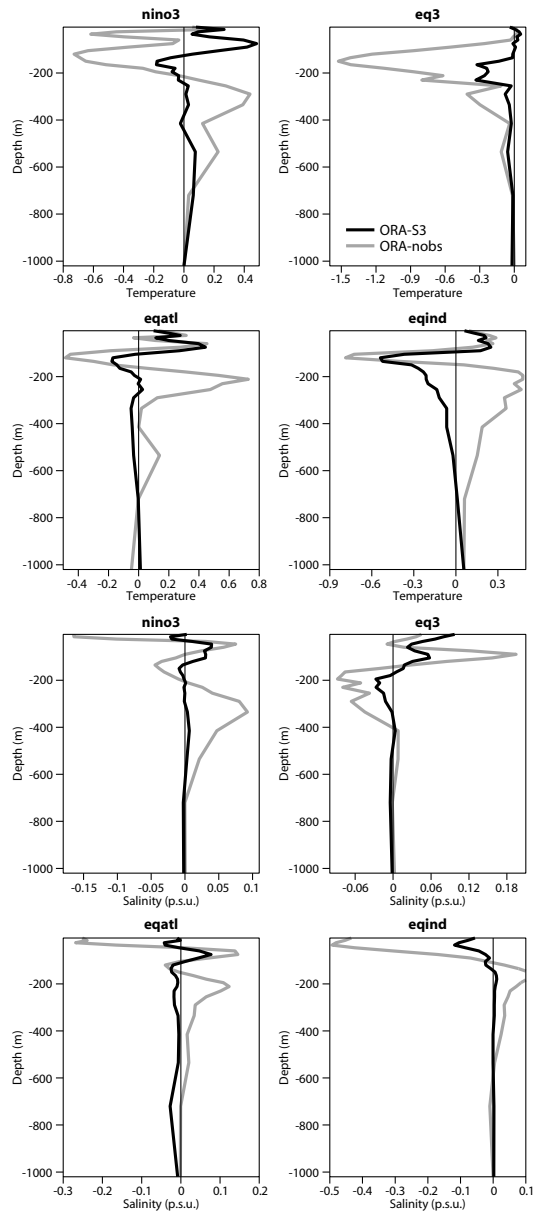


FIG. 11. Mean difference between analyses and observations in the equatorial oceans. The top two rows show the differences for temperature (in $^{\circ}C$) and the lower rows for salinity (in psu). ORA-S3 is shown in black and ORA-nobs in grey.

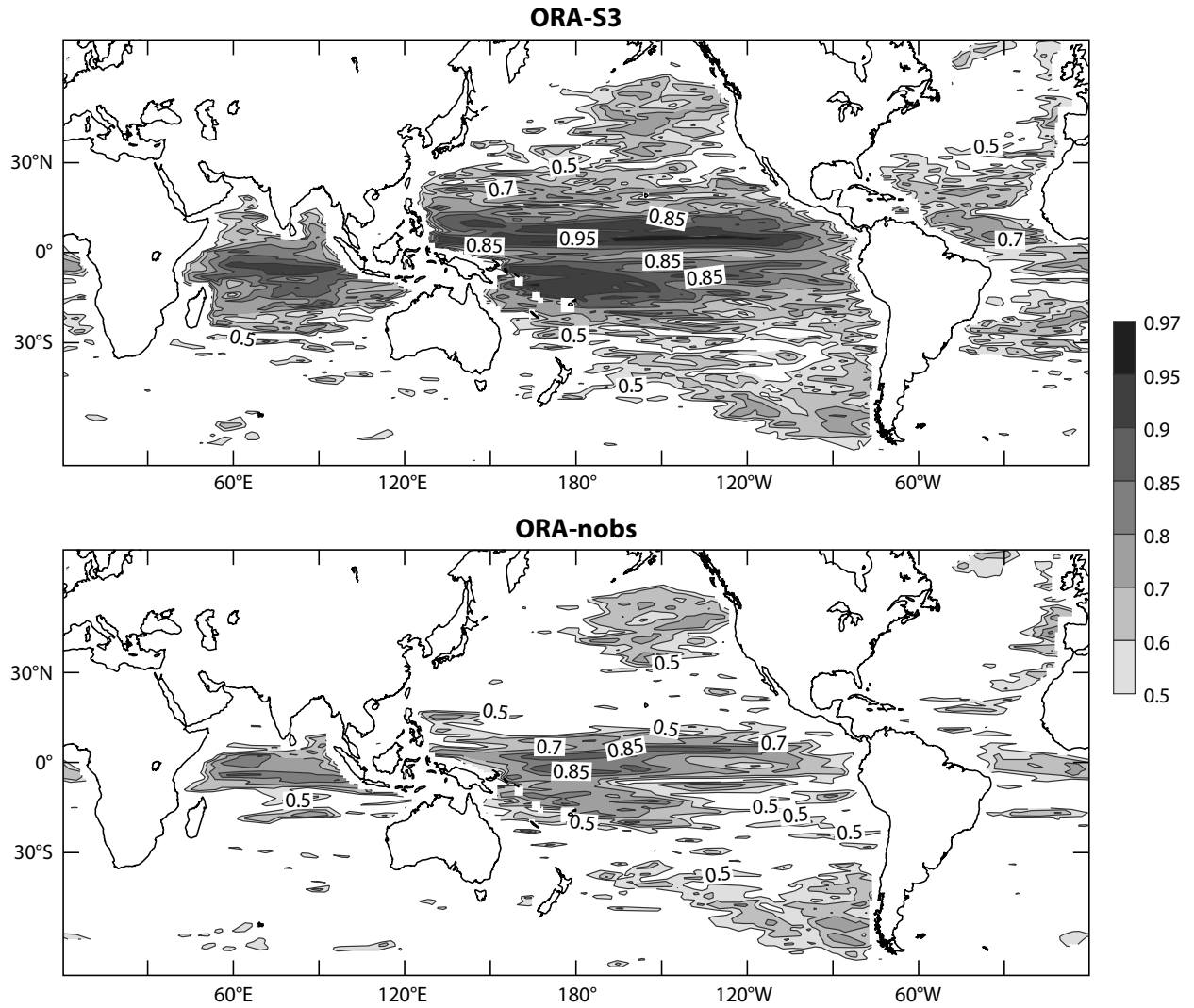


FIG. 12. Correlations with Oscar zonal component of the surface currents for ORA-S3 (top), and ORA-nobs (bottom). The contour interval is irregular (0.5,0.6,0.7,0.8,0.85,0.9,0.95,0.97)