

**Validation of an EnKF system for OGCM  
initialisation assimilating temperature, salinity and  
surface height measurements**

Olwijn Leeuwenburgh

Royal Netherlands Meteorological Institute (KNMI)

P.O. Box 201

3730 AE, De Bilt, The Netherlands

[olwijn.leeuwenburgh@knmi.nl](mailto:olwijn.leeuwenburgh@knmi.nl)

*Submitted to Monthly Weather Review*

May 17, 2005

## Abstract

Results are presented from a decade-long assimilation run with an 64-member OGCM ensemble in a global configuration. The intended purpose of the assimilation system is to produce ocean initial conditions for seasonal forecasts. The model ensemble is constructed with the Max Planck Institute Ocean Model, where each member is forced by differently perturbed ERA40 atmospheric fields over sequential 10-day intervals to produce ensemble hindcasts. Along-track altimetric data from the ERS and TOPEX/POSEIDON satellites, as well as quality-controlled subsurface temperature and salinity profiles are subsequently assimilated using the standard formulation of the EnKF. The applied forcing perturbation method, and data selection and processing procedures are described, as well as a framework for the construction of appropriate data constraint error models for all three data types. The results indicate that the system is stable, does not experience a tendency towards ensemble collapse, and provides smooth analyses that are closer to the data than an unconstrained control run. Both the bias and time-dependent errors are reduced by the assimilation but not entirely removed. Time series of assimilation and ensemble statistics also indicate that the model is not very strongly constrained by the data because of an overspecification of the data errors. A comparison with insitu current meter data shows a deterioration in the equatorial zonal velocity profiles in the central Pacific. This is primarily associated with a shift in the time-mean profile to lower values resulting from an assimilation-induced bias.

## 1. Introduction

One of the main aims of climate centers over the previous decade has been the development of operational seasonal forecast systems. The primary focus during these initial stages of development has been on improving forecasts of the dominant mode of global climate variability, ENSO (El Niño/Southern Oscillation). Both ECMWF (European Centre for Medium-Range Weather Forecasts) and NCEP (National Centers for Environmental Prediction) currently run seasonal forecast systems on an operational basis. The NCEP system is based on the 3DVar assimilation scheme developed by Derber and Rosati (1989). It has been used to assimilate TAO (Tropical Atmosphere-Ocean) buoy and XBT (expendable bathythermograph) temperature profiles (Ji et al., 1995), as well as TOPEX/POSEIDON altimetry data (Ji et al., 2000). For the assimilation of altimetric sea level, a regression between sea level anomalies and the principal components of the EOFs of the vertical temperature profile, as determined from a model control run (Fischer et al., 1997), is used to update the subsurface temperature. Vossepoel and Behringer (2000) subsequently adjusted the system to allow for additional correction of the salinity field. The ECMWF seasonal forecasting system assimilates temperature profiles and merged T/P and ERS sea level anomaly maps (Segsneider et al., 2000), where sea level innovations are projected on the model subsurface density field using the method developed by Cooper and Haines (1996) which effectively lifts and lowers the T/S profile. Balanced velocity increments are calculated following Burgers et al. (2002). Alternative schemes for correction of the subsurface density and velocity fields have also been

used. For example, Fukumori et al. (1999) assimilated T/P data into a global ocean model using the Extended Kalman Filter and projected all sea level differences onto the barotropic and first baroclinic vertical modes to correct T, S, U and V simultaneously. While the model error-covariance matrix did account for the time-varying data distribution, a time-asymptotic limit of the covariances was assumed. In 3DVar systems the background error-covariances also remain fixed during the assimilation window, while in incremental 4DVar the propagation of the model error-covariances by the linear tangent model is implicit. Error variances can be diagnosed at the start of each new window using the previous analysis increments (Weaver et al., 2003). The variational system described by Weaver et al. (2003) uses a univariate background covariance matrix, but current development of such systems is towards implementation of multivariate constraints (T-S, geostrophy) into the error-covariances. Multivariate covariances between T, S, U, and V were estimated from ensemble runs by Borovikov et al. (2004) for use in an optimal interpolation scheme. Model error statistics are typically kept fixed over time in these applications. An assimilation method that utilizes time-evolving multivariate statistics is the Ensemble Kalman Filter (Evensen, 1994). Keppenne and Rienecker (2003) developed an ocean data assimilation system based on EnKF and presented initial results from the assimilation of temperature profiles. Identical-twin experiments where the EnKF was used to assimilate sea level in the tropical Pacific section of a general circulation model ensemble were recently presented by Leeuwenburgh (2005). It was shown that subsurface corrections could be obtained in all state variables and that the analyzed fields were closer to the truth than the unconstrained control run. In this paper an EnKF system

is presented which assimilates subsurface T and S profiles and altimetric sea level. It was developed to produce analyzed ocean states that can be used as initial conditions for seasonal forecasts, and is tested here during a 10-year hindcast run covering the 1987-1996 period. The assimilation of real sea level data requires the construction of an error model that incorporates representation errors resulting from the absence of small-scale eddy variability in the model. A similar approach is needed for the T and S data. These error models are constructed here for the OGCM MPIOM-1, which is run in a global configuration with increased resolution ( $0.5^\circ$  in latitude) in the tropics and in the high latitude sinking regions. Since the truth is only known approximately from the data, the quality of the analyses needs to be determined from comparisons with independent data, and from statistical tests of the consistency between model and data uncertainty estimates. Practical problems associated with the use of such a complex system are highlighted and discussed.

In Section 2. the forcing perturbation method will first be explained. Sections 3 and 4 are concerned with the description of data sources and processing, and with the estimation of appropriate data constraint errors respectively. The assimilation method is discussed in Section 5. Results from the 10-year assimilation run are presented in Section 6. Section 7 ends with a summary and some concluding remarks.

## 2. Representation of uncertainty in the surface forcing

The ocean model is forced using daily ERA40 fields of surface wind stress, 10m wind speed, 2m air temperature, dew point temperature, precipitation, cloud cover and net incoming solar radiation. Together with the sea surface temperature and sea ice characteristics, which are evolved by the model equations, these fields determine the surface exchange of momentum, heat and fresh water between ocean and atmosphere, which are the main mechanisms responsible for driving the ocean circulation and setting the upper ocean stratification. Since their exact values are not known, the forcings should be treated as stochastic variables. It is assumed that the ERA40 estimates are unbiased but contain a certain Gaussian distributed error. It is furthermore assumed that the statistics of differences between daily ERA40 and NCEP/NCAR reanalysis (Kalnay et al, 1995) fields can be used as proxies for those of the true errors in the ERA40 reanalysis. However, because of differences in the treatment of ice covered parts of the ocean, the difference between the reanalyses may not be an appropriate proxy for the true error in polar regions. The presence of unrealistic differences in air temperature in places that are known to contain significant amount of sea ice during winter indicates that this is the case.

Oceanographic observations are very sparse in the Southern Ocean south of 40°S (see next section) and it is expected that they will be insufficient to constrain the ocean model state to any significant extent. Also, since at this stage seasonal forecast attempts will focus primarily on the tropics, which may be assumed to be insensitive to the exact magnitude of ocean-atmosphere exchanges at high latitudes, it is decided not to make an

effort to represent forcing uncertainty at high latitudes. The ERA40-NCEP differences are therefore slowly relaxed to zero within a 20° latitude band south of 40°S. Given the much larger amount of data in the northern hemisphere, the same procedure is applied here north of 50°N.

A combined EOF decomposition of the daily difference fields of zonal and meridional surface stress, air and dew point temperature, and short-wave radiation over the period 1992-1993 is used to obtain the dominant spatial patterns and their time evolution. This is an extension of the approach taken by Robert and Alves (2004) who used EOF analysis to identify the leading orthogonal modes of intra-seasonal ERA40-NCEP wind stress differences only. The combined EOF decomposition produces different, but temporally covarying spatial patterns for the individual variables. This approach is chosen to decrease the possibility that perturbations to the various forcing variables have opposing effects on sea surface temperature, which would reduce their effectiveness in maintaining ensemble spread.

The spectrum of the daily differences is dominated by the seasonal cycle (representing more than 40% of the variance) and by short time scales associated with weather systems. The regular part of the seasonal cycle is perfectly predictable and therefore not of interest for ensemble forecasting. Weather systems on the other hand tend only to induce local changes in the ocean state (e.g. mixed-layer deepening) but do not directly affect the large-scales. Variability on intermediate time-scales has been suggested as a possible trigger for ENSO events. The MJO for example has been associated with westerly wind bursts in the western Tropical Pacific on 30-60 day time scales. In order to ensure that the

EOF patterns are representative of variability on such time scales, all variability on time scales shorter than 20 days is removed from the record prior to the EOF decomposition.

The first 80 EOFs represent 99% of the total variance of the resulting normalized difference fields. The forcing perturbations are constructed as random combinations of these 80 EOFs where temporal correlation between perturbations is enforced. The temporal decorrelation scales are determined from each associated principal component timeseries separately.

### **3. Data sources and selection criteria**

All observations are taken from the data set prepared especially for the ENACT project at the UK MetOffice (Ingleby and Huddleston, 2005). The primary source for this set is the World Ocean Database 2001, but for the periods covered in this study it is supplemented with XBT (expendable bathythermograph) data from BMRC/CSIRO, CTD (conductivity temperature depth) casts from PMEL, and data from the GTSP (global temperature and salinity profile program). All these data are checked for duplicity and quality controlled. Quality flags are based on checks for spikes, location errors, and vertical consistency, and on background comparisons and buddy checks. Super-observations are created for moorings with high-frequency output. Only data which are flagged 'good' or 'probably good' are assimilated.

Observations below 800m depths are discarded, since no serious effort is made to simulate the dynamics of the deeper layers. In order to increase the speed of the calculations,

the profiles are subsampled in the vertical, retaining only 2 samples per model layer. From the moment of introduction of altimetry data in 1992, the temperature and salinity profile data set is further subsampled in the horizontal, retaining no more than 1 profile per model grid box. Assimilation takes place in waters deeper than 500m. Only observations located within a limited region surrounding each analysis grid point are selected for the updates, as will be discussed in Section 5.. The assigned data error variance is the greater value of the assumed processed data error (1°C for temperature and 0.17 psu for salinity) and the effective data constraint error estimated as described in the next section.

#### 4. Data constraint error

The estimation of appropriate data errors follows the method and terminology of Fu et al. (1993), later extended by Fukumori et al. (1999). First of all it is assumed that the observations  $\mathbf{y}$  are related to the true ocean state  $\mathbf{w}$  as

$$\mathbf{y} = \mathcal{H}(\mathbf{w}) + \boldsymbol{\epsilon} \quad . \quad (1)$$

The first rhs term is a symbolic notation for the physical relationship between the observed quantity and the true ocean state, and  $\boldsymbol{\epsilon}$  contains all errors associated with the measurement process, conveniently termed instrument error. A model is only able to represent a limited fraction of the total number of degrees of freedom of the real ocean. For example, only variability on spatial scales greater than a few grid box lengths can typically be reliably reproduced by finite difference schemes. Even on these scales, however, the model  $\mathbf{x}$  is likely to differ from the real ocean as a result of errors in parameterizations or

in the model forcing. This can be expressed as

$$\mathbf{x} = \mathcal{G}(\mathbf{w}) + \boldsymbol{\mu} \quad . \quad (2)$$

The first rhs term constitutes the ‘true’ model state  $\mathbf{x}^* = \mathcal{G}(\mathbf{w})$ , containing only those features of the real ocean which the model is able to represent. The operator  $\mathcal{G}$  can be seen as mapping the true ocean state to the space spanned by all realizable model solutions.

The second rhs term is the model state error. Considering the intended application of the ocean analyses as initial states for seasonal forecasts, the aim of data assimilation here is estimation of the true model state. (An alternative view of data assimilation, perhaps appropriate in now-casting applications, is that of physics-based interpolation of the data.)

With the above definitions the true model state can be related to the observations by

$$\mathbf{y} = \mathbf{H} \mathbf{x}^* + \mathcal{H}(\mathbf{w}) - \mathbf{H} \mathcal{G}(\mathbf{w}) + \boldsymbol{\epsilon} = \mathbf{H} \mathbf{x}^* + \mathbf{r} \quad . \quad (3)$$

For convenience the measurement operator  $\mathbf{H}$  relating the observations to the model state is assumed linear here, but this is not required. The second and third rhs terms together constitute the ‘representation error’ which is associated with the incomplete representation of the observed quantity by the model. This equation shows that the effective data constraint error  $\mathbf{r}$  is the sum of representation and instrument errors. In many ocean applications the former can be much larger than the latter.

Fu et al. (1993) suggested a method to obtain estimates for the data constraint error and applied it within the context of the assimilation of altimetric sea level data. It is based on equation (3) for  $\mathbf{y}$  and a corresponding expression for the model equivalent of

this observation, given by

$$\mathbf{m} = \mathbf{H}\mathbf{x}^* + \mathbf{p} = \mathbf{s} + \mathbf{p} \quad . \quad (4)$$

where  $\mathbf{m}$  is a particular model state realisation and  $\mathbf{p}$  is the associated model error. Given a long unconstrained model integration, in conjunction with the observations available over the same time period, the data and model variances and the cross-covariances between them can be determined according to

$$\langle \mathbf{y}\mathbf{y}^T \rangle = \langle \mathbf{s}\mathbf{s}^T \rangle + \langle \mathbf{r}\mathbf{r}^T \rangle \quad (5)$$

$$\langle \mathbf{m}\mathbf{m}^T \rangle = \langle \mathbf{s}\mathbf{s}^T \rangle + \langle \mathbf{p}\mathbf{p}^T \rangle \quad (6)$$

$$\langle \mathbf{y}\mathbf{m}^T \rangle = \langle \mathbf{s}\mathbf{s}^T \rangle \quad , \quad (7)$$

from which the following expressions for the data constraint error variance and the error variance of the control state can be derived

$$\langle \mathbf{r}\mathbf{r}^T \rangle = \langle \mathbf{y}\mathbf{y}^T \rangle - \langle \mathbf{y}\mathbf{m}^T \rangle \quad (8)$$

$$\langle \mathbf{p}\mathbf{p}^T \rangle = \langle \mathbf{m}\mathbf{m}^T \rangle - \langle \mathbf{y}\mathbf{m}^T \rangle \quad . \quad (9)$$

In deriving these equations, it is assumed that all errors are additive, have zero mean, and are uncorrelated with each other. By removing the mean of each record prior to estimation of the variances it is ensured that they only represent the time-dependent part of the data error.

The above approach is used here to obtain estimates of data constraint error variances for altimetric sea level measurements as well as for insitu measurements of temperature and salinity. All insitu data between January 1987 and December 1999 are averaged

on a monthly basis in 1-degree bins and within standard depth layers. Corresponding monthly mean model values over the same period are obtained from the output of the control run. Figure 1 shows the resulting estimates of the data constraint error  $\sqrt{\langle \mathbf{r}\mathbf{r}^T \rangle}$  for temperature and salinity observations at different depth levels. Error estimates are given for every grid box that contains more than one observation over the 13 year period. Increased values are associated with regions characterized by strong internal variability. In these regions the model contains too little energy, which can primarily be related to the fact that the model does not properly resolve small scales. However, large errors may also point to regions where the model forcing has large errors. Large values for the data constraint error variances extend to deeper layers only in regions where the ocean variability has a strong barotropic component, such as the western boundary currents. The errors in the Tropics, on the other hand, are largely confined to the surface and thermocline depths, illustrated by the fact that the highest temperature error values are found at shallower depths when going eastward along the equator, following the shoaling thermocline.

Observations of salinity are much sparser than observations of temperature. The WOCE sections clearly stand out as the main source for salinity observations away from the busy shipping routes, which are primarily located in the northern hemisphere.

While the above method appears to produce useful estimates for the time varying error component, it does not account for constant biases between observations and model. While this is no problem when assimilating anomalies, as is the usual practise with sea level data, significant biases between model and data exist for subsurface quantities such

as temperature and salinity. As with the time-varying error components, such biases can often be related to the use of a coarse grid, a typical example being an incorrect Gulf Stream separation point, or to systematic errors in the forcing.

It is difficult to reproduce and maintain the correct mean structure and position of a strong boundary current with a coarse-grid model which will tend to assume its own preferred circulation structure. It would therefore make sense to relax the data constraint in such regions further. One way of doing this is to lower the relative weight given to the data by increasing the error variances (an infinite error variance is equivalent with not using the data at all). Weights can for example be assigned 'on-line' as a function of the innovation magnitude. Tests with this approach in a 4Dvar system were found to lead to undesired features in the analysis (A. Weaver, personal communication). Given also that the use of such weighting will be somewhat arbitrary, an alternative method is used here that consists of adding the squared bias to the data constraint error variance defined above. Ideally one would like to have an adaptive measure of the evolution of the bias during the assimilation run (e.g. Dee and Da Silva, 1998). This would allow for reduction of the error variance when the mean misfit between data and ensemble mean becomes smaller. No such adaptive bias correction scheme is used here, but the bias contribution to the total data error variance is reduced to a tenth of its original value after a few years in the run without apparently causing large problems.

## 5. Analysis algorithm

The Ensemble Kalman Filter (Evensen, 1994) applies the Kalman Filter equation simultaneously to a model ensemble of finite size  $N$ . The forecast error covariance matrix  $\mathbf{P}$  is the same for all ensemble members and is calculated as the spread around the ensemble mean,

$$\mathbf{P} = \frac{\mathbf{A}'\mathbf{A}'^T}{N-1}$$

where  $\mathbf{A}_{n \times N} = (\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N)$  holds the ensemble of  $n$ -dimensional state vectors, and primes indicate anomalies. For the resulting analysis ensemble to have the correct spread, the observations should also be treated as stochastic quantities which is achieved by using an ensemble of randomly perturbed observations (Burgers et al., 1998). In practise the use of a finite size ensemble has two important consequences; first, only  $N$  directions of the solution space can be sampled whereas in many practical applications the number of effective degrees of freedom of the model is many times greater, and secondly, since errors in ensemble-based variance estimates decrease proportional to  $1/\sqrt{N}$ , the accuracy of the Kalman gains and model state error covariances will be limited for practical ensemble sizes. An approach that addresses both problems to some extent is commonly known as ‘local analysis’. In calculating the analysis for each individual grid column, the weights (covariances) given to observations which lie far away from the grid point are set to zero (the ‘global analysis’ uses all observations for each grid point). The proper way to apply this ‘covariance localisation’ involves the Schur product of the empirical covariances with an analytical correlation function  $\mathcal{C}$  with local support (Gaspari and Cohn, 1999). This

method was applied by Houtekamer and Mitchell (2001) and by Keppene and Rienecker (2002) in atmospheric and oceanographic applications respectively. Including the Schur product, defined by

$$(\mathcal{C} \circ \mathbf{P})_{ij} = \mathcal{C}(i, j) \mathbf{P}_{ij}$$

the final form of the standard EnKF equation becomes

$$\mathbf{A}_a = \mathbf{A} + \mathcal{C} \circ (\mathbf{P}\mathbf{H}^T) [\mathcal{C} \circ (\mathbf{H}\mathbf{P}\mathbf{H}^T) + \mathbf{R}]^{-1}(\mathbf{D} - \mathbf{H}\mathbf{A}) \quad (10)$$

where  $\mathcal{C}$  is Eq. (4.10) of Gaspari and Cohn (1999), and  $\mathbf{D} = \mathbf{d} + \mathbf{E}$  is the sum of the data vector  $\mathbf{d}$  and an ensemble of observation perturbations  $\mathbf{E}$ . The observation error covariance matrix  $\mathbf{R}$  has the data constraint error variances determined in the previous section as diagonal elements and zero values off the diagonal (twin experiments with different formulations for  $\mathbf{R}$  showed that a diagonal form yielded the best results). Following Houtekamer and Mitchell (2001) the order of application of the Schur product and measurement operator has been interchanged. Following Keppene and Rienecker (2002) the half axes of the local support ellipse are chosen as  $30^\circ$  in the zonal direction and  $15^\circ$  in the meridional direction respectively for sea level, while twice these values were used for temperature and salinity data (Figure 2). A vertical decorrelation scale of 500m is chosen. The three types of observations are assimilated sequentially, which is justified when the data errors are uncorrelated. This condition is met for the instrument errors, but the situation is less clear for the data constraint errors estimated here. For example, resolution issues are found to affect all 3 observed variables in the mid-latitude western boundary currents. However, as a first step it will be assumed in the following that all data errors

are uncorrelated.

## 6. Results

### 6.1. Introduction

The control run for this hindcast experiment is started from a resting state and a prescribed density distribution taken from the Polar science center Hydrographic Climatologies (Steele et al., 2001). A model spinup is performed by integrating the model forward for 5 years (1978-1982) using ERA40-based climatological forcing, followed by 4 years (1983-1986) using daily ERA40 forcing. A model ensemble spinup is run for 2 years using the January 1985 control state as an initial condition for all members. Each member is subsequently forced by differently perturbed ERA40 fields, as described in section 2.

The first analysis is produced for 8 January 1987, and successive assimilation steps follow each 10-day forward integration (referred to as 'forecasts'). During the first 5 years and 9 months only temperature and salinity profiles are assimilated. Sea level observations from TOPEX/POSEIDON and ERS-1 (replaced by ERA-2 in May 1995) are additionally assimilated from 10 October 1992 onwards. The final analysis is produced for December 28 1996.

The verification and analysis of the results follows two separate approaches; the consistency of data and model, given their uncertainties, will be verified using the instantaneous ensemble statistics, as well as the average statistics obtained by comparing the ensemble mean with the data over the 10 years. The analyses are furthermore compared with the

control run, in order to judge whether the assimilation has brought the model closer to the data. Finally, the results are compared to independent data which are not used in the assimilation.

### *6.2. Time series of ensemble statistics*

The Kalman Filter equations show that the uncertainty in the analysis will be smaller than that in the forecast if the assimilated observations have a finite error variance. When using an ensemble, these uncertainties are assumed to be proportional to the spread of the ensemble before and after assimilation. Figure 3 shows time series over the 10 year run of the mean ensemble spread of the forecast and analysis ensembles at the locations of the observations (the values for the Tropical Pacific are shown). The figure shows that ensemble spread in temperature and sea level is reduced at each assimilation step throughout the run (the analysis spread is consistently smaller than the forecast spread), as expected from the theory. Furthermore, the forecast spread tends to be larger than the analysis spread at the previous time step, and shows no signs of ensemble collapse, suggesting that the forcing perturbations are effective in maintaining the forecast uncertainty at a reasonable level. The figure for salinity is much more noisy than that for temperature and sea level because of the small number of observations over which to average.

Three periods can be distinguished in the temperature time series. From 1987 to mid 1992 the ensemble maintains an approximately constant level of spread. A clear reduction of ensemble spread at the locations of the temperature observations is seen

after the introduction of sea level data in October 1992. The spread increases again in the middle of 1993. This increase was found to be due to a misspecification of the data error variances by a factor 2 (the standard error by a factor 1.4) from this point onwards. While this has brought the analysis error diagnosed from the ensemble spread back to the pre-altimetry level, the forecast spread typically remains at lower levels than those of the 1987-1992 period. The ensemble spread timeseries for temperature (all subsurface observations) strongly resembles that of the spread in sea level. This suggests that the sea level assimilation has the strongest impact on the ensemble, with the temperature and salinity assimilation only adding relatively small additional corrections.

The character of the corrections themselves is illustrated in Figs. 4 and 5. Figure 4 shows the typical magnitude of the innovations (determined as the root mean-square of the observation-model misfits) which were determined for control, forecast and analysis at each assimilation time step. Again the focus is on the Tropical Pacific. Looking at the temperature innovations first, it is clear that the differences with the data are largest for the control and smallest for the analysis, with the forecast values somewhere inbetween. This picture is consistent throughout the run, although the differences between control and analyses become increasingly smaller towards the end of the run. Also noticeable is the fact that the magnitude remains more or less at a constant level, where one might expect a drop in the analysis innovation magnitude during the first stages of the run, in which the system adjusts to a state which is more consistent with the data. Something like this is seen for the sea level record, although the initial drop is followed by a steady increase towards the end of the run. The fact that this behaviour is also seen in the

control misfits, suggests that the analysis ensemble is not very strongly constrained by the data.

A further measure of the performance of the ensemble is presented in Fig. 5, which shows the time series of the average of the innovations themselves (as opposed to the average of the RMS values). This quantity may point towards systematic large-scale differences between the model and the data. The Tropical Pacific can contain smaller regions with persistent biases of opposing sign which may cancel in the average over the entire domain, but the time series may still identify long-term trends in offsets between the model and data. In fact, it was used to identify a bias in the altimetry data (later corrected), which should be removed if one assimilates sea level anomalies. A persistent bias is seen in the control temperatures, which is not as prominent in the ensemble results, suggesting that the assimilation has produced analyses with a better overall mean subsurface state than the control. After the introduction of sea level data however, a slight bias of opposing sign appears in the analysis temperature, perhaps indicating an overadjustment of subsurface temperatures as a result of a sea level mismatch between model and altimetry. Curiously, this bias results from the assimilation of sea level anomalies rather than absolute values, and is not reproduced in the sea level innovation time series. A possible explanation is discussed in section 7.

### *6.3. Rank histograms*

A further indication of the consistency between model and data can be obtained using so-called rank histograms (see e.g. Hamill, 2001). This diagnostic was originally developed

for the verification of ensemble forecasts and provides an indication of the reliability of the ensemble; if an event has a certain probability of occurring in reality, the ensemble forecast should suggest the same probability. This can be verified by repeatedly tallying the rank of the verification (in principle the true state) relative to the sorted (from low to high) ensemble. If the ensemble is reliable, in the sense defined above, the observation and ensemble members can be considered random samples of the same probability distribution, which would result in an equal chance for any rank to occur, reflected by a flat histogram. Deviations from a flat histogram are an indication of problems with the ensemble. For example, a bias in the ensemble will cause the verification to appear at one side of the sorted ensemble more often than on the other side, resulting in a shifted and sloped rank histogram. Undervariability of the ensemble has a similar effect, but with equal chance on the verification appearing on both ends of the sorted ensemble, leading to a U-shaped histogram. Figure 6 shows rank histograms based on the ensemble forecasts, using the data as verification. Hamill (2001) showed that when the verification has a known uncertainty, a random error should be added to each ensemble member before determining the rank. In this case the estimated data constraint errors for temperature, salinity and sea level anomalies were used. The histograms were determined separately for the three types of observations and for different geographical regions. Figure 6 shows that in all cases there is a relative overpopulation of the middle ranks, typically indicating an excess of variability in the ensemble. However, when a random data error is added to the ensemble members, as is the case here, it may also indicate an overly pessimistic data error estimate. No indication of bias is found for the sea level verification, which is

not surprising since sea level anomalies rather than absolute values are compared. For temperature, and in particular salinity, however, there are indications of offsets between the model ensemble and the data (in the North Atlantic in particular). There are also regions, such as the Tropical Indian Ocean, where a U-shaped appearance of the histogram is found in combination with an apparent excess of ensemble variability. The high number of extreme ranks may indicate different biases in subregions or changes in a bias over the period of the run.

#### *6.4. Subsurface states*

A complementary picture of some of the apparent improvements and inconsistencies identified in the previous section can be obtained by comparing the three-dimensional ocean states from the control and assimilation run (the analyses) with the data. Time-constant and time-dependent model-data differences are considered.

First of all, the temperature biases in control and analyses are shown in Fig. 7. Differences between the data and the control run show a fully three-dimensional pattern. Surface temperatures are predominantly too low in the control run. Larger biases are associated with displacements or the strength of strong currents, most notably the North Equatorial Current and the North Equatorial Counter Current in the Eastern Tropical Pacific, and several of the major western boundary currents, such as the Gulf Stream, the Kuroshio, and the confluence of the Malvinas and Brazil Currents. These latter displacements are also found in the salinity signal and extend to large depth, reflecting the strongly barotropic character of these currents. The salinity bias is mostly confounded

to the surface (only the plot for 50m depth is shown here), and has a very large-scale character. Of particular interest are the strong biases in salinity of the East Greenland Current, the Tropical and North Pacific, and the subpolar gyre of the North Atlantic. Figure 7 also shows the analysis-control bias. In many places the patterns have the same sign and shape as for the data-control bias, but with smaller amplitude, indicating that the assimilation has reduced the bias in the correct direction, but only partly. The correction of mid- to high-latitude biases is less effective than of those in low latitudes, since only data equatorward of  $50^\circ\text{N}$  and  $40^\circ\text{S}$  have been assimilated. Also weaker, but large scale patterns appear to have been properly adjusted, with the exception of the large data-control bias maximum in the central North Pacific. The assimilation has increased the salinity in the North Atlantic with respect to the control, but has done this just east and southward of where the data suggests it should have been increased. The distribution of salinity data is insufficient to judge whether the strong increase in surface salinity just west of Mexico’s Baja California peninsula is realistic. Curiously, it appears to be strongly related to a temperature bias correction at 150m depth instead of 50m depth.

The time-dependent improvements in the analyses are investigated in Fig. 8, which compares the model error  $\sqrt{\langle \mathbf{p}\mathbf{p}^T \rangle}$  (see Eq. (9)) from the control with the corresponding estimate based on the analyzed states. The reader is reminded here that these errors do not reflect differences with the real ocean, but with the true model state, as discussed in section 4. Time-dependent temperature and salinity errors in the control (as far as they can be detected from the limited data) are primarily associated with thermocline variability in the Tropics, and to a lesser extend with the western boundary currents.

Increased error levels are found at the depth of the Tropical Pacific thermocline which shoals from west to east. The corresponding figures for the analysis error estimates show only a weak reduction of the major errors found in the control on the order of  $0.5^{\circ}$  to  $1^{\circ}\text{C}$ . The improvement in the sea level simulation is much more significant however. Time-dependent sea level errors in the control run have typically been reduced by half through the assimilation. The large scale patterns of sea level error in the Southern Ocean can be associated with high-frequency wind-driven barotropic motions in the ocean (Stammer et al., 2000). Since these signals are not resolved by the altimetry, which has a relatively low-frequency sampling pattern, the data error estimation method as described by Eqs. (8) and (9) incorrectly ascribes this variability to errors in the model.

### *6.5. Independent data*

A final test of the performance of the assimilation system is a comparison with independent data. Almost all available temperature, salinity and altimetric sea level data have been used in the assimilation, which leaves very few sources of independent data. Zonal velocity profiles measured by current meters at several TAO moorings are used here to check whether the analyses are better than the control. Figure 9 shows comparisons for the zonal velocity component at two sites on the equator. The first of these two, at  $165^{\circ}$  east in the western Pacific, contains a significant seasonal signal and relatively little high-frequency variability, as opposed to the second site at  $220^{\circ}$  east in the central Pacific, where velocities are typically 1.5 to 2 times higher. Both the control and analyses capture the main characteristics of the current variability, which is strongly controlled by large-scale

wind stress variability. A clearer picture of the differences between the data, control and assimilation is obtained from estimates of the constant bias and time-dependent differences between the model states and the data. These are summarized in Fig. 10 for all depths at which data are available. The assimilation has shifted the entire velocity profile at 165°E to lower values. In general this has led to offsets which are smaller in the assimilation than in the control, with the exception of the surface and 200m depth. The uniform character of the shift suggests that the change is perhaps more strongly related to large scale adjustments of the equatorial circulation than to the impact of assimilation near this particular site. The assimilation has left the time-dependent differences between model and data more or less unaltered, as illustrated by the RMS plot. A very similar picture appears for the site at 220°E, the main difference being that only velocities below 50m are affected by the assimilation. The changes at lower depths represent a reduction in the strength of the Equatorial Under Current (EUC). In this case, the mean assimilated state is further away from the data than the mean control state. Assimilation of temperature profile data is known to affect the position and strength of the EUC negatively in many systems because of its dependence on the delicate balance between the zonal pressure gradient and surface wind stress (e.g. Huddleston et al., 2004). Bias corrections schemes as developed by Bell et al. (2004) may be able to counter unwanted negative impacts of assimilation on the equatorial current profile.

## 7. Summary and conclusions

A global ocean data assimilation system is presented and tested during a 10-year long hindcast experiment. In this system the MPI ocean model is used to produce a 64-member ensemble of ocean forecasts which is combined with insitu profiles of temperature and salinity and along-track altimetric sea level observations using the Ensemble Kalman Filter. Three-dimensional data constraint error fields have been estimated for all three observables using a statistical method that compares the model states of an unconstrained control run with all available data over the same period. This method is successful in identifying representation errors in regions of the global ocean characterized by small scale variability that can not be resolved by the model. It also allows the estimation of the time-dependent error variance of the control run. Systematic differences between control run and data are accounted for by adding a term to the data constraint error, which is reduced during the run. A forcing perturbation method is used that represents uncertainty in four surface forcing fields by quasi-random combinations of the dominant orthogonal modes of ERA40-NCEP/NCAR reanalyses. The quasi-random character is due to imposed temporal correlations between daily perturbations. The assimilation scheme is a localized version of the standard EnKF. Results from the hindcast experiment shows that the forcing perturbations are effective at maintaining ensemble spread throughout the run. Time series at the observation locations show a reduction and subsequent rebound of ensemble spread which can be associated with the introduction of sea level data and a misspecification of data error

by a factor  $\sqrt{2}$  respectively. RMS values of the temperature innovations are smaller for the analyses than for the control run but this difference is deduced from 1993 onwards. The similarity in the time series of RMS sea level innovations between control and analyses suggests that the analyses are not very strongly constrained by the insitu data. This may be due to the relatively small number of profile data in the Tropical Pacific compared to the number of sea level data. The subsampling of the profile data set may contribute as well. The impact of a weak data constraint can also be seen in a comparison of the constant and time-dependent three-dimensional error fields in the control and analyses. The systematic differences with the data are reduced by the assimilation but not removed entirely. The reduction of time-dependent errors, defined as differences with a true model state, is rather modest for temperature but significant for sea level, particularly in the tropics. Rank histograms verifying the reliability of the ensemble based on the observations suggest an excess of variability in the ensemble in several geographical regions and for all three observables. It is argued, however, that this may also be due to an overly pessimistic estimate of the data error. This would be in agreement with the earlier observation that the assimilation has only produced fairly modest corrections, and with the overspecification of errors during the last few years of the assimilation run. The rank histograms also suggest the presence of remaining model-data biases, which is consistent with the time-mean three-dimensional analysis error fields. Final comparisons with independent velocity data on the equator again show limited improvement in the time-dependent part of the velocity profile, but reveal a shift in the mean profile between control and analysis, indicative of an assimilation-induced

bias. While the model error estimation method had already been used previously in sea level assimilation applications, it produces very reasonable patterns and magnitudes for temperature and salinity profiles as well, and provides a good alternative for the ad hoc methods that are common practise. The apparently dominant impact of the sea level assimilation on the ensemble will need to be re-evaluated. One may prefer to place more weight on the insitu data. The strong indications that data constraint errors are too large in this hindcast experiment suggest that some tuning of the system should be done before it is applied to operational seasonal forecasting. A short run with reduced error values should suffice to detect remaining inconsistencies. The procedure that has been adopted here to account for biases between data and model is not entirely satisfactory. Assimilation is known to create a bias in the equatorial region through disturbance of the balance between the near-surface pressure gradients and the surface wind stress. This is the most likely cause for the increase of the bias in the velocity profile as indicated in the comparison with TAO current meter data from the central Pacific. Such biases were also found in identical-twin experiments with the EnKF where model error is not an issue (Leeuwenburgh, 2005). The implementation of recently developed bias correction schemes (Dee and Da Silva, 1998; Bell et al., 2004) with ensemble methods should therefore be considered. The assimilation scheme that has been presented here could provide a convenient extension to seasonal forecast systems. The use of an ensemble allows for a probabilistic forecast rather than a deterministic one based on a single model. The next step would therefore be to test the impact of the improved ocean initialisation of coupled model forecast skill.

## **Acknowledgements**

This work was supported by the EU through its funding for the ENACT project. The author would like to thank Gerrit Burgers and Geert Jan van Oldenborgh (KNMI), Laurent Bertino (NERSC), Helmuth Haak (MPIfM), Noel Keenlyside (IfM), and Alberto Troccoli and Paul Dando (ECMWF) for their suggestions and support.

## References

Bell, M. J., Martin, M. J., and Nichols, N. K., Assimilation of data into an ocean model with systematic errors near the equator, *Q. J. R. Meteorol. Soc.*, 130, 873–893, 2004.

Borovikov, A., M. M. Rienecker, C. L. Keppenne, and G. C. Johnson, Multivariate error covariance estimates by Monte-Carlo simulation for assimilation studies in the Pacific Ocean, *submitted to Mon. Wea. Rev.*, 2004.

Burgers, G., P. J. van Leeuwen, and G. Evensen, Analysis scheme in the Ensemble Kalman Filter, *Mon. Wea. Rev.*, 126, 1719-1724, 1998.

Burgers, G., The El Nio Stochastic Oscillator, *Clim. Dyn.*, 15, 521-531, 1999.

Chen, D., Cane, M. A., Kaplan, A., Zebiak, S. E., and Huang, D., Predictability of El Nino over the past 148 years, *Nature*, 428, 733–734, 2004.

Cooper, M., and K. Haines, Altimetric assimilation with water property conservation, *J. Geophys. Res.*, 101, 1059-1077, 1996.

Dee, D. P. and Da Silva, A. K., Data assimilation in the presence of forecast bias, *Q. J.*

*R. Meteorol. Soc.*, 117, 269–295, 1998.

Derber, J., and Rosati, A., A global oceanic data assimilation system, *J. Phys. Oceanogr.*, 19, 1333–1347, 1989.

Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, C99, 10143-10162, 1994.

Evensen, G., The ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dyn.*, 53, 343-367, doi:10.1007/210236-003-0036-9, 2003.

Fischer, M., M. Flügel, and M. Ji, The impact of data assimilation on ENSO simulations and predictions, *Mon. Wea. Rev.*, 125, 819-829, 1997.

Fu, L.-L., Fukumori, I., and R. N. Miller, Fitting dynamic models to the Geosat sea level observations in the Ropical Pacific Ocean. Part II: A linear, wind-driven model, *J. Phys. Oceanogr.*, 23, 2162–2181, 1993.

Fukumori, I., R. Raghunath, L.-L. Fu, and Y. Chao, Assimilation of TOPEX/POSEIDON altimeter data into a global ocean circulation model: How good are the results, *J. Geo-*

*phys. Res.*, 104, 25647-25665, 1999.

Gaspari, G., and S. E. Cohn, Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, 125, 723-757, 1999.

Hamill, T. M., Interpretation of rank histograms for verifying ensemble forecasts, *Mon. Wea. Rev.*, 129, 550-560, 2001.

Houtekamer, P. , and H. Mitchell, A sequential Ensemble Kalman Filter for atmospheric data assimilation, *Mon. Wea. Rev.*, 129, 123-137, 2001.

Ingleby, B., and M. Huddleston, Quality control of ocean profiles - historical and real time data, *submitted to J. Marine Sys.*, 2004.

Ji, M., Leetma, A., and J. Derber, An ocean analysis system for seasonal to interannual climate studies, *Mon. Wea. Rev.*, 123, 460-481, 1995.

Ji, M., R. W. Reynolds, and D. W. Behringer, Use of TOPEX/Poseidon sea level data for ocean analyses and ENSO prediction: Some early results, *J. Climate*, 13, 216-231, 2000.

Kalnay, E., and co-authors, The NCEP/NCAR 40 -year reanalysis project. *Bull. Amer. Meteorol. Soc.*, 77, 437-471, 1996.

Keppenne, C. L., and M. M. Rienecker, Initial testing of a massively parallel Ensemble Kalman Filter with the Poseidon isopycnal ocean general circulation model, *Mon. Wea. Rev.*, 130, 2951-2965, 2002.

Keppenne, C. L., and M. M. Rienecker, Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter, *J. Mar. Sys.*, vol. 40-41, 363-380, 2003.

Leeuwenburgh, O., Assimilation of along-track altimeter data in the Tropical Pacific region of a global OGCM ensemble, *Q. J. R. Meteorol. Soc.*, accepted for publication, 2005.

Robert, C., and O. Alves, Tropical Pacific ocean model error covariances from Monte-Carlo simulations, to appear as a BMRC research report, 2004.

Stammer, D., Wunsch, C. and Ponte, R., De-aliasing of global high-frequency barotropic motions in altimeter observations, *Geophys. Res. Lett.*, 27, 1175–1178, 2000.

Steele, M., R. Morley, and W. Ermold, PHC: A global ocean hydrography with a high quality Arctic Ocean, *J. Climate*, 14, 2097-2087, 2001.

Vossepoel, F. C., and D. W. Behringer, Impact of sea level assimilation on salinity variability in the western equatorial Pacific, *J. Phys. Oceanogr.*, 30, 1706-1721, 2000.

Weaver, A. T., Vialard, J., and D. L. T. Anderson, Three and four-dimensional variational assimilation with a general circulation model of the Tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks, *Mon. Wea. Rev.*, 131, 1360–1378.

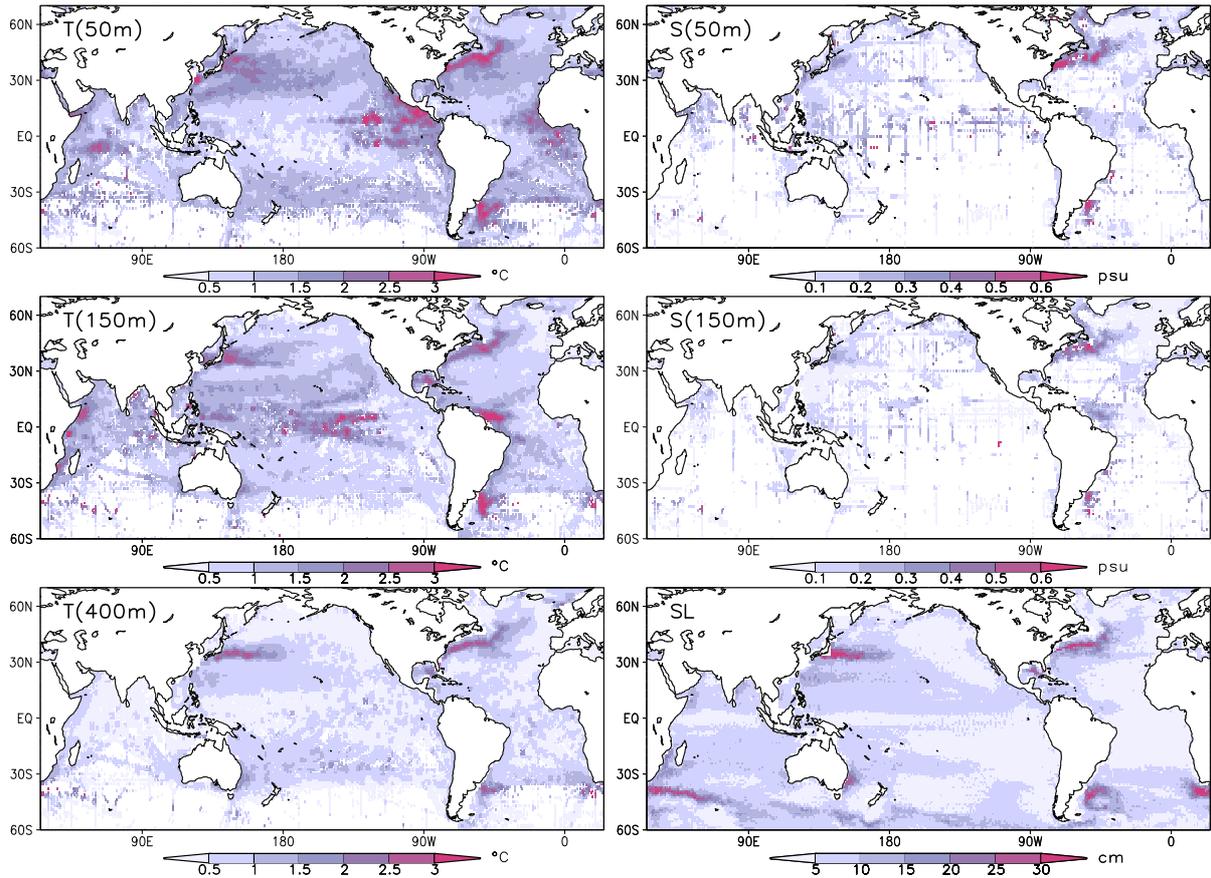


Figure 1: Data constraint errors, calculated using Eq. (8), for temperature at 50m, 150m, and 400m (left), and for salinity at 50m and 150m and sea level (right).

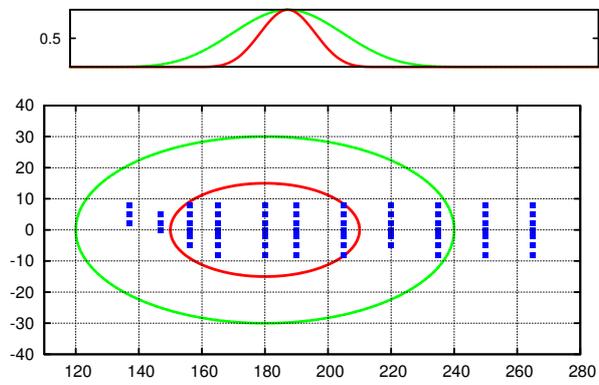


Figure 2: Schematic of the observation influence ellipse and localisation function applied to ensemble covariances for the analysis grid point located at (180E, 0N). The dots indicate the distribution of the TAO array buoys.

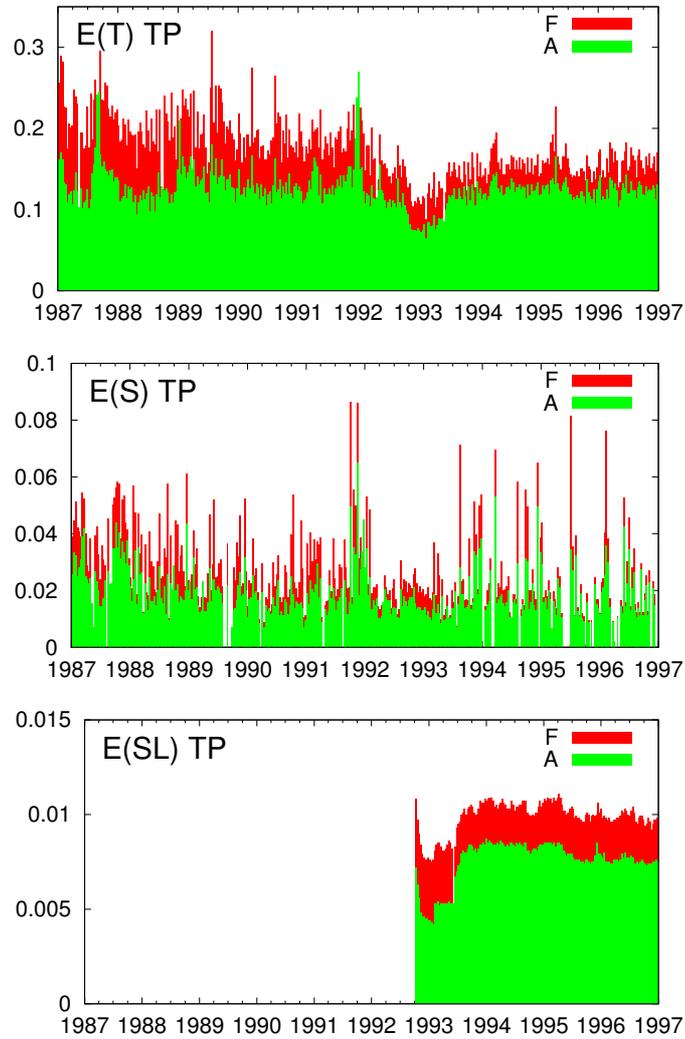


Figure 3: Time series of ensemble spread before (F) and after (A) assimilation, averaged over all observations of temperature ( $^{\circ}\text{C}$ ), salinity (psu), and sea level (m) respectively, in the Tropical Pacific.

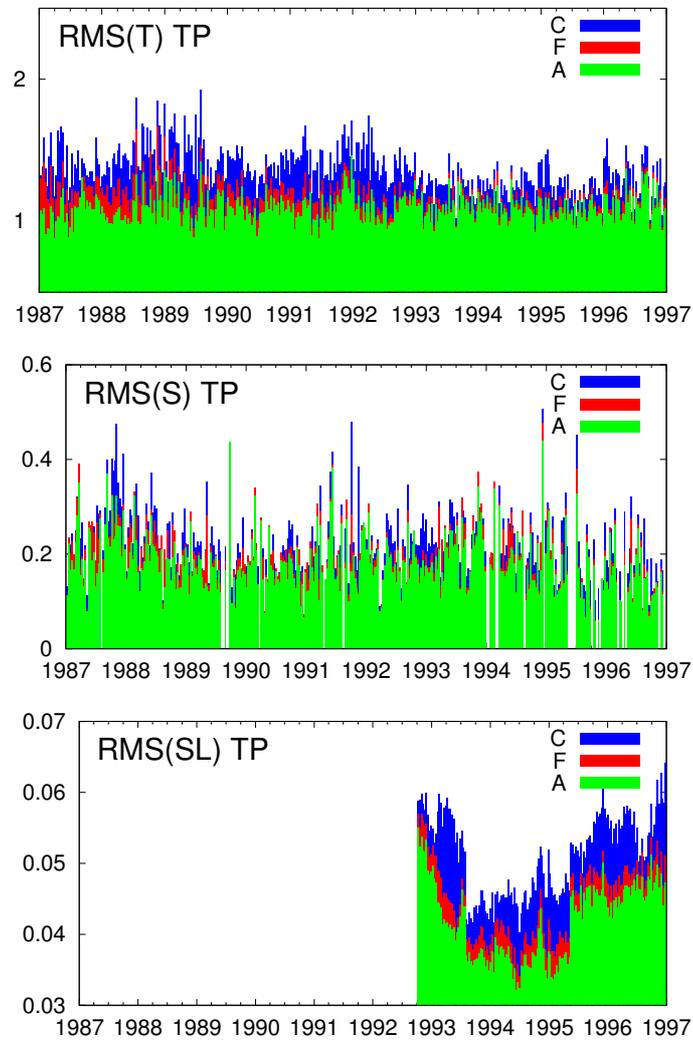


Figure 4: Time series of the root-mean-square of model-data misfits before (F) and after (A) assimilation, and for the control (C), averaged over all observations of temperature ( $^{\circ}\text{C}$ ), salinity (psu), and sea level (m) respectively, in the Tropical Pacific.

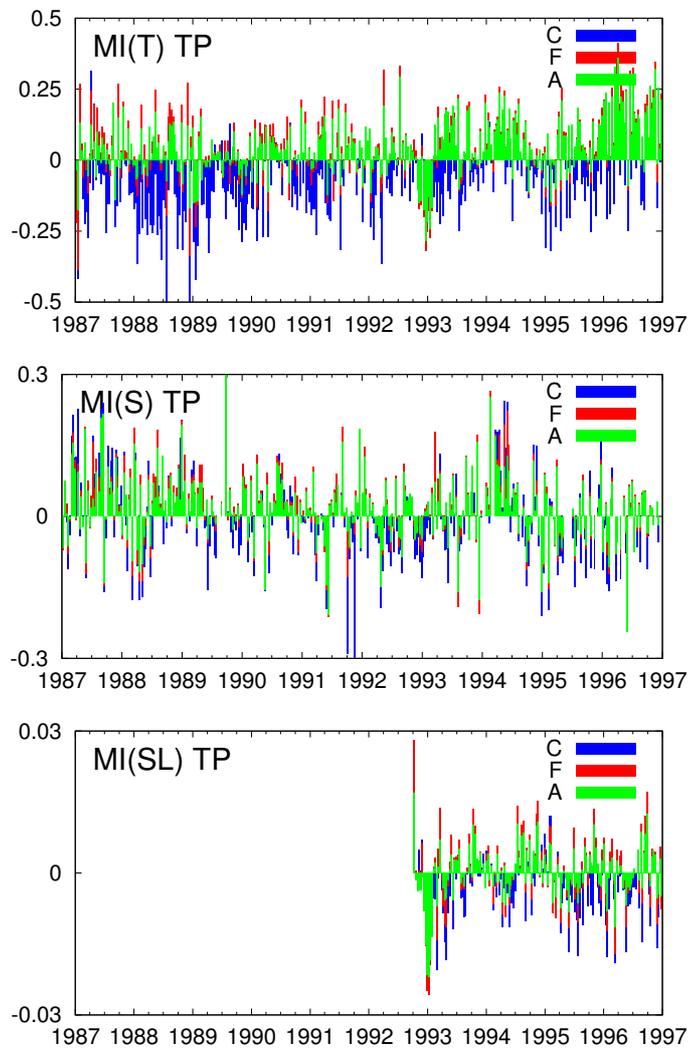


Figure 5: Time series of model-data misfits before (F) and after (A) assimilation, and for the control (C), averaged over all observations of temperature ( $^{\circ}\text{C}$ ), salinity (psu), and sea level (m) respectively, in the Tropical Pacific.

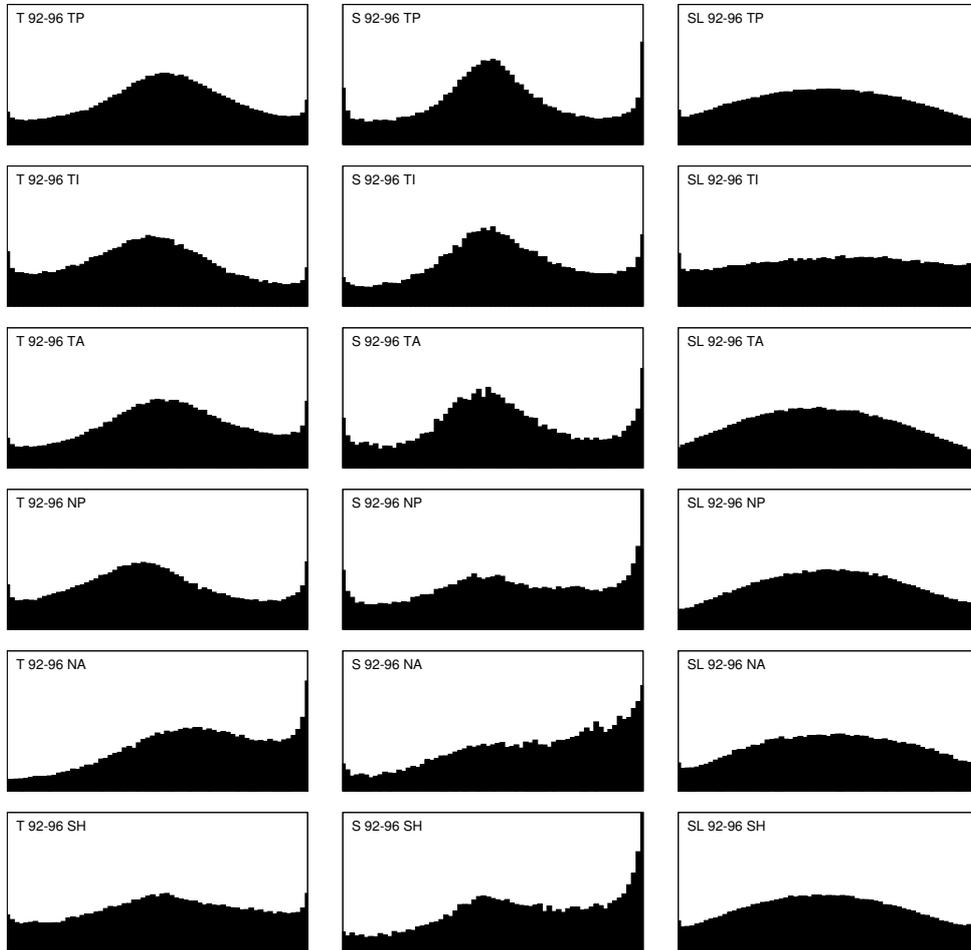


Figure 6: Rank histograms for temperature, salinity and sea level in 6 regions; TP (120-275E,23S-23N), TI (30-110E,23S-23N), TA (290E-20E,23S-23N), NP (120-260E,23N-50N), NA (290E-360E,23N-50N), and SH (0-360E,23S-40S). The horizontal axes represent the rank, ranging from 1 to 65, and the vertical axes represent the probability  $P(\text{rank})$ , ranging from 0 to 0.04.

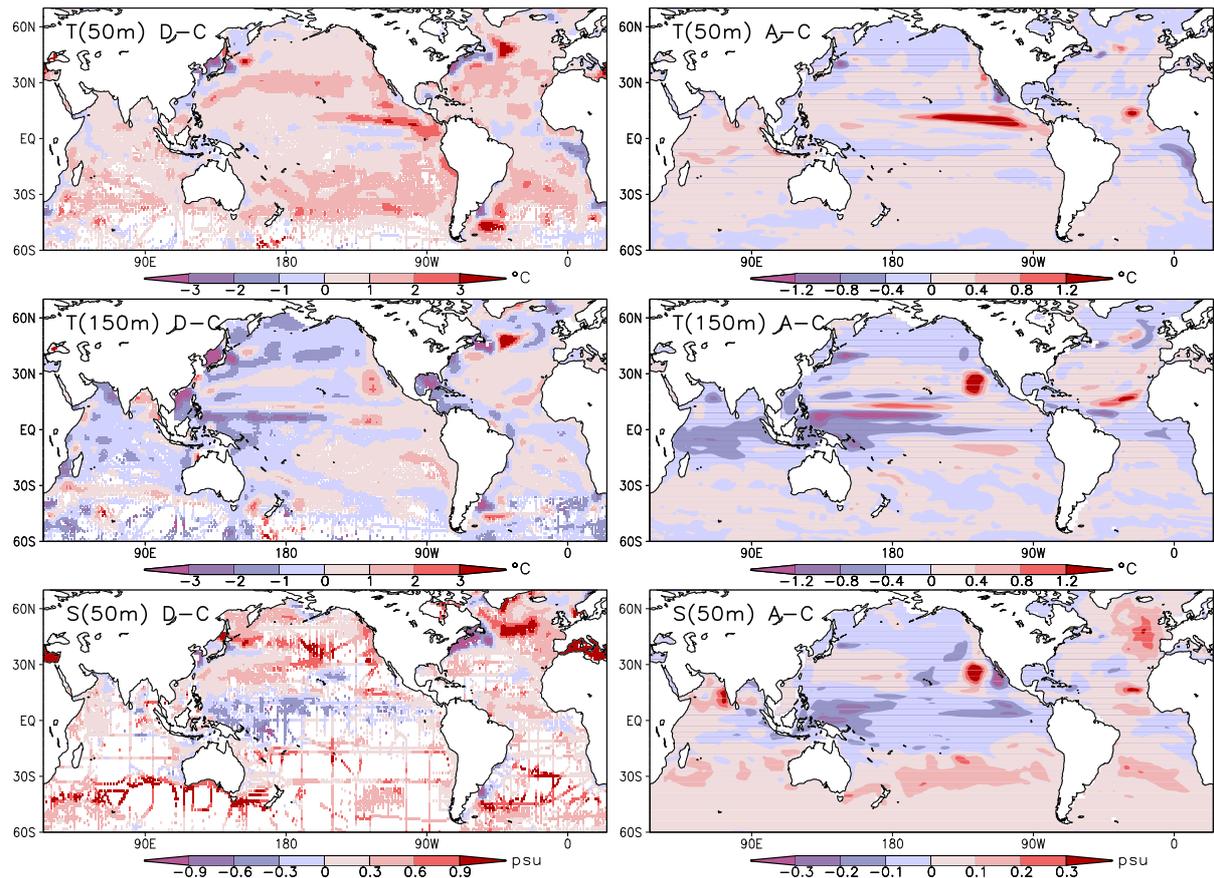


Figure 7: Mean difference over experiment period between data and control (left) and between analyses and control (right) for temperature at 50m and 150m depth and for salinity at 50m depth.

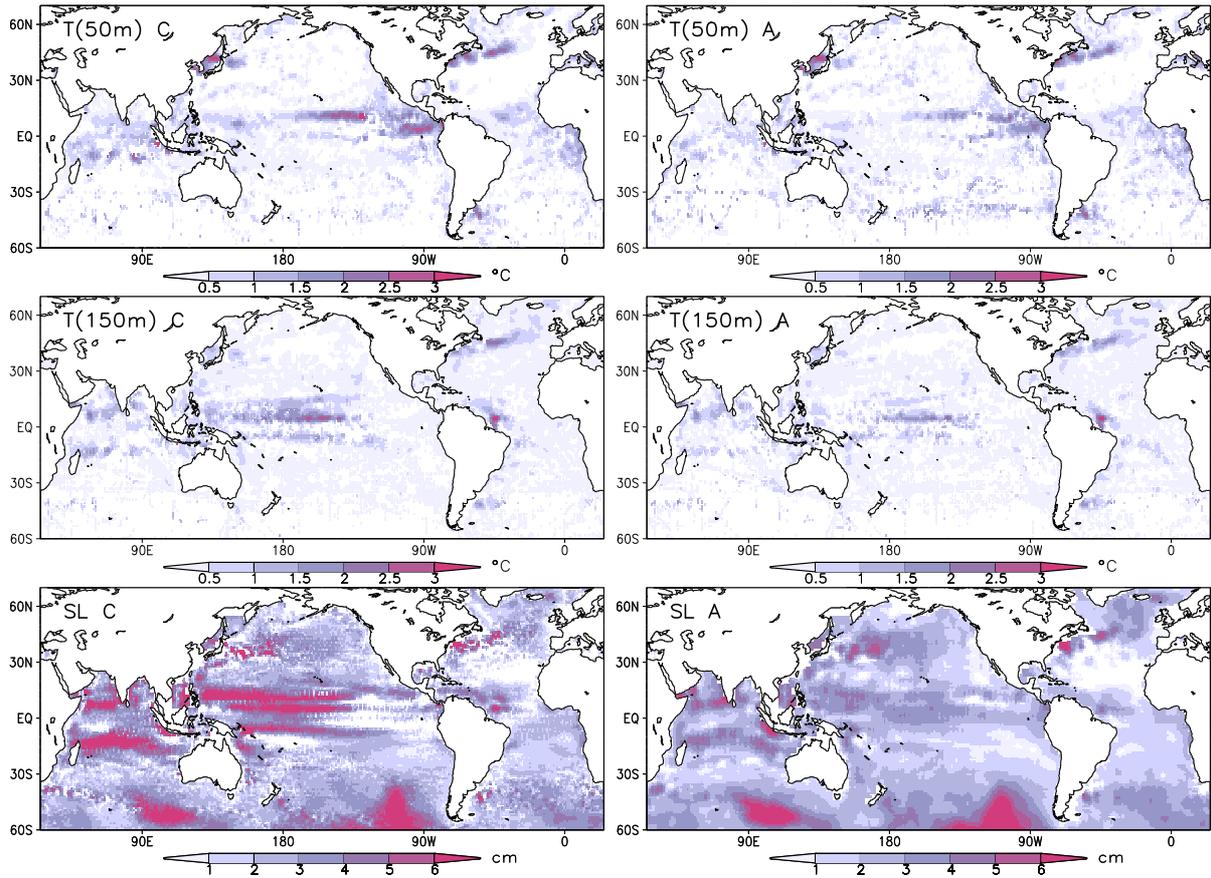


Figure 8: Time-dependent error (standard deviation) for the control (left) and analyses (right) for temperature at 50m and 150m depth, and for sea level, calculated from Eq. (9).

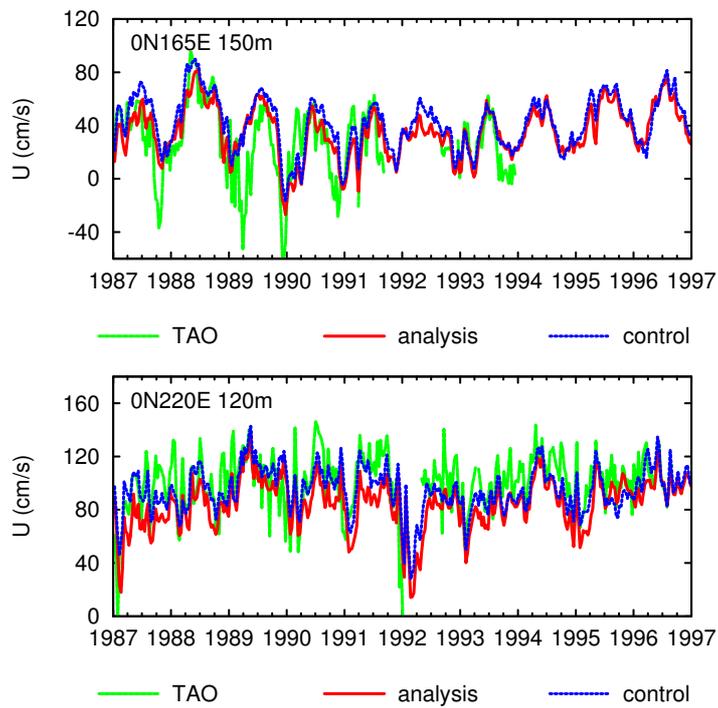


Figure 9: Time series of zonal velocity at  $165^{\circ}\text{E}$ ,  $0^{\circ}\text{N}$  and 150m depth, and at  $220^{\circ}\text{E}$ ,  $0^{\circ}\text{N}$  and 120m depth from the control and analyses, compared with current meter measurements.

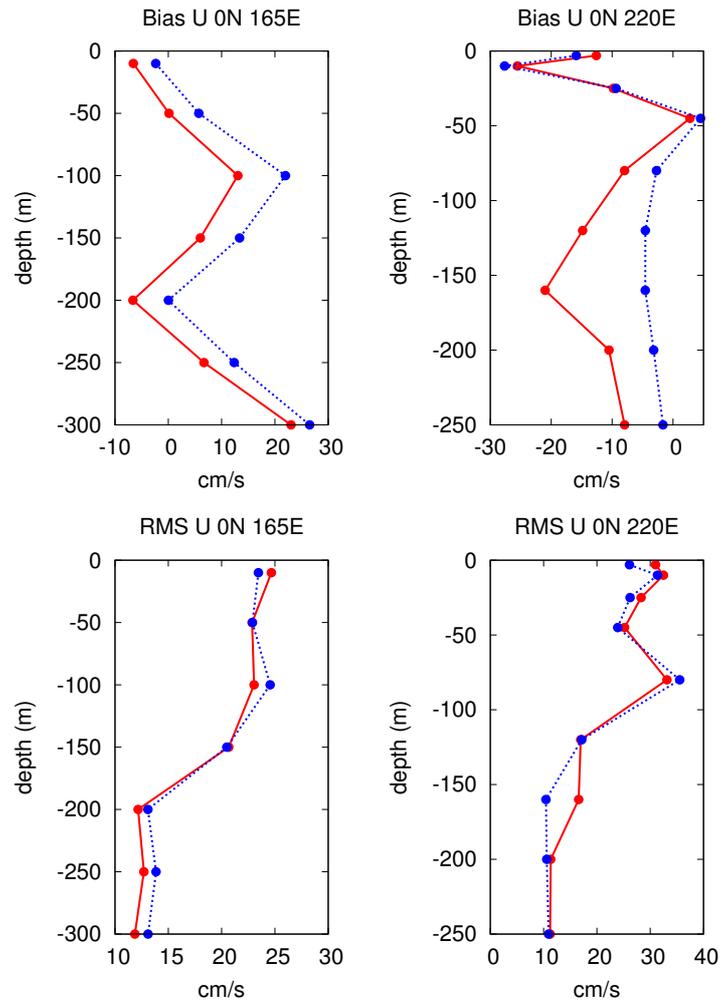


Figure 10: Bias and RMS difference in the velocity profile between control and data (blue), and between analyses and data (red) at 165°E and 22°E.