The Global Ocean Circulation During 1992 –1997, Estimated from Ocean Observations and a General Circulation Model

D. Stammer[1], C. Wunsch[2], R. Giering[3] C. Eckert[2], P. Heimbach[2], J. Marotzke[4], A. Adcroft[2], C.N. Hill[2], and J. Marshall[2]

Submitted to

Journal of Geophysical Research

Revised Manuscript

---

[1] Corresponding Author: Detlef Stammer, Scripps Institution of Oceanography, La Jolla, CA 92093, U.S.A; dstammer@ucsd.edu

[2] Massachusetts Institute of Technology

[3] FastOpt, GBR, Hamburg

[4] Southampton Oceanography Centre

Abstract

We discuss the three-dimensional oceanic state estimated for the period 1992 - 1997 as it results from bringing together large-scale ocean data sets with a general circulation model. To bring the model into close agreement with ocean data, its initial temperature and salinity conditions where changed as well as the time-dependent surface fluxes of momentum, heat and freshwater. Resulting changes of those control fields are largely consistent with accepted uncertainties in the hydrographic climatology and meteorological analyses. Our results show that the assimilation procedure is able to correct for the traditional shortcomings of the flow field by changing the surface boundary conditions. Changes of the resulting flow field are predominantly on the gyre scale and affect many features which are often poorly simulated in traditional numerical simulations, such as the strengths of the Gulf Stream and its extension, the Azores Current and the anticyclonic circulation associated with the Labrador Sea.

A detailed test of the results and their consistency with prior error assumptions shows that the constrained model has moved considerably closer to those observations which have been imposed as constraints, but also to independent data from the World Ocean Circulation Experiment not used in the assimilation procedure. In some regions where the comparisons remain indeterminate, not enough ocean observations are available. And in such situations, it is difficult to ascribe the residuals to either the model or the observations.

We conclude from this experiment that we can find an acceptable solution to the global time-dependent ocean state estimation problem. As the estimates improve through the evolution of numerical models, computer power increases, and better assimilation schemes, improved and routine estimates will become possible.

1 Introduction

Because the ocean circulation shows vigorous variability on a wide range of temporal and spatial scales, a substantial emphasis has to be put on adequately observing it. Many of the most important scientific properties of the circulation, e.g. its fluxes of heat, carbon, or potential vorticity, are never actually measured, but can only be calculated from estimates of the circulation. A combination of regional or global ocean data sets with a state-of-the-art numerical circulation model is therefore required to exploit the diverse data types, and to obtain the best estimate of the time-varying ocean circulation. This process, known as ocean state estimation, has some commonality with ongoing analysis and reanalysis activities in the atmospheric community. But there are substantial differences from meteorological procedures because the present focus is much less on forecasting and much more on estimating the oceanic state to improve our understanding of ocean dynamics. Furthermore, the technologies of oceanic observation are sometimes radically different from those used in the atmosphere, and this consideration also leads to differing estimation problems and solutions. Finally, the much smaller dynamical scales present in the ocean present a more difficult computational burden.

A complete ocean state estimation system will eventually combine the entire suite of large-scale ocean observations of any type, with the dynamics of an ocean circulation model. The aim of this paper is to introduce a prototype of such an estimation system, but which uses only a subset of available ocean data, and not the most complete model physics. Our primary focus here is on the time-evolving global circulation as it emerges from the monthly mean *Levitus et al.* (1994) hydrographic climatology, monthly mean *Reynolds and Smith* (1994) sea surface temperature (SST) fields, the altimetric measurements from TOPEX/POSEIDON (T/P) and ERS-1,2 and daily surface forcing over the time interval 1992 through 1997. These prototype results are sufficiently conclusive to be scientifically useful and demonstrate both that a complete World Ocean Circulation Experiment (WOCE)-type data synthesis is now possible, and that ongoing, near-operational ocean state estimates are at hand.

Results are presented in several parts. Here, in Part 1, we describe details of the model and the optimization, and provide an overall description of the estimated mean state and the seasonal variability in the solution. This includes a test of the results through comparison with independent information available from large-scale WOCE data sets, including the global hydrography, the XBT data. A more thorough analysis of the estimated state for oceanic property transports and pathways will be found in a separate paper (*Stammer et al.*, 2001a). The estimated surface fluxes and are being evaluated in detail by *Stammer et al.*, (2001b). *Ponte et al.*, (2001) have already shown that the results have improved skill in predicting the earth's angular momentum balance, in what is a globally integrated test of these results.

2 The Formalism and Models

2.1 The Formalism

State estimation is a major part of control theory, in which one combines the dynamics embedded in numerical circulation models, statistical information in form of prior data and error covariances and observational information from ocean data. Formally we bring a numerical model into consistency with ocean observations by solving a constrained least-squares optimization problem. Recent textbook accounts of ocean applications are provided by *Bennett* (1992) and *Wunsch* (1996), and there are numerous collections of research papers on this subject, e.g., *Malanotte-Rizzoli* (1996). We therefore summarize here only briefly the present discrete-time methodology, using a notation adopted from *Wunsch* (1996).

Algebraically, a GCM can be written in canonical form as,

$$\mathbf{x}(t+1) = \mathcal{L}[\mathbf{x}(t), \mathbf{B}\mathbf{q}(t), \mathbf{\Gamma}\mathbf{u}(t)], \tag{1}$$

where $\mathbf{x}(t)$ is the state vector, $\mathcal{L}$ represents the full non-linear operator stepping the model forward in time starting from a prescribed initial condition $\mathbf{x}(t_0)$. $\mathbf{q}(t)$ represents all externally specified boundary conditions, and sources and sinks. The state vector contains all the physical variables (here three components of velocity, pressure, temperature and salinity) necessary to calculate the system one time step into the future, when given necessary boundary conditions. Matrices $\mathbf{B}$, $\mathbf{\Gamma}$ are used to map the known surface forcing fields $\mathbf{q}(t)$, and the unknown "controls" $\mathbf{u}(t)$ onto the model grid. Formally we think of $\mathbf{u}(t)^T = [\mathbf{u}_0^T, \mathbf{u}_f^T, \boldsymbol{\epsilon}^T]$ as including separate components comprising errors in the initial conditions $\mathbf{u}_0$, the external forcing fields $\mathbf{u}_f(t)$, and the internal model physics $\boldsymbol{\epsilon}(t)$. In the present experiment, we keep $\boldsymbol{\epsilon}(t) = 0$; but this restriction will be relaxed explicitly in future applications.

Most oceanographic measurements are approximately a linear combination of the model state vector, e.g. velocity, temperature, and salinity, but are contaminated by noise,

$$\mathbf{y}(t) = \mathbf{E}(t)\mathbf{x}(t) + \mathbf{n}(t). \tag{2}$$

The "observation matrix" $\mathbf{E}$ relates the model state vector to observables and is normally very sparse, because observables usually involve only local subsets of $\mathbf{x}(t)$. Examples include along-track altimetric observations of the dynamically induced surface elevation, moored velocities time series at one point, or tomographic ray paths.

In general terms, we seek an estimate, $\tilde{\mathbf{x}}(t)$, of the state vector and its uncertainty, $\mathbf{P}(t)$, that is consistent with the observations, $\mathbf{y}(t)$ and their uncertainties, $\mathbf{R}(t)$, and with the model dynamics (1) and its uncertainty, $\mathbf{Q}(t)$. Reflecting the structure of $\mathbf{u}$, $\mathbf{Q}$ contains contributions from the initial conditions, the internal model physics, and the external forcing, i.e., $\mathbf{Q}^T = [\mathbf{Q}_0^T, \mathbf{Q}_f^T, \mathbf{Q}_\epsilon^T]$. The control variables are modified from their initial values so as to produce the change in the estimated $\mathbf{x}\,(t)$ that minimizes the model-data misfit measured by the quadratic "objective" or "cost" function,

$$J = \sum_{t=1}^{t_f} (\mathbf{y}(t) - \mathbf{E}\tilde{\mathbf{x}}(t))^T \mathbf{W}(t)(\mathbf{y}(t) - \mathbf{E}\tilde{\mathbf{x}}(t)), \tag{3}$$

subject to the model physics constraints. The model is enforced by appending it to the cost function in the form,

$$\begin{aligned} L \;&= J + \mathbf{u}_0^T \mathbf{Q}_0^{-1} \mathbf{u}_0 + \textstyle\sum_{t=0}^{t_f-1} [\boldsymbol{\epsilon}^T(t) \mathbf{Q}_\epsilon^{-1}(t) \boldsymbol{\epsilon}(t) + \mathbf{u}_f^T(t) \mathbf{Q}_f(t)^{-1} \mathbf{u}_f(t)] \\ &\quad -2 \textstyle\sum_{t=0}^{t_f-1} \boldsymbol{\mu}(t+1)^T \{\mathbf{x}(t+1) - \mathcal{L}[\mathbf{x}(t), \mathbf{B}\mathbf{q}(t), \boldsymbol{\Gamma}\mathbf{u}(t)]\}. \end{aligned} \tag{4}$$

with the $\boldsymbol{\mu}(t)$ being Lagrange multipliers and t_f being the final time step. By setting the derivatives of L with respect to $\mathbf{u}$, $\boldsymbol{\mu}$, $\mathbf{x}$ to zero, we find that the conditions for a stationary value of L (a constrained minimum of J) must satisfy the "normal equations" (see *Wunsch*, 1996). We note that in no sense is this optimization a "hard" or "strong" constraint one (see *Wunsch*, 1996 for details), a confusing terminology that refers to estimation problems which include *only* the initial conditions [here?].

2.2 The Models

For the dynamics, we use the M.I.T. ocean general circulation model (GCM) (*Marshall et al.*, 1997a) and its adjoint, which have been developed at the Massachusetts Institute of Technology. This model is based on the primitive equations on a sphere under the Boussinesq approximation. It consists of prognostic equations for horizontal velocity, heat, and salt, and an equation of state which are integrated forward in time on a staggered "C"-grid (*Arakawa and Lamb* 1977). Each time step the internal pressure is calculated from the hydrostatic relation and the vertical velocity is being diagnosed from continuity. Spatial coordinates are longitude, latitude, and height. A detailed description of the model is provided by *Marshall et al.* (1997a,b). For present purposes, we use a hydrostatic version with an implicit free-surface. A full surface mixed layer model is used (called KPP, *Large et al.*, 1994) and convective adjustment is used to remove gravitational instabilities underneath the planetary boundary layer.

The forward model is used to compute the model-data misfit. Without any adjustment to control variables, the initial "best-guess" model-data misfit $\mathbf{y}(t) - \mathbf{E}(t)\,\mathbf{x}(t)$ can be large. The adjoint model is used to provide the gradient of the cost function J with respect to the model variables; this gradient is then used to modify the control variables so as to reduce the value of J in an iterative scheme. A standard optimizing descent algorithm (here a quasi-Newton method, see *Gilbert and Lemaréchal,* 1989) is used with this information to determine the correction to the control terms.

Coding the adjoint of a complex numerical model is time consuming and difficult, comparable in effort to development of the forward code itself. Some care was therefore taken in writing the MIT model, making it possible to obtain the adjoint code from the forward code in a semi-automatic way through automatic differentiation (*Giering and Kaminski*, 1998). *Marotzke et al.*, (1999) give a full account for the construction of the adjoint model from the forward GCM code.

In practice, this system of automatic adjoint code generation has proven to be extremely flexible, and it permits relatively easy regeneration of the adjoint code whenever

Level	Thickness	Layer Center	Level	Thickness	Layer Center
1	10	5	12	200	610
2	10	15	13	275	847.5
3	15	27.5	14	350	1160
4	20	45	15	415	1542.5
5	20	65	16	450	1975
6	25	87.5	17	500	2450
7	35	117.5	18	500	2950
8	50	160	19	500	3450
9	75	222.5	20	500	3950
10	100	310	21	500	4450
11	150	435	22	500	4950

Table 1: Model layer thicknesses and center depths (meters)

a change in the forward code, or objective function, is necessary. We use the adjoint model here for the single purpose of estimating the ocean circulation through finding a minimum of a constrained optimization problem. But the adjoint solution has also the important property of providing a measure of sensitivity of the cost function to the physical variables of the system; see *Marotzke et al.* (1999) for details.

The GCM is configured globally with 2° horizontal resolution over ± 80° in latitude with 22 levels in the vertical (see Table 1). The model is being run with free-slip bottom boundary conditions and non-slip boundary conditions at lateral walls. We use Laplacian background viscosity ν and diffusivity κ horizontally and in the vertical with values of $\nu_h = 10^3$ and $\kappa_h = 10^3$ and $\nu_v = 10^{-3}$ and $\kappa_v = 10^{-5}$, respectively. Near the surface the vertical coefficients are specified by the KPP mixed layer model and can therefore be higher by an order of magnitude or more above the surface planetary boundary layer. To allow a time step of one hour, an implicit scheme is being implemented for the vertical mixing. Initial conditions were obtained from the *Levitus et al.* (1994) January potential temperature and salinity fields, with the velocity field then adjusted over a 1 month period. For surface heat and freshwater fluxes, daily fields are used from the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) re-analysis project. Twice-daily fields are used for the windstress.

2.3 Testing the Forward Model

Before attempting to combine a model with observations, one must compare it to the observations to assure that they are consistent within estimated uncertainties of both. Prior to the present work, the GCM was tested in a variety of configurations summarized in *Marshall et al.* (1997b), *Ponte et al.* (1998), and *Ponte and Stammer* (1999). Here we have simulated, as a control run, the ocean circulation from 1992 through 1997. In the estimation procedure that follows later, a pure surface flux boundary condition is estimated that leads to adjusted surface flux fields consistent with the ocean data. However, in the control run, use of the NCEP flux fields without the adjustments determined by the optimization, leads to extremely poor results. For this reason, in the reference solution, the model θ, S fields in the surface layer are relaxed towards climatological monthly mean fields with a 30-day time scale in addition to NCEP fluxes.

Observations of the SSH anomalies as observed by T/P are shown in Fig. 1 along various time-longitude sections across the North Pacific Ocean. Qualitatively, the model compares reasonably well with the T/P observations (Fig. 2): Spatial anomaly scales and amplitudes are similar and several wave-like anomalies are successfully simulated, especially those associated with El Niño along the equator. However, quantitatively there exist differences, especially in middle and high latitudes, where the model fails to simulate the observed seasonal SSH cycle. In addition, a clear model drift is obvious owing to inconsistent initial temperature and salinity conditions and surface flux fields. The goal is that the state estimation corrects these deficiencies so that the resulting state lies within the ocean data uncertainties and the estimated control bounds.

Fig. 1

Fig. 2

3 The Optimization

The solution described below required about 100 iterations to become acceptable. "Acceptable" means that the total value of J was consistent with the normalized average term having magnitude unity and that to a first approximation, the distribution of individual terms approaches a χ^2 distribution. The latter requirement is only partially met, which we interpret as meaning that our initial estimates of observational and model noise are only approximations to the truth.

A plausible first estimate for the control values would be zero, but the poor behavior of the control run in these circumstances suggested that a simple non-zero value could do significantly better. We thus diagnosed a traditional relaxation term required to keep the model temperature and salinity history on track with Levitus monthly mean fields. The time-mean of the local relaxation term were used as the initial guess of the required changes to surface heat and salt fluxes, respectively. This step decreases the initial misfit substantially and thus reduces the number of required iterations (see also *Sirkes et al.*, 1996).

3.1 Data Constraints

In setting up J, mean and time-dependent components of surface elevation were separated, thus isolating errors owing to the geoid from the distinctly different ones in the time-evolving components. Surface forcing fields (wind stress, heat and fresh water fluxes) and the initial hydrography were required to stay acceptably close to their starting values. In addition, monthly mean θ and S fields were constrained by the monthly mean *Levitus et al.* (1994) climatology, which thus are used as though they were very noisy observations available every month. The explicit form of J is then,

$$\begin{aligned}
J \;=\; & \frac{1}{2}[(\overline{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}_{tp})^T \mathbf{W}_{geoid}(\overline{\boldsymbol{\eta}} - \overline{\boldsymbol{\eta}}_{tp}) \\
& +(\boldsymbol{\eta}' - \boldsymbol{\eta}'_{tp})^T \mathbf{W}_{\boldsymbol{\eta_{tp}}}(\boldsymbol{\eta}' - \boldsymbol{\eta}'_{tp}) + (\boldsymbol{\eta}' - \boldsymbol{\eta}'_{ers})^T \mathbf{W}_{\boldsymbol{\eta_{ers}}}(\boldsymbol{\eta}' - \boldsymbol{\eta}'_{ers}) \\
& +(\boldsymbol{\delta\tau}_x)^T \mathbf{W}_{\boldsymbol{\tau}_x}(\boldsymbol{\delta\tau}_x) + (\boldsymbol{\delta\tau}_y)^T \mathbf{W}_{\boldsymbol{\tau}_y}(\boldsymbol{\delta\tau}_y) \\
& +(\delta\mathbf{H}_Q)^T \mathbf{W}_{H_Q}(\delta\mathbf{H}_Q) + (\delta\mathbf{H}_F)^T \mathbf{W}_{H_F}(\delta\mathbf{H}_F) \\
& +(\delta\mathbf{T}_0)^T \mathbf{W}_T(\delta\mathbf{T}_0) + (\delta\mathbf{S}_0)^T \mathbf{W}_S(\delta\mathbf{S}_0) \\
& +\sum_i(\overline{\theta}_{1i} - \overline{\theta}_{iSST})^T \mathbf{W}_{SST}(\overline{\theta}_{1i} - \overline{\theta}_{SSTi}) \\
& +\sum_i(\overline{\theta}_i - \overline{\theta}_{Levi})^T \mathbf{W}_T(\overline{\theta}_i - \overline{\theta}_{Levi}) \\
& +\sum_i(\overline{\mathbf{S}}_i - \overline{\mathbf{S}}_{Levi})^T \mathbf{W}_S(\overline{\mathbf{S}}_i - \overline{\mathbf{S}}_{Levi})].
\end{aligned} \tag{5}$$

Here $\bar{\theta}, \bar{S}$ indicate monthly mean potential temperature and salinity fields, δT and δS changes in initial conditions; δH terms represent corrections to daily averages of surface momentum, heat and freshwater flux fields. T/P and ERS altimeter anomalies are evaluated on a daily basis with along-track data averaged over 2° grid cells, and the mean T/P SSH field minus the EGM96 geoid model (*Lemoine et al.* 1997) is imposed over the entire period. The terms containing monthly mean *Levitus et al.* (1994) climatological fields are important ingredients and are required to adjust surface forcing fields that preserve the climatological water mass structures. The number of elements in each term is: 5041; 2,601,148 (1,839,957 for T/P; 761,191 for ERS-1/2); 20,371,380; 20,371,380; 20,371,380; 20,371,380; 169,778; 169,778; 669,744; 3,056,004; 3,056,004 for a total of 91,213,017 terms. Note that the hydrography term was down-weighted by a factor of 4, assuming that not every layer is statistically independent, equivalent to assuming that only about five vertical modes are necessary to describe the ocean.

A schematic of the optimization setup is given in Fig. 3. In the present calculations, the control parameters include adjustments to the initial-condition potential temperature (θ) and salinity (S) fields, as well as the daily surface forcing fields over the full six years (see lower part of Fig. 3), i.e., we assume that the model uncertainties reside entirely in the initial conditions and surface forcing fields. Over the six year assimilation period, the control vector has 130,753,664 elements. Fig. 3

There is no guarantee that that a descent procedure has found the "global minimum" as opposed to local or other equal minima. But the first, and most important goal of the procedure is to produce an *acceptable* solution. To the extent that there might exist other equally valid fits of the model to the data, they must differ from the present solution by changes in one or more elements of the state vector that the present observations cannot distinguish. That is, unobserved elements of the general circulation could in principle, be radically different in some other solution; for example, it is conceivable that there are large-scale recirculations at depth producing no surface signature detectable by the altimeter and remaining consistent with the climatologies. Whether such flows actually exist, and whether they are of any kinematical or dynamical importance, remains to be determined. An elaborate comparison presented below shows that the solution is found in general to be much closer to the independent ocean observations than is the first guess solution.

Finally, the extent to which the present suite of observations would permit a radically different (on the resolved, large-scale) general circulation of the ocean remains obscure. It seems unlikely that a truly different solution would be possible (e.g., a large-scale monsoonal shift at great depth in the Indian Ocean) given the geographical breadth both of our climatologies and time-continuing observation systems. We have no formal proof of impossibility and none may be forthcoming. Should however, the possibility be proven theoretically, it would have very great implications for the design of ocean observing systems.

3.2 Error Covariances

The weight matrices $\mathbf{W}$ in each term of J determine the solution to the minimization problem. In principle, one should specify the inverse of the full a priori error covariance matrix for each data type. In practice however, this information is unavailable and various ad hoc estimates must be used. For example, the mean hydrography terms were weighted using uncertainties ranging in the vertical from 0.5° C near the surface to 0.05° C at depth for potential temperature, and 0.13 to 0.01 in salinity (Fig. 4a), reflecting the global uncertainties stated by *Levitus et al.* (1994); no covariances were provided by them.

Fig. 4

The only non-diagonal inverse error covariance matrix used in the present calculation is $\mathbf{W}_{geoid}$, as provided with the EGM96 geoid model (*Lemoine et al.*, 1997). But because of the large size of $\mathbf{W}_{geoid}$ when specified in geographical coordinates (it would have $14400^2 = 207{,}360{,}000$ elements), this term of the cost function is evaluated in spherical harmonic space up to degree and order 70 (producing the much smaller number $71^2 \times 71^2 = 25{,}411{,}681$ for the number of terms). In Fig. 4b the square root of the diagonal of the geoid error covariance is displayed. Amplitudes over the ocean range from about 15 cm close to the equator to less than 5 cm in high latitudes. Enhanced errors can be found along major topographic structures such as ocean trenches or ridges. An additional error exists from the terms (spherical harmonic degree greater than 70) for structures omitted from the geoid height estimate; like other such errors in the present system, these are described in the present computation as a unstructured *model* error.

If a model cannot reproduce some physics present in observations, one can regard the discrepancy as either a model error, described in the model error covariance **Q**, or as a noise in the data, represented by data error covariance **R**. (For example, internal waves present in hydrography that a GCM does not resolve can be deemed either a data noise, or a model deficiency. The structure of explicit solutions to the normal equations shows that the final state estimate is the same, no matter which view is taken. In the present case, for example T/P data are specified along-track without any smoothing applied. The observed eddy variability in the data is treated here as observational error, as the model does not resolve these features and the SSH variance was down-weighted therefore by a factor of 1/2 of its total variance (compare Fig. 8a in *Wunsch and Stammer*, 1998).

Windstress components are weighted by the rms difference between NSCAT observations and ECMWF fields (Figs. 4c, d (*D. Chelton*, personal communication, 1998). NSCAT did not produce data from a full seasonal cycle, but we believe the fields shown in Figs. 4c and 4d produce the best available error estimates at the time of writing. For surface heat and fresh water fluxes, in the absence of any available information, a fraction of the local rms variability was used as an estimate of the surface heat flux and fresh water flux fields. A weighting of $(1/3)^2$ of their reciprocal local variances over the six year period was used respectively (Figs. 4e,f).

4 The Estimated Ocean State

We now turn to the results of the optimization procedure. These differ from either the model, or the data alone, and illustrate a three-dimensional time-evolving model that is mostly consistent with all data shown in Fig. 3. The consistency is the result of modifications in the initial conditions and the daily NCEP surface forcing fields, and these adjusted fields must be regarded as part of the oceanic state. In a fully rigorous statistical context, one would refer to the changes we estimated as "corrections". But because of the lack of a formal internal model error covariance and other remaining inconsistencies, we refer here to them only as being "adjustments". A thorough analysis of them in terms of errors in the ocean model and the atmospheric analysis will be described by *Stammer et al.*, (2001b). Because of remaining uncertainties in model physics, the surface forcing and associated model drifts, one anticipates a priori, a more accurate and precise estimate of the oceanic variability than of the absolute (time-average) state. The ultimate adjustment time of the GCM is far longer than the 6-year interval over which data have been provided, and formally, we here provide no information to the optimization about the extent to which for example, the real ocean is undergoing century-long drifts.

Adjustments to the initial *Levitus et al.* (1994) January mean θ and S fields are illustrated in Fig. 5 from layer 1, 7 and 13 (5m, 118m, and 848m depth, respectively). Fig. 5 Many structures in the changes can be associated with real interannual ocean variability, which renders the hydrography at the beginning of 1992 different from the climatological January values. The fact that temperatures are changed in the model at 850m and much deeper, is important: it supports the inference that surface elevation (altimetry) reflects processes occurring deep within the ocean and that the altimeter provides a window into the ocean abyss.

Temperature changes in the optimization can be summarized as a general warming of the sub-tropical gyres and in many of the boundary currents (the climatology is much too smooth there), while the tropical regime and high-latitudes show predominant cooling. Note the strong salinity increase near the surface over high latitudes, especially the Southern Ocean, while most of the remaining near-surface ocean is freshened. The opposite tendency can be found at depth.

Mean changes of net surface heat and freshwater flux fields relative to the prior NCEP fields as they result from the optimization are displayed in the upper row of Fig.6. Fig.6 Modifications of the net NCEP heat fluxes are of the order of ± 20 W m^{-2} over large parts of the interior oceans. Maximum changes occur along the boundary currents in the Northern Hemisphere where shifts of up to ± 80 W m^{-2} can be found. Most of the eastern boundary currents now show a significant heat uptake. The same is true in the Arabian Sea where similarly, the off-shore Ekman transport brings up cold water from below that is being heated by the atmosphere (compare with Fig.11). Strong warming occurs over Flemish Cap, and in the North Pacific and along most of the ACC. Note that the optimization removes some of the small-scale Gibbs effects known to be present in the initial estimate buoyancy flux fields as a result of mountain ranges such as the Andes, visible in the eastern Pacific.

Positive net evaporation minus precipitation $(E - P)$ values are directed into the atmosphere over the eastern side of all subtropical gyres. Large net precipitation can be found in the tropics, especially over the Pacific warm pool, and over most of the high latitudes. The biggest variations of $E - P$ occur near the boundaries, and can often be associated there with river discharge (e.g., in the Gulf of Bengal or near the Amazon Delta) or with ice import and ice melting (e.g., over parts of the Labrador Sea) not properly represented in the prior $E - P$ fields (see the discussion in *Stammer et al.,* 2001b).

The wind stress fields also adjust so as to better reproduce the observed oceanic fields (lower part of Fig. Fig.6). As compared to the net heat and freshwater changes, smaller scales are visible in the modifications of the stress fields. The fact that their largest modifications exist close to intense boundary current systems indicates the problem that a model with 2° horizontal resolution has in producing the proper current separation without extra vorticity input by the modified wind stress. On the other hand, the increase of the trade winds over the tropical Pacific is consistent with the prior knowledge of NCEP errors there (*Milliff et al.,* 1999), and the adjustment are a true correction in that region (see also discussion in *Stammer et al.,* 2001b).

In summary, the changes in the initial θ, S fields and amplitudes of changes in the external forcing fields are, overall, consistent with accepted uncertainties in the hydrographic climatology and meteorological analyses. By applying the estimated surface forcing fields to the model, the total model kinetic energy increases by about 15%, as compared to the reference run.

4.1 The Mean State

As already noted, there are reasons to anticipate that the mean state of the model could not be fully in accord with known physics. Nonetheless, the constrained model is closer to the SSH and hydrographic observations than is the control run (see below), and there is no evidence that any significant element has become less realistic than the hydrographic climatology. In Fig. 7 we show (upper panel) the estimated mean sea surface height field as it results from the six-year period. The associated velocity velocity fields from 27.5m and 1975m depth are shown in Fig. 8, respectively.

Fig. 7

Fig. 8

All major current systems are present, but with the present low model resolution, they are necessarily overly smooth. On smaller scales, eddy-like features (at the long wavelength extreme) become visible, e.g., in the Gulf of Bengal, as the Great Whirl in the Arabian Sea, and even in the Caribbean. In the North Atlantic, a deep western boundary current is present along the entire meridional extent of the Atlantic Ocean, with significant sources coming from the east of the Mid-Atlantic Ridge over large parts of the North Atlantic basin. Maximum strength of the deep western boundary current in Atlantic Ocean is located below the layer shown, which was chosen to provide a more complete picture in the Pacific and Indian Oceans. At that larger depth in the Pacific there is a much weaker, but quite similar deep western boundary current in the Northern Hemisphere and a quite strong, deep southward flowing western boundary current in the Southern Hemisphere. However, these flows are discontinuous being separated by a westward flow along the equator at this depth. But the equatorial resolution in the model is not adequate for providing a detailed picture there. Note also the deep western boundary current present in the Southern Hemisphere of the Indian Ocean.

To illustrate the estimated flow field and its change relative to the first guess in more detail, Fig. 9 shows results from the North Atlantic Ocean. Quite noticeably, the estimate has a much enhanced flow field associated with the Gulf Stream and its extension, including the Azores Current and the anticyclonic circulation associated with the Northwest corner. Those features are very poorly simulated in most coarse-resolution numerical models. Here the assimilation procedure is able to largely correct for the traditional shortcomings by changing the surface boundary conditions. A marked increase in the volume of slope water north of the Gulf Steam leads to a better current separation there. At depth, it appears that the deep western boundary current is mostly weakened, as is the sub-polar gyre strength. Fig. 9

In many previous model simulations the resulting θ, S structures were quite different from those observed (cf., *Klinck* 1995). However, the estimated θ, S structures here preserve the basic water mass distributions within uncertainty limits and with uncorrelated the θ and S residuals (Fig. 10). Difference between both fields are uncorrelated in θ and S and stay within error bounds. A further test of the consistency of rms differences between estimated and climatological fields with prior θ, S errors is described below. Fig. 10

The Ekman pumping velocity, $w_E = \hat{\mathbf{k}} \cdot curl(\tau)/f$, is displayed in the upper panel of

Fig. 11 as it results from the mean estimated wind field. For comparison, the mean vertical velocity in the model is shown in the lower panels at 37.5m and 1750m depth, respectively. In the tropics, there are clear indications of Ekman convergence and equatorial upwelling, and the general pattern of the near-surface vertical velocity over the bulk of the ocean shows large-scale Ekman pumping and suction regimes. At depth, however, the up- and down-welling patterns are complex, far from spatially uniform, showing the clear impact of topographic features. In particular, we find the largest vertical velocity amplitudes along all boundaries, irrespective of depth, especially in the North Atlantic and in the Antarctic Circumpolar Current (ACC) region. We will discuss elsewhere the relationship to theories of the abyssal circulation. Fig. 11

An important indicator of exchange processes with the atmosphere and of internal water mass conversion mechanisms, is the winter-time mixed layer depth which is shown in Fig. 12 as the annual maximum. Deepest-reaching convection occurs in the North Atlantic, where levels around 1500m are reached. General mixed-layer structures are qualitatively consistent with observations (e.g., Woods, 1984), but the mixing and 18° water mass formation in the vicinity of the Gulf Stream extends somewhat too far south. Note the deepest convection occurring over most of the eastern North Atlantic and over large parts of the Norwegian Sea. The area of the eastern subtropical Atlantic shows enhanced winter deepening as compared to the surrounding areas. Estimated mixed layer depth values of 200m are consistent with *Spall et al.*, (199?) in that region. Other regions of enhanced convection can be found near the Kuroshio, over most subtropical gyres in the southern hemisphere, and along the ACC, especially upstream of Drake Passage and in the Indian Ocean. Most of those regions coincide with areas of of net surface cooling (not shown, but see also *Stammer et al.*, 2001b)), and points to the net surface buoyancy forcing as a primary (but not necessarily local) cause of deep convection in the model. Enhanced winter time mixing occurs also in the Sea of Japan and in the eastern Mediterranean, where the Levantine Intermediate Water is formed. Fig. 12

4.2 Meridional Mass Transport Stream function

The time-mean meridional mass transport stream function,

$$\Psi(y,z) = \int_{-H}^{\eta} \int_{0(y)}^{L(y)} v(x,y,z) dz dx \qquad (6)$$

is plotted in Fig. 13 for the Atlantic, Pacific and Indian Oceans. Fields were evaluated the last 5 years of the assimilation period thus omitting the first year to avoid initialization transients in Ψ arising from the adjustment of the flow field to estimated θ, S initial conditions. Fig. 13

In the North Atlantic (upper panel) about 15 Sv of NADW are produced, of which about 4 Sv are being upwelled around 40° N at the western boundary (compare Fig. 11). The remaining 11 Sv leave the Northern Hemisphere, but entrain more water on their way south leading to 16-18 Sv being injected into the ACC area. About 2 Sv of AABW flows northward in the Atlantic originating south of the ACC.

A substantially larger inflow exists for the Pacific and Indian Oceans (middle and bottom panel of Fig. 13) both of which show a large upwelling cell below 2000m and 1000m depth, respectively with most of the upwelling occurring south of the equator. In the North Pacific, a similar but reversed cell is present. The temporal variability of Ψ is very large, and is of the same order or greater than its mean value. Maximum variability can be found in low latitudes and around the ACC.

5 Seasonal to Interannual Variability

Turning to the inferred temporal variability, we begin with the seasonal and lower frequencies. Longitude-time plots along the same section as shown in Figs. 1 and 2 are given in in Fig. 14, but from the constrained model. Results are visually similar to the T/P data. Fig.14 But the general variability level is lower than is observed by T/P, as one expects with a non-eddy resolving model. Here variability is primarily associated with the seasonal cycle in heat storage and circulation changes (compare with Plate 3 in *Wunsch and Stammer*, 1995). Differences are associated with an apparent eastward moving pattern along 30° N, which is only weakly represented in the altimeter data.

The bulk of the energy on the annual cycle is associated with the seasonal cycle of heating. The amplitude and phase of the annual harmonic of SSH and the vertically integrated heat content in the model are shown in Fig.15. Both amplitude and phase are similar to the observations, especially in mid-latitudes and in the Indian Ocean. Differences between the harmonic amplitudes and the rms field are largest in the eastern tropical Pacific, owing to the interannual and ENSO-related variability there, and in the ACC due to fast barotropic signals. Note that in mid-latitudes, phase fields show mostly the seasonal heating and cooling cycle. Low latitudes are very different however, where the changes in the wind field introduce a rapid dynamical response of SSH and temperature at this period.

Fig.15

There is a noteworthy difference in the phases of the SSH and heat content annual harmonics in the Southern Ocean upstream of Drake Passage that appears also but less pronounced in the annual harmonic amplitudes. This region has been identified before as one of vigorous barotropic variability in the T/P data (e.g., *Fukumori et al.*, 1998; *Stammer et al.*, 2000). The differences in the amplitudes and phases of the annual harmonics imply that much of the SSH variability on annual period is associated there with mass-redistribution rather than steric (heat content) changes. *Ponte et al.* (2001) provide a discussion of the annual harmonic or the bottom pressure and barotropic circulation of the global ocean.

Of considerable interest is also the annual harmonic of the sub-surface flow or temperature and salinity fields; especially so in low latitudes and the Indian Ocean. We show in Fig. 16 the anomalies of the velocity and temperature fields as they emerges on the annual period at 610m depth. Shown are the September anomaly only; March shows similar anomalies but with opposite signs. The low latitude dominate the changes in the flow field, but variations of comparable amplitude are also present in mid-latitudes, especially in the Kuroshio/Oyashio regime where previous studies indicated a measurable contribution of changes in barotropic transports to annual SSH anomalies (e.g., *Chelton and Mestas-Nuñez*, 1996; *Stammer*, 1997). Note also the relatively strong response of the flow field all along the south-eastern coast of Australia and connecting from there well into the tropical Indian Ocean.

Fig. 16

Temperature anomalies on annual harmonic coincide mostly with those in the flow field, especially in the eastern tropical Pacific and Indian Oceans where they show with

Rossby-wave structures with amplitudes as large as 0.1°C or more at 600m depth. Those structures appear mostly consistent with advective processes of the anomalous flow field. But even in sub-tropics and higher latitudes we encounter similar amplitude anomalies. Some of those will arise from deep convection events, e.g., near the Gulf Stream extension or over the Agulhas region. However, some of them can also be associated with changes in frontal structures, e.g. near the Kuroshio Extension.

6 Testing Results

Any quantitative estimation procedure requires a detailed posterior test of the results and their consistency with prior error assumptions. This step does include the computation of formal estimation uncertainties. But because the calculation of those error bars is currently beyond of what is computationally possible, we undertake here such a test by systematically comparing the results given above with some of the data that were used as constraints and with additional information that was withheld from the assimilation calculation. The purpose of these current comparisons is to provide a quantitative understanding of the extent to which the estimated state is realistic, and where remaining model/data inconsistencies may lie. In future calculations, the withheld data, if statistically consistent with the present assimilated state, will be added to the calculation, thus further constraining the system.

6.1 Mean Surface Height Residuals

The estimated mean residual $\overline{\eta}_e - \overline{\eta}_{TP}$, which is the difference between EGM96 and the geoid implied by the time average ocean circulation and T/P, displayed in the lower panel of Fig. 7, shows amplitudes of the order of ±10 cm over large parts of the ocean. Residuals up to ± 50 cm, exist however, in some locations, notably along most island arcs (e.g., the Aleutian Trench, in the Caribbean and the Indonesian Archipelago, Hawaii and the Emperor Sea Mount Chain, near Bermuda, etc.) and in the vicinity of topographic features along the Antarctic Circumpolar Current (ACC), e.g., south of New Zealand. Although residuals are within estimated EGM96 geoid error bars, these extreme values exceed prior error statistics, especially in high latitudes, where the geoid error is supposed be about 5 cm or less (see Fig. 4)

Some structures in the residuals mimic the flow field itself, e.g., the Gulf Stream, the Kuroshio, the sub-tropical gyre in the South Atlantic, etc., and could therefore point towards problems in the simulated mean circulation. However, the close association of most large residuals with topographic features and their resemblance to independent estimates of the EGM96 error (*N. Pavlis et al.*, 1999) suggests that we are seeing, primarily, geoid errors in excess of those formally estimated for EGM96. To test this hypothesis, we have used the model results to construct a new geoid by subtracting the mean estimated SSH field (of an earlier but nearby, solution) from the mean T/P absolute SSH observations. *N. Pavlis et al.* (personal communication, 2000) used the resulting estimate to calculate the T/P and other satellite orbits. The resulting orbit ephemerides are at least as good as those based on more conventional geoids, and for some satellite missions, they are actually superior to those obtained from EGM96. Our tentative conclusion here is that the more extreme residuals we see between the a priori estimated sea level (η_{TP}) and that resulting from the assimilation are consistent with geoid errors and the hypothesis of a dominating model error is not required to explain the SSH residuals.

6.2 Mean Temperature and Salinity Fields

The model was constrained to stay close to the *Levitus et al.* (1994) monthly climatological hydrographic θ and S fields within error bounds (see Fig. 4a). Here we will first test the resulting mean θ, S fields against the "Levitus climatology" and its a priori error. Below, we will further test the assimilation against the withheld WOCE hydrography.

Differences between the mean-over-six-years estimated θ,S fields and those from the climatology are shown in Figs. 17 from meridional sections along 180° E and 330° E longitude. Above 100m, the assimilation produces temperatures generally lower than in the climatology. Below the region of reduced temperatures, the values are slightly increased in the assimilation, although not uniformly so. Taken together, these results hint at a weakening of the thermocline in the model by downward diffusion of heat. Further down in the water column, differences relative to the climatology are substantially smaller, but still significant over the entire Southern Ocean and along lateral boundaries. Fig.17

For salinity, the residuals show more structure visually correlated with the flow and gyre structures. Note particularly the signature of the tropical Pacific and Atlantic, the sub-polar North Pacific and along the ACC. The northern North Pacific is colder and saltier near the surface, but fresher and warmer at depth relative to the climatology. The North Atlantic is fresher and colder almost throughout (with smaller scale deviations), while near the surface, the ACC is fresher and saltier north and south of the ACC axis, respectively, with a reversal of this pattern at depth.

To compare the estimation results against independent information not employed in the estimation process, we use here a large part of the global WOCE and pre-WOCE one-time hydrography depicted in the top panel of Fig. 18. Many of those sections will be displayed in the WOCE Hydrography Atlases (*Talley*, 2001a,b, *Koltermann et al.*, 2001). Because several of the sections shown in the figure are from outside our estimation period, we compare them with the time-mean estimated θ,S fields.

Fig. 18

Differences between the WOCE data on the one hand, and the model results and climatology on the other hand, are summarized in the lower panels of Fig.18 in terms of rms differences computed along all sections. Red lines represent the rms differences obtained from the unconstrained model, while the blue and green lines show results obtained relative to the constrained model and the climatology, respectively. For θ and S, a clear improvement can be found in the top 500-1000 m. Here the the rms differences between WOCE fields and climatology are as large as those obtained relative to the constrained run, and both are significantly larger than the prescribed error information (shown in magenta). Below about 1000 m depth, all three curves remain close, and are consistent with the prior error field. Note that for salinity the errors of the constrained model are slightly larger than those from Levitus and the unconstrained model implying that the estimated initial salinity fields at depth are moving away from the WOCE sections. They are still consistent within error bounds, however. Note also that near the surface the climatological temperature is actually closer to the WOCE data than both model solutions. Nevertheless, the near-surface temperature differences are larger than prior error estimates by a factor of 3. However, part of that error is due to the deviation of the WOCE sections from annual mean conditions. Moreover, a significant eddy variability present in the WOCE measurements has a substantial impact over most the upper ocean.

In summary, we obtain a clear improvement of the estimated model state as compared to the full WOCE hydrography measurements over the top 1000m depth range. While differences below that depth are consistent with prior error estimates, uncertainties in the climatologies over the top 1000m may have been underestimated by a factor of three in the estimation procedure.

6.3 Comparison with Time-Varying XBT and TAO Data

Although monthly mean SST fields are assimilated, most of the time-varying signal in the solution is being provided by the T/P and ERS altimeter data. To test the impact of these latter fields on the model state, we shift here to the time-dependent portions of the estimated temperature field, and which can be compared directly to the XBT and temperature timeseries obtained by the TAO array (McPhaden et al., 1998; see Behringer, 1994, for an earlier comparison between altimetry and XBT-derived estimates). The XBT/TAO sampling during 1992- 1997 is shown in the upper panel of Fig.19. Fig.19

To obtain a measure of improvement in the estimates compared to the unconstrained model, Fig. 19 shows in its lower panel the rms difference as a function of depth, evaluated over the entire model domain and the full 6 year period. Root-mean-square values of the XBT data minus the estimate (blue line) and minus the unconstrained forward model results (green line) are displayed. For this purpose, model fields were interpolated from monthly mean fields to XBT positions both in space and time. The difference between XBT and estimated fields has decreased compared to the unconstrained model, over the entire depth range by a factor of about 20%, and nearer the surface, by about 30%.

As expected, the rms differences between the estimate and the XBT/TAO data show a substantial spatial structure (Fig.20). The fields were computed as rms differences within 10° regions horizontally and within various 100 m thick depth levels in the vertical. Near the surface, differences indicate deficiencies in the seasonal mixed layer of the present model. Further down in the water column however, enhanced rms differences can be found along the path of major current systems. While the error is quite small in the quiet eastern regions (less than 0.5° C), the larger errors over the western basins are likely indications of mesoscale motions present in the XBT observations. However, a lateral shift or weakening of frontal structures in the model would appear as a similar structure. Fig.20

To illustrate the degree by which the error is being reduced by the data assimilation, we show in the right column of Fig.20 the difference of the fields provided in the left column with similar ones based on the unconstrained model results. The reduced misfit upon data assimilation is negative in a few locations, indicating that the results from the constrained model are degraded there. However, the misfit reduction is mostly positive and in many places as big as 1.5°C near the surface, and even as big as 0.5°C at 600m depths.

7 Summary and Conclusions

The most important result from this experiment is the demonstration that it is now feasible to obtain estimates of the time-dependent ocean circulation by combining ocean observations with a numerical model. Through this procedure, it is possible to obtain three-dimensional oceanic state fields every few days and the associated surface forcing fields that are consistent with the global in situ and altimetric observations, SST data, atmospheric estimates of air/sea fluxes, and the equations of motion governing the ocean circulation.

Results presented here can be summarized as estimates of a remarkably rapidly time-varying flow field that, although still too smooth due to a lack of spatial resolution, simulates many realistic features of the large-scale circulation, and includes some which are traditionally missing in numerical coarse-resolution simulations. These latter include vigorous barotropic motion, a more realistic Gulf Stream separation, the correction of the path of the North Atlantic Current, and the presence of an Azores Front in the North Atlantic, among many other features of the global ocean. *Käse et al.* (2001) discuss an application of the surface forcing discussed here in a regional high-resolution model of the North Atlantic.

As anticipated, the constrained model has moved considerably closer to the observations, both those imposed as constraints and those withheld. Also as anticipated, the failure of the model to fully reproduce the observations of both types leads immediately to sometimes difficult issues of determining whether the model is in error, or whether the data errors were accurately specified. It is intrinsic to the nature of any statistical estimation procedure (which is what data assimilation consists of) that the process of attribution of misfits between model errors and data errors can remain ambiguous and unresolved until further information is obtained. In several cases, the comparisons remain indeterminate over large areas of the world ocean because there are too few observations to reduce the data error to levels where it would truly test the model.

As the estimates improve through the evolution of numerical models, increasing computer power, and better assimilation schemes, more realistic estimates of the time-evolving flow field will become available that will enable the computation of a host of oceanic processes. But even with the preliminary results at hand, we can now start to estimate the global heat and freshwater fluxes and divergences as a function of time, calculate the day-by-day variability of the global current system, and diagnose upwelling, potential vorticity fluxes, etc. all in ways that exploit a great variety of data types as well as the diverse physical processes embodied in the GCM code. There are many other uses of the present results (e.g., in biological studies), and among the important ones are the possibility of studying the impact and efficiency of various possible oceanic observation systems.

Several immediate improvements in the system can be made. Beyond the incorporation of more data, two steps are particularly important: (1) Improved representation of eddy transfers (eddy parameterization), and (2) Improved model resolution to 1° or better. In the long run, we expect to introduce more formal model error estimates, as well as examining the system sensitivity to the observations by analysis of the adjoint solution (*Marotzke et al*, 1999); ultimately, we expect to generate formal error bars for the estimated state.

In summary, we have completed a prototype global ocean state estimate from which we see no fundamental obstacles to moving quickly toward quasi-operational estimates and products as envisioned in GODAE and other programs.

All model fields described here are available through the internet. For details see the ECCO Project web page at http://www.ecco-group.org where animations of the model results are also provided.

Acknowledgments. We thank D. Behringer and Ming Ji at NCEP for providing us with their edited XBT data set. Diana Spiegel and Charmaine King helped with processing of the model fields and WOCE data sets. The help of Lynne Talley and the WOCE SAC (Hamburg) in obtaining pre-processed WOCE hydrographic data is gratefully acknowledged. Re-analysis surface forcing fields from the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) are obtained through a computational grant at NCAR. The computational support through a NRAC grant from the National Partnership for Computational Infrastructure (NPACI) is acknowledged. Supported in part through ONR (NOPP) ECCO grants N00014-99-1-1049 and N00014-99-1-1050, through NASA grant NAG5-7857, through NSF grant OCE 9730071 and through two contracts with the Jet Propulsion Laboratory (958125 and 1205624).

References

[1] Arakawa, A. and V.R. Lamb, 1977: Methods of Computational Physics. 17, Academic Press, 174-265.

[2] Bennett, A. F., 1992. Inverse methods in physical oceanography. Cambridge Monographs on Mechanics and Applied Mathematics, Cambridge University Press.

[3] Behringer, D.W., 1994: Sea surface height variations in the Atlantic Ocean: a comparison of TOPEX altimeter data with results from an ocean data assimilation system. *J. of Geophys. Res.*, 99(C12), 24685-24690.

[4] Chelton, D.B. and A.M. Mestas-Nuñez, 1996: The large-scale, wind-driven response of the North Pacific. Int. WOCE Newsletter, 26, 3-6.

[5] Fukumori, I., R. Raghunath, and L.-L. Fu, 1998: Nature of global large-scale sea level variability in relation to atmospheric forcing: a modeling study, *J. Geophys. Res., 103*, 5493-5512.

[6] Giering, R., and T. Kaminski, 1998: Recipes for adjoint code construction, *Association for Computing Machinery Transactions on Mathematical Software*, **24**, 437–474.

[7] Gilbert, J.C., and C. Lemarechal, 1989: Some numerical experiments with variable-storage quasi-Newton algorithms, *Mathematical Programming* 45, p. 407-435.

[8] Käse, R.H., A. Biastoch and D. Stammer, 2001, On the mid-depth circulation in the Labrador and Irminger Seas, in press.

[9] Klinck, J.M., 1995: Thermohaline structures of an eddy-resolving North-Atlantic model: the influence of boundary conditions. *J. Phys. Oceanogr.*, 25, 1174–1195.

[10] Koltermann, P., et al. 2001: WOCE Atlas of the Atlantic Ocean. In preparation.

[11] Large, W. G., J.C. McWilliams, and S.C. Doney, 1994: Oceanic vertical mixing: a review and a model with nonlocal boundary layer parameterization, *Rev. Geophys.*, 32, 363–403.

[12] Lemoine, F. and 17 others, 1997: The development of the NASA GSFC and NIMA Joint Geopotential Model. in, *Proceedings of the International Symposium on Gravity*, Geoid and Marine Geodesy, IAG Symposium: in press H. Fujimoto, ed. Springer-Verlag.

[13] Levitus, S., R. Burgett, and T. Boyer, 1994: *World Ocean Atlas 1994*, vol. 3, *Salinity*, and vol. 4, *Temperature*, *NOAA Atlas NESDIS 3 & 4*, U.S. Dep. of Comm., Washington, D.C..

[14] Malanotte-Rizzoli, P., 1996. *Modern Approaches to Data Assimilation in Ocean Modeling*, Elsevier, Amsterdam, The Netherlands, 455pp.

[15] Marotzke, J., R. Giering, Q. K. Zhang, D. Stammer, C. N. Hill, and T. Lee, 1999: Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity, *J. Geophys. Research*, 104, 29,529 - 29,548.

[16] McPhaden MJ, Busalacchi AJ, Cheney R, Donguy JR, Gage KS, Halpern D, Ji M, Julian P, Meyers G, Mitchum GT, Niiler PP, Picaut J, Reynolds RW, Smith N, Takeuchi K, 1998: The tropical ocean global atmosphere observing system: A decade of progress. *J. of Geophys. Res.*, 103, 14169-14240.

[17] Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, 1997a, A finite-volume, incompressible navier-stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.*, 5753–5766.

[18] Marshall, J., C. Hill, L. Perelman, and A. Adcroft, 1997b, Hydrostatic, quasi-hydrostatic and non-hydrostatic ocean modeling. *J. Geophys. Res.*, 5733–5752.

[19] Milliff, R. F., W. G. Large, J. Morzel, G. Danabasoglu, and T. M. Chin, cean general circulation model sen sitivity to forcing from scatterometer winds, *J. Geophys. Res.,* 104, 11337-11358.

[20] Pavlis, N.L., C.M. Cox and F.G. Lemoine, 1999: Comparison of Dynamic Ocean Topography Solutions Combining Geodetic and Oceanographic Information, AGU Spring 99 abstract.

[21] Ponte, R. M., and D. Stammer, and J. Marshall, 1998: Oceanic signals in observed motions of the Earth's pole, *Nature*, 391, 476–479.

[22] Ponte, R. M., and D. Stammer, 1999: The role of regional mass and current changes in the ocean on seasonal polar motions fluctuations, *J. Geophys. Res.,* 104, 23,393–23,409.

[23] Ponte, R., D. Stammer and C. Wunsch, 2001: Improved ocean angular momentum estimates using an ocean model constrained by large-scale data, *Geoph. Res. Letters*, in press.

[24] Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analysis using optimum interpolation. *J. Climate*, 7, 929–948.

[25] Sirkes, Z., E. Tziperman, and W. C. Thacker, 1996: Combining data and a global primitive equation ocean general circulation model using the adjoint method. In: Modern Approaches to Data Assimilation in Ocean Modeling, P. Malanotte-Rizzoli, ed., 119-145.

[26] Spall, M. et al., 199?

[27] Stammer, D., C. Wunsch, and R. Ponte, 2000: De-Aliasing of Global High Frequency Barotropic Motions in Altimeter Observations, *Geophysical Res. Letters*, 27, 1175-1178.

[28] Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C.N. Hill, and J. Marshall, 2001: Volume, Heat and Freshwater Transports of the Global Ocean Circulation 1992 –1997, Estimated from a General Circulation Model Constrained by WOCE Data, submitted for publications.

[29] Stammer. D., K. Ueyoshi, W. Large and C. Wunsch, 2001b: Estimates of Surface Momentum, heat and freshwater fluxes, obtained from combining an ocean circulation model with global ocean data sets, to be submitted for publications.

[30] Stammer, D. 1997: Steric and wind-induced changes in TOPEX/POSEIDON large-scale sea surface topography observations, *J. Geophys. Res.*, 20,987–21,010.

[31] Talley, L., 2001a: WOCE Hydrographic Office Atlas: The Pacific Ocean, in preparation.

[32] Talley, L., 2001b: WOCE Hydrographic Office Atlas: The Indian Ocean, in preparation.

[33] Woods, J.D., 1984: The Warmwatersphere of the North Atlantic — A Miscellany. Ber. Inst. f. Meereskunde Kiel, 128, 39pp.

[34] Wunsch, C., 1996: *The Ocean Circulation Inverse Problem*, 442 pp., Cambridge Univ. Press, New York.

[35] Wunsch, C. and D. Stammer, 1995: The global frequency-wavenumber spectrum of oceanic variability estimated from TOPEX/POSEIDON altimeter measurements, *J. Geophys. Res.*, 100, 24,895–29,910.

[36] Wunsch, C., and D. Stammer, 1998: Satellite Altimetery, the Marine Geoid and the Oceanic General Circulation, *Annual Reviews of Earth and Planetary Sciences,*.26, 219-254.

Figure Captions

Fig. 1: Longitude-time sections for TOPEX/POSEIDON SSH anomalies. Individual panels show SSH anomalies from 0°N, 10°N, 20°N and 30°N, respectively.

Fig. 2: Longitude-time sections for SSH anomalies in the unconstrained model. Individual panels show SSH anomalies from 0°N, 10°N, 20°N and 30°N, respectively.

Fig. 3: Schematic of the optimization. The middle part of the figure shows the data constraints imposed on the model. The lower part summarizes the control variables which are being modified to lead to the optimal solution (although the horizontal velocity fields (u, v) are not part of the control vector, they are adjusted geostrophically at the beginning to the new density field. The top part, summarizes the withheld data sets which we use to test this current solution.

Fig. 4: (a) Uncertainty profiles prescribed for potential temperature (red) and salt (blue) as a function of level number. (b) Diagonal elements of the EGM96 error covariance matrix (in meters). (c) and (d): rms uncertainties for the zonal and meridional wind stress components. Fields were obtained as rms differences between NSCAT scatterometer wind stress measurements and simultaneous ECMWF wind stress analyses (*D. Chelton*, pers. comm., 1998). (e),(f): rms uncertainties for heat and freshwater fluxes estimated from 30% and 100% of the local rms variabilities of the NCEP forcing fields.

Fig. 5: Adjustments of the initial Levitus et al. (1994) January mean θ (left) and S (right) fields at 5m and 435m depth, respectively).

Fig. 6: Mean changes of net surface heat (in W/m^2) and freshwater flux fields (in m/yr) as they result from the optimization relative to the prior NCEP fields are displayed in the upper row. The net heat flux fields are positive into the ocean, and a positive net fresh water field means net freshwater flux into the atmosphere. The lower two panels show the mean change of surface wind stress components as they result from the optimization relative to NCEP first guess fields. (N/m^2, positive eastward and northward).

Fig. 7: The estimated mean sea surface height field (in cm) as it results from the 6 year assimilation period is shown in the upper panel. The lower panel shows the estimated mean residual $\overline{\eta}_e - \overline{\eta}_{tp}$ in cm.

Fig. 8: Mean estimated velocity field from 27.5m and 1975 m depth, respectively (in cm/s) as they result from the the 6 year assimilation period.

Fig. 9:The mean estimated velocity (left column) and its differences relative to the unconstrained models (right column) plotted from the Atlantic at 27.5m and 1975 m depth, respectively.

Fig. 10:$\theta - S$ diagram of the 6 year mean model (blue) and the climatological annual mean Levitus θ and S fields (red).

Fig. 11: Mean vertical velocity at 37.5m and 1750 m respectively. The Ekman pumping velocity, $w_E = curl(\tau)/f$, is displayed in the upper panel as it results from the mean estimated wind field. The lower two panels show the mean vertical velocity at 37.5m and 1750 m respectively. The contour interval is 0.2x10^{-5} m/s in all three panels.

Fig. 12: Annual maximum mixed layer depth (m) as it results from the KPP (*Large et al.* (1994) mixed layer model.

Fig. 13: (Top) The mean meridional mass transport stream function evaluated over the Atlantic sector. (middle panel) Mean meridional mass transport stream function evaluated over the Pacific sector. (bottom panel) Mean meridional mass transport stream function evaluated over the Indian Ocean sector.

Fig. 14: Longitude-time sections for SSH anomalies in the constrained model. Individual panels show SSH anomalies from 0°N, 10°N, 20°N and 30°N, respectively.

Fig. 15: The amplitude and phase of the annual harmonic of SSH (left, in centimeters and ° relative to January 1) and the vertically integrated heat content in the model (right, in Joules and ° relative to January 1).

Fig. 16: (top) Velocity (cm/s) and temperature (°C) anomalies on the annual frequency plotted for September at 610 m depth, respectively.

Fig. 17: (left) Differences between the mean-over-six-years estimated θ,S fields and those from the *Levitus et al.* (1994) climatology from meridional sections along 180° E (top) and 330° E longitude (bottom). (right) Differences between the mean estimated S field and that from the *Levitus et al.* (1994) climatology from meridional sections along 180° E (top) and 330° E longitude (bottom).

Fig. 18: Top:Section locations of the WOCE one-time hydrography used in this study. Bottom: Root-mean-squares differences computed along all WOCE sections and between the WOCE data and the unconstrained model run (red), the constrained model (blue), and the Levitus annual mean fields (green), respectively. The dashed blue curve shows the prior error prescribed for the Levitus fields.

Fig. 19: Top: Available XBT data sampled over the world ocean during 1992-1997; also included are the data from the TOGA-TAO buoy network. Note that the TOGA-TAO buoy data are included in the figure in the near-equatorial Pacific. Bottom: Global and time averaged rms difference as a function of depth of the XBT data minus the estimate (blue line) and minus the control run (green line).

Fig. 20: (a) Root-mean-squares differences within 10° regions horizontally and within various 100 m thick depth levels in the vertical. Shown are resulting fields for the top 100m, 200-300m and 400-500 m depth range, respectively. (b) Difference field of rms misfits shown in (a) minus the similar field but based on the first guess.

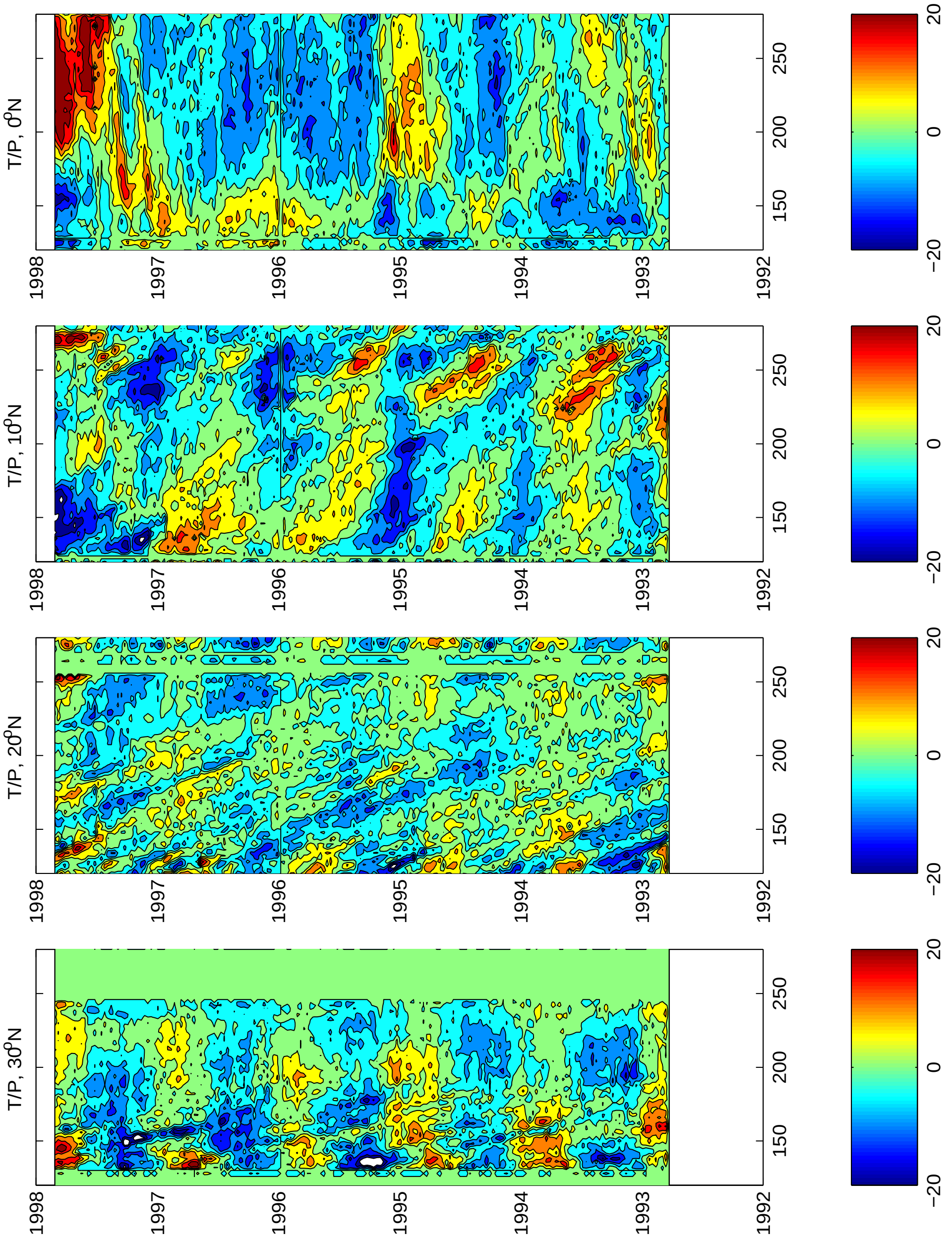

T/P, 0^0N
T/P, 10^0N
T/P, 20^0N
T/P, 30^0N
1998
1997
1996
1995
1994
1993
1992
150
200
250
20
0
−20

Figure 1:

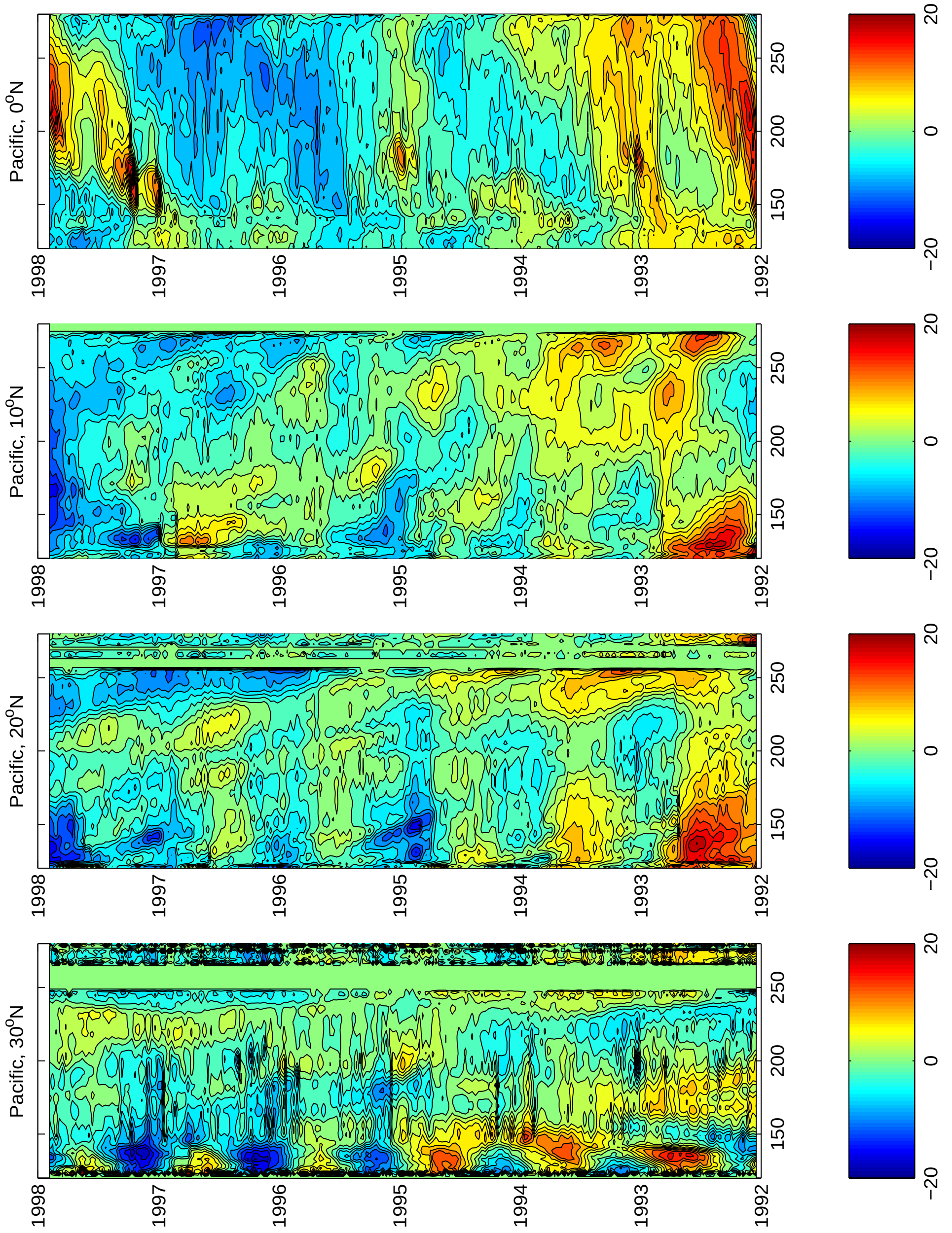

Pacific, 0^0N
Pacific, 10^0N
Pacific, 20^0N
Pacific, 30^0N
1998
1997
1996
1995
1994
1993
1992
150
200
250
20
0
−20

Figure 2:

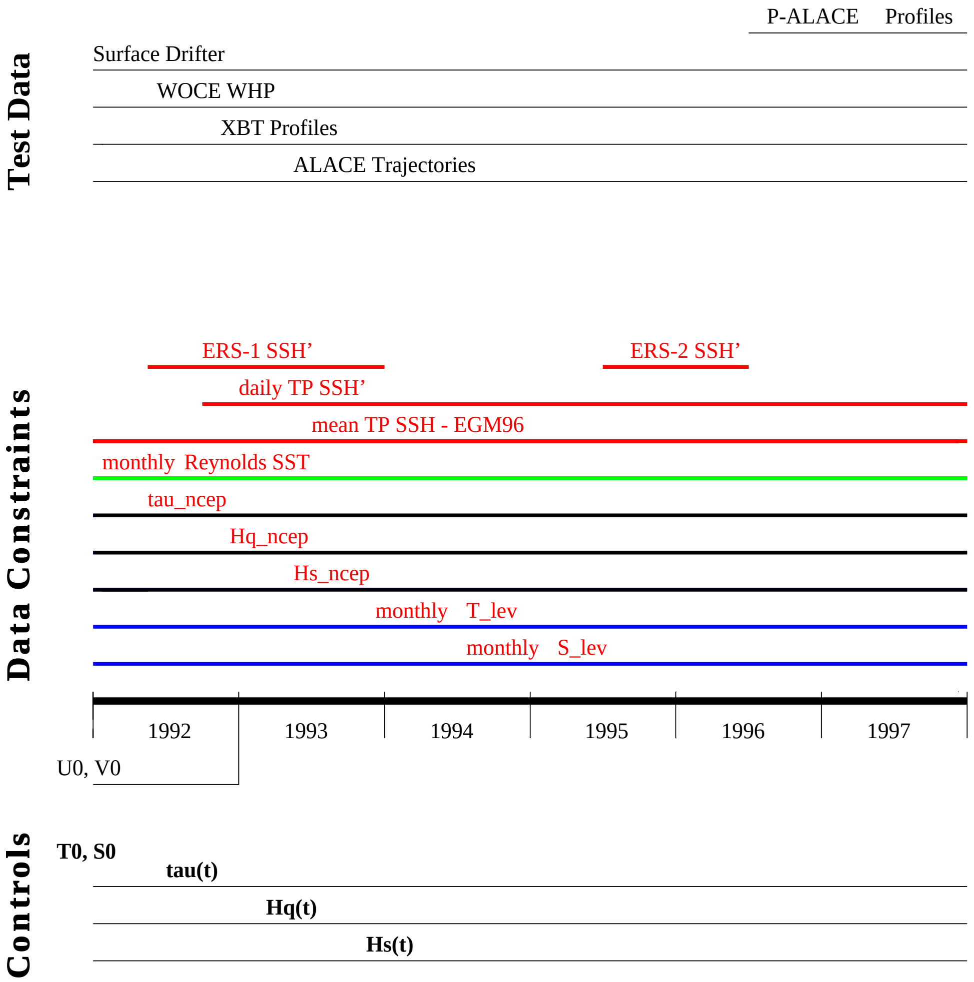

P-ALACE Profiles
Surface Drifter
WOCE WHP
XBT Profiles
ALACE Trajectories
Test Data
ERS-1 SSH'
ERS-2 SSH'
daily TP SSH'
mean TP SSH - EGM96
monthly Reynolds SST
tau_ncep
Hq_ncep
Hs_ncep
monthly T_lev
monthly S_lev
Data Constraints
1992
1993
1994
1995
1996
1997
U0, V0
T0, S0
tau(t)
Hq(t)
Hs(t)
Controls

Figure 3:

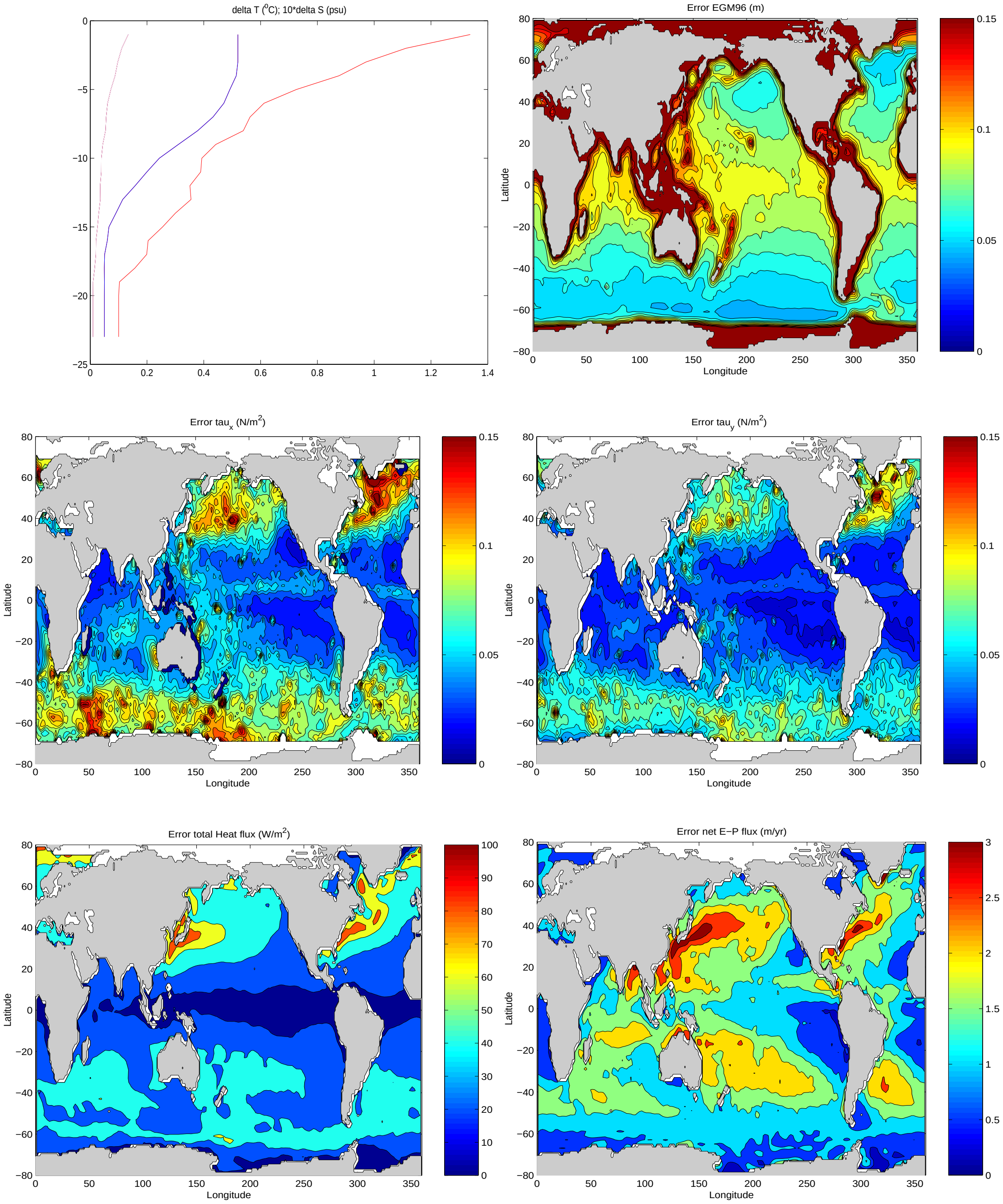
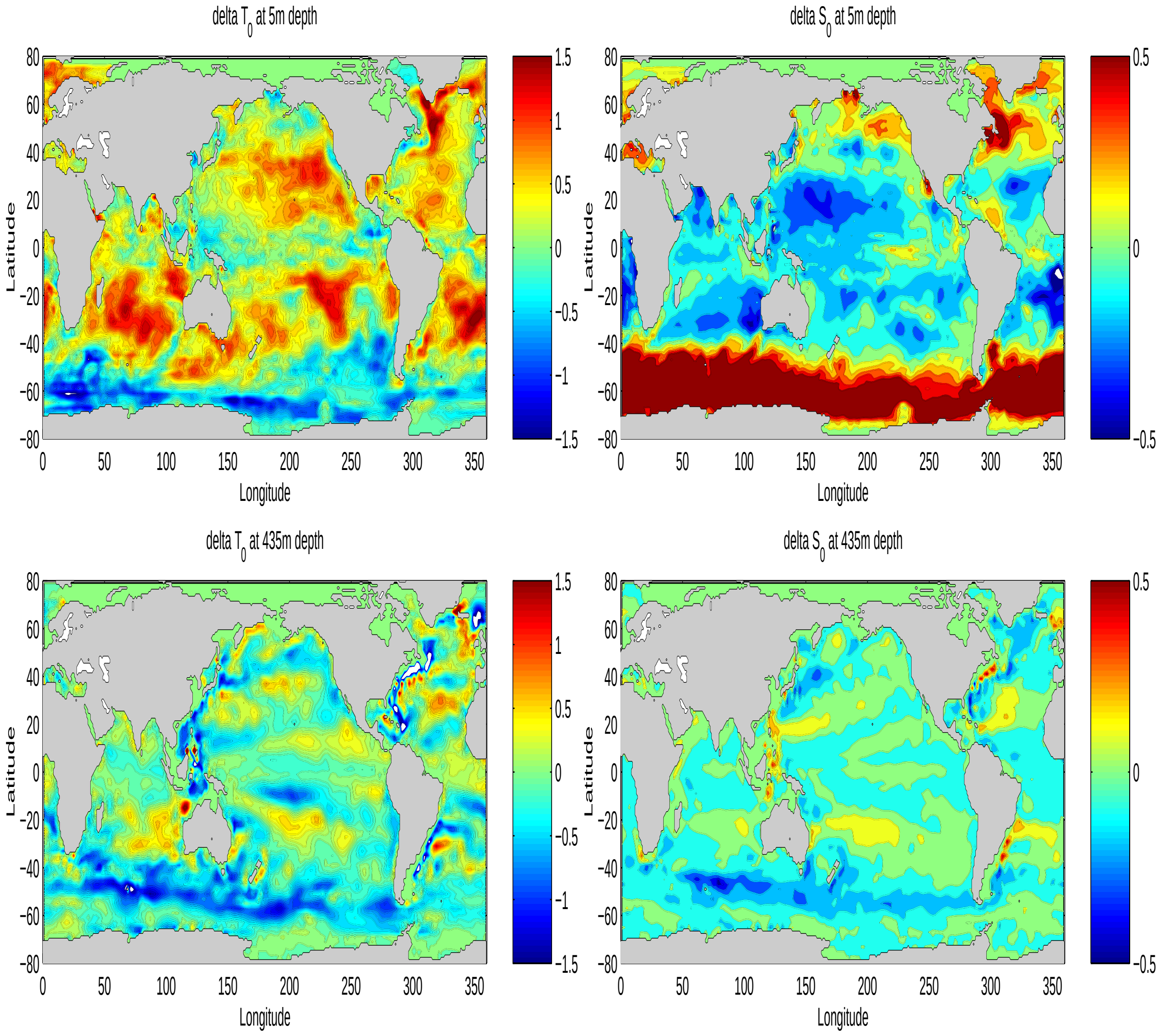

delta T (^{0}C); 10*delta S (psu)
0
-5
-10
-15
-20
-25
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Error EGM96 (m)
Error tau$_x$ (N/m^2)
Error tau$_y$ (N/m^2)
Error total Heat flux (W/m^2)
Error net E−P flux (m/yr)
Latitude
Longitude
80
60
40
20
0
−20
−40
−60
−80
0
50
100
150
200
250
300
350
0.15
0.1
0.05
0
100
90
80
70
60
50
40
30
20
10
3
2.5
2
1.5
1
0.5

Figure 4:



delta T$_0$ at 5m depth
delta S$_0$ at 5m depth
delta T$_0$ at 435m depth
delta S$_0$ at 435m depth
Latitude
Longitude
80
60
40
20
0
-20
-40
-60
-80
0
50
100
150
200
250
300
350
1.5
1
0.5
0
-0.5
-1
-1.5
0.5
0
-0.5

Figure 5:

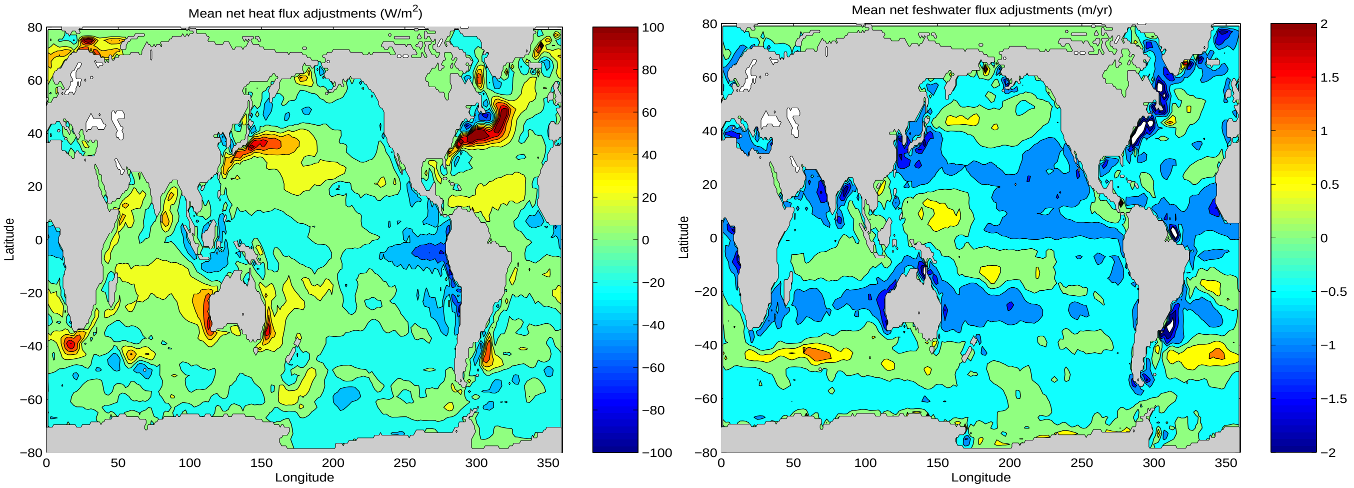

Mean net heat flux adjustments (W/m^2)
Latitude
Longitude
80
60
40
20
0
−20
−40
−60
−80
0
50
100
150
200
250
300
350
100
80
60
40
20
0
−20
−40
−60
−80
−100
Mean net feshwater flux adjustments (m/yr)
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2

Figure 6:

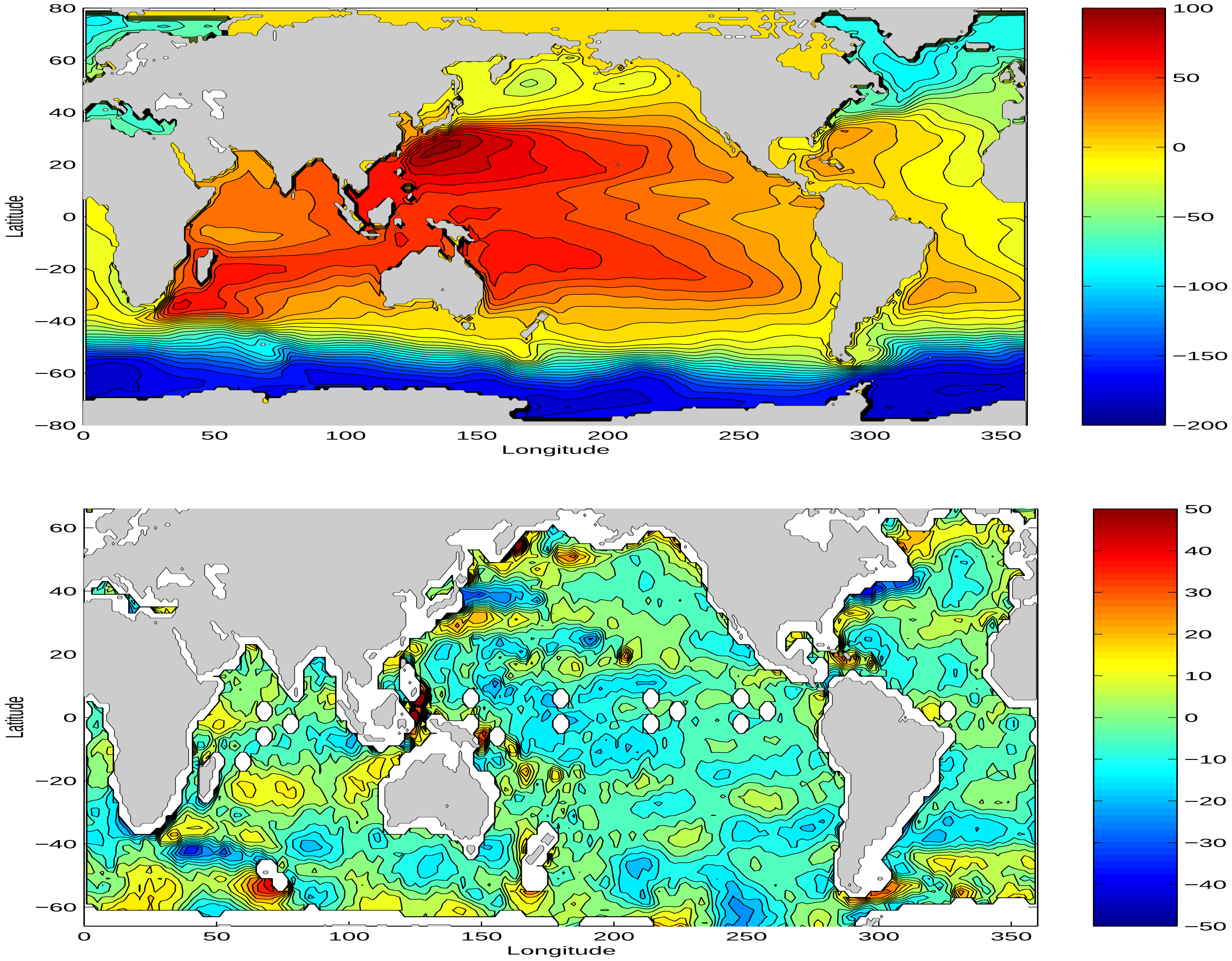

80
60
40
20
0
−20
−40
−60
−80
Latitude
0
50
100
150
200
250
300
350
Longitude
100
50
0
−50
−100
−150
−200
60
40
20
0
−20
−40
−60
Latitude
0
50
100
150
200
250
300
350
Longitude
50
40
30
20
10
0
−10
−20
−30
−40
−50

Figure 7:

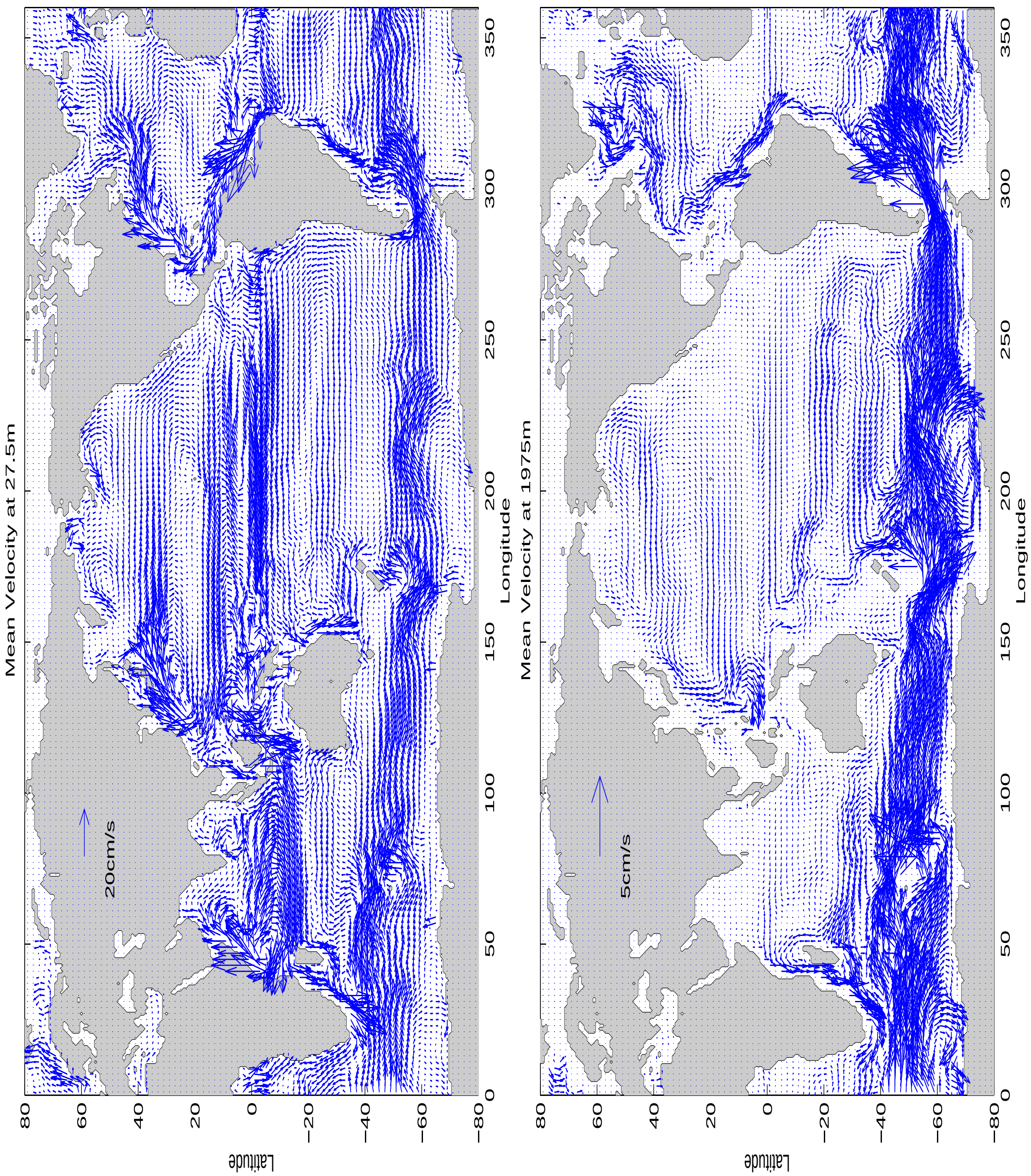

Mean Velocity at 27.5m
20cm/s
Mean Velocity at 1975m
5cm/s
Longitude
Latitude
0
50
100
150
200
250
300
350
80
60
40
20
0
−20
−40
−60
−80

Figure 8:

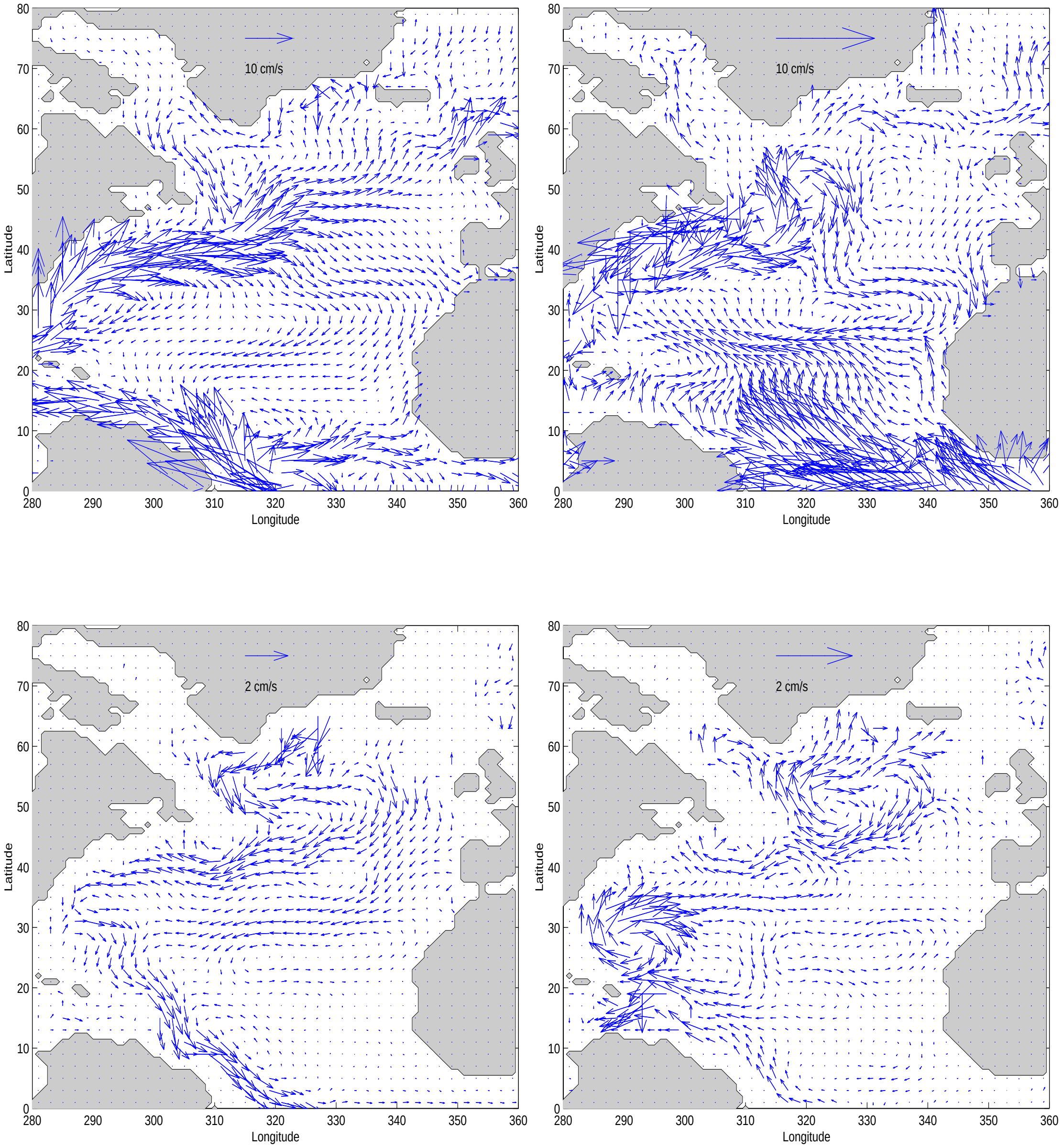

10 cm/s
10 cm/s
2 cm/s
2 cm/s
Latitude
Longitude
0
10
20
30
40
50
60
70
80
280
290
300
310
320
330
340
350
360

Figure 9:

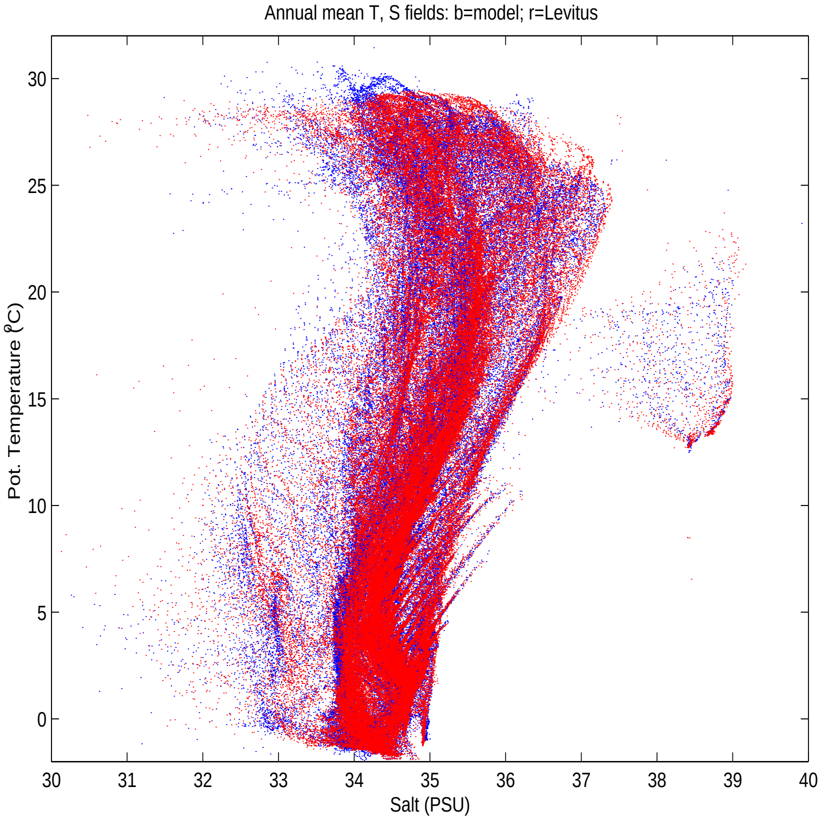

Annual mean T, S fields: b=model; r=Levitus
Pot. Temperature (ºC)
Salt (PSU)
30
25
20
15
10
5
0
30
31
32
33
34
35
36
37
38
39
40

Figure 10:

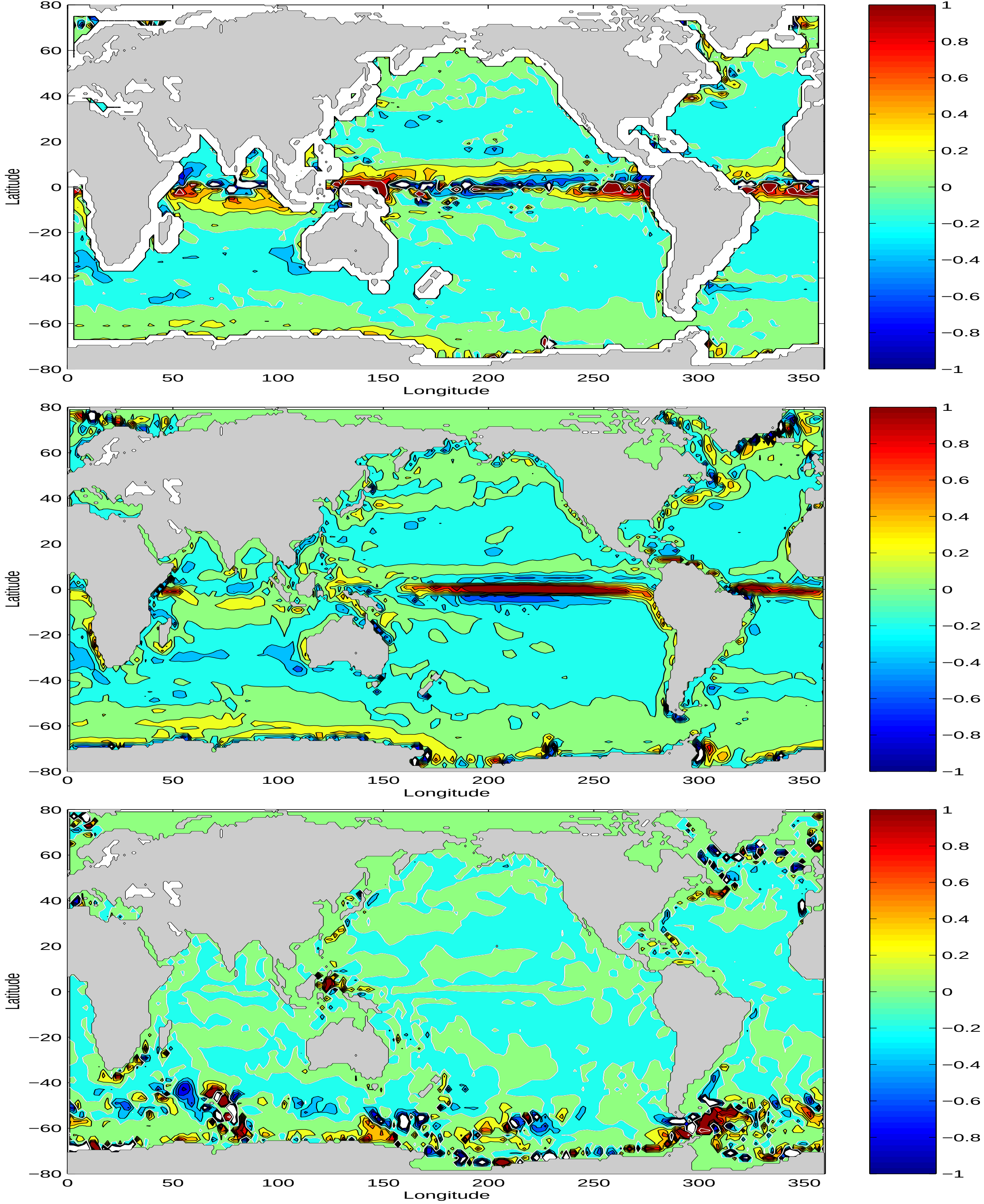

Latitude
Longitude
80
60
40
20
0
−20
−40
−60
−80
0
50
100
150
200
250
300
350
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

Figure 11:

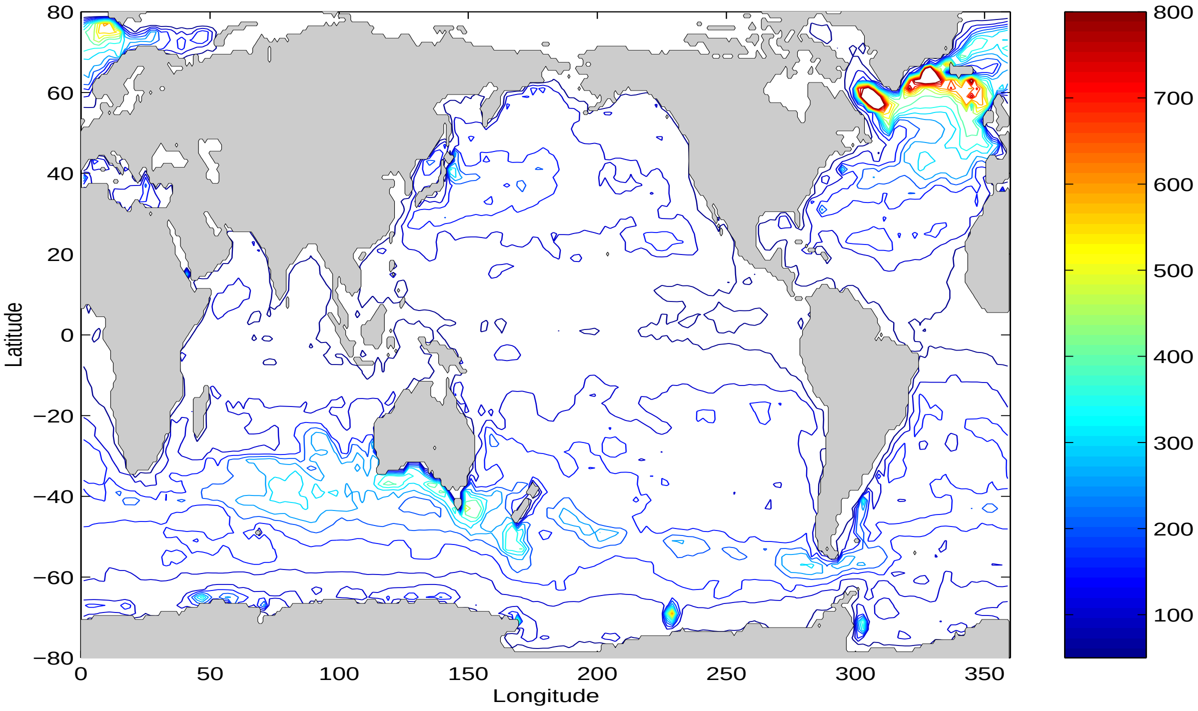

80
60
40
20
0
−20
−40
−60
−80
Latitude
0
50
100
150
200
250
300
350
Longitude
800
700
600
500
400
300
200
100

Figure 12:

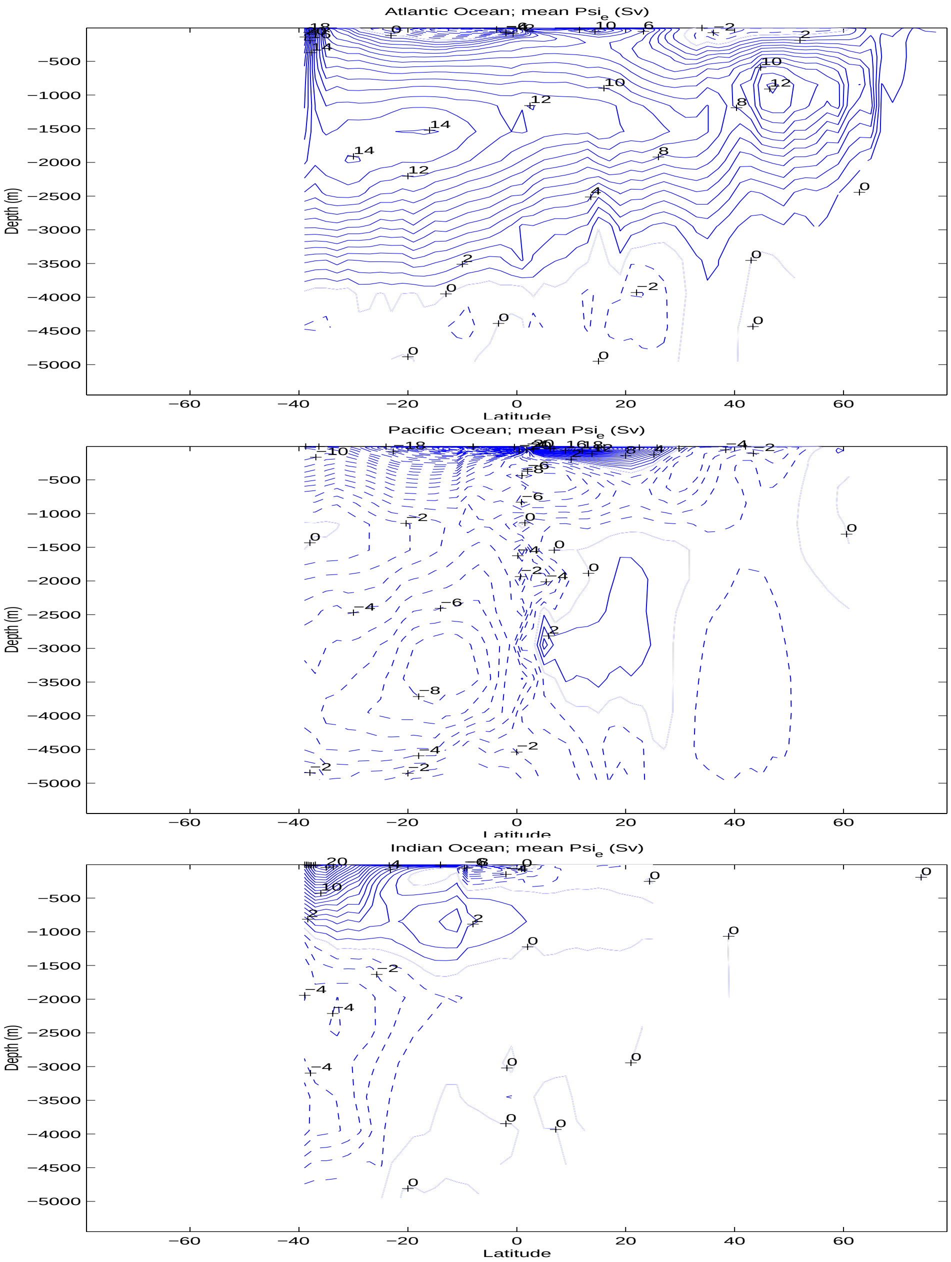

Atlantic Ocean; mean Psi$_e$ (Sv)
Depth (m)
−500
−1000
−1500
−2000
−2500
−3000
−3500
−4000
−4500
−5000
−60
−40
−20
0
20
40
60
Latitude
Pacific Ocean; mean Psi$_e$ (Sv)
Depth (m)
Latitude
Indian Ocean; mean Psi$_e$ (Sv)
Depth (m)
Latitude

Figure 13:

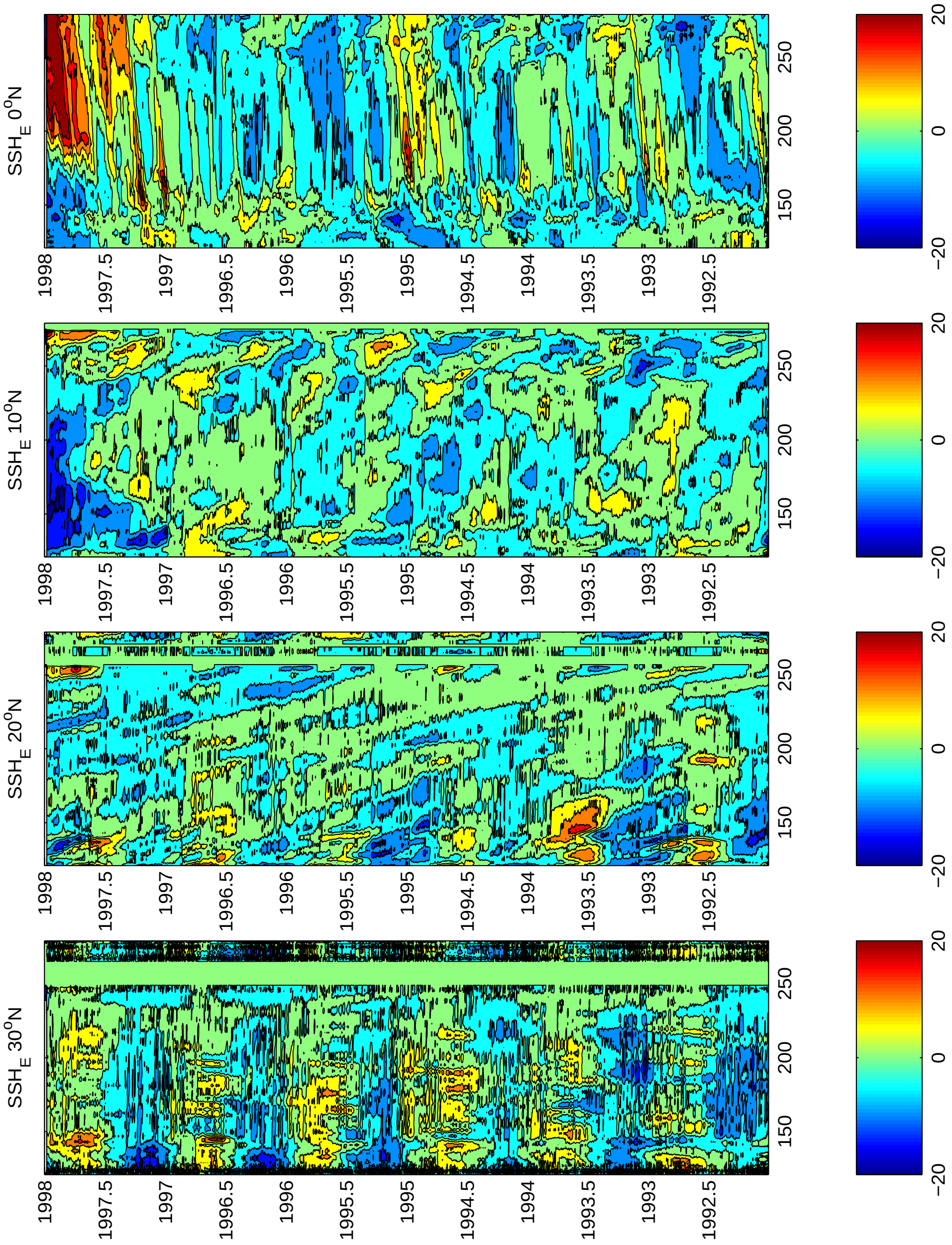

SSH$_E$ 0^{o}N
SSH$_E$ 10^{o}N
SSH$_E$ 20^{o}N
SSH$_E$ 30^{o}N
1998
1997.5
1997
1996.5
1996
1995.5
1995
1994.5
1994
1993.5
1993
1992.5
250
200
150
20
0
−20

Figure 14:

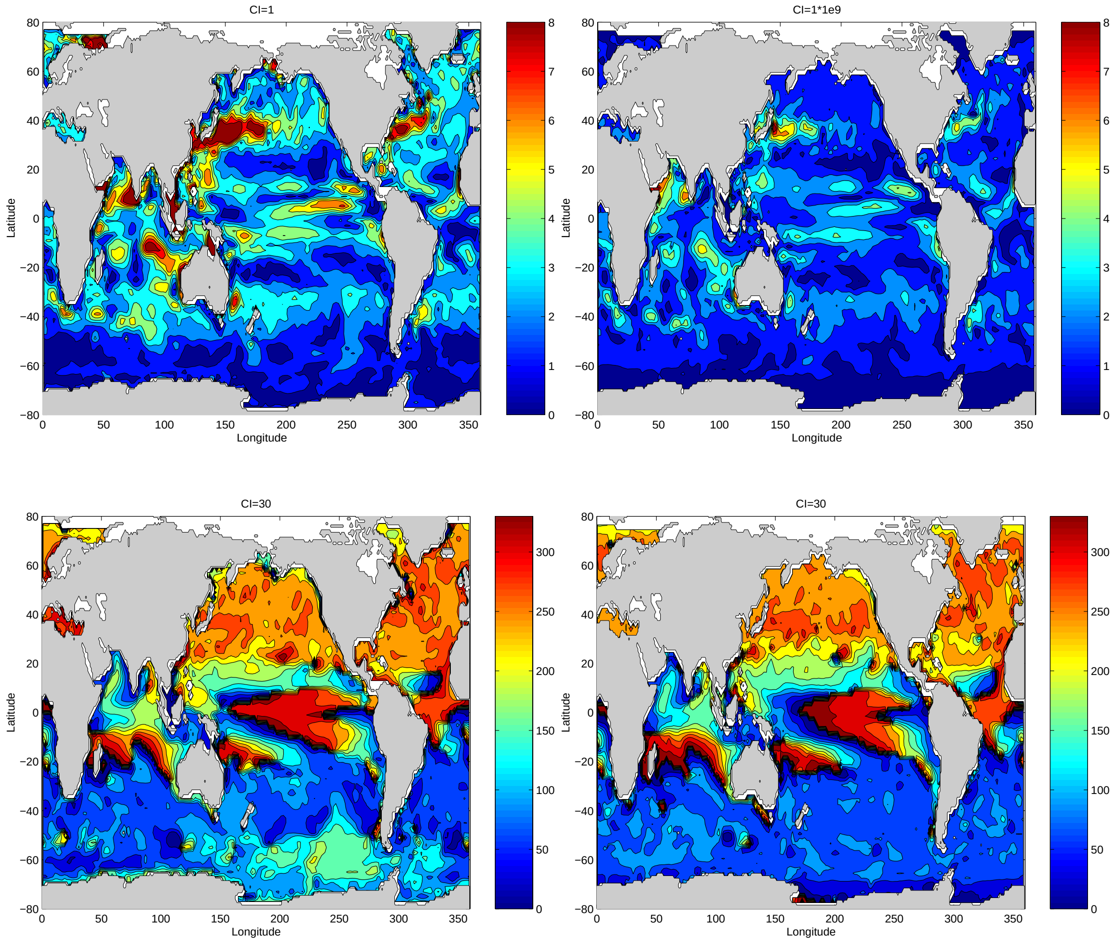

CI=1
CI=1*1e9
CI=30
CI=30
Latitude
Longitude
80
60
40
20
0
−20
−40
−60
−80
0
50
100
150
200
250
300
350
8
7
6
5
4
3
2
1
0
300
250
200
150
100
50
0

Figure 15:

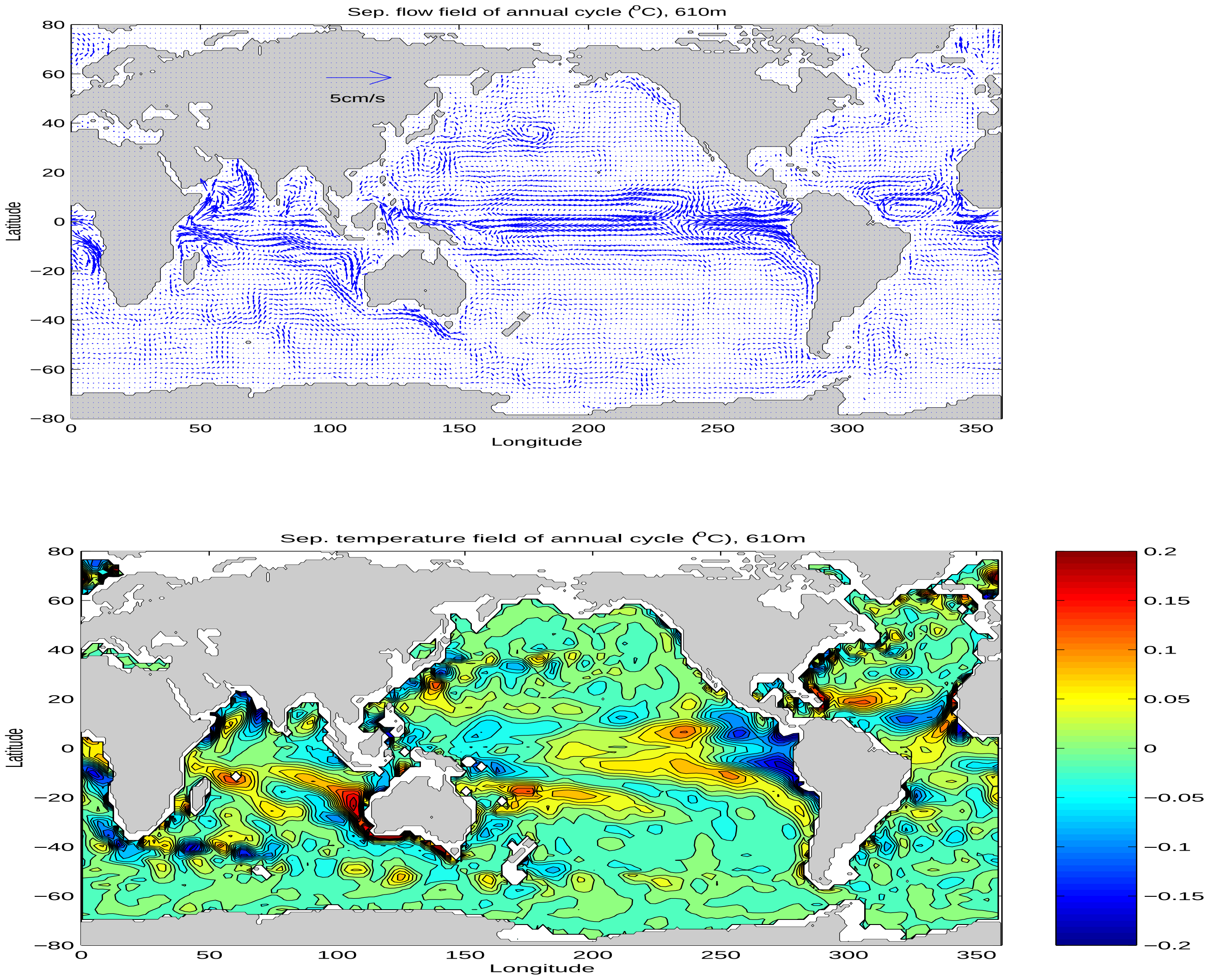

Sep. flow field of annual cycle (^oC), 610m
5cm/s
Latitude
Longitude
80
60
40
20
0
−20
−40
−60
−80
0
50
100
150
200
250
300
350
Sep. temperature field of annual cycle (^oC), 610m
0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2

Figure 16:

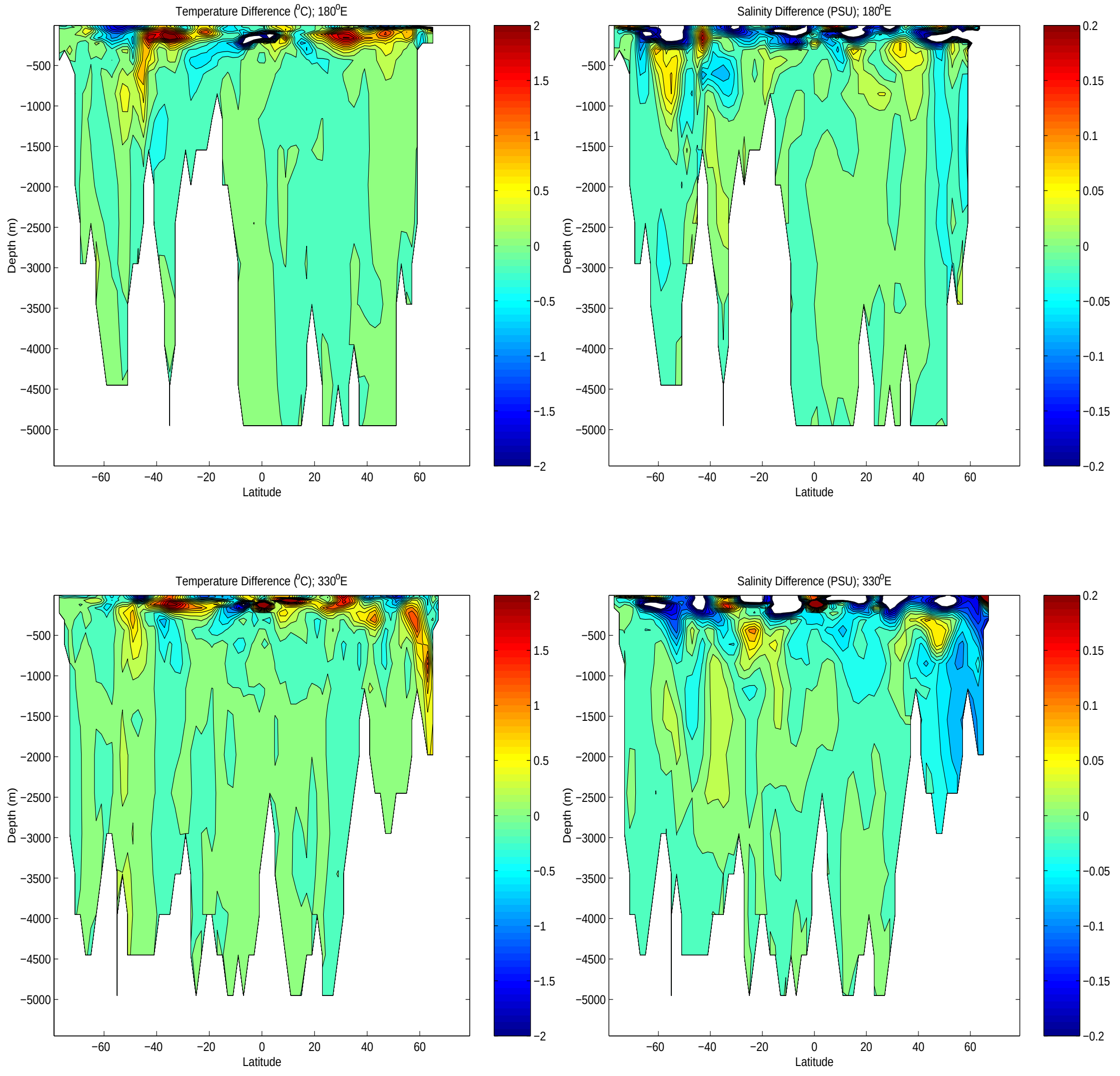

Temperature Difference (^{0}C); 180^{0}E
Salinity Difference (PSU); 180^{0}E
Temperature Difference (^{0}C); 330^{0}E
Salinity Difference (PSU); 330^{0}E
Depth (m)
Latitude
−500
−1000
−1500
−2000
−2500
−3000
−3500
−4000
−4500
−5000
−60
−40
−20
0
20
40
60
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2

Figure 17:

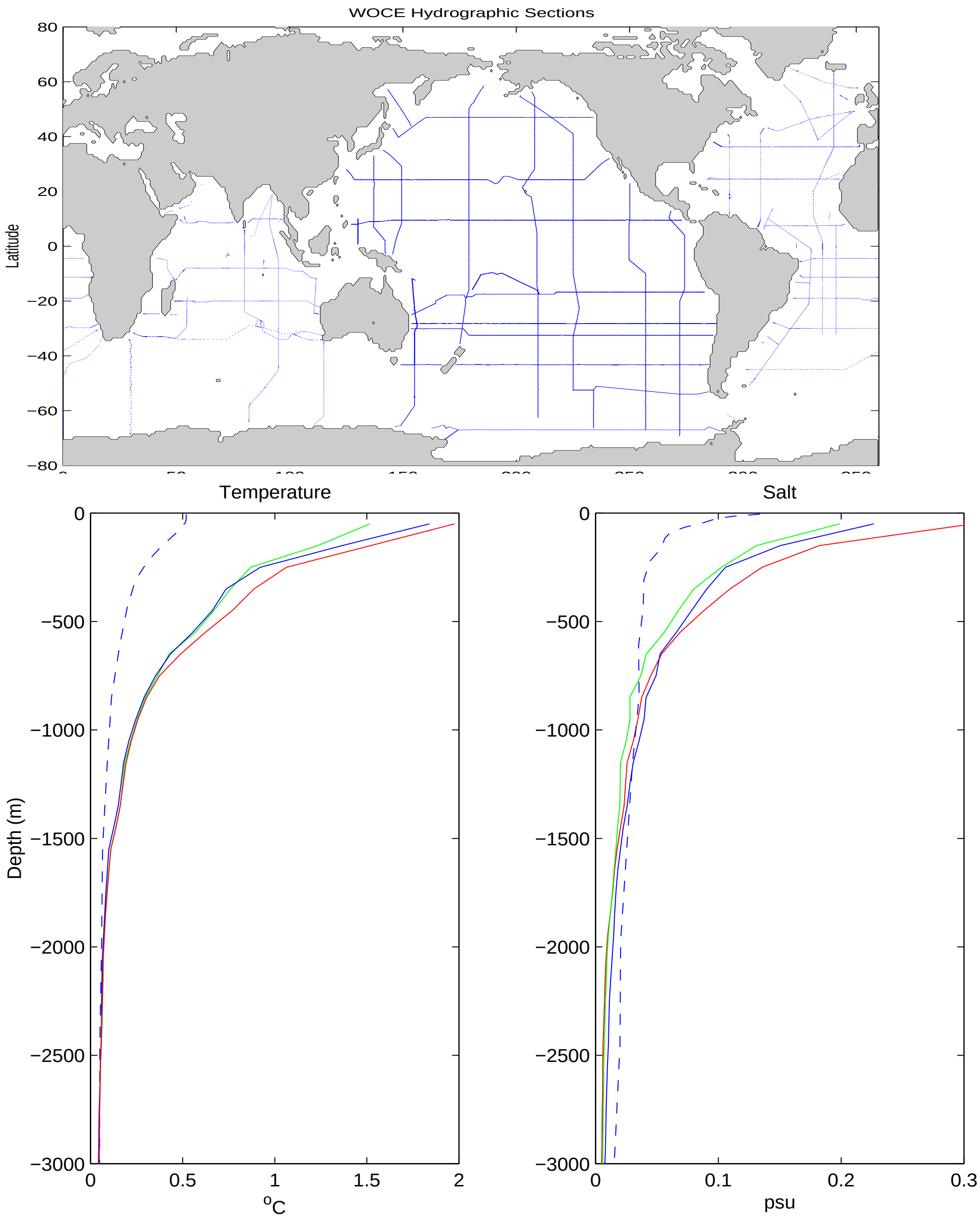

WOCE Hydrographic Sections
Latitude
80
60
40
20
0
−20
−40
−60
−80
Temperature
Salt
Depth (m)
0
−500
−1000
−1500
−2000
−2500
−3000
0
0.5
1
1.5
2
°C
0
0.1
0.2
0.3
psu

Figure 18:

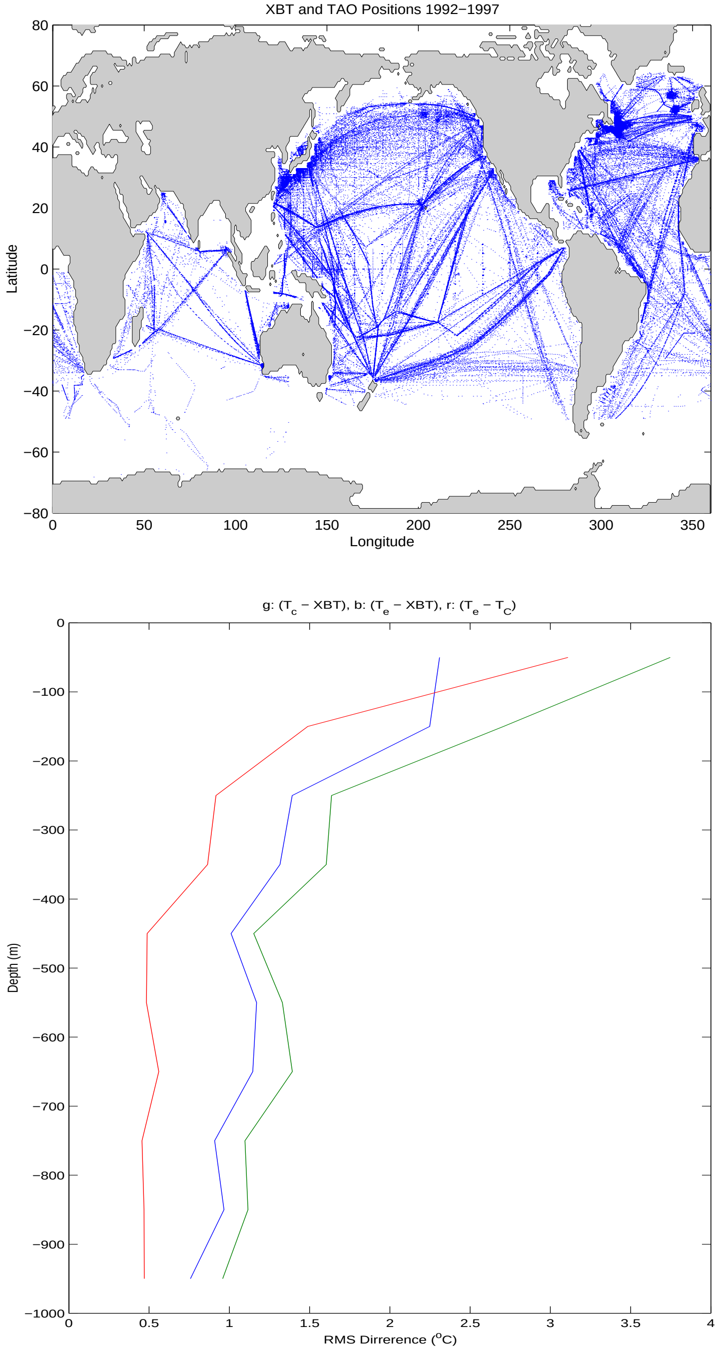

XBT and TAO Positions 1992−1997
Latitude
80
60
40
20
0
−20
−40
−60
−80
Longitude
0
50
100
150
200
250
300
350
g: (T_c − XBT), b: (T_e − XBT), r: (T_e − T_C)
Depth (m)
0
−100
−200
−300
−400
−500
−600
−700
−800
−900
−1000
RMS Dirrerence (°C)
0
0.5
1
1.5
2
2.5
3
3.5
4

Figure 19:

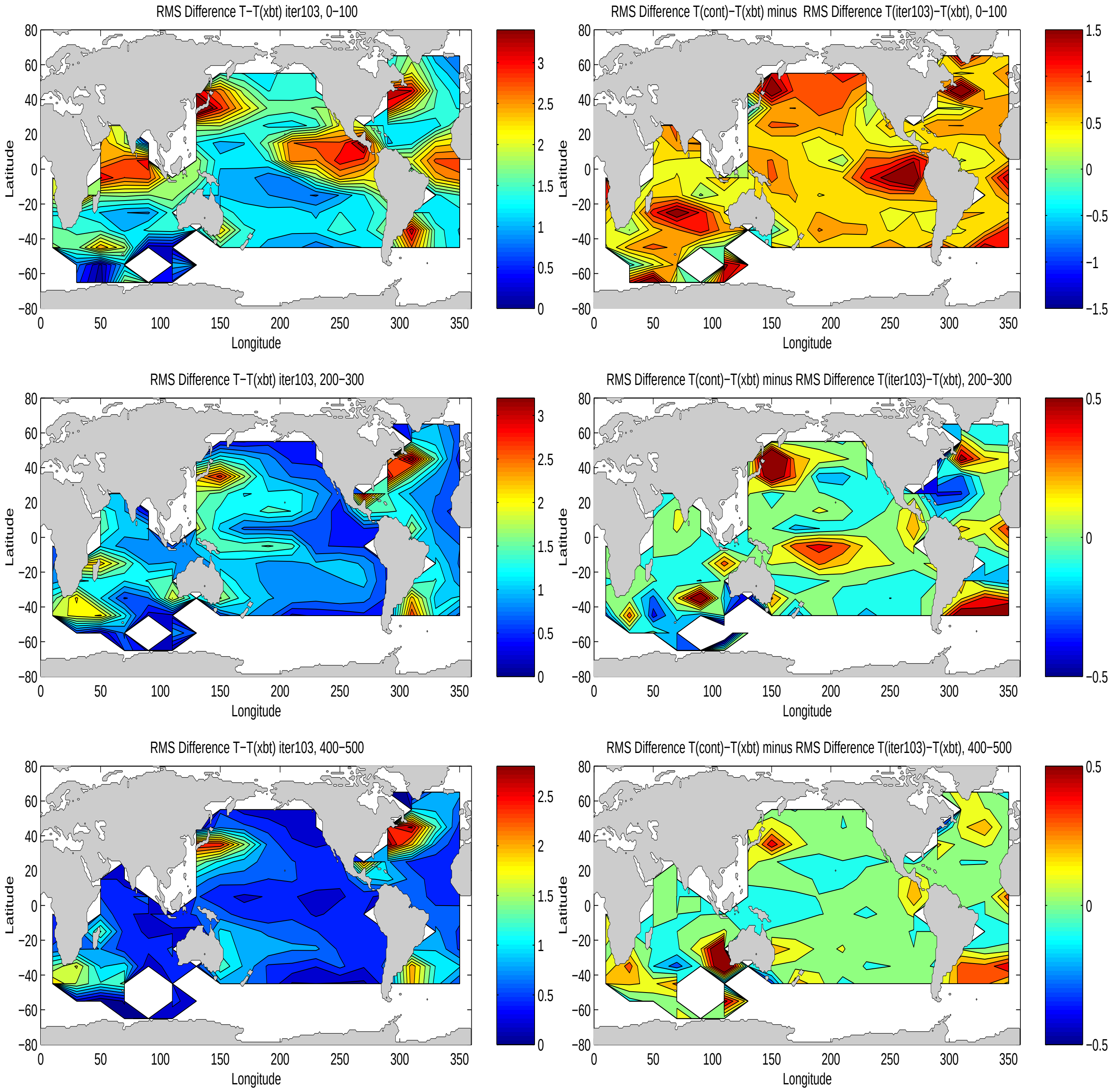

RMS Difference T−T(xbt) iter103, 0−100
RMS Difference T(cont)−T(xbt) minus RMS Difference T(iter103)−T(xbt), 0−100
RMS Difference T−T(xbt) iter103, 200−300
RMS Difference T(cont)−T(xbt) minus RMS Difference T(iter103)−T(xbt), 200−300
RMS Difference T−T(xbt) iter103, 400−500
RMS Difference T(cont)−T(xbt) minus RMS Difference T(iter103)−T(xbt), 400−500
Latitude
Longitude
80
60
40
20
0
-20
-40
-60
-80
0
50
100
150
200
250
300
350
3
2.5
2
1.5
1
0.5
0
1.5
1
0.5
0
−0.5
−1
−1.5
0.5
0
−0.5

Figure 20: