

**An Ensemble Circulation-Dependent Inflation Filter  
in a Coupled GCM – Coping with Deep Ocean Biases,  
Part II: Impact on Oceanic Climate Detection**

S. Zhang\* and A. Rosati

*Geophysical Fluid Dynamics Laboratory, Princeton University*

*Princeton, NJ 08542, USA*

(submitted to QJ)

\* *Corresponding author address:* Shaoqing Zhang, GFDL/NOAA, Princeton University,  
P.O. Box 308, Princeton, NJ 08542, USA. Email: Shaoqing.Zhang@noaa.gov Tel: (609)452-  
6540 Fax: (609)987-5063

## SUMMARY

As the second part of ‘biased’ oceanic data assimilation (ODA) twin experiment studies using coupled GCMs (CGCMs), here we examined the impact of ensemble circulation-dependent inflation filter (EcdiF) on oceanic climate detection. Two CGCMs – GFDL’s CM2.0 and CM2.1 are used. Observations are produced by projecting the CM2.0’s simulation onto the 2005 Argo network and assimilated into CM2.1. Because of model bias and limitation of the representation of a finite ensemble for the low frequency deep ocean variability, a standard ensemble filter (ENSF) fails to construct a coherent vertical structure of ocean and develops spurious velocities. EcdiF uses pre-computed standard deviations of anomalies to inflate the covariance for deep ocean data constraints. The EcdiF improves substantially the tropical undercurrent and upwelling and the associated Western Boundary Current systems as well as subtropical gyre structure. Consequently, the estimates of the world ocean’s hydrographic features such as global overturning, pycnocline depth and its correlation with sea surface height, are improved dramatically. Decadal trends of the basin scale heat content and salinity and the seasonal-interannual variability of tropical oceans are constructed coherently.

Results also showed that the Indian Ocean, especially the North Indian Ocean, is the most sensitive basin to the covariance formulation used in the assimilation, probably because of the stronger atmospheric feedbacks. The local thermohaline structure plays a leading-order important role for estimating the decadal trend of the North Atlantic meridional overturning circulation (NA MOC) but more accurate estimate of the NA MOC’s variability requires better external forcings and internal heat and salt transports.

# 1 Introduction

Viewing the evolution of climate states as a continuous stochastic dynamical process, the coupled ensemble data assimilation (CDA) system at GFDL (Geophysical Fluid Dynamics Laboratory, NOAA) solves for a temporally-varying probability density function (PDF) of climate states by combining data and models (Zhang et al. 2007). The assimilation-generated temporally-varying PDF of climate states is a complete solution for climate estimation, given atmospheric and oceanic observations. Due to incomplete understanding of radiative effects in various components of the earth system and inaccurate numerical implementation of physical processes, coupled climate models are “biased,” in which the model always drifts away from the real world. Such model bias usually appears systematically colder/warmer and/or fresher/saltier in the deep oceans, and could induce computational modes in the results of oceanic data assimilation (ODA) due to the shocks between data and model background. It is difficult to identify the real data-sampled signals from the bias-generated artifacts in ODA products.

In order to examine the impact of coupled model’s bias on ODA, previous study (Zhang and Rosati 2008) has already designed an imperfect ODA twin experiment using two coupled general circulation models (CGCMs) that are biased with respect to each other (Delworth et al. 2006; Gnanadesikan et al. 2006; Zhang et al. 2008; Zhang and Rosati 2008), in which observations taken from one CGCM based on the 2005 Argo network are assimilated into the other. Under this imperfect model twin experiment framework, both the model bias and the true solution of data assimilation problem are unambiguously defined *a priori* so that all aspects of the bias’s impact on oceanic climate estimation can be quantatified.

First it was found that the assimilating imperfect model with a standard ensemble filter (ENSF hereafter) can successfully recover the upper ocean temperature and salinity from observations but fails to converge in the deep ocean. This failure occurs because in deep

ocean the model bias is relatively large compared to the ocean’s intrinsic variability which is represented by the spread of a finite ensemble so that the filtering analysis based on the ensemble variance cannot produce a sufficient data constraint for deep ocean. Furthermore, the inconsistency between well-constrained upper and poorly-constrained deep oceans generates spurious currents and vertical motions due to incorrect pressure gradients derived from incoherent vertical structure. Then to cope with this problem an ensemble circulation-dependent inflation filter (EcdiF hereafter) was designed based on the covariance inflation theory of *filtering* (Chapter 8, Jazwinski 1970). EcdiF uses pre-computed standard deviations of anomalies to inflate the covariance wherever a small ensemble spread would otherwise make the model over-confident and rejecting data. Verification shows that EcdiF dramatically improves the estimate of oceanic currents and vertical motions from ENSF due to its improved vertical structure of ocean.

As a follow-up of the previous study (Zhang and Rosati 2008) that focused on the EcdiF’s algorithm design and tuning, this study addresses the impact of EcdiF on oceanic climate detection (estimation). Three questions will be answered: 1) What is the impact of EcdiF on hydrographic features of the world ocean? 2) What is the impact of the EcdiF-refined undercurrent and vertical motions on the tropical ocean’s climatology and ENSO’s variability? 3) What is the impact of the EcdiF-improved heat and salt transports on the North Atlantic meridional overturning circulation? After a summary on methodology in section 2, section 3 examines the time mean global fields and section 4 examines decadal trends of each basin as well as the seasonal-interannual variability of tropical oceans. The analyses and diagnostics of the North Atlantic meridional overturning circulation are given in section 5. Conclusions and discussions are given in section 6.

## 2 Methodolgy

### 2.1 Two ‘biased’ CGCMs at GFDL

Two fully-coupled models, CM2.0 and CM2.1 were developed at GFDL, which use two different atmospheric models to couple with the same ocean (Fourth Version of Modular Ocean Model, MOM4) model, sea ice simulator (SIS) and land model. The two atmospheric models have different dynamical cores but the same vertical (24 levels) and horizontal ( $2.5^\circ$  longitude by  $2^\circ$  latitude) resolution, identical physics and land processes but with their own tuned values for parameters. The atmospheric dynamical cores in CM2.0 and CM2.1 are the B-grid finite-difference scheme (Wyman 1996; GAMDT 2004) and the finite-volume scheme (Lin 2004).

The MOM4 is configured with 50 vertical levels (22 levels of 10 m thickness for each in the top 220 m),  $1^\circ \times 1^\circ$  horizontal B-grid resolution telescoping to  $1/3^\circ$  meridional spacing near the equator. The model has an explicit free surface with freshwater fluxes exchanged between atmosphere and ocean. Parameterized physical processes include K-profile parameterization (KPP) vertical mixing, neutral physics, a spatially-dependent anisotropic viscosity, a shortwave radiative penetration depth that depends on a prescribed climatological ocean color. Insolation varies diurnally and the wind stress at the ocean surface is computed using the velocity of the wind relative to the surface currents. An efficient time-stepping scheme (Griffies 2005) is employed. The SIS in the coupled model is a dynamical ice model with three vertical layers (one snow and two ice) and five ice-thickness categories. The elastic-viscous-plastic technique (Hunke and Dukowicz 1997) is used to calculate ice internal stresses, and the thermodynamics is a modified Semtner three-layer scheme (Winton 2000). Four major model components (ocean/atmosphere/land/sea-ice) in the coupled system interact through the exchange fluxes (see Fig. 2 in Zhang et al. 2008).

Detailed description for CM2.0 and CM2.1 and their major features for climate simulation can be found in Delworth et al. (2006) and Gnanadesikan et al. (2006). A specific comment we want to make here is that these two models produce ‘biased’ oceanic states with respect to each other, e.g., CM2.1 is overall 0.2°C colder and 0.01 PSU fresher than in CM2.0 for the ocean below 1 km (see Zhang and Rosati 2008). Given this characteristic, the next section uses these two models to conduct imperfect model assimilation twin experiments.

## 2.2 Imperfect model assimilation twin experiments

“Twin experiment”, or called “observing system simulation experiment” (OSSE) is used in this study. However, different from a perfect model study, this study uses two fully-coupled general circulation models (CGCMs) that produce different climate states (‘biased’ with respect to each other): one can be thought of as the “true” oceanic states along with their samples (observations) and the other tries to recover the “truth” by assimilating those observations. It could be therefore called a “biased twin experiment.”

Specifically, we use CM2.0 to produce a “true” climate variation and corresponding “observations,” and CM2.1 as assimilation model to attempt to retrieve the “truth” based on the “observed” data. Same as in Zhang et al. (2008) and Zhang and Rosati (2008), the GFDL’s IPCC historical simulation produced by CM2.0 is set as the target (called TRUTH hereafter) of assimilation experiments. As described there, the standard IPCC simulation is initialized by a 300-year spinup from the previous integration (Stouffer et al. 2004) and forced by the temporally-evolving greenhouse gas and natural aerosol (GHGNA) during 1861 to 2000, which is referred to as the model calendar. The other set GFDL’s IPCC model integration starting from the same initial conditions and using the same radiative forcings as above, but produced by CM2.1, is set as a free model control, called CTL, a reference for evaluating the assimilation’s quality.

The OSSE in this study attempts to simulate the 21<sup>st</sup>-century Argo observing system. First the daily data of oceanic temperature and salinity in the CM2.0’s IPCC simulation from 1 January 1976 up to 31 December 2000 are projected onto the Argo network by a tri-linear interpolation sampling process based on the 2005 Argo’s locations and depth (Zhang et al. 2008). Then a whitenoise is superimposed to simulate random observational errors. Note no explicit surface data are used in this study to address how to process *in situ* measurements to get a coherent ocean vertical structure.

Based on the CTL states, the ensemble initial conditions ( $IC_s$ ) for assimilation are formed by imposing yearly-separated atmospheric (including land) states centered at 00 UTC 1 January 1976 at which the oceanic (including sea-ice) state is taken. For example, the 6-member ensemble initial conditions that are used in the next section assimilation experiments are produced by combining the atmospheric and land states at 00 UTC 1 January of 1973-1978 and the oceanic and sea-ice states at 00 UTC 1 January 1976. For a coupled system, once an initial disturbance occurs in the atmosphere or other coupled model components, the strong internal variability of atmospheric circulations and oceanic state’s responses to atmospheric forcings will eventually produce inter-ensemble variations through feedbacks among these coupled components. Then the ensemble CGCM’s integrations form a simulation for the stochastic feature of climate evolution.

### **2.3 Ensemble circulation-dependent inflation filter (EcdiF) vs. standard ensemble filter (ENSF)**

*Filtering* theory (e.g. Jazwinski 1970) is based on the probabilistic nature of a dynamical system such as the atmosphere-ocean coupled system of the earth. An ensemble filter uses a Monte Carlo approach to simulate the model-described prior PDF by finite-ensemble model realizations to project an observational signal onto model space. In this context, CDA solves for the problem of sampling the probability of states of a coupled dynamical system given

noisy and sparse measurements, which have multi-variate joint-distribution.

First, at an observational location the estimated variance from model ensemble is compared to the observational variance to determine the shift of the prior ensemble mean and the adjustment of the ensemble spread. This process produces eventually an observational increment for each ensemble member. Then the estimated covariance from model ensemble is used to distribute the observational increment onto model grids and/or transform information between different physical spaces by the following linear regression process (for ENSF):

$$\begin{aligned}\Delta x_{i,j} &= \Omega_{j,k} \frac{Cov(x_j, y_k)}{\sigma_{y_k}^2} \Delta y_{i,k}^o = \Omega_{j,k} r(x_j, y_k) \Delta y_{i,k}^o \\ &= \Omega_{j,k} \rho(x_j, y_k) \frac{\sigma_{x_j}}{\sigma_{y_k}} \Delta y_{i,k}^o = \Omega_{j,k} \rho(x_j, y_k) \kappa(x_j, y_k) \Delta y_{i,k}^o.\end{aligned}\quad (1)$$

Here  $\Delta y_{i,k}^o$  and  $\Delta x_{i,j}$  represent respectively the observational increment at the  $k$ th observational location and the adjustment of the oceanic state variable at the  $j$ th grid-point for the  $i$ th ensemble member.  $\rho(x_j, y_k)$  and  $r(x_j, y_k)$  represent the correlation coefficient and the linear regression coefficient between the  $k$ th observational datum,  $x_j$  and  $y_k$ , where  $y_k$  is the model's estimate for the observation  $y_k^o$ .  $\kappa(x_j, y_k)$  is the ratio of estimated standard deviations from model ensemble for  $x_j$  and  $y_k$ .  $\Omega$  is the covariance localization function [ $\Omega(a, d)$  in Zhang et al. (2005)] which is determined only by the distance between  $j, k$  locations. Note that since all error statistics here are instantaneously evaluated by ensemble model integrations,  $Cov(x_j, y_k)$ ,  $\rho(x_j, y_k)$ ,  $\sigma_{y_k}$ ,  $\sigma_{x_j}$  and  $r(x_j, y_k)$  are functions of time. All terms in Eq. (1) (including the observation) have however the common time level,  $t$ , so for all of them the subscript  $t$  is dropped.

The EcdiF algorithm modifies the linear regression equation (Zhang and Rosati 2008) as

$$\Delta x_{i,j} = \begin{cases} \Omega_{j,k} \rho_t(x_j, y_k) \left[ \frac{0.5 \times (1-\alpha) \kappa_t(x_j, y_k)}{(1-\alpha) \kappa_t(x_j, y_k) + \alpha \kappa_0(x_j, y_k)} \kappa_t(x_j, y_k) \Delta y_{i,k,t}^o \right. \\ \left. + \frac{0.5 \times \alpha \kappa_0(x_j, y_k)}{(1-\alpha) \kappa_t(x_j, y_k) + \alpha \kappa_0(x_j, y_k)} \kappa_0(x_j, y_k) \Delta y_{i,k,0}^o \right] \end{cases}$$

and

$$\Omega_{j,k} = \begin{cases} \Omega(a^h, d_{j,k}^h)\Omega(a^v, d_{j,k}^v), & D^o \neq D_{bottom}^o \text{ or } D^o = D_{bottom}^o \text{ but } Z \leq D_{bottom}^o \\ \Omega(a^h, d_{j,k}^h)\Omega(a_b^v, d_{j,k}^v), & D^o = D_{bottom}^o \text{ and } Z > D_{bottom}^o \end{cases} \quad (2)$$

Here  $\kappa_0(x_j, y_k)$  is the ratio of the pre-computed standard deviations ( $\sigma_0$ ) for  $x_j$  and  $y_k$  using the historical dataset of oceanic states, and  $\Delta y_{i,k,0}^o$  is the observational increment computed by  $\sigma_0$  [the prior background standard deviation,  $\sigma^p$ , in Eqs. (2)-(5) of Zhang et al. (2007) is replaced by the corresponding  $\sigma_0$ ].  $a^h$  and  $a^v$  are the e-folding horizontal and vertical scales in the covariance localization function respectively, and  $d_{j,k}^h$  and  $d_{j,k}^v$  is the horizontal and vertical distance between  $x_j$  and  $y_k$ .  $D^o$  and  $D_{bottom}^o$  represent respectively the current observation depth and the depth at the end of an observed profile, and  $Z$  is the vertical coordinate.  $a_b^v$  is a vertical impact scale that controls the impact depth to which the adjustment from the bottom of an observed profile is extended.  $\alpha$  is added in this study and used to set an on-off switch of the inflation.

Except for the following two aspects in this study, all other parameters are the same as Zhang and Rosati (2008) (see Table 1):

- 1) The temperature's vertical extension scale  $a_b^v$  is the half of the salinity's.
- 2) For this 6-member experiment, in the North Atlantic beyond 20°N,  $\alpha$  is set to be 0 to sustain active convections in this region (see section 5) (as the ensemble size increases,  $\alpha$  could be set to be 0 poleward of 20° to sustain convections and gyre structure there).

With the ensemble initial conditions and the oceanic “observations” described in section 2.2, ENSF and EcdiF are run for 25 years (up to 31 December 2000) using 6-member ensemble with the covariance localization and observation smoothing technique (Zhang et al. 2005) on a daily assimilation cycle. Next, starting from examining major hydrographic features produced by EcdiF, the impact of EcdiF on oceanic climate detection will be analyzed and discussed thoroughly.

### 3 Global annual mean fields

#### 3.1 Time mean of heat content, salinity, velocities and wind stresses

We leave off the first 5 years as the assimilation spinup period; all diagnostics and analyses next are based on the last 20 years of assimilation data, i.e. from 1 January 1981 to 31 December 2000.

First Fig. 1 gives a brief summary of the EcdiF’s assimilation improvement over the ENSF in the tropical oceans (20°S-20°N average). Compared to CTL, although ENSF reduces greatly the potential temperature (temperature hereafter) and salinity errors in upper oceans, it produces extra errors for currents, and vertical velocity. EcdiF further reduces the error of temperature and salinity, especially for the deep ocean’s below 1 km, by using the pre-computed standard deviation of anomalies to inflate the covariance and an appropriate  $a_b^v$  value. This refinement of temperature and salinity appears critically important for eliminating the spurious currents and vertical motions in ENSF (see lower three panels at right column).

The 20-year time mean of assimilation errors of oceanic heat content (averaged temperature) (left) and salinity (right) over 0-4 km is presented in Fig. 2. Compared to CTL, both ENSF and EcdiF make the oceanic temperature and salinity converge to TRUTH to a large extent. However, in ENSF, due to its inconsistent convergent rate in the upper and deep oceans the vertical structure of the assimilating ocean does not coherently converge to TRUTH. This develops spurious currents and vertical velocity (see the lower 3 panels at the middle column in Fig. 1). These spurious vertical motions can develop computational Kelvin waves. The eastward-propagating Kelvin waves and their reflection at the east costal boundary of the Pacific Ocean maintain the time mean error centers over the tropical east Pacific Ocean. The extra error reduction created by EcdiF is critical for constructing the

watermass's structure so that the spurious currents and vertical velocity in ENSF are eliminated (lower three panels at the right column in Fig. 1). Thus, the error centers of the tropical east Pacific Ocean vanish. The improvement of currents and vertical velocity also gives rise to the changes in thermohaline properties (most of them are improved) at middle and high latitudes. These changes are, on one hand, directly related to the improvement of hydrography due to the oceanic data constraint. They are, on the other hand, also associated with the changes of other coupled components (atmospheric conditions, for instance) due to the improvement on sea surface temperature (SST) in EcdiF (left panels of Fig. 3). Although the new adjustment method in EcdiF can also improve the data constraint on upper oceans, the significant improvement of the SST shall be attributed to more consistent velocities. The major SST's error reduction occurs at the tropical Pacific Ocean, the Northern Atlantic Ocean, and the high latitude Southern Pacific Ocean. The former one reflects the sensitivity of the tropical Pacific SSTs to the undercurrent and upwelling/dowwelling while the latter two reflect the improvement of the estimate of gyres' location and structure due to the improvement of velocity estimates. The improved tropical SSTs must improve precipitation in the atmosphere, which leads to the improvement of the sea surface salinity (SSS) in tropics (right panels of Fig. 3). A noticeable phenomenon is that the surface coastal areas of the Atlantic-Antarctic Ocean in both assimilations become even colder (compared to CTL), especially for EcdiF. This must be associated to the ice activities in Antarctic caused by the changes of atmospheric and oceanic conditions in assimilations, and the discovery of the mechanism requires further studies.

This study is focusing on oceanic climate detection and not addressing the impact of the EcdiF-improved oceanic states on the atmospheric circulations. Here we check the time mean wind stress, a synthesis product of the atmospheric feedback ODA. The time mean of zonal ( $\tau_x$ , left) and meridional ( $\tau_y$ , right) wind stress errors is shown in Fig. 4. Basically, the EcdiF's wind stress shares most of features of the ENSF's. Relative to the error reduction

from the CTL to ENSF, except for the Indian Ocean, especially the North Indian Ocean, the improvement of  $\tau_x$  produced by EcdiF from ENSF is in a negligible range while the error reduction of  $\tau_y$  produced by EcdiF from ENSF is greater (nearly double) than the error reduction produced by ENSF from CTL, especially for tropics. For example, while ENSF reduces the  $\tau_y$  Rms error by 11% (5%) from CTL, EcdiF reduces the  $\tau_y$  error by 22% (11%) from ENSF for the tropics/globe. (the North Indian Ocean is the most sensitive basin to the covariance formulation used in the assimilation. This will be expanded next.) This can be understood as  $\tau_x$  is basically determined by the meridional gradient of SSTs, which is the leading order information from ODA’s data constraint, while  $\tau_y$  is more tied with the zonal gradient of SSTs. Correctly estimating the zonal gradient of SSTs requires a refined local structure of circulations. Thus, the improved currents and vertical velocity in EcdiF have stronger impact on  $\tau_y$  than on  $\tau_x$ . Consistently with the improved SSTs, the major improved regions of  $\tau_y$  are the tropical Pacific Ocean and the Northern Atlantic Ocean.

### 3.2 Meridional mass, heat and salt transport

This section examines the capability of EcdiF to sustain the general hydrographic features established in a coupled system, which is critically important for our motivation of initializing numerical climate predictions from seasonal-interannual (SI) up to multi-decadal time scales.

The global overturning stream function in depth space reflects the pole-to-pole circulation associated with the North Atlantic Deep Water. As described in Gnanadesikan et al. (2006), while CM2.1 (CTL) and CM2.0 (TRUTH) share a common character – most of water downwelling in the northern oceans travels all the way to the Southern Ocean, the CTL shows a stronger and deeper-penetrating overtruning at high latitudes (see their difference in panel *a* of Fig. 5, called “CTL errors”). Figs. 5*bc* present the assimilation errors of the global overturning streamfunction produced by ENSF and EcdiF. In ENSF, due to

too strong spurious upwelling/downwelling at the tropical deep oceans induced by incorrect vertical structure, the pole-to-pole circulation is broken and appears a reverse circulation at tropics (see panel *b*). More analyses in the next will show that it is the North Indian Ocean that makes the tropical reverse circulation, where the worst destruction of coupled balances occurs due to the inconsistent oceanic data constraints in ENSF. The improved vertical structure of data constraints in EcdiF eliminates the spurious velocity mostly so as to reduce the global overturning errors greatly. The fact that the EcdiF's global overturning errors stay in the same level as CTL's (or a little smaller) (compare panel *c* to panel *a*) means that the EcdiF's adjustment overall sustains the model balance in a global view.

In order to show how differently ENSF and EcdiF assimilations maintain the oceanic circulations balanced with sea surface wind forcings, the global overturning in potential density space ( $\sigma_2$ ) is presented in Figs. 5*def*. In a balanced model simulation, the overturning stream function in  $\sigma_2$  space represents the surface wind-driven circulation which plays an important role in heat/salt transport (Gnanadesikan et al. 2006; Bocaletti et al. 2005). Again, the difference of model simulated overturning in  $\sigma_2$  space (panel *d*) reflects the wind stress's difference in CM2.0 and CM2.1 (Gnanadesikan et al. 2006). However, the ENSF-produced spurious velocities in tropics damage seriously the wind-driven circulations below mixing layer ( $\sigma_2 > 1032$ ) and generates too much stronger watermass transformation crossing the isopycnal surfaces. This leads that the northward heat/salt transport at the equatorial region is completely destroyed (red lines in Fig. 6*bd*) and a reverse circulation (deep water travels from south to north) generated. While the spurious velocities induced from inconsistent data constraints is eliminated by EcdiF, the watermass transformation (panel *f* of Fig. 5) and the heat/salt transport (blue lines in Fig. 6*bd*) are dramatically improved.

The global distributions of the northward heat and salt transport errors in ENSF and EcdiF are presented in Figs. 7*bcef*. Figures 7*be* show that the worst destruction of the

northward heat/salt transport at the equatorial region occurs in the Indian Ocean. The further analyses and diagnostics in following sections will demonstrate that the Indian Ocean is the most sensitive basin on whether an ODA algorithm performs the vertically-consistent data constraint or not. This can be explained by the mechanism of the formation of the Indian Ocean’s circulations, which requires that more subtle balance set by the atmospheric conditions be sustained as the oceanic state is modified by ODA. This point will be expanded in sections 4.1 and 5. The northward heat/salt transport is mainly carried out by the Western Boundary Current (WBC) systems and gyre systems (Figs. 7*ad*). Besides the Indian Ocean, naturally, the improvement of the heat/salt transport produced by EcdiF occurs mainly at the WBC systems and gyre regions (compare panels *cf* to panels *be* of Figs. 7).

The zonal-depth integrals of the world ocean heat content (Fig. 6*a*) and salinity (6*c*) show that although the assimilated temperature and salinity in both ENSF (red lines) and EcdiF (blue lines) in a whole column converges to TRUTH (black lines) from the model control (green lines) in nearly the same amount, the interior structure of circulations they generate has significant difference. This phenomenon says that we have to be cautious when we see the convergence of temperature and salinity in evaluating an ODA product. In addition, from Figs. 5 and 6 as well as Fig. 7, we still notice some EcdiF errors greater than the CTL’s or the ENSF’s, especially for the transport at high latitudes of the Southern Ocean. We may attribute those errors to the discrepancy between the constrained ocean and the free atmosphere, and expect more improvement on oceanic state estimate when both ocean atmosphere are constrained by data, which will be explored in follow-up studies.

### 3.3 Pycnocline and sea surface height (SSH)

Pycnocline represents a sharp discontinuity boundary layer between light and dense water. The physical property of light (dense) waters above (below) the boundary is set in the low

latitude surface (high latitudes). Thus, the time mean pycnocline depth at middle and low latitudes is another interesting synthesis measure of the general transport of heat, salt and other tracers. It also is an interesting measure of the assimilation-estimated hydrography. Following Gnanadesikan (1999) and Park and Bryan (2000), the pycnocline depth  $Z_\sigma$  is defined as

$$Z_\sigma = \int_{z=-H}^0 \Delta\sigma_2 z dz / \int_{z=-H}^0 \Delta\sigma_2 dz, \quad (3)$$

where  $\sigma_2$  is potential density refereneeced to 2 km and  $\Delta\sigma_2 = \sigma_2(z) - \sigma_2(z_{max})$  ( $z_{max} = 2.5$  km).

Considering that the water property at high latitudes is strongly influenced by external forcings there – atmospheric fluxes, ice melting, run-off from land etc., in this ODA study we only examine the estimated pycnocline between 40°S - 40°N. First, referred to TRUTH (black line in Fig. 8), the zonal mean of the 20-year averaged pycnocline depth produced by EcdiF (blue line) shows a dramatical improvement on the estimate of pycnocline depth from the ENSF’s (red line). The assimilation error distributions (Figs. 9bc) show although the assimilation errors in both ENSF and EcdiF are somewhat reduced (9% and 45% of Rms by ENSF and EcdiF respectively) compared to the difference of two model (CM2.1-CM2.0, called CTL errors in Fig. 9a), ENSF overshoots the correction in most of basins [mean error changes from -46 m (CTL) to 31 m (ENSF) and 10 m (EcdiF)]. The negative values of CTL errors in most regions reflect stronger overturning and its deeper penetration in CM2.1 (than in CM2.0) (Gnanadesikan et al. 2006). The overshooting caused by the spurious vertical motions due to inconsistent data constraints in upper and deep oceans in ENSF occurs mainly in the tropical oceans and the North Atlantic where correctly-estimated upwelling and convection play important roles for obtaining a correct water property. The low correlation of the ENSF’s pycnocline depth with the TRUTH (0.92) relative to the CTL’s (0.96) means that due to the existence of these spurious velocities, ENSF cannot produce

a correct structure of light/dense waters. After eliminating these spurious velocities, the EcdiF assimilation recovers the water property well.

In the self-balanced oceanic states produced by model simulation, the pycnocline depth is a mirror of the time mean sea surface height (SSH). It is expected that the ODA-generated adjustment for oceanic states also maintains this kind of physical balances. As the contrast of the pycnocline depth, Figs. 9*def* present the SSH errors in CTL, ENSF and EcdiF. The SSH's differences of two models are basically represented by the differences of their pycnocline depths (compare panel *d* to panel *a*). Here we define a correlation of innovations of pycnocline depth and SSH as a measure of how much an ODA algorithm to sustain the internal balance of oceanic states. For example, the correlation computed by 2-decade's tendencies (the 10-year mean of the 1990's minus the 10-year mean of the 1980's) of pycnocline depth and SSH in model is around -0.7 (-0.73 for CM2.0 and -0.63 for CM2.1). However, persistently being contaminated by spurious upwelling and convection, the ENSF's SSH tendency completely loses the correlation with the pycnocline depth's (-0.07) although the ENSF assimilation reduces the SSH's errors from CTL, while the EcdiF assimilation enhances the correlation of SSH's tendency with the pycnocline depth's up to -0.43 while it reduces further the SSH's errors.

## 4 Decadal trends and seasonal-inteannual (SI) variability

### 4.1 Trends of basin scale heat content and salinity

Time series of anomalies of basin or global averaged temperature (salinity) over top 4 km are presented in Fig. 10 (Fig. 11). (Note for comparison all anomalies are computed using the TRUTH's climatology.) Both models show a roughly  $0.002^{\circ}\text{C}/\text{year}$  warming trend in the top 4 km world Ocean although they have a  $-0.35^{\circ}\text{C}$  (CM2.1-CM2.0) bias in climatological

level, and the Atlantic and Indian Oceans are the major contributors of the warming trend (black lines for CM2.0, called TRUTH; green lines for CM2.1, called CTL). The warming trend can be attributed to their common GHGNA external forcings. No significant trend is found in the integral of salinity in individual basins and the World Ocean but the bias of salinity (CM2.1 vs. CM2.0) is obvious in all basins.

Except for the North Atlantic Ocean, in all other basins the assimilating model (CM2.1) shows a cold bias to TRUTH (CM2.0). The biggest cold bias of CM2.1 relative to CM2.0 is found in the North Indian, close to  $-1^{\circ}\text{C}$ . Opposit to a nearly-uniform cold bias, every basin has its own fresh or salty bias. A net result is that the top 4 km World Ocean maintain a 0.002 PSU tiny fresh bias. Overall speaking, except for a little overshooting in the temperature of the South Atlantic Ocean and the salinity of the North Pacific Ocean, both ENSF (red lines) and EcdiF (blue lines) show converging to TRUTH from CTL, i.e. reducing the biases dramatically (the North Indian's temperature bias is reduced to  $-0.35^{\circ}\text{C}$  and its salinity bias is reduced to  $-0.1$  PSU from  $-0.16$  PSU, for instance). Compared to ENSF, EcdiF speeds up the convergence of heat content and salinity in terms of their integrals in the basin water volume, and also EcdiF further reduces the biases except for the South Atlantic Ocean. Due to the effect of cancellation, the net World Ocean's salinity bias is dominant by the overshooting of certain individual basin. This leads that the assimilation's benefit cannot be seen in the world Ocean's salinity bias (either from CTL to ENSF or from ENSF to EcdiF) although the salinity's bias is apparently reduced in all other basins in either case.

Interestingly, the greatest improvement made by EcdiF is found in the North Indian Ocean and the Arctic Ocean for both temperature and salinity. In the North Indian Ocean the temperature's bias is further reduced to  $-0.05^{\circ}\text{C}$  while the salinity's bias is eliminated almost completely. Another interesting phenomenon in the North Indian Ocean is that the anomalies (both temperature and salinity) produced by ENSF shows a computational

seasonal-cycle-like fluctuation while EcdiF does not. Given the fact that the variability of the North Indian Ocean’s circulations are strongly influenced by the Indian monsoon system, ones might attribute this artifact variability to a different seasonal-cycle phase of the Indian monsoon driven by the ENSF-generated SST. However, neither SSTs and wind stresses over the North Indian Ocean in ENSF present any noticeable seasonal-cycle-like oscillation. Instead, the SST distributions in ENSF and EcdiF are very similar (see Fig. 3). In fact, Fig. 10 tell us that this seasonal-cycle-like artifact variability in ENSF is induced by the vertically-inconsistent data constraint in the ODA algorithm, which persistently conflicts with the monsoon system’s driving. These phenomena suggest that atmospheric data constraint in a CDA framework is particularly important for estimating the North Indian Ocean’s state. And also, improving ocean model errors is very important to make a CDA approach successful. These will be discussed in more details by the follow-up studies about the impact of atmospheric data constraint on oceanic climate estimation.

While the ENSF-generated heat content in other basins converges to TRUTH, the Arctic Ocean’s heat content diverges (the bias is bigger than the CTL’s and has a growing trend), but the EcdiF-generated Arctic Ocean’s heat content tends to converge. Since there is no data constraint in the Arctic Ocean, the ENSF’s divergence or the EcdiF’s convergence has to be the consequence of the model’s responding to oceanic data constraints in other basins, either due to ocean interior heat/salt transports or owing to the changes of atmospheric conditions based different SSTs, or both. This aspect will also be further explored in follow-up studies.

## 4.2 El Nino – Southern Oscillation (ENSO) variability

Figure. 1 has shown that EcdiF produces quite different upwelling and undercurrents in tropical oceans. This section discusses how these different tropical upwelling and undercurrents

influence the seasonal-interannual variability of tropical oceans.

First, the seasonal cycles of the domain-averaged temperature and salinity over Nino3.4 (170°W-120°W, 5°S-5°N) in 4 cases – the free model (CM2.1) control (CTL) (top), the ENSF (middle-upper) and EcdiF (middle-lower) assimilations, and the “truth” (TRUTH) (bottom) model (CM2.0) run – are presented in Figs. 12 and 13. We can see a stronger annual-cycle of both temperature and salinity in TRUTH than in CTL for the top ocean. The difference of temperature’s annual cycle in CM2.0 and CM2.1 reflects the different wind-driven thermocline seasonal oscillations in the two models (Wittenberg et al. 2006). The stronger annual cycle of the top ocean’s salinity in TRUTH than in CTL can mainly be attributed to stronger seasonal oscillation of precipitation in CM2.0 relative to CM2.1. With oceanic data constraint, relative to CTL, while ENSF and EcdiF strengthen/weaken the subsurface cold/warm peak in boreal spring/winter, both assimilations weaken the surface warm peak in boreal later spring and cold peak in boreal later summer. While both ENSF and EcdiF underestimate the strength of the temperature’s annual cycle, the EcdiF’s estimate is improved from the ENSF’s. In contrast, compared to the CTL’s, the salinity’s annual cycle in ENSF and EcdiF is progressively strengthened approaching to TRUTH. While ENSF overestimates the strength, the EcdiF’s is closer to TRUTH although with a little underestimate. The comparative different effect of assimilation in upper ocean temperature and salinity may be related to the fact that the atmosphere responds to SST but not to SSS, while the latter is influenced by the atmosphere-generated precipitation.

In contrast to the annual cycle, consistent with the spectrum analyses in Wittenberg et al. (2006), the time series of the anomalies of the Nino3.4 temperature and salinity (Figs. 14 and 15) show much stronger ENSO variability in CTL (CM2.1, top) than in TRUTH (CM2.0, bottom). Compared to the TRUTH’s, the CTL’s ENSO events have too strong intensity and too short period, and thus they have entirely different ENSO phases in the whole 25 years [see

the green (CTL) and black (TRUTH) lines in the bottom panel]. Note that all anomalies are referenced to the TRUTH’s climatology, and then due to the existence of annual cycle errors as shown in Figs. 12 and 13, the anomaly’s time series of CTL, ENSF (middle-upper) and EcdiF (middle-lower) show a noticeable seasonal oscillation, especially around the thermocline layer. Through ODA, overall speaking both ENSF and EcdiF reduce the intensity and prolong the period of ENSO events, and the phases of ENSO events converge to to the TRUTH’s [see the red (ENSF) and blue (EcdiF) lines in the bottom panel], decreasing the Rms exceeding 50% (64% for temperature and 54% for salinity). However, due to the artifact vertical motions induced by the inconsistent upper and deep ocean data constraints, the ENSF assimilation still remains too strong cold/warm phases near surface/thermocline. By improving vertical structure of data constraints, the EcdiF assimilation further reduces the errors of the phase and intensity of ENSO events, reducing extra 20%/12% Rms error for the temperature/salinity anomalies. Again, due to the existence of the seasonal-cycle’s error of the EcdiF assimilation (compare the middle-lower panels to bottom panels in Figs. 12 and 13), the variability of the temperature and salinity anomalies in EcdiF still shows a seasonal-cycle pattern but it is much weaker than that of ENSF.

It is worth to mention that in the coupled model assimilation framework the improvement of the estimate of the tropical ocean’s variability made by EcdiF represents a balance result between oceanic data constraint and surface forcings provided by the atmosphere. Further understanding about the contributions of better atmospheric conditions due to improved SSTs to improving the ENSO variability will be given in follow-up studies. In addition, from diagnostics and analyses in this and previous sections we can see it’s relatively easier for an ensemble filter to capture the tropical ocean’s variability than to construct the ocean’s hydrography. This is because the temporally-varying error statistics used in the filtering computation (linear regression, for instance) represent short timescale activities very well but it’s rather difficult representing slow processes of the heat and salt transport (the formation

of water property) by a finite ensemble.

## 5 North Atlantic meridional overturning circulation (NA MOC)

The North Atlantic Ocean is a major basin that should be concerned when we reconstruct the oceanic states by an ODA approach since the well-known North Atlantic meridional overturning circulation (NA MOC) is an important part of global overturning. Thus, the NA MOC has been recognized as one of the most important oceanic circulations that have significant impacts on the global climate (e.g. Delworth and Greatbatch 2000, Gent and Danabasoglu 2004).

The NA MOC is a synthesis product of ocean, atmosphere and hydrological cycle. 4 major factors that influence the NA MOC's phase and variability are listed in Fig. 16 as a schematic illustration. For example, while its active/inactive regimes tie with the large-scale heat/salt transport in a global thermohaline circulation structure, the sea surface forcings from atmosphere and the fresh water forcings from ice and land runoff as well as their interaction with oceanic transport may play a more important role to form its interannual variability. All these working with the complex NA topographic features set a particular challenge for the estimate of the NA MOC through a data assimilation approach. This section serves as a very initial step in our long term efforts to reconstruct the NA MOC's phase and variability using data and models: examining the impact of the improved-estimate of thermohaline structure on the NA MOC's estimate.

The errors of the Atlantic MOC's stream functions produced by the CM2.1 free model control (CTL) (top) and the ENSF (middle) and EcdiF (bottom) assimilations are presented in Fig. 17 in depth space (left) and potential density space (right). Same as the global overturning, CM2.1 produces a stronger overturning in the Atlantic basin than CM2.0 does.

In fact, the overturning in the Atlantic Ocean is a major part of global overturning, which can be seen from the similar error patterns of overturning stream function in the northern hemisphere NA channel (panel *a* of Fig. 17) and the whole northern hemisphere (panel *a* of Fig. 5).

However, comparing Fig. 17 to Fig. 5, we find a different picture of the Atlantic overturning stream function from global overturning produced by ENSF. While ENSF reduces the errors of the stream function at the NA high latitudes, due to spurious velocities induced from inconsistent vertical structure of data adjustments, ENSF produces strong spurious recirculations at the tropical Atlantic Ocean (Fig. 17*b*), instead of a strong reverse circulation of global overturning at tropics (Fig. 5*b*). This is consistent with too strong watermass transformation crossing the isopycnal lines (panel *e* of Fig. 17) and northward heat transport (panel *b* of Fig. 7). Figure 7 has shown it is the Indian Ocean (middle panel) that makes the strong southward heat/salt transport at tropics and contributes the reverse circulation to the global overturning in ENSF.

Unlike other basins where the inflation of filtering adjustments by pre-computed standard deviations of anomalies is very important to estimate the vertical structure of ocean, due to the existence of deep convections, the North Atlantic Ocean prefers the use of a pure temporally-varying error covariance for effectively extracting observational signals. (Test experiments have shown that the introduction of any stationary error statistics into the filtering computation in the NA domain degrades the estimate of the NA thermohaline structure.) Thus, as described in section 2.3, the use of pre-computed standard deviations of anomalies in Eq. (2) for EcdiF is not beyond 20°N. As shown before, in EcdiF the vertically-consistent data constraint corrects the vertical structure of ocean at low latitudes of the Atlantic Ocean and therefore eliminates the spurious velocity there. As the result, the strong spurious recirculations of meridional overturning at the tropical Atlantic are eliminated mostly (panels *cf*

of Fig. 17) and the heat/salt transport is coherently improved by EcdiF (Fig. 7).

Checking the NA's water property produced by ENSF and EcdiF can further our understanding of the formation mechanism of the NA water so as to obtain some clues to improve the state estimate of the NA Ocean in the future. Figures 18*abc* presents the time mean potential density ( $\sigma_0$ ) layers (shaded) and isohalines (contours) at 35°N in CTL (panel *a*), ENSF (panel *b*), EcdiF (panel *c*) and TRUTH (panel *d*). The corresponding thickness of intermediate water layer (shaded) and potential vorticity (contours) distributions over the NA domain are shown in right column (panels *efgh*). Compared to the CTL's and TRUTH's, the EcdiF's water property has been improved from the ENSF's greatly. While EcdiF substantially improves the NA's subtropical gyre, it fails to improve the subpolar gyre. In fact the EcdiF's intermediate water layer over the Labrador Sea is too shallow so that the  $\sigma_0(27)$  isopycnal surface vanishes in part of the Sea. This must be associated with the incorporation of oceanic heat/salt transport and external forcings provided by the atmosphere, ice and land over the far north NA regions, which shall be explored in follow-up studies.

Finally, the time series of the maximum value of the NA MOC stream functions in 40-65°N in 2 model simulations and 2 data assimilations are shown in Fig. 19. Again, CM2.1 (CTL) (green) simulates a stronger overturning at the NA high latitudes than CM2.0 (TRUTH) (black). The trend that both ENSF's (red) and EcdiF's (blue) curves tend to converge to the truth shows the leading-order importance of the local thermohaline structure built by oceanic data constraint for the estimation of the NA's overturning. However, the improvement of heat/salt transport in low latitudes also tend to refine the NA MOC's variability. Previous studies (Delworth and Greatbatch 2000; Delworth and Dixon 2006) have shown that the surface forcings provided by the atmosphere is important to determine the NA MOC's interannual variability. It is expected that a fully-coupled data assimilation including atmospheric data constraint could provide a more self-consistent NA MOC estimate. From a view

of initializing numerical climate prediction, it is extremely important to obtain an estimate of NA MOC's phase and variability which is consistent with the large scale heat/salt transport and external forcings from other coupled components. These require long term efforts along the line of coupled data assimilation with the advance of coupled modeling.

## 6 Conclusion and discussions

Following Zhang and Rosati (2008), this study examines the impact of the ensemble circulation-dependent inflation filter (EcdiF) on oceanic climate detection in 'biased' ODA twin experiments. Two coupled GCMs – GFDL's CM2.0 and CM2.1 – are used. Observations are drawn from the CM2.0's IPCC simulation based on the 2005 Argo network and then assimilated into CM2.1. Due to the existence of model bias and limitation of the representation of a finite ensemble for the low frequency deep ocean variability, a standard ensemble filter (ENSF) fails to construct a consistent vertical structure and develops spurious velocities. EcdiF uses pre-computed standard deviations of anomalies to inflate the deep ocean's covariance for improving the consistency of data constraints in upper and deep oceans.

EcdiF improves substantially the tropical undercurrent, upwelling, and the associated Western Boundary Current (WBC) systems as well as subtropical gyre structure. Consequently, the world ocean's hydrographic features such as global overturning, pycnocline depth and the correlation of SSH and pycnocline are constructed better. Coherently, the estimates of the large time scale trends of basin scale heat content and salinity, and seasonal-interannual variability of tropical ocean states are improved. Results also showed that the Indian Ocean, especially the North Indian Ocean, is the most sensitive basin on the covariance formulation used in the assimilation, in which the stronger atmospheric feedbacks are involved. The local thermohaline structure plays a leading-order role for estimating the large scale phase of the North Atlantic meridional overturning circulation (NA MOC) but a more

accurate estimate of the NA MOC’s variability requires refined external forcings and internal transports of heat and salt upon the observing system.

This study only focuses on the ODA side of the CDA system under a twin experiment framework, directly using the dataset of the “truth” to compute the anomaly’s standard deviation, a key to implement EcdiF. On one hand, the results from the twin experiment do serve as a mirror for us to understand the results of the real data assimilation. GFDL’s coupled reanalyses by ENSF using real observed atmospheric (NCEP/NCAR reanalysis) and oceanic (*in situ* measurements and SSTs) data do not present any strong reverse circulation in global overturning at tropics but do present the recirculation at the tropical Atlantic Ocean. The former suggests that the atmospheric data constraint in a full CDA experiment relax greatly the destruction of coupled balances in the North Indian Ocean, while the latter means that even though in a full CDA circumstance, the inconsistency of oceanic data constraints in upper and deep oceans still exists. On the other hand, due to the use of the “true” oceanic states in computing the inflation standard deviation, the results presented in this study probably represent the most optimistic case. An open question is how to obtain the pre-computed standard deviation in real data assimilation. Although a primary implementation of the EcdiF in the GFDL’s coupled reanalysis, which uses model simulated oceanic states to compute the inflation standard deviation, has eliminated most of the recirculation at the tropical Atlantic Ocean, the sensitivities of EcdiF on the accuracy of the pre-computed standard deviation still needs to be examined. If strong sensitivity exists, an iteratively-refined procedure may be necessary for future coupled reanalyses. In addition, a bias correction assimilation scheme implemented in both atmospheric and oceanic data assimilations (Dee and Silva 1998; Dee 2005) is expected to further improve the CDA performance. Continuously enhancing the quality of the estimate of oceanic circulations within an improved hydrographic environment is the direction of our long term efforts since this is critically important for our understanding of climate changes and the initialization of

seasonal to decadal numerical predictions.

### *ACKNOWLEDGEMENT*

The authors would like to thank Drs. Qian Song and Andrew Wittenberg for their comments on earlier version of this manuscript. Thanks also go to Drs. Guijun Han and You-Soon Chang for their suggestions in processing observation data during their visiting at GFDL. Authors want to pay their thanks to F. Zeng who frequently provides suggestions on data processing and visualization. The authors thank two anonymous reviewers for their thorough examination and comments that are very useful for improving the manuscript.

## REFERENCES

- Anderson, J. L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Bocaletti, G., R. Ferrari, A. Adcroft, D. Ferreira, and J. Marshall, 2005: The vertical structure of ocean heat transport. *Geophys. Res. Lett.*, **32**, L10603, doi:10.1029/2005GL022474.
- Carton, J., G. Chepurin, X. Cao, and B. Giese, 2000: A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part I: Methodology. *J. of Physical Oceanography*, **30(2)**, 294–309.
- Curry, R. and C. Mauritzen, 2005: Dilution of the Northern North Atlantic Ocean in Recent Decades. *Science*, **308**, 1772–1774.
- Dee, D. P. and A. M. D. Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269–295.
- Dee, D. P., 2005: Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3323–3343.
- Delworth, T. L. and R. J. Greatbatch, 2000: Multidecadal thermohaline circulation variability driven by atmospheric surface flux forcing. *J. Climate*, **13**, 1481–1495.
- Delworth, T. L. and K. Dixon, 2000: Implications of the recent trend in the Arctic/North Atlantic oscillation for the North Atlantic thermohaline circulation. *J. Climate*, **13**, 3721–3727.
- Delworth, T. L., A. Rosati, Ronald J. Stouffer, Keith W. Dixon, John Dunne, Kirsten L. Findell, Paul Ginoux, Anand Gnanadesikan, C. T. Gordon, Stephen M. Griffies, Rich Gudgel, Matthew J. Harrison, Isaac M. Held, Richard S. Hemler, Larry W. Horowitz,

- Stephen A. Klein, Thomas R. Knutson, Shian-Jiann Lin, V. Ramaswamy, M. Daniel Schwarzkopf, Joseph J. Sirutis, Michael J. Spelman, William F. Stern, Michael Winton, Andrew T. Wittenberg, Bruce Wyman, A. J. Broccoli, R. J. S. V. Balaji, Joellen Russell, Rong Zhang, John A. Beesley, Jian Lu, William F. Cooke, Jeffrey W. Durachta, Amy R. Langenhorst, Hyun-Chul Lee, Fanrong Zeng, K. A. Dunne, P. C. D. Milly, Paul J. Kushner, Sergey L. Malyshev and Elena Shevliakova, 2006: GFDL CM2 global coupled climate models, Part I: Formulation and simulation characteristics. *J. Climate*, **19**(5), 643–674.
- Dukowicz, J. K., R. D. Smith, 1994: Implicit free surface method for the Bryan-Cox-Semtner ocean model. *J. Geophys. Res.*, **99**, 7991–8014.
- Ezer, T. and G. L. Mellor, 1994: Continuous assimilation of Geosat altimeter data into a three-dimensional primitive equation Gulf Stream model. *Journal of Phy. Ocean.*, **24**, 832–847.
- Gandin, L. S., 1963: Objective analysis of meteorological fields. *Gidrometeorologicheskoe Izdatelstvo*, Leningrad. English translation by: Israel Program for Scientific Translations, 242pp. [NTIS N6618047, Library of Congress QC996.G3313].
- Fedorov, A. V. and S. G. Philander, 2001: Is El Niño changing? *Science*, **288**, 1997–2001.
- 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.
- Fukumori, I., R. Raghunath, L. L. Fu and Y. Chao, 1999: Assimilation of TOPEX/Poseidon altimeter data into a global ocean circulation model: How good are the results? *J. Geophys. Res.*, **103**, 25647–25666.
- Gent, P. R., and G. Danabasoglu, 2004: Heat uptake and the thermohaline circulation in

- the Community Climate System Model, version 2. *J. Climate.*, **17(10)**, 4058–4069.
- Gill, A. E., 1982: Atmosphere-Ocean Dynamics, 662 pp., *Academic*, San Diego, California, USA
- Gnanadesikan, A., Keith W. Dixon, Stephen M. Griffies, Thomas L. Delworth, Matthew J. Harrison, Isaac M. Held, William J. Hurlin, Ronald C. Pacanowski, Anthony Rosati, Bonita L. Samuels, Michael J. Spelman, Ronald J. Stouffer, Michael Winton, Andrew T. Wittenberg, John P. Dunne, V. Balaji, Marcelo Barreiro, Joellen Russell, Qian Song, Colm O. Sweeney, Rong Zhang, J. Anthony Beesley, Gabriel Vecchi, William F. Cooke, Hyun-Chul Lee, Zhi Liang, Giang Nong, Fanrong Zeng, Rudiger Gerdes, 2006: GFDL CM2 global coupled climate models, Part II: The baseline ocean simulation. *J. Climate.*, **19(5)**, 675–697.
- Gnanadesikan, A., 1999: A simple predictive model for the structure of the oceanic pycnocline. *Science*, **283**, 2077–2079.
- Griffies, S. M., M. J. Harrison, R. C. Pacanowski and A. Rosati, 2003: A Technical Guide to MOM4. *GFDL Ocean Group Technical Report*, No. 5, Princeton, NJ 08542, NOAA/Geophysical Fluid Dynamics Laboratory, 295 pp.
- Griffies, S. M. and R. W. Hallberg, 2000: Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models. *Mon. Wea. Rev.*, **128**, 2935–2946.
- Hurlburt, H. E., 1986: Dynamic transfer of simulated altimetric data into subsurface information by a numerical ocean model. *J. Geophys. Res.*, **91**, 2372–2400.
- Hunke, E. C., and J. K. Dukowicz, 1997: An elastic-viscous-plastic model for sea ice dynamics, *J. Phys. Oceanogr.*, **27**, 1849–1867.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, New

York, 376pp.

- Lin, S.-J., 2004: A “vertically Lagrangian” finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307.
- Mellor, G., C. R. Mechoso and E. Keto, 1982: A diagnostic calculation of the general circulation of the Atlantic Ocean. *Deep-Sea Research*, **Vol. 29, No. 10A**, 1171–1192.
- Mellor, G. and T. Ezer, 1991: A Gulf Stream model and an altimetry assimilation scheme. *J. Geophys. Res.*, **96**, 8779–8795.
- Myers, P. G., A. F. Fanning and A. J. Weaver, 1996: JEBAR, bottom pressure torque, and Gulf Stream separation. *J. of Physical Oceanography*, **26**, 671–683.
- Park, Y.-G., and K. Bryan, 2000: Comparison of thermally driven circulations from a depth-coordinate model and an isopycnal-layer model. Part I: Scaling-law sensitivity to vertical diffusivity. *J. Phys. Oceanogr.*, **30**, 590–605.
- Pinardi, N., A. Rosati, R. C. Pacanowski, 1995: The sea surface pressure formulation of rigid lid models: Implications for altimetric data assimilation studies. *J. of Marine Systems*, **6**, 109–119.
- Sarkisyan, A. S, and V. F. Ivanov, 1971: Joint effect of baroclinicity and bottom relief as an important factor in the dynamics of sea currents. *Izv., Atmospheric and Oceanic Physics*, **Vol. 7, No. 2**, 172–188.
- Sarkisyan, A. S, and V. F. Ivanov, 1971: Joint effect of baroclinicity and bottom relief as an important factor in the dynamics of sea currents. *Izv., Atmospheric and Oceanic Physics*, **Vol. 7, No. 2**, 172–188.
- Verron, J., L. Gourdeau, D. T. Pham, R. Murtugudde and A. J. Busalacchi, 1999: An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical Pacific Ocean: Method and validation. *J. Geophys. Res.*, **104**,

5441–5458.

- Wittenberg, A. T., A. Rosati, N.-G. Lau and J. J. Ploshay, 2006: GFDL’s CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. *J. Climate.*, **19(5)**, 675–697.
- Winton, M., 2000: A reformulated three-layer sea ice model. *J. Atmos. Ocean. Tech.*, **17**, 525–531.
- Wyman, B. L., 1996: A step-mountain coordinate general circulation model: Description and validation of medium-range forecasts. *Mon. Wea. Rev.*, **124**, 102–121.
- Zhang, S. and J. L. Anderson, 2003: Impact of spatially and temporally varying estimates of error covariance on assimilation in a simple atmospheric model. *Tellus*, **55A**, 126–147.
- Zhang, S., M. J. Harrison, A. T. Wittenberg, and A. Rosati, 2005: Initialization of an ENSO forecast system using a parallelized ensemble filter. *Mon. Wea. Rev.*, **133**, 3176–3201.
- Zhang, S., M. J. Harrison, A. Rosati and A. T. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.*, **135**, 3541–3564.
- Zhang, S., A. Rosati and M. J. Harrison, 2008: Detection of multi-decadal oceanic variability within a coupled ensemble data assimilation system. *J. of Geophys. Res.*, **in press**, manuscript available at [http://www.gfdl.noaa.gov/~snz/gfdl\\_cda2.pdf](http://www.gfdl.noaa.gov/~snz/gfdl_cda2.pdf).
- Zhang, S. and A. Rosati, 2008: An ensemble circulation-dependent inflation filter in a coupled GCM – coping with deep ocean biases, Part I: Algorithm design and tuning. **Under review**.

## LIST OF TABLES

**Table 1** Values of parameters in Eq. (2).

## FIGURE CAPTIONS

**Fig. 1** Time mean (1981-2000) errors of oceanic temperature (top), salinity (middle-upper), u-component (middle), v-component (middle-lower) and vertical velocity (bottom) on the x-z plane, averaged over 20°S-20°N, in a free model control integration (left, denoted by CTL), a standard ensemble filter assimilation (middle, denoted by ENSF) and an ensemble circulation-dependent inflation filter assimilation (right, denoted by EcdiF). The contour interval is 0.05°C for temperature between 0.5 and -0.5°C, 0.02 PSU for salinity between 0.1 and -0.1 PSU,  $0.2 \times 10^{-2} \text{ m s}^{-1}$  for u- and v-components and  $0.2 \text{ m day}^{-1}$  for vertical velocity.

**Fig. 2** Time mean errors of global oceanic temperature (left) and salinity (right) assimilation errors over 0-4 km produced by ENSF (middle) and EcdiF (bottom). The CM2.1 free model control integration (CTL) is also plotted in top panels as the reference. The contour interval is 0.1°C for temperature and 0.01PSU for salinity.

**Fig. 3** Time mean errors of sea surface temperature (SST, left) and salinity (SSS, right) produced by ENSF (middle) and EcdiF (bottom) assimilations. The CM2.1 free model control integration (CTL) is also plotted in top panels as the reference. The contour interval is 0.2°C for SST and 0.05PSU for SSS.

**Fig. 4** Time mean errors of the zonal ( $\tau_x$ , left) and meridional ( $\tau_y$ , right) wind stress produced by ENSF (middle) and EcdiF (bottom) assimilations. The contour interval is 0.01 (for  $\tau_x$ ) and 0.005 (for  $\tau_y$ )  $\text{N m}^{-1}$ .

**Fig. 5** Time mean errors of the global overturning stream function in depth space (left) and potential density space (right) produced by the free model control (CTL) (*ad*), the ENSF (*be*) and EcdiF (*cf*) assimilations. The contour interval is 5/10 Sv as the absolute value of the stream functions is less/greater than 30 Sv.

**Fig. 6** Variation of the zonal-depth averaged oceanic temperature (a) and salinity (c) and northward heat (b) and salt (d) transport with latitudes in the CTL (black-dashed), ENSF (red), EcdiF (green) and TRUTH (black).

**Fig. 7** Time mean of the northward heat (left, *abc*) and salt (right, *def*) transports in TRUTH (top, *ad*) and the errors of these transports in ENSF (middle, *be*) and EcdiF (bottom, *cf*). The contour intervals are 0.2 PW ( $10^{15}$  Watts) for values between -2 – 2 PW, otherwise 2 PW for *a*); 1 PW for values between -1 – 1 PW, otherwise 0.1 PW for *b*) and *c*);  $0.1 \times 10^8$  kg/s for values between -1 –  $1 \times 10^8$  kg/s, otherwise  $1 \times 10^8$  kg/s for *d*), *e*) and *f*).

**Fig. 8** Variation of the zonal mean pycnocline depth with latitude in TRUTH (solid-black), CTL (green), ENSF (red) and EcdiF (blue). The corresponding root mean squared (Rms) error and mean error (Mer) are marked in parenthesis.

**Fig. 9** Time mean errors of the pycnocline depth (left) and the sea surface height (SSH) (right) produced by the CTL model simulation (top) and the ENSF (middle) and EcdiF (bottom) assimilations. Contour interval is 20 m for pycnocline depth and 0.02 m for SSH.

**Fig. 10** Time series of the top 4 km heat content (averaged temperature) in individual basins and the World Ocean in CTL (green), ENSF (red), EcdiF (blue) and TRUTH (black).

**Fig. 11** Same as Fig. 10 but for salinity.

**Fig. 12** Annual cycle of the Nino3.4's temperature in CTL (top), ENSF (middle-upper), EcdiF (middle-lower) and TRUTH. The contour interval is  $0.2^\circ\text{C}$ .

**Fig. 13** Same as Fig. 12 except for salinity and the contour interval is 0.01PSU.

**Fig. 14** Time series of the anomalies of the Nino3.4's temperature in CTL (top), ENSF (middle-upper), EcdiF (middle) and TRUTH (middle-lower), and their vertical integrals (bottom). The contour interval is  $0.5^{\circ}\text{C}$  and the values in parenthesis is the corresponding Rms error.

**Fig. 15** Same as Fig. 14 except for salinity and the contour interval is  $0.05\text{PSU}$  in top 4 panels.

**Fig. 16** Schematic illustration of 4 factors influencing the North Atlantic meridional overturning circulation.

**Fig. 17** Time mean errors of the North Atlantic meridional overturning circulation stream function in depth space for CTL (a), ENSF (b), EcdiF (c) and potential density space for CTL (d), ENSF (e), EcdiF (f). The contour interval is  $2\text{ Sv}$ .

**Fig. 18** Upper-ocean potential density (color-shaded) and isohaline at  $35^{\circ}\text{N}$  (left, panels *abc*) and the thickness of intermediate water layer ( $\sigma_0$  between 27 and 27.5) (color-shaded) and the PV's distribution (contour) over the North Atlantic domain (right, panels *def*). The contour interval is  $0.2\text{ PSU}/10^{-7}\text{ s}^{-1}\text{m}^{-1}$  for panel *abc/def*.

**Fig. 19** Time series of the maximum value of the North Atlantic meridional overturning circulation stream function over  $40^{\circ}\text{N}$ - $65^{\circ}\text{N}$  in CTL (green), ENSF (red), EcdiF (blue) and TRUTH (black). The thick lines are the corresponding 13-point running smooth for each case. The number in parenthesis is the corresponding Rms error.

Table 1: Values of parameters in Eq. (2).

name	physical meaning	value range	value in this study
$\alpha$	on-off switch for inflation	0 or 1	0 for the North Atlantic ( $> 20^\circ\text{N}$ ) 1 elsewhere
$Z_0$	inflation starting depth	$[0, +\infty)$	0
$a^h$	e-folding horizontal scale	$[0, +\infty)$	1000 km
$a^v$	e-folding vertical scale	$[0, +\infty)$	$2 \times \text{gridbox's thickness}$
$a_b^v$	e-folding vertical scale for expanding the adjustment at the end of obs	$[0, +\infty)$	$4a_b^{v,ENSF} \approx 4$ km for salinity $2a_b^{v,ENSF} \approx 2$ km for temperature

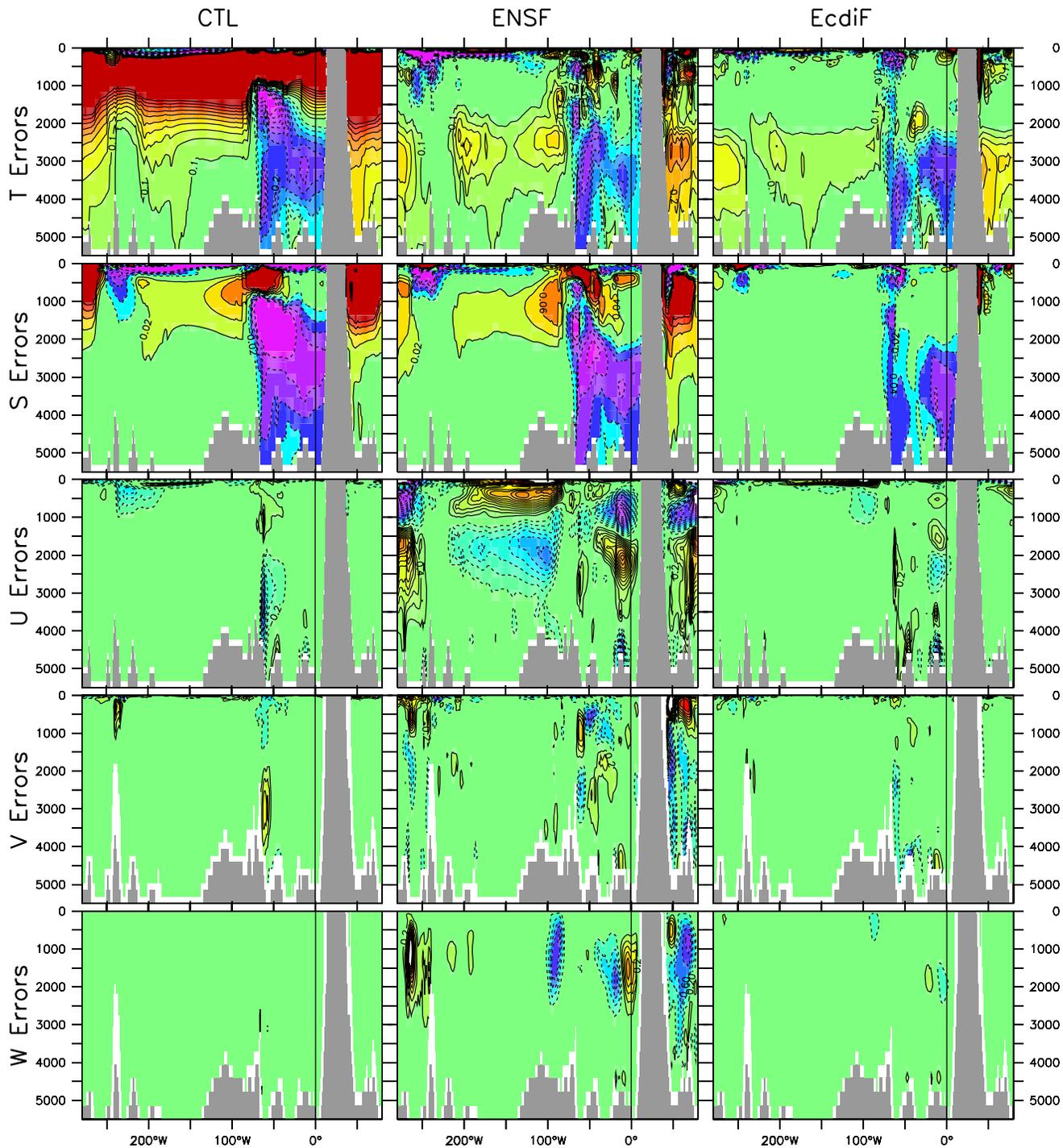


Figure 1: Time mean (1981-2000) errors of oceanic temperature (top), salinity (middle-upper), u-component (middle), v-component (middle-lower) and vertical velocity (bottom) on the x-z plane, averaged over 20°S-20°N, in a free model control integration (left, denoted by CTL), a standard ensemble filter assimilation (middle, denoted by ENSF) and an ensemble circulation-dependent inflation filter assimilation (right, denoted by EcdiF). The contour interval is 0.05°C for temperature between 0.5 and -0.5°C, 0.02 PSU for salinity between 0.1 and -0.1 PSU,  $0.2 \times 10^{-2} \text{ m s}^{-1}$  for u- and v-components and  $0.2 \text{ m day}^{-1}$  for vertical velocity.

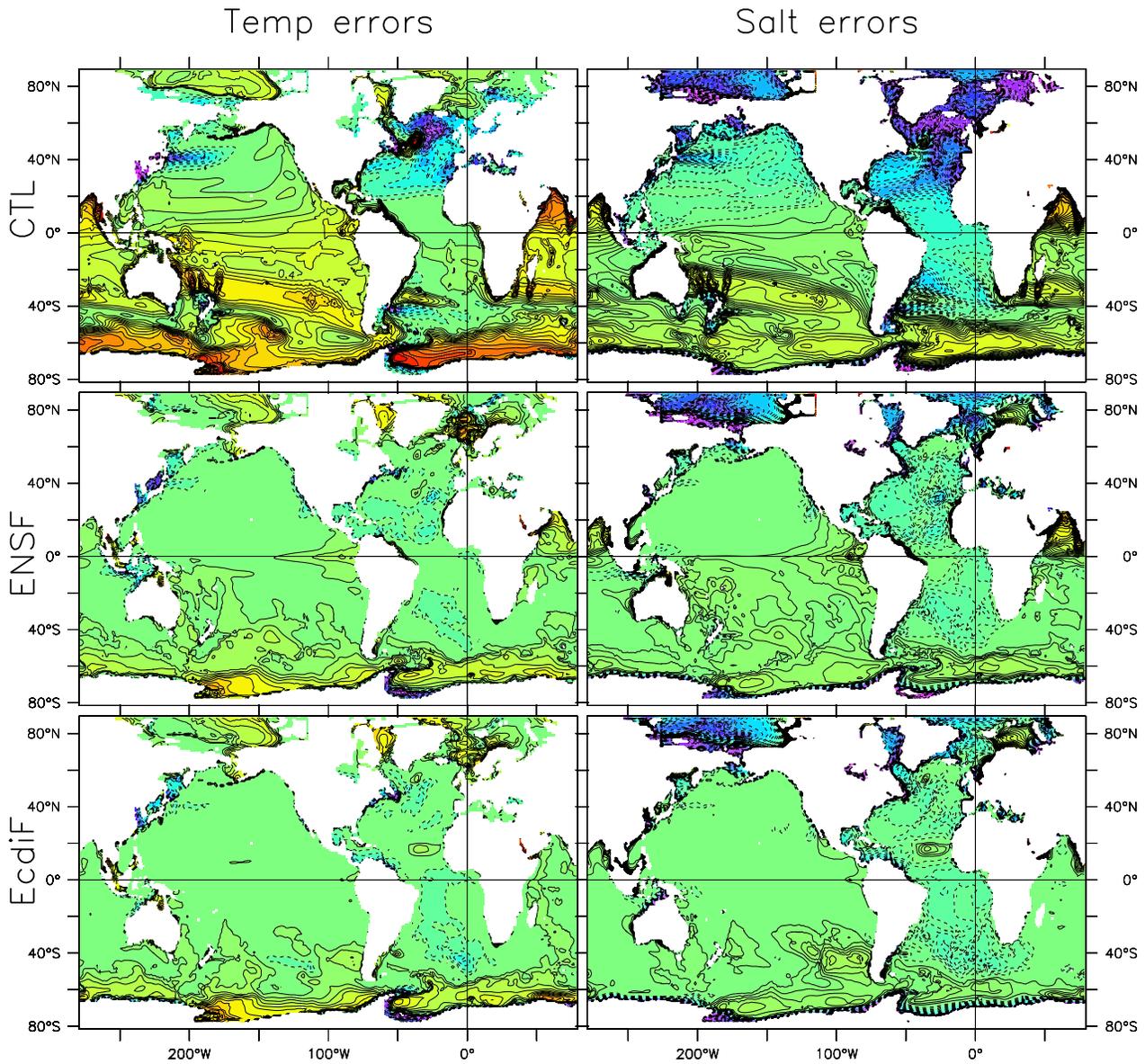


Figure 2: Time mean errors of the global oceanic temperature (left) and salinity (right) over 0-4 km produced by ENSF (middle) and EcdiF (bottom). The CM2.1 free model control integration (CTL) is also plotted in top panels as the reference. The contour interval is  $0.1^{\circ}\text{C}$  for temperature and  $0.01\text{PSU}$  for salinity.

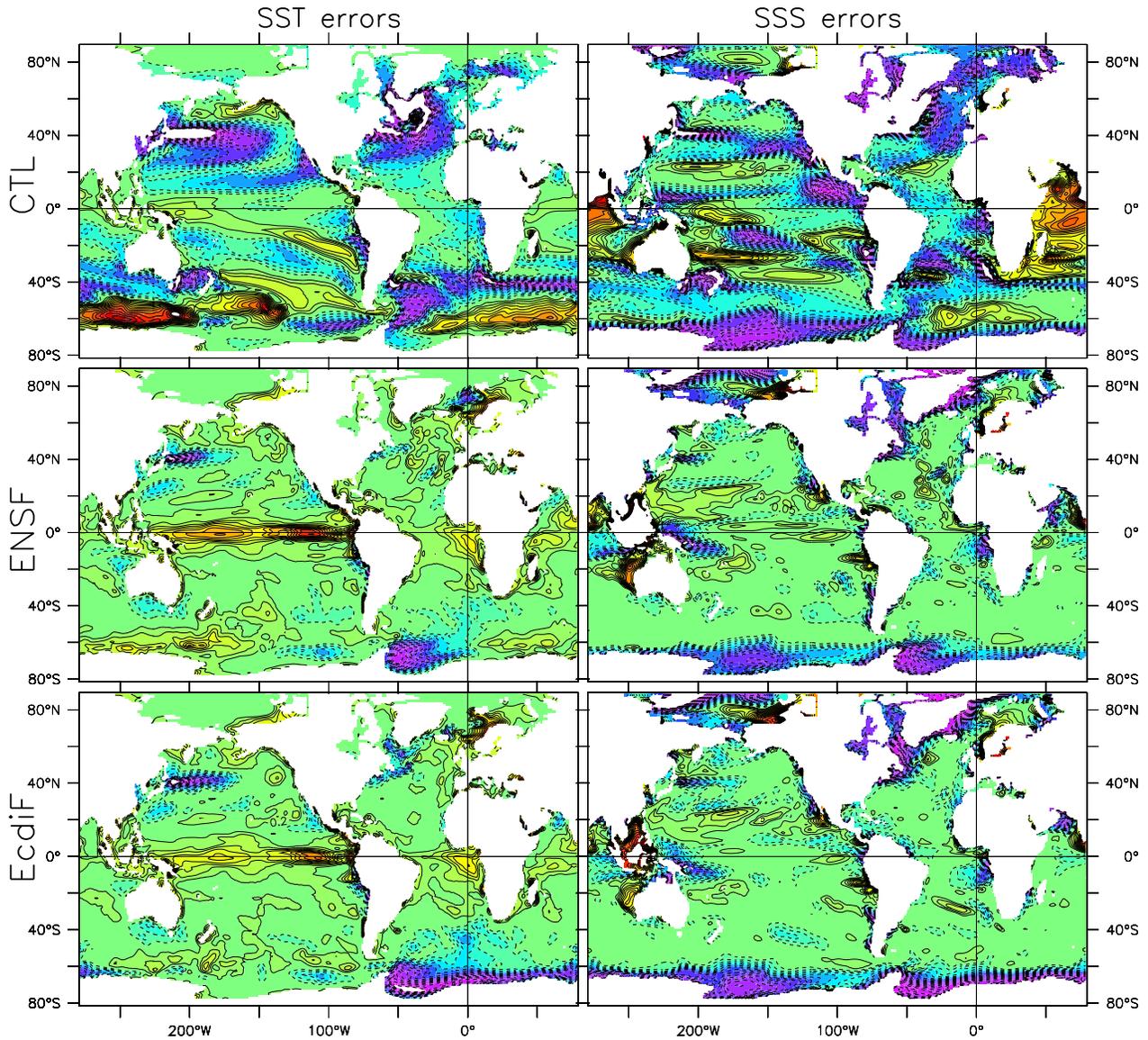


Figure 3: Time mean errors of sea surface temperature (SST, left) and salinity (SSS, right) produced by ENSF (middle) and EcdiF (bottom) assimilations. The CM2.1 free model control integration (CTL) is also plotted in top panels as the reference. The contour interval is  $0.2^{\circ}\text{C}$  for SST and  $0.05\text{PSU}$  for SSS.

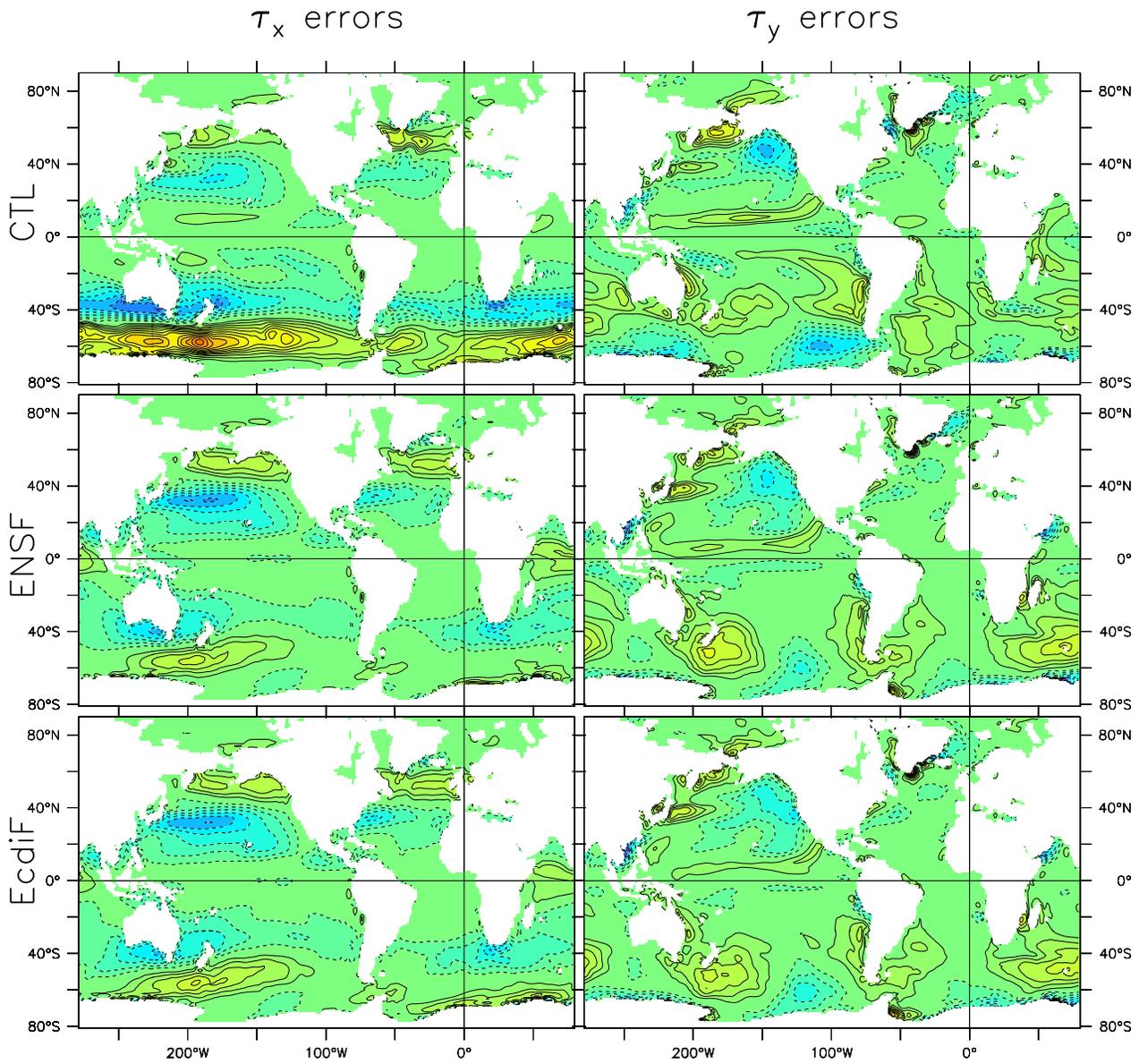


Figure 4: Time mean errors of the zonal ( $\tau_x$ , left) and meridional ( $\tau_y$ , right) wind stress produced by ENSF (middle) and EcdiF (bottom) assimilations. The contour interval is 0.01 (for  $\tau_x$ ) and 0.005 (for  $\tau_y$ )  $\text{N m}^{-1}$ .

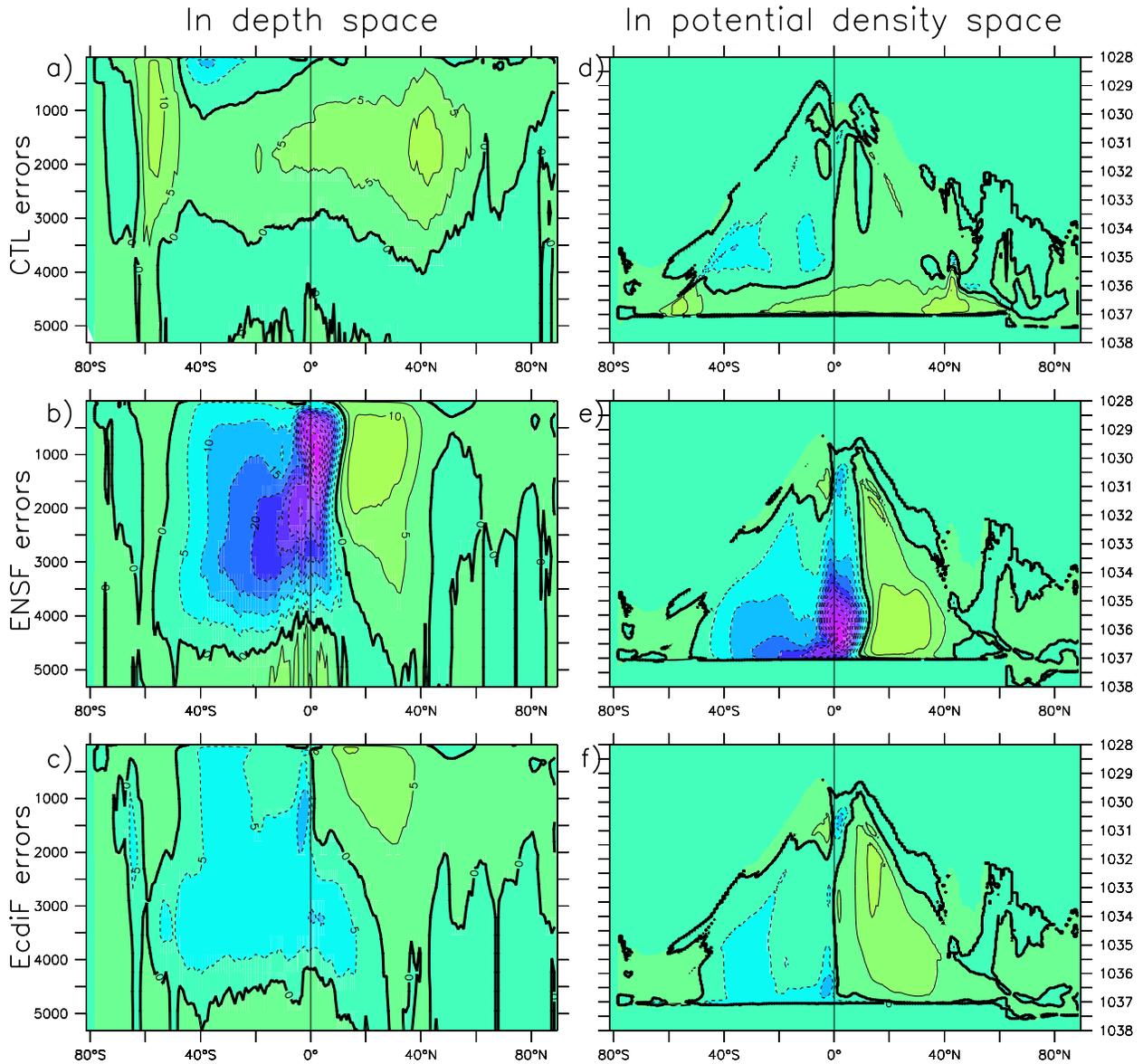


Figure 5: Time mean errors of the global overturning stream function in depth space (left) and potential density space (right) produced by the free model control (CTL) (*ad*), the ENSF (*be*) and EcdiF (*cf*) assimilations. The contour interval is 5/10 Sv as the absolute value of the stream functions is less/greater than 30 Sv.

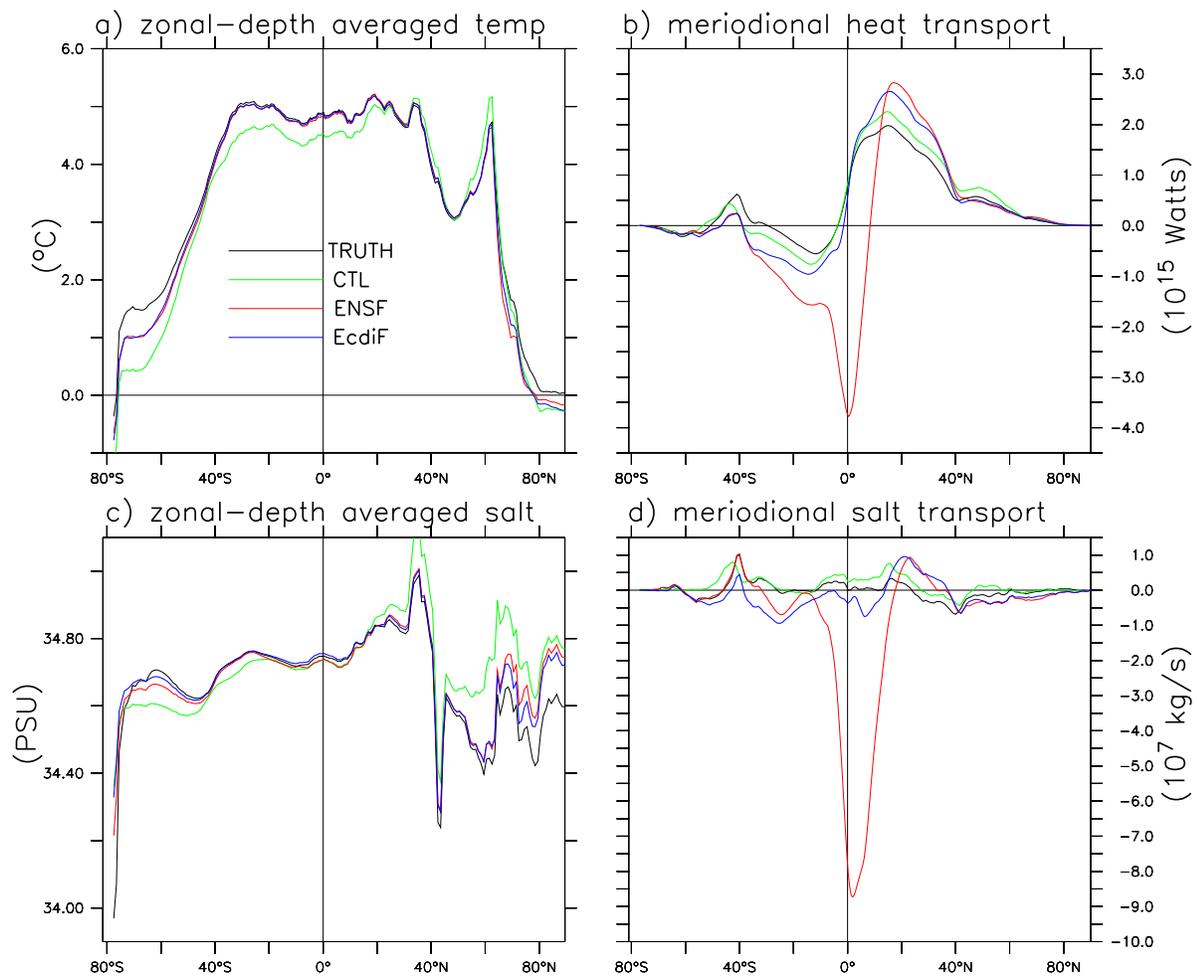


Figure 6: Variation of the zonal-depth averaged oceanic temperature (a) and salinity (c) and northward heat (b) and salt (d) transport with latitudes in CTL (green), ENSF (red), EcdiF (green) and TRUTH (black).

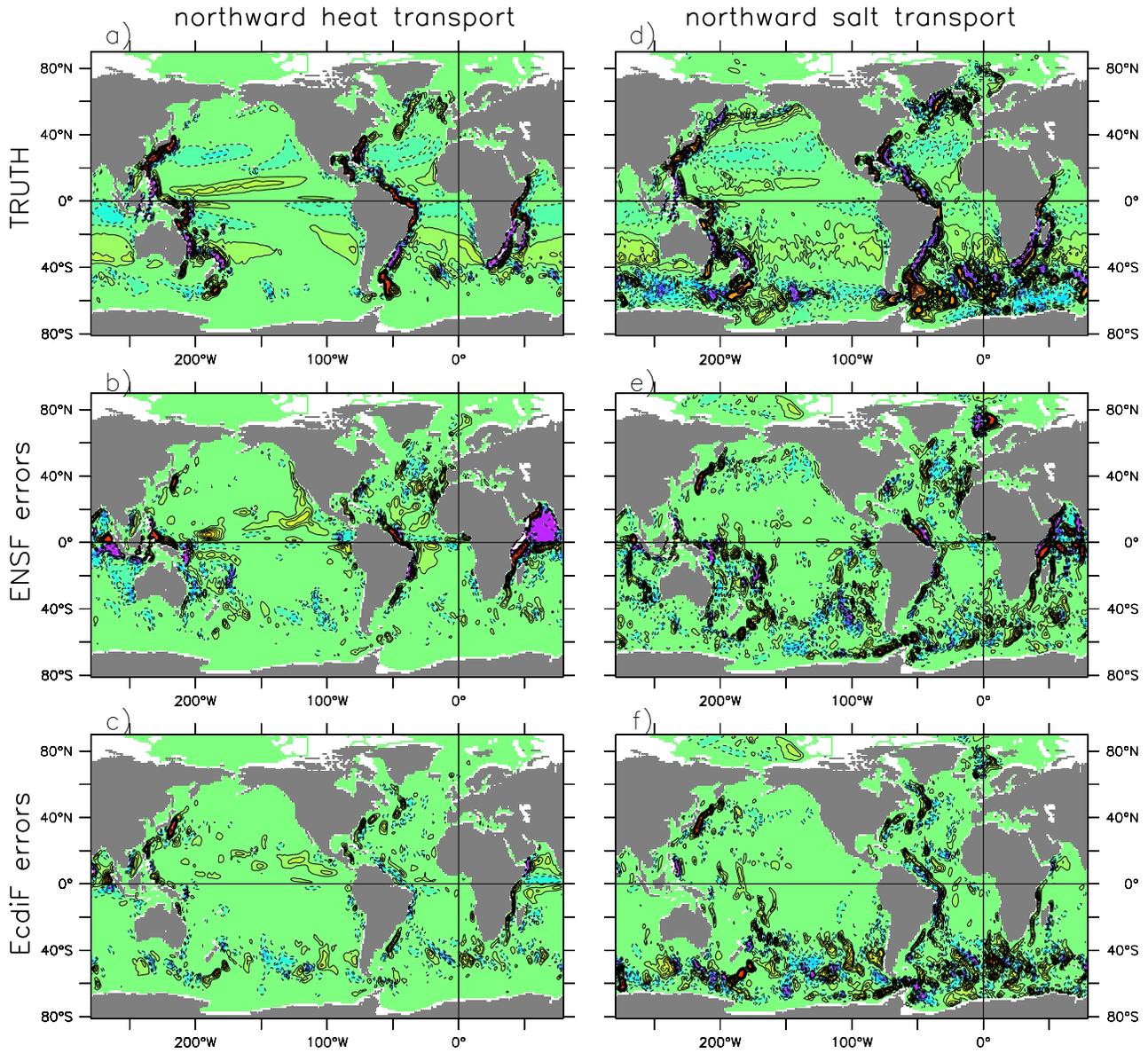


Figure 7: Time mean of the northward heat (left, *abc*) and salt (right, *def*) transports in TRUTH (top, *ad*) and the errors of these transports in ENSF (middle, *be*) and EcdiF (bottom, *cf*). The contour intervals are 0.2 PW ( $10^{15}$  Watts) for values between  $-2 - 2$  PW, otherwise 2 PW for *a*); 1 PW for values between  $-1 - 1$  PW, otherwise 0.1 PW for *b*) and *c*);  $0.1 \times 10^8$  kg/s for values between  $-1 - 1 \times 10^8$  kg/s, otherwise  $1 \times 10^8$  kg/s for *d*), *e*) and *f*).

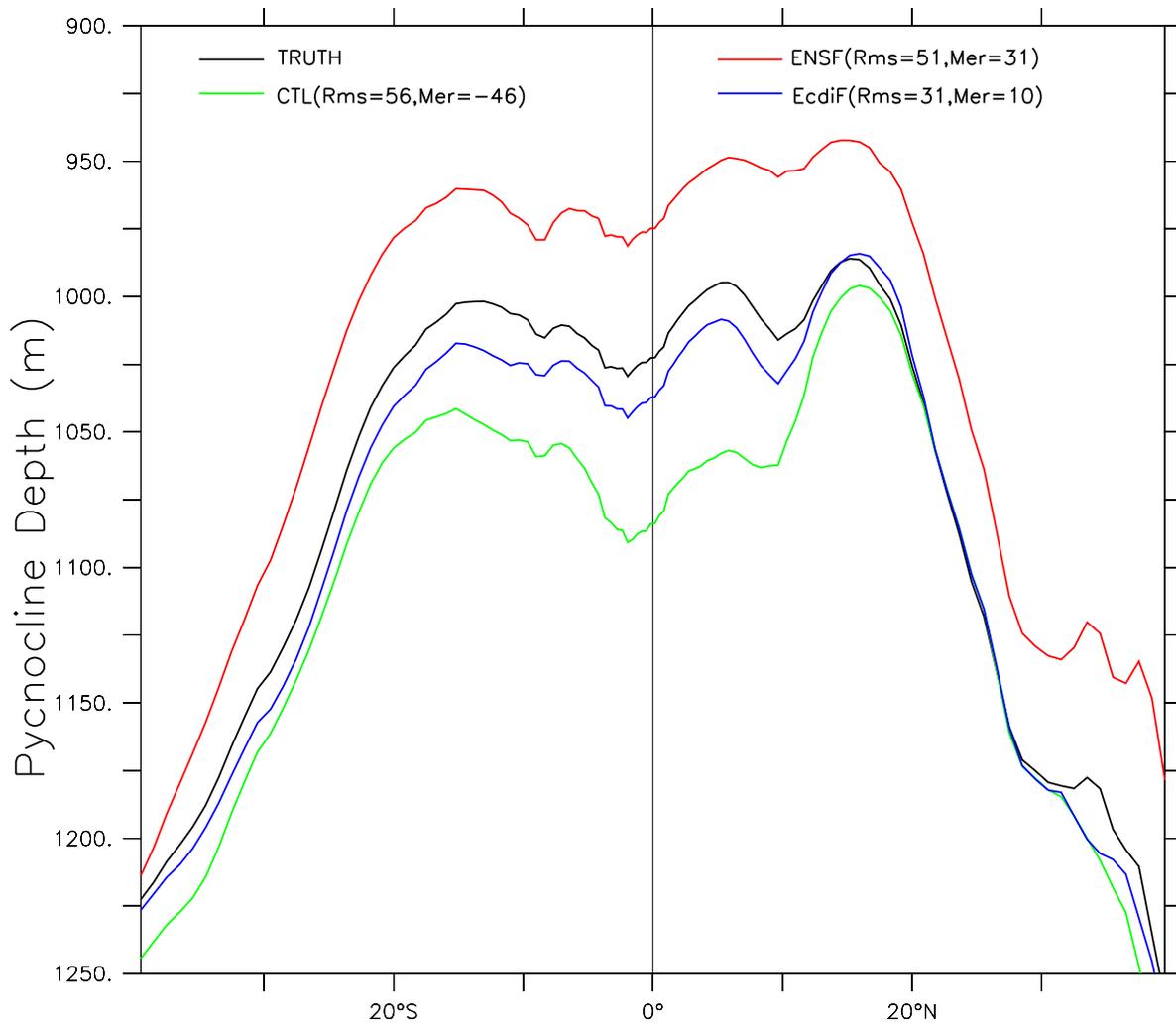


Figure 8: Variation of the zonal mean pycnocline depth with latitude in TRUTH (solid-black), CTL (green), ENSF (red) and EcdiF (blue). The corresponding root mean squared (Rms) error and mean error (Mer) are marked in parenthesis.

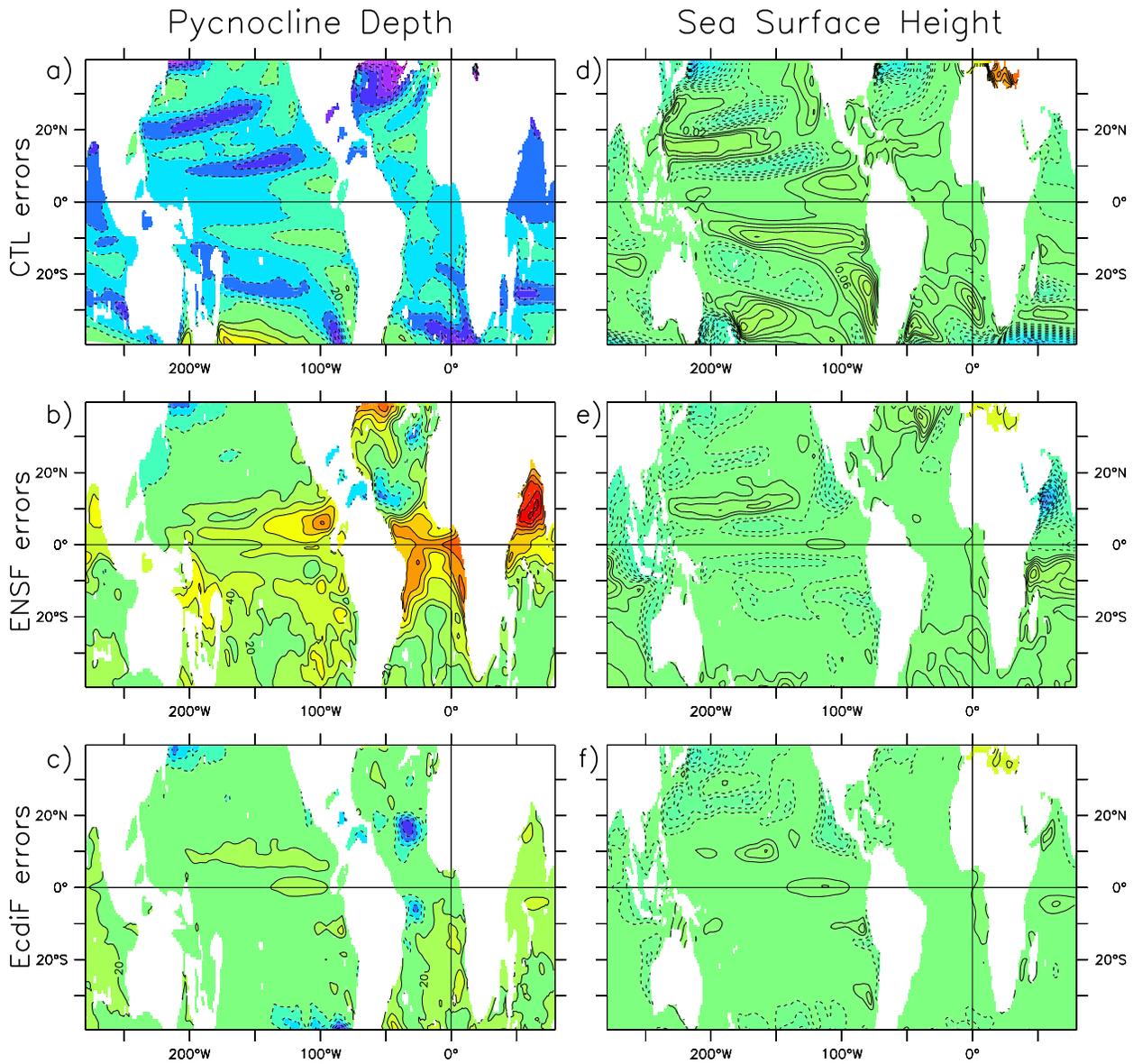


Figure 9: Time mean errors of the pycnocline depth (left) and the sea surface height (SSH) (right) produced by the CTL model simulation (top) and ENSF (middle) and EcdiF (bottom) assimilations. Contour interval is 20 m for pycnocline depth and 0.02 m for SSH.

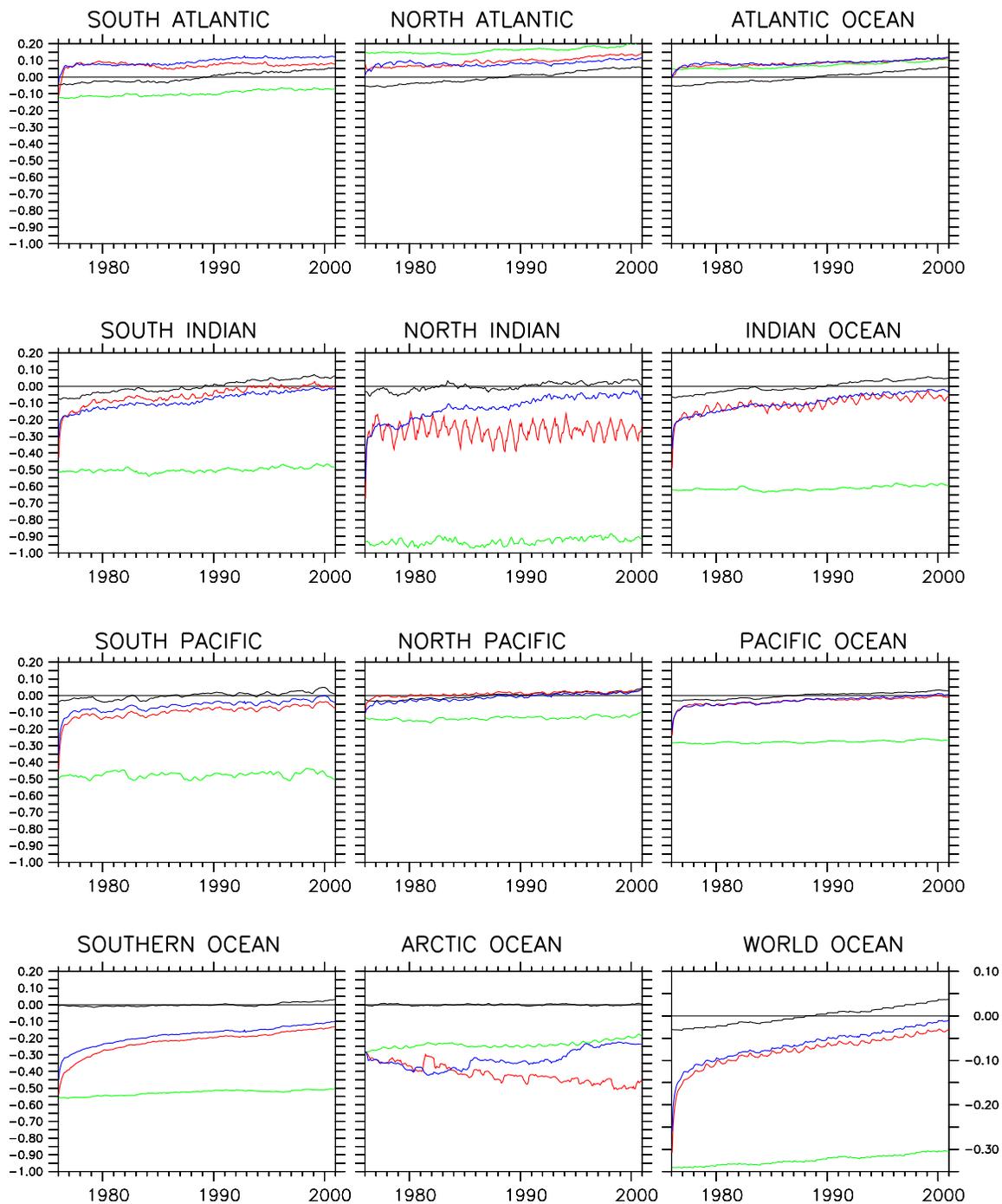


Figure 10: Time series of the top 4 km heat content (averaged temperature) in individual basins and the world ocean in CTL (black-dashed), ENSF (red), EcdiF (green) and TRUTH (black).

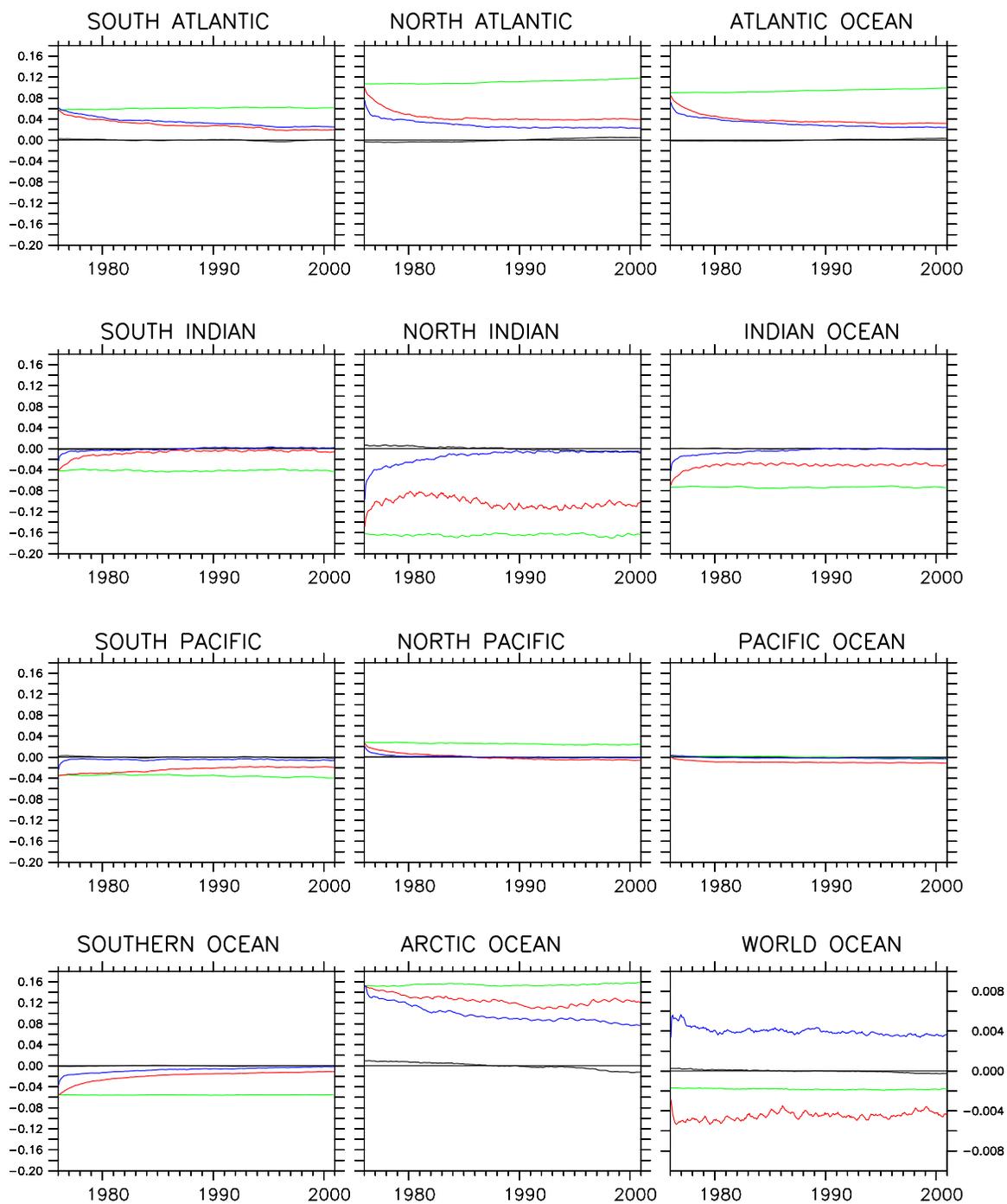


Figure 11: Same as Fig. 10 but for salinity.

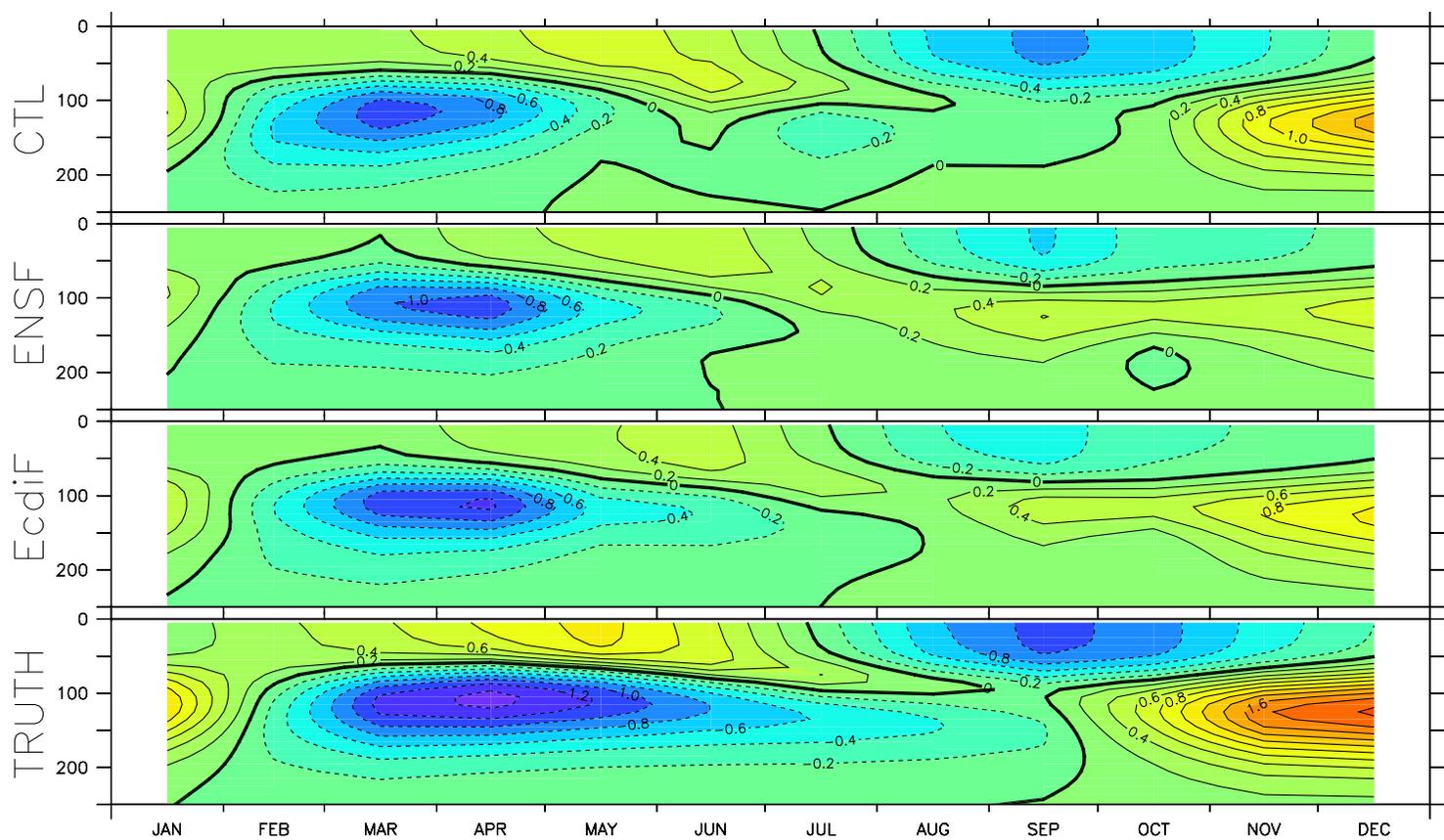


Figure 12: Annual cycle of the Nino3.4's temperature in CTL (top), ENSF (middle-upper), EcdiF (middle-lower) and TRUTH. The contour interval is  $0.2^{\circ}\text{C}$ .

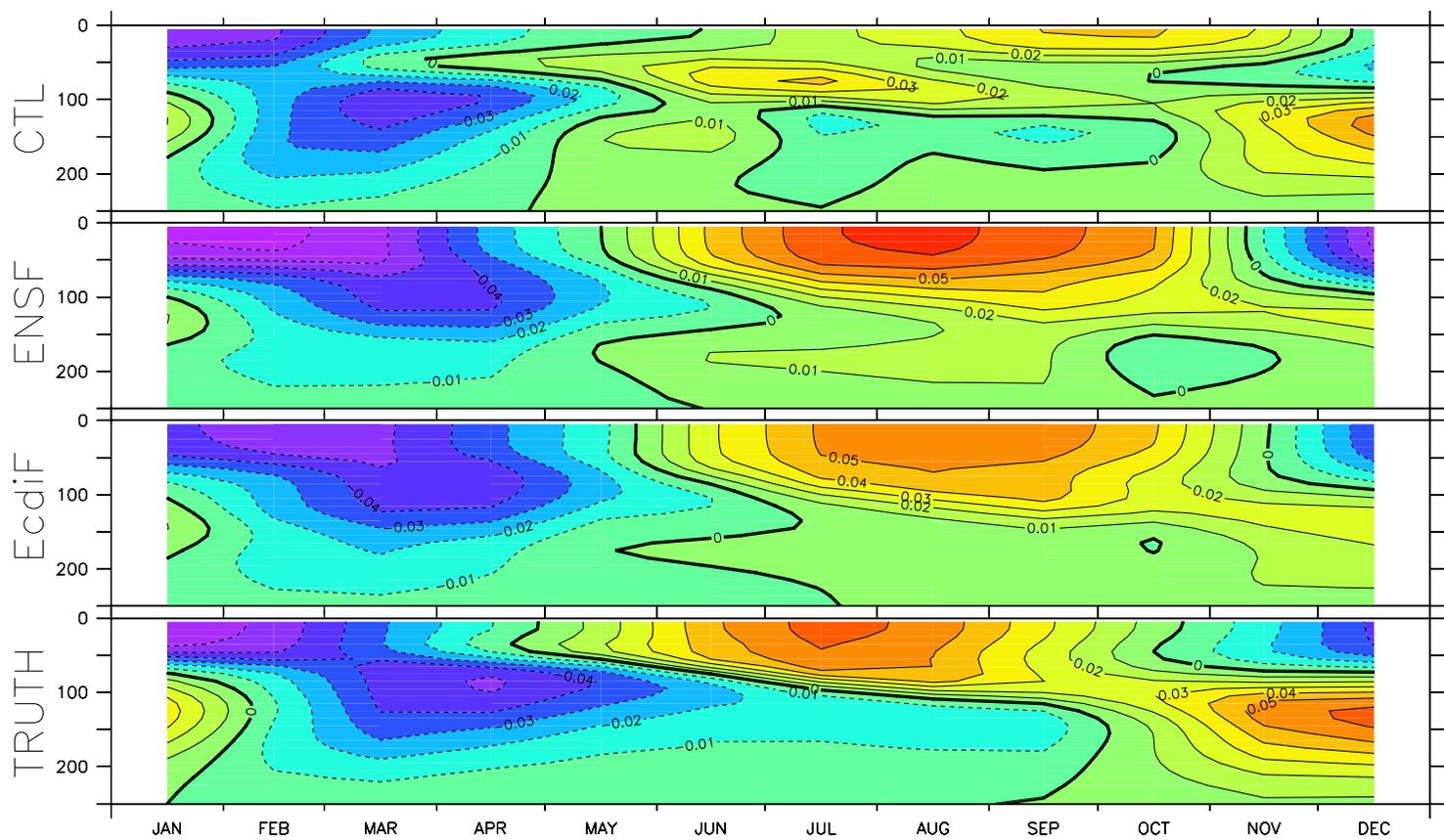


Figure 13: Same as Fig. 12 except for salinity and the contour interval is 0.01PSU.

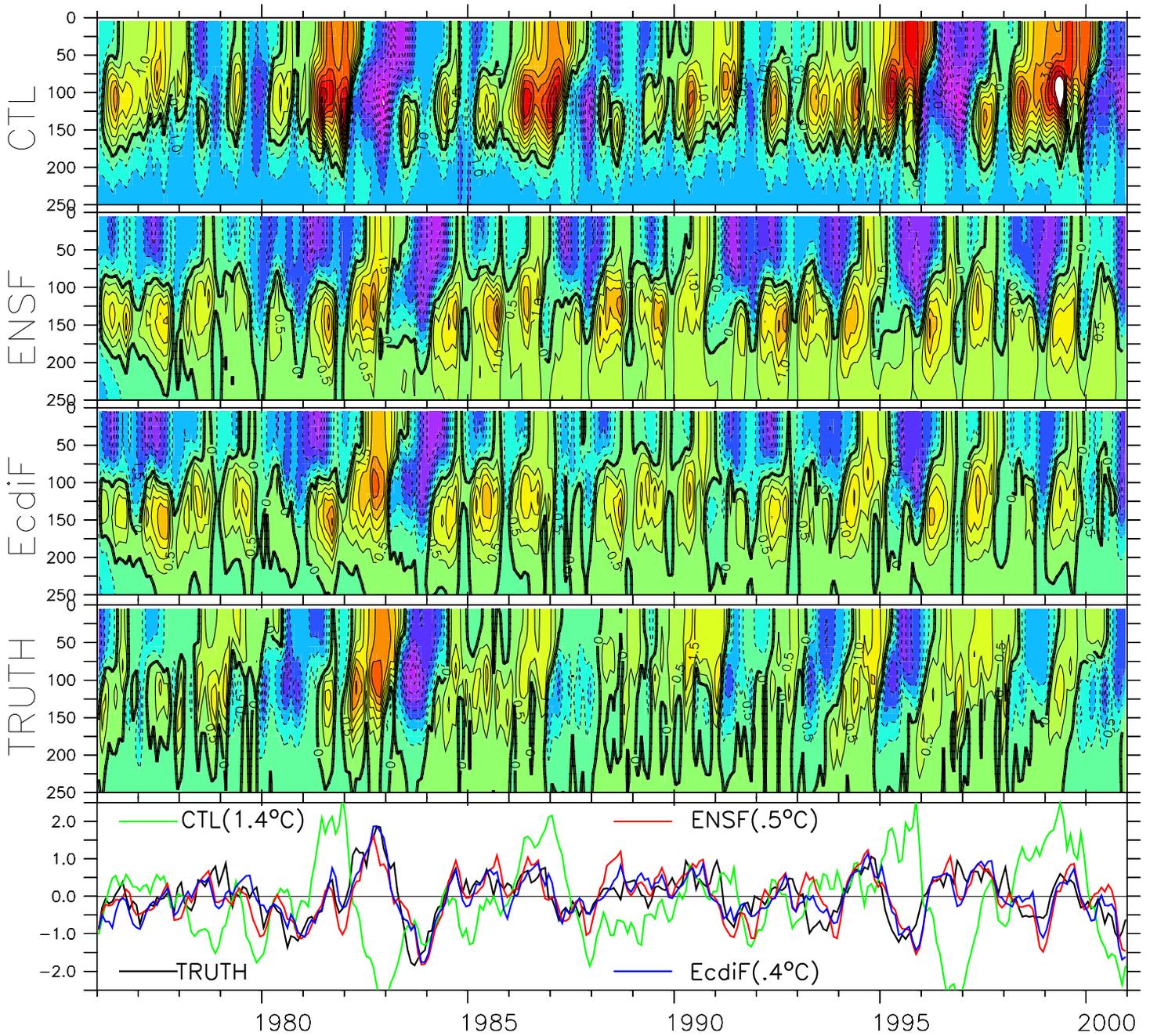


Figure 14: Time series of the anomalies of the Nino3.4's temperature in CTL (top), ENSF (middle-upper), EcdiF (middle) and TRUTH (middle-lower), and their vertical integrals (bottom). The contour interval is  $0.5^{\circ}\text{C}$  and the number in parenthesis is the corresponding Rms error.

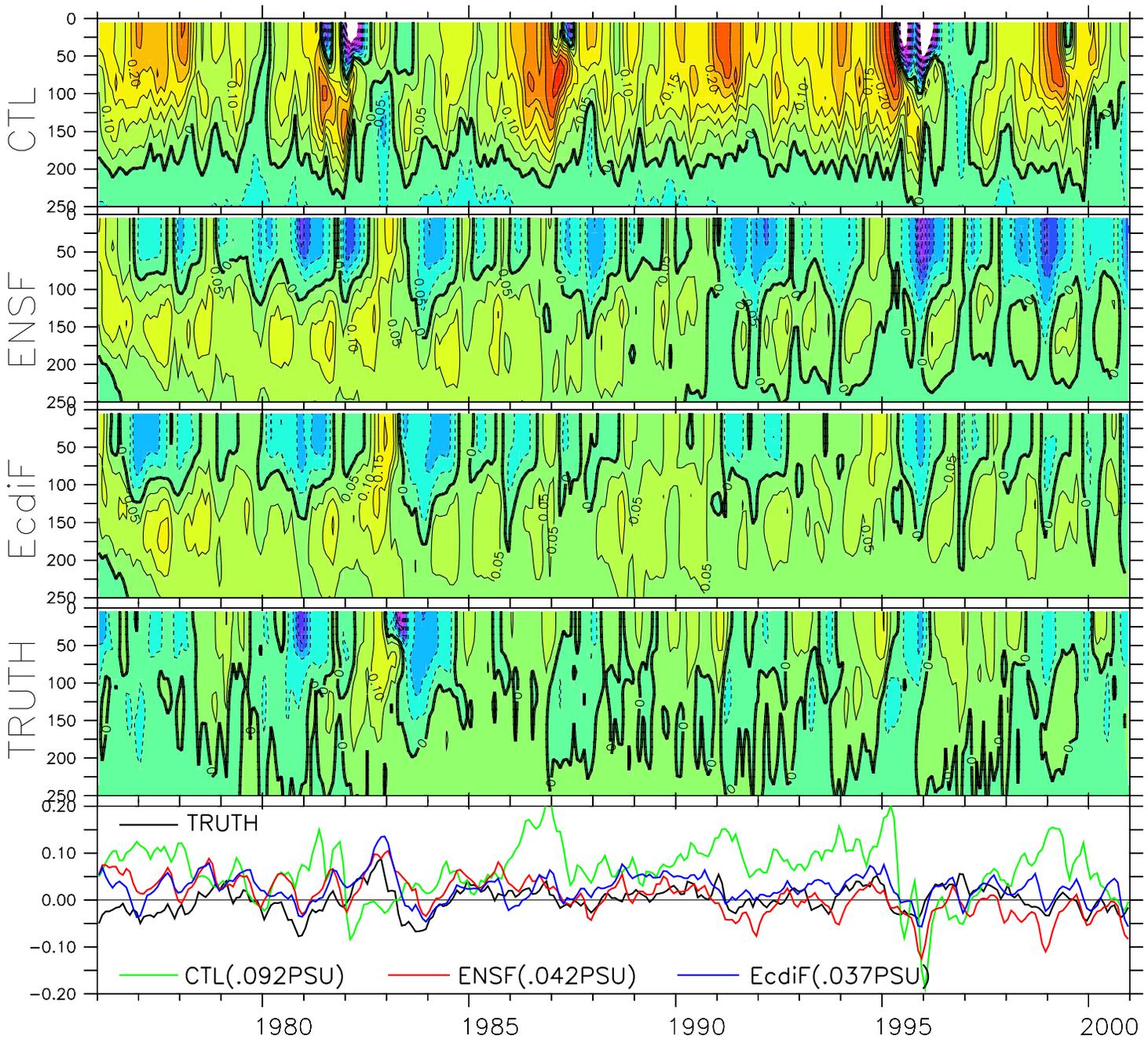


Figure 15: Same as Fig. 14 except for salinity and the contour interval is 0.05PSU in top 4 panels.

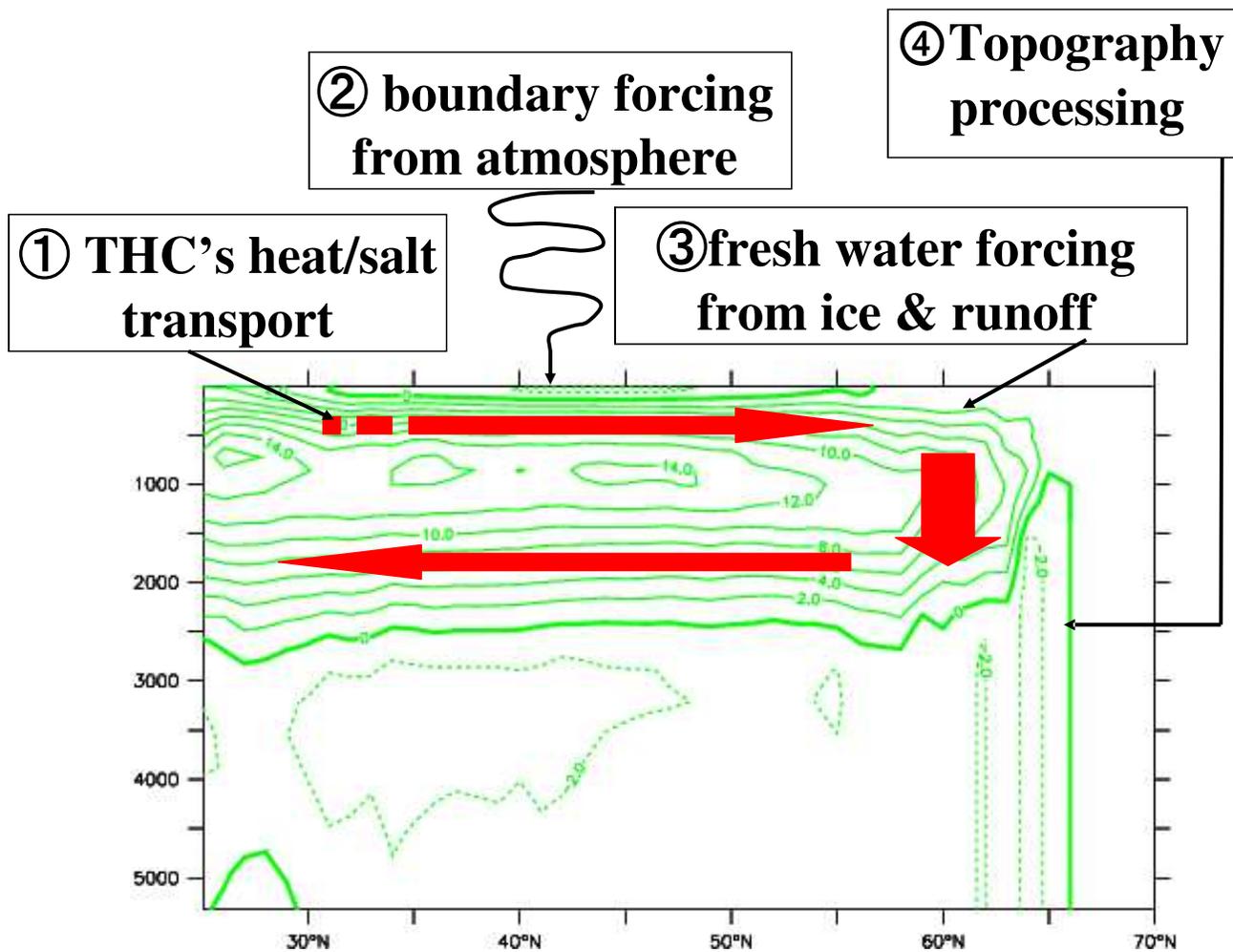


Figure 16: Schematic illustration of 4 factors influencing the North Atlantic meridional overturning circulation. The background green contours are the 25-year time mean of the overturning circulation stream function in TRUTH (the CM2.0 model simulation).

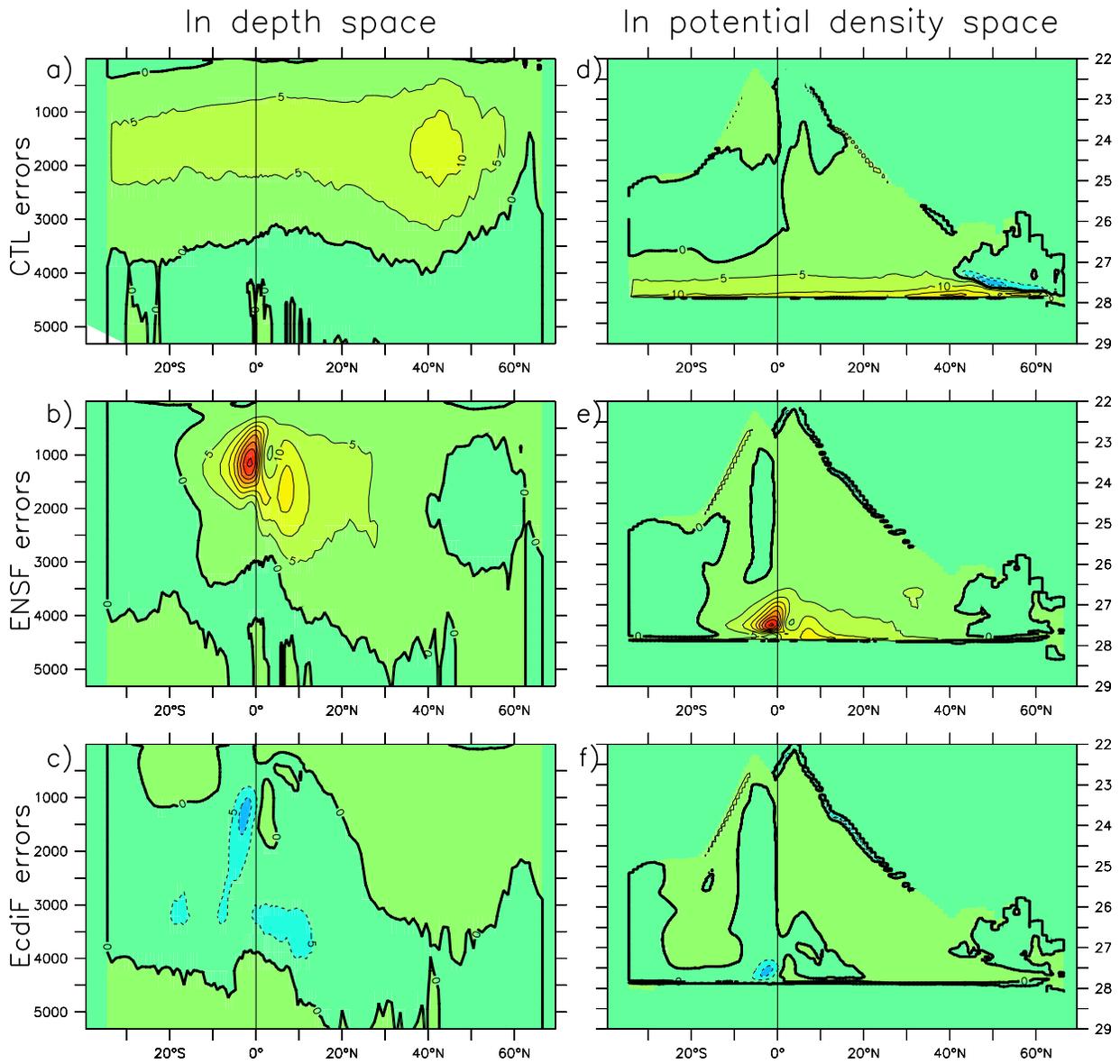


Figure 17: Time mean errors of the North Atlantic meridional overturning circulation stream function in depth space for CTL (a), ENSF (b), EcdiF (c) and potential density space for CTL (d), ENSF (e), EcdiF (f). The contour interval is 2 Sv.

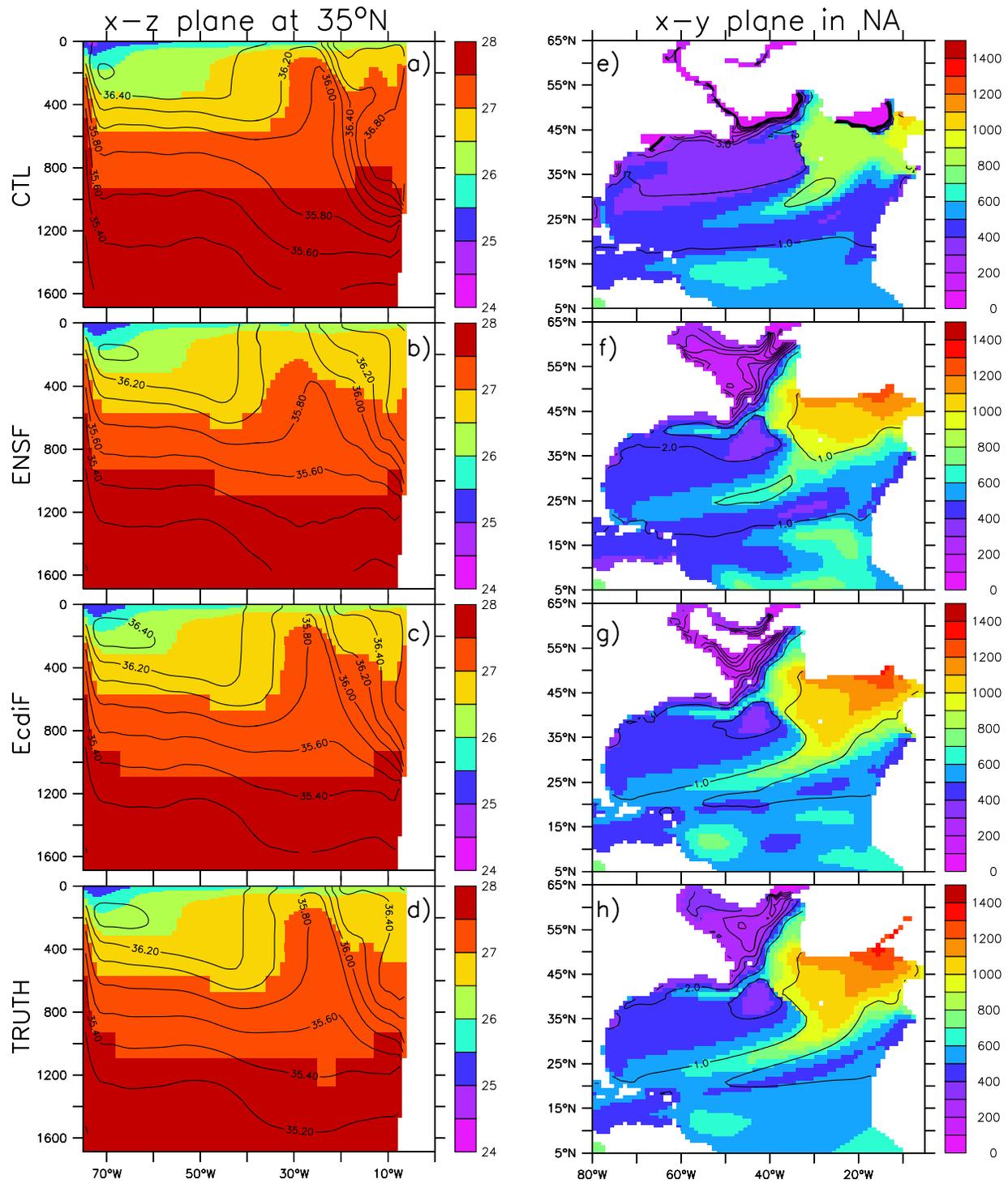


Figure 18: Upper-ocean potential density (color-shaded) and isohaline at 35°N (left, panels *abcd*) and the thickness ( $\zeta$ ) of intermediate water layer [between  $\sigma_0(27)$  and  $\sigma_0(27.5)$ ] (color-shaded) and the PV ( $f/\zeta$ ) distribution (contours) over the North Atlantic domain (right, panels *efgh*). The contour interval is 0.2 PSU/ $10^{-7}$  s $^{-1}$ m $^{-1}$  for panels *abcd/efgh*.

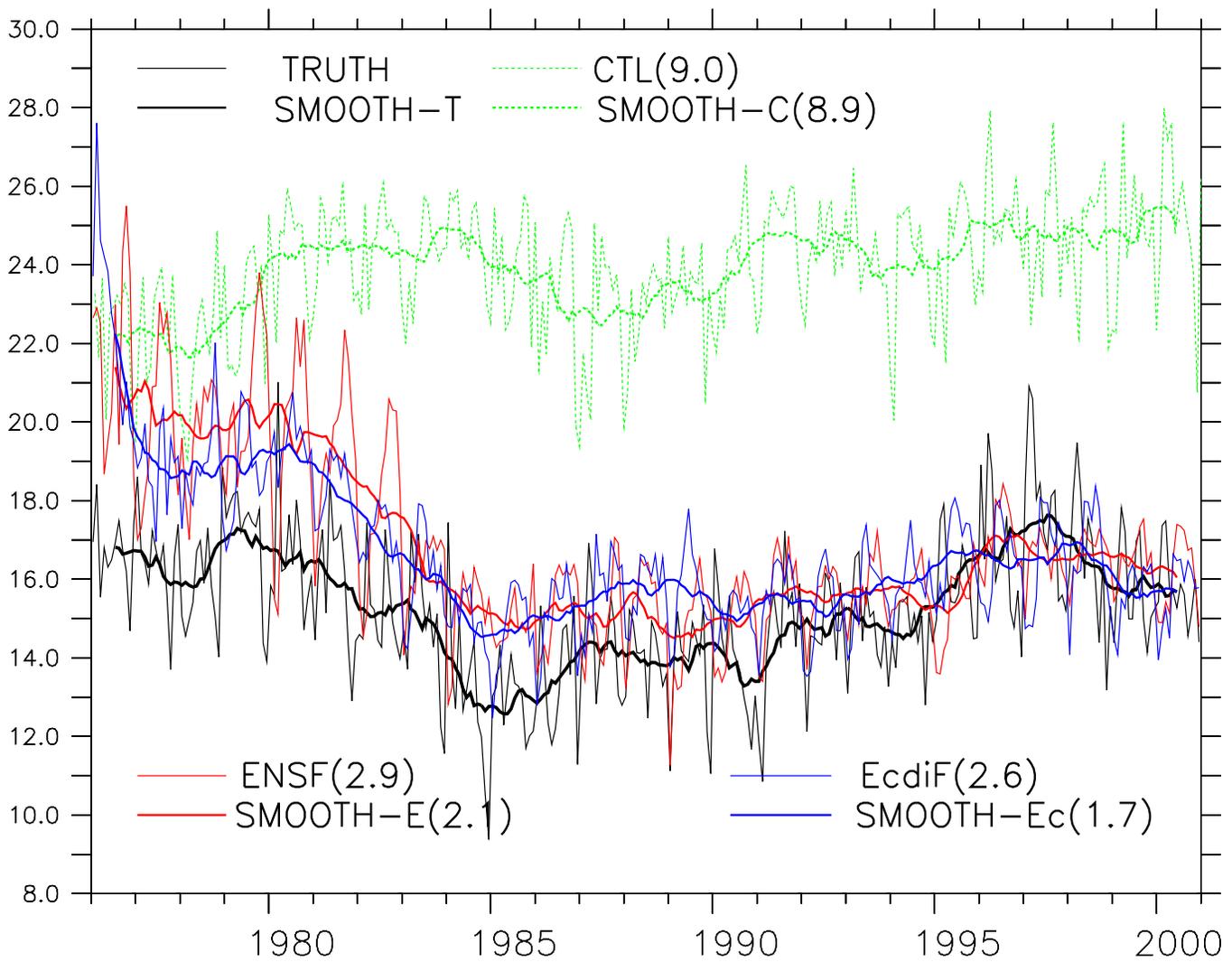


Figure 19: Time series of the maximum value of the North Atlantic meridional overturning circulation stream function over 40°N-65°N in CTL (green), ENSF (red), EcdiF (blue) and TRUTH (black). The thick lines are the corresponding 13-point running smooth for each case. The number in parenthesis is the corresponding Rms error.