Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation

Yiguo Wang, François Counillon, Ingo Bethke, Noel Keenlyside, Marc Bocquet, Mao-Lin Shen

Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation

Yiguo Wang[1,*], François Counillon[1], Ingo Bethke[2], Noel Keenlyside[1,3], Marc Bocquet[4], and Mao-lin Shen[3]

[1]Nansen Environmental and Remote Sensing Center, Bjerknes Centre for Climate Research, Bergen, Norway

[2]Uni Research Climate, Bjerknes Centre for Climate Research, Bergen, Norway

[3]Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway

[4]CEREA, Joint laboratory of École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

Abstract

Hydrographic profiles are crucial observational datasets for constraining ocean models and their vertical structure. In this study, we investigate a key implementation setup for optimising their assimilation into isopycnal ocean models. For this purpose, we use the Norwegian Climate Prediction Model (NorCPM), which is a fully-coupled climate prediction system based on the Norwegian Earth System Model and the ensemble Kalman filter. First, we revisit whether it is more accurate to assimilate observations in their original coordinate (z-level coordinate) or to transform them into isopycnal coordinates prior to assimilation. The analysis is performed with a single assimilation step using synthetic observations that mimic the characteristic properties of hydrographic profiles: varying vertical resolutions,

*Corresponding author (yiguo.wang@nersc.no)

profiles of only temperature and observations only in the top 1000 m. Assimilating profiles in their native coordinate (z-level coordinates) performs best because converting observations into isopycnal coordinates is strongly non-linear which reduces the efficiency of the assimilation. Secondly, we investigate how to set the horizontal localisation radius for our system. A radius that varies with latitude following a bimodal Gaussian function fits the system well. Thirdly, we estimate observation error, which consists of both instrumental error and representativeness error. In the proposed formulation only the instrumental error decreases with the number of observations during superobing, because the representativeness error is dominated by model limitation. Finally, we demonstrate the impact of assimilating hydrographic profiles from the observational EN4 dataset into NorCPM. An analysis of 10 years with monthly assimilation is performed with special focus on assessing the accuracy and the reliability of our analysis. The assimilation of hydrographic profiles into NorCPM is found to efficiently reduce the model bias and error, and the ensemble spread is found to be a reliable estimator for the forecast error in most regions.

Keywords: Data assimilation, EnKF, hydrographic profiles, isopycnal coordinate ocean models

1 Introduction

There is a high societal need for better understanding natural and anthropogenic driven climate variations. Current observations are too sparse and inhomogeneously distributed to provide a complete and dynamical picture of the climate and thus we rely on **data assimilation** (DA) to fuse scarce observations into dynamical models (Zhang et al., 2009, 2010; Brune et al., 2015; Laloyaux et al., 2016; Mochizuki et al., 2016). DA has been widely used to produce long-term reanalyses of the climate and to improve predictions ranging from short-term up to a decade (Meehl et al., 2009). While accurate reconstructions of the atmosphere have been available for over a decade (Kalnay et al., 1996; Dee et al., 2011), the paucity of observations in the ocean, in particular below the surface of the ocean, makes the long-term reconstruction of the ocean more challenging. Hydrographic profiles are the main source of available observations for monitoring the vertical structure of the ocean. They have been shown to be a crucial ingredient in constraining ocean models (Oke and Schiller, 2007; Zhang et al., 2009; Brune et al., 2015). In

particular, Zhang et al. (2009) showed that profile data can lead to significant improvement in reconstructing the thermohaline structure and that the Argo system (Roemmich et al., 2009), the first global float array for observing the subsurface ocean, is very useful for global oceanic climate studies.

Isopycnal coordinate models are a specific type of ocean models that are discretised vertically with potential densities (Bleck et al., 1992). They allow for an excellent conservation of water mass properties and have thus become popular among the Earth system modelling community, e.g. GISS-HYCOM (Sun and Bleck, 2006), FSU-HYCOM, ESM2G (Dunne et al., 2012), and the Norwegian Earth System Model (NorESM, Bentsen et al., 2013). However, hydrographic profiles are measured in geopotential height (or z-level coordinate) and their assimilation in such models becomes non-trivial because the model coordinate system varies with time and the equation of sea water non-linearly relates potential density to **potential temperature** (T), **salinity** (S) and pressure (Thacker and Esenkov, 2002; Xie and Zhu, 2010).

In this paper, we investigate three key practical aspects for the assimilation of hydrographic profiles into isopycnal ocean models: choice of innovation coordinate (z-level or isopycnal coordinates), localisation radius, and observation error settings. Here, we use the Norwegian Climate Prediction Model (NorCPM, Counillon et al., 2014), which combines the NorESM with the ensemble Kalman filter (EnKF, Evensen, 2003) for the purpose of climate predictions and reanalyses. A prototype version of NorCPM (Counillon et al., 2014, 2016; Wang et al., 2016) shows encouraging results in constraining the variability of the upper ocean heat content by assimilating sea surface temperature (SST). It is expected that complementing the system with the assimilation of T-S profiles will further improve its accuracy. Although the assimilation of T-S profiles would likely improve the climate prediction skill of the system, the current study focuses on the reanalysis capability of NorCPM.

This paper is organised as follows. The NorCPM system - namely the model system, the DA method and the DA implementation for hydrographic profiles - is described in section 2. In section 3, we compare the assimilation of hydrographic profiles observations when they are used in their native coordinate system as opposed to when they are transformed in isopycnal coordinate system. We then identify the localisation radius for our particular application in section 4. We estimate the observation error variance in section 5. Finally, in section 6, we

perform an analysis for 10 years based on the outcomes of section 3, 4 and 5, and assess the stability of our system regarding accuracy and reliability.

2 The Norwegian Climate Prediction Model

The NorCPM (Counillon et al., 2014) combines NorESM (Bentsen et al., 2013, presented in Section 2.1) and the EnKF (Evensen, 2003, presented in Section 2.2) for the purpose of seasonal-to-decadal climate predictions and long-term reanalyses. So far, DA has only been performed in the ocean part of NorESM, since the ocean is the compartment of the Earth system where predictability up to decadal time scales is expected (Meehl et al., 2009). The adjustment of the other compartments occurs dynamically during the system integration. The system combines uniquely a global isopycnal ocean model with an advanced error flow-dependent DA method.

2.1 The Norwegian Earth System Model

The NorESM (Bentsen et al., 2013) is a global fully-coupled system for climate simulations. It is based on the Community Earth System Model version 1.0.3 (CESM1, Vertenstein et al., 2012), a successor to the Community Climate System Model version 4 (CCSM4, Gent et al., 2011). As in the CESM1, the NorESM combines the Community Atmosphere Model (CAM4, Neale et al., 2010), the Community Land Model (CLM4, Oleson et al., 2010; Lawrence et al., 2011) and the Los Alamos sea ice model (CICE4, Gent et al., 2011; Holland et al., 2012) with the version 7 coupler (CPL7, Craig et al., 2012). The ocean component is an updated version (Bentsen et al., 2013) of the isopycnal coordinate ocean model MICOM (Bleck et al., 1992). With potential density as vertical coordinate, the model layer interfaces are good approximations to neutral surfaces and allow for minimising spurious mixing compared to other choices of vertical coordinates. This ensures an excellent conservation of water mass properties. The reference potential densities are selected to best represent characteristic water masses. Potential densities are referenced here to 2000 dbar to maximise neutrality of the isopycnal surfaces (McDougall and Jackett, 2005). When a layer's potential density falls outside the range of its reference densities in the water column, it becomes empty or massless. The model uses a bulk surface mixed layer that is divided into two layers with freely evolving density. The first isopycnal layer (below the mixed layer)

is not required to stay close to its prescribed reference potential density. A diapycnal diffusion scheme adjusts the isopycnal layer's potential density to its reference potential density when the two differ by exchanging water with adjacent layers. For further model details see Bentsen et al. (2013).

In this study, the atmosphere component CAM4 has a horizontal resolution of 1.9° at latitude and 2.5° at longitude. The ocean component MICOM and sea-ice component CICE4 are configured on a tripolar grid with a nominal resolution of 2° that is equatorward enhanced to 0.5° between 22° S and 22° N. The resolution measured by the squared-root of grid cell area is lower in subtropics where it is close to 200 km and higher towards the poles (e.g., 75 km at the North Pole). MICOM uses 2 layers with varying density representing the mixed layer and 51 isopycnal layers respecting the chosen reference potential densities in the range 1028.202–1037.800 kg m^{-3} with reference pressure set to 2000 dbar. NorESM was initialised with data from the Polar Hydrographic Climatology version 3.0 (PHC 3.0, Steele et al., 2001) and then spun up for 1500 years using constant preindustrial external forcing. The January states from the last year of the spin-up simulation were used to initialise a 30-member historical simulation ensemble, which was integrated from 1850 to 2010 using transient external forcing from Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al., 2012), allowing the internal spread to become saturated by 1980 on most time scales.

2.2 Data assimilation method

The traditional EnKF (Evensen, 2003) is a recursive DA method that consists of a Monte Carlo integration of the model and that allows for a flow-dependent error estimation and a variance-minimising update. In this paper, a deterministic variant of the traditional EnKF (DEnKF, Sakov and Oke, 2008) is used. The DEnKF updates the ensemble perturbations around the updated mean using an expansion in the expected correction to the forecast. This yields an approximate but deterministic form of the traditional stochastic EnKF.

Let the ensemble of model states $\mathbf{X}_\mathrm{f} = [\mathbf{x}_\mathrm{f}^1, \mathbf{x}_\mathrm{f}^2, ..., \mathbf{x}_\mathrm{f}^m] \in \mathbb{R}^{n \times m}$, the ensemble mean be $\overline{\mathbf{x}}_\mathrm{f} \in \mathbb{R}^n$ and the ensemble anomalies or perturbations $\mathbf{A}_\mathrm{f} = \mathbf{X}_\mathrm{f} - \overline{\mathbf{x}}_\mathrm{f}\mathbf{1}_m \in \mathbb{R}^{n \times m}$, where the subscript 'f' denotes forecast, n is the size of model states, m is the ensemble size, and $\mathbf{1}_m = [1, 1, ..., 1] \in$

$\mathbb{R}^{1\times m}$. The DEnKF update can be written as follows:

$$\overline{\mathbf{x}}_{\mathrm{a}} = \overline{\mathbf{x}}_{\mathrm{f}} + \mathbf{K}\left(\mathbf{y} - H\left[\overline{\mathbf{x}}_{\mathrm{f}}\right]\right), \tag{1}$$

$$\mathbf{A}_{\mathrm{a}} = \mathbf{A}_{\mathrm{f}} - \frac{1}{2}\mathbf{K}\mathbf{H}\mathbf{A}_{\mathrm{f}}, \tag{2}$$

$$\mathbf{X}_{\mathrm{a}} = \overline{\mathbf{x}}_{\mathrm{a}}\mathbf{1}_m + \mathbf{A}_{\mathrm{a}}, \tag{3}$$

where the subscript 'a' denotes analysis, $\mathbf{y}$ is the observation vector, H is the observation operator that maps model states to the observation space and $\mathbf{H}$ is the tangent linear operator of H. The Kalman gain $\mathbf{K}$ is defined by

$$\mathbf{K} = \mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}, \tag{4}$$

where $\mathbf{P}_{\mathrm{f}}$ is the forecast error covariance matrix estimated by the ensemble perturbations:

$$\mathbf{P}_{\mathrm{f}} = \frac{1}{m-1}\mathbf{A}_{\mathrm{f}}\mathbf{A}_{\mathrm{f}}^{\mathrm{T}} \tag{5}$$

and $\mathbf{R}$ is the observation error covariance matrix.

We do not expect that using the traditional EnKF changes any of the conclusions of this paper. Therefore, in the following we do not distinguish the DEnKF from the traditional EnKF (hereafter the **EnKF**).

2.3 Hydrographic profiles

Hydrographic (T-S) profile observations are downloaded from the EN4 dataset of global quality controlled ocean T and S profiles (`http://www.metoffice.gov.uk/hadobs/en4/download-en4-1-1.html`, Gouretski and Reseghetti, 2010; Good et al., 2013). The EN4 consists of data from all types of ocean profiling instruments, including from the World Ocean Database, the Arctic Synoptic Basin Wide Oceanography project, the Global Temperature and Salinity Profile Program, and Argo. The EN4 data is available from 1900 to present and contains quality flag. Here, only observations tagged with the quality flag of '1' (excellent quality) are assimilated. However, the uncertainty of observed T-S profile is not provided with the dataset. We assume the observation

error covariance matrix, $\mathbf{R}$, to be diagonal (sections 3 and 5), implying that the observation errors are assumed to be uncorrelated.

2.4 Assimilation of hydrographic profiles: practical implementation

Hydrographic profiles from observations are in z-level coordinates but the model variables are in isopycnal ones. Thus, to calculate the innovation vector we must either interpolate the observations to isopycnal coordinates or the model data to z-level coordinates. Previous studies with regional isopycnal models (Thacker and Esenkov, 2002; Xie and Zhu, 2010) found it best to interpolate the profile observations to the model coordinates. In addition, they recommended updating S and isopycnal layer thickness, but diagnosing T from the equation of state of seawater below the mixed layer. In Srinivasan et al. (2011), it was found that updating both T and S performs best and Wang et al. (2016) found that the artificial caballing resulting from the update of both T and S is negligible. Here, we revisit this approach in a global domain and using an idealised framework. Unlike in Xie and Zhu (2010), we do not diagnose T or S and propose an improved formulation of the observation error for assimilation of T-S profiles in isopycnal coordinates.

In MICOM, T and S are defined at depths (z), which correspond to the centre of the isopycnal layers, that can be calculated from layer thicknesses (DP). In the first scheme (EnKF-z), we interpolate the discrete T-S profile estimates from the model to the geopotential depths of the observations. The innovation vector is constructed in z-level coordinates (i.e., $\mathbf{y} - H\left[\overline{\mathbf{x}}_\mathrm{f}\right]$) and the observation error covariance matrix $\mathbf{R}$ is diagonal with depth dependent entries (i.e. in z-level coordinates). Note that using a linear interpolation and/or a cubic Hermite spline interpolation for the observation operator H lead to very similar results (not shown). We choose a cubic Hermite spline interpolation as the operator H, because it produces a smooth profile that matches well high resolution observations without overshooting values. The state analysis minimisation problem of the EnKF-z can be formulated as follows:

$$\overline{\mathbf{x}}_\mathrm{a} = \mathrm{argmin}\ J_\mathrm{z}(\mathbf{x}), \tag{6}$$

$$J_\mathrm{z}(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}}_\mathrm{f})^\mathrm{T}\mathbf{P}_\mathrm{f}^{-1}(\mathbf{x} - \overline{\mathbf{x}}_\mathrm{f}) + (\mathbf{y} - H[\overline{\mathbf{x}}_\mathrm{f}])^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y} - H[\overline{\mathbf{x}}_\mathrm{f}]). \tag{7}$$

In the second scheme (EnKF-ρ), we first estimate potential density profile (ρ) from observed T-S profiles via the equation of state for seawater (ICES et al., 1981). The unstable profiles (i.e., when density is not increasing with depth) are corrected locally using linear fitting between the neighbouring stably stratified parts of the profile. The mixed layer depth is estimated based on the density gradient criterion ($\partial\rho/\partial z > 0.01$ kg m^{-4}, Table 1 in Thomson and Fine, 2003). Other diagnostics have been tested but were found to perform slightly poorer for our application. As in previous studies (Thacker and Esenkov, 2002; Xie and Zhu, 2010), the interface depths of isopycnal layers are defined as the mid-distance between two reference densities (i.e., $(\rho_i^{\mathrm{ref}} + \rho_{i+1}^{\mathrm{ref}})/2$). From the initial profiles of observed T and S, we can then construct the T-S profiles in isopycnal coordinates with corresponding layer thicknesses (Thacker and Esenkov, 2002). Figure 1 shows an example of converting z-level coordinate observations into isopycnal coordinates. Horizontal dashed lines represent the interfaces of isopycnal layers that are used to create layer-averaged T and S observations (i.e., red dots in top right and bottom panels) and layer thickness 'observations' (i.e., the distances between two adjacent dashed lines and red dots in the bottom right panel). The top horizontal dashed line is the base of the first two layers that represent the mixed layer in MICOM. In the EnKF-ρ scheme, layer thickness 'observations' are first assimilated to update DP. Then T and S observations formulated in isopycnal coordinates are assimilated to update T and S. The state analysis minimisation problem of the EnKF-ρ can be formulated as follows:

$$\overline{\mathbf{x}}_{\mathrm{a}} = \mathrm{argmin}\ J_\rho(\mathbf{x}), \tag{8}$$

$$J_\rho(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}}_{\mathrm{f}})^{\mathrm{T}} \mathbf{P}_{\mathrm{f}}^{-1} (\mathbf{x} - \overline{\mathbf{x}}_{\mathrm{f}}) + (F_\rho[\mathbf{y}] - \overline{\mathbf{x}}_{\mathrm{f}})^{\mathrm{T}} {\mathbf{R}_\rho}^{-1} (F_\rho[\mathbf{y}] - \overline{\mathbf{x}}_{\mathrm{f}}), \tag{9}$$

where F_ρ stands for the operator converting z-level coordinate observations into isopycnal coordinates, $\mathbf{x}$ stands for DP when assimilating layer thickness 'observations' and consists of T and S when assimilating T and S observations.

In EnKF-ρ, the observations of T-S profiles are "superobed" so that there is at most one observation per isopycnal layer (a super observation); the new super observation is set equal to the average of observations falling within the layer and the observation uncertainty reduced accordingly (Sakov et al., 2012). Note that in Xie and Zhu (2010), it was assumed that $\mathbf{R}_\rho$ is

the same as $\mathbf{R}$ for T and S.

Estimating the observation error for observations of layer thickness is more challenging. In Xie and Zhu (2010), it was estimated by the standard deviation of potential density. Here, the observation error of layer thickness is estimated by a stochastic process. Hence, an ensemble of profiles $\{\mathbf{y}^i\}_{i=1,30}$ is perturbed with white noises (drawn from the observation distribution $\mathcal{N}(\mathbf{y}, \mathbf{R})$) and is used to estimate an ensemble of layer thicknesses $\{\mathbf{DP}^i_{\mathrm{o}}\}_{i=1,30}$. Then the observation error of layer thickness is defined by the spread of the ensemble $\{\mathbf{DP}^i_{\mathrm{o}}\}_{i=1,30}$ [1]. An example of observation errors of layer thickness is shown in the bottom right panel of Fig. 1.

The rest of the practical implementation follows that of Counillon et al. (2014) and Wang et al. (2016). Horizontal superobing is employed, meaning that profiles at the same grid cell are combined and its accuracy adjusted (superobing, Sakov et al., 2012). This technique reduces the number of observations and its refinement is discussed in section 5. The ensemble spread is sustained by using a moderation technique and a pre-screening method (Sakov et al., 2012). The moderation technique consists of increasing observation error variance (here by a factor of 2) for the update of the ensemble anomalies while the original observation error variance is used for the update of the ensemble mean. In the pre-screening method, the observation error is inflated so that the analysis remains within two standard deviations of the forecast error from the ensemble mean of the forecasts. It should also be noted that the DEnKF overestimates the analysed error covariance by adding a semi-definite positive term to the theoretical error covariance given by the Kalman filter, which reduces the amount of inflation needed. We use the **local analysis framework** (Evensen, 2003), meaning that the analysis is carried for each local domain. The observation errors are weighted by a distance-dependent localisation function (Gaspari and Cohn, 1999; Hamill et al., 2001; Sakov and Bertino, 2010) to avoid discontinuities at the edge of the local domain. We identify the localisation radius that best fits our system in section 4.

3 Comparing EnKF-z and EnKF-ρ

It would be too costly to test the two schemes (EnKF-z and EnKF-ρ) with the full system (assimilation and model integration) and for different observational configurations. Therefore, we

[1] It was found that the error estimate converges when sampled from an ensemble of at least 30 members (not shown).

only perform a single assimilation with the two schemes at an arbitrary time (January 1980), and identify the scheme that minimises the error, when assimilating a set of idealised observations perturbed about the truth. Each profile is updated by only its observations from the same water column (i.e., single column localisation). The verification is based on about 15000 profiles updated for January 1980. We have repeated the performance for July 1980 and come with very similar results (not shown). The accuracy of the assimilation should be representative of the performance of both schemes in forecast mode. The discrepancies between the two schemes solely differs in the estimation of the linear weights based on a given prior and observations (i.e. the weight of the transform matrix; $\mathbf{X}_5$ in Evensen, 2003).

In this section, the forecast ensemble $\mathbf{X}_\mathrm{f}$ consists of T, S and DP profiles from 30 ensemble members of NorESM in January 1980 (an arbitrary time). The EN4 objective analysis (42 z-levels) in January 1980 is defined as the truth, $\mathbf{x}_\mathrm{t}$, and the synthetic observations are drawn from the distribution $\mathcal{N}(\mathbf{x}_\mathrm{t}, \mathbf{R})$. The observation errors for T and S are set as a function of depth from Stammer et al. (2002) and Xie and Zhu (2010), which is based on Levitus et al. (1994a,b).

In order to have a broader overview of the two schemes, we test them for several idealised observational network configurations that mimic the different characteristics of real observations:

- Hydrographic profiles have very different vertical resolutions, depending on the instrument and the objective. Thus in the first case, we consider three different vertical resolutions. The original dataset (EN4 objective analysis) from which the synthetic datasets are constructed is provided at 42 standard depth levels with a higher discretisation near the surface. The low resolution uses only half of the original levels of the truth (42 z-levels) and is referred to as **21 z-levels**; the medium resolution (referred to as **42 z-levels**) uses all z-levels of the truth, and the high resolution has 210 z-levels (referred to as **210 z-levels**) that are computed from the truth using cubic Hermite spline interpolation.

- Many instruments measure only temperature (e.g., the Expendable Bathythermograph, XBT). Thus in the second case, we assume that only observations of T are available and the experiment is referred to as **T only**. Since the EnKF-ρ relies on the profile of S to estimate the potential density profile, the profile of S is taken from the climatology. The monthly climatology is constructed from the EN4 objective analysis from 1980 to 1999.

The uncertainty of the climatology profile is estimated by the sum of the observation error variance for the observed profile (Stammer et al., 2002; Xie and Zhu, 2010) and the time variability of the EN4 objective analysis. Although the assimilation of climatology S profiles is not necessary with EnKF-z, we assimilate it to ensure that the two schemes are directly comparable.

- Many instruments are capable of measuring only the upper ocean (e.g., Argo and XBT). Thus in the third case, we consider observations of T and S only in the upper 1000 m (referred to as $<$ **1000 m**).

Experiment labels distinguish the vertical coordinate system used for assimilation. For example, the experiments assimilating the **42 z-level** observations in the EnKF-z and EnKF-ρ are referred to as **EnKF-z (42 z-levels)** and **EnKF-ρ (42 z-levels)**, respectively.

Figure 2 shows the maps of differences between vertical root mean squared errors (RMSEs) for EnKF-ρ (42 z-levels) and EnKF-z (42 z-levels). For T, the EnKF-z outperforms EnKF-ρ over most of the oceans, with a global average differences = +0.16 °C. For S, it is unclear which scheme performs best. EnKF-z performs better than EnKF-ρ in the North Atlantic, North Pacific and Arctic, while EnKF-ρ seems to perform better than EnKF-z in other regions.

Figure 3 shows the horizontal area-weighted RMSEs (appendix A) for T and S for different synthetic networks. The accuracy of the forecast is added to visualise the vertical impact of assimilation with depth. The upper panels show the sensitivity studies on the vertical resolution. Both schemes show a pronounced reduction of RMSE from the forecast. For all three observation networks, EnKF-z outperforms EnKF-ρ in almost every depth for T, with larger differences in the subsurface and deep oceans. The reduction of errors with increasing resolution is more pronounced with EnKF-z than with EnKF-ρ, except for the top 200 m. For S, EnKF-z outperforms EnKF-ρ in the subsurface and deep oceans, but leads to similar RMSEs to EnKF-ρ at intermediate depth.

When only T is observed (experiment **T only**, Fig. 3 lower panels), both schemes show large improvements both for T and S. We find that the dataset **T only** leads to similar results to the dataset **42 z-levels** (T and S observations). The benefit of DA for S in the upper 1000 m is mostly due to S climatology and the benefit of DA for S below 1000 m is due to T observations.

It is mostly because the water properties in the deeper ocean are stable and the T-S correlation is high. The impact of T observations propagates efficiently to S via their covariance. Again, it is found that EnKF-z outperforms EnKF-ρ.

When observations are limited to the upper 1000 m (Fig. 3 lower panels), we notice that assimilation yields a weak degradation between 1500-2500 m and between 4000-5500 m. This degradation is likely due to spurious correlations and would suggest that vertical localisation should be preferable. However, the degradation is small and not visible when deep observations are available. Furthermore, using the vertical localisation has also drawbacks with respect to mass conservation. In isopycnal coordinates, it is important to ensure that the sum of layer thicknesses (defined by differential pressures for isopycnal layers) matches the bottom pressure, which is satisfied with a linear analysis such as the EnKF. Applying a postprocessing to handle their mismatch induces a drift (Wang et al., 2016) while using the full depth covariance performs well with 30 ensemble members (Counillon et al., 2014, 2016). Overall, EnKF-z outperforms EnKF-ρ once again.

The main conclusion of this section is that both schemes are found to be quite successful for assimilating hydrographic profiles. Although there are regional differences in performance, EnKF-z performs overall better than EnKF-ρ. In Xie and Zhu (2010), two schemes (i.e. EXP1A and EXP2T) are compared. They are similar to EnKF-z and EnKF-ρ but there are some differences between EXP2T and EnKF-ρ. In EXP2T of Xie and Zhu (2010), they updated separately T and S with their corresponding observations. Below the mixed layer, they diagnosed T from the updated S in order to preserve the density of the layer. In EnKF-ρ, we update simultaneously both variables and refine the formulation of the observation errors of T, S and DP in EnKF-ρ. Their work is also based on the ensemble optimal interpolation and tested for a regional system of the Pacific Ocean. They found that assimilating observations in isopycnal coordinates performed best for both T and S. In that region, we find that EnKF-z performs better than EnKF-ρ for T but both schemes lead to similar scores for S.

The relative performance of the two schemes is rooted to the non-linearity of the observation operator H and of the conversion operator F_ρ (section 2.4). While the non-linearity of H only depends on the choice of the vertical interpolation (that can also be linear), the operator F_ρ is strongly non-linear and non-unique as we try to construct three state variables (T, S and DP)

from the observed T-S profiles (Thacker and Esenkov, 2002; Xie and Zhu, 2010). The resulting isopycnal profiles are sensitive to key choices, i.e. how to estimate the mixed layer depth and how to extract layer thicknesses from the density profile. The non-uniqueness of the solution adds uncertainties during the minimisation process, which leads to a reduced efficiency of the assimilation algorithm and a suboptimal minimisation of the cost function (Eq. 9).

4 Horizontal localisation radius

The implementation of localisation is one of the critical setting of ensemble DA methods. As the ensemble size is often too small to span the dimension of the whole model subspace, an ad-hoc approach (known as **localisation**, Hamill et al., 2001) is used to limit the influence of observations within a given radius of influence. Localisation discards correlations at longer distance, considering them spurious. However, localisation may introduce discontinuities at the edge of the local domain. To counteract this effect, it is common to taper the precision matrix or the weights to the observations with a smooth distance-dependent (typically quasi-Gaussian) localisation function. The localisation radius needs to be tuned for each system, because it varies with the model space, ensemble size and observation spatial distribution.

In this section, we investigate the choice of the horizontal localisation radius for NorCPM. A common and simple way to estimate this radius is to assume that the errors have isotropic and Gaussian correlations (Hamill et al., 2001). Under this hypothesis, the localisation radius ensures that the amplitude of the spurious correlations remains smaller than the meaningful part (large signal-to-noise ratio). In this paper the localisation radius is set equal to the correlation length scale. However, note the localisation radius and correlation length do not have to coincide, although they should intuitively be related. It was recently shown on theoretical grounds that both were closely related in geophysical fluids dominated by advection or convection and should vary concurrently in time and space (Bocquet, 2016).

Here the horizontal localisation radius is estimated from a typical ensemble-based correlation length scale at arbitrarily chosen time (January 1980) but based on sufficiently many samples to provide a robust climatological estimator. The quantity L_{h} is the correlation length scale that best fits the Gaspari and Cohn function (Gaspari and Cohn, 1999) with least-square minimisation.

We rely on the ensemble-based auto-correlations in each layer of 400 arbitrary water columns to obtain L_h samples (about 53×400 samples) for both T and S. Based on these L_h samples, we will investigate whether the localisation radius varies with latitude under the influence of the Coriolis force or with depth.

We first investigate whether the localisation radius varies with isopycnal layer (in the vertical). This is somewhat expected because when we go deeper the structure of the dynamics gets larger (less energy input and weaker stratification). Figure 4 shows the mode (maximum probability value, $\hat{L}_\mathrm{h}$) of L_h in each layer for T and S. As expected, the correlation length scale increases with the potential density. There are some layers at intermediate depths and in the deep ocean, where a large $\hat{L}_\mathrm{h}$ is found, but this estimation suffers from sampling issue.

Secondly, we investigate whether the localisation radius varies with latitude. This estimate is only based on the top 30 layers of T and S to limit the undersampling of the bottom layers (total L_h samples are here about $30 \times 400 \times 2$). The blue line in Fig. 5 shows the mode (maximum probability value, $\hat{L}_\mathrm{h}$) of L_h with latitude. As expected, $\hat{L}_\mathrm{h}$ decreases with latitude as a consequence of increasing the Coriolis force. For this reason, Zhang et al. (2005) defined the localisation radius as a cosine function of latitude with a maximum value of 2000 km at the Equator. However, we find that the localisation radius follows a bimodal Gaussian function (red line in Fig. 5), which has local maximum of approximately 2300 km at mid-latitudes consistently with cross-basin inter-gyre barotropic flow and reduces to 1500 km near the Equator. The local minimum near the Equator is due to the singularity of the Coriolis effect at the Equator that causes a strong anisotropy of the structure there and thus a reduction of the correlation length that best fits the isotropic Gaspari and Cohn function.

Hamrud et al. (2015) tested the sensitivity of localisation radius in an EnKF implementation of the forecasting system at the European Centre for Medium-Range Weather Forecasts (ECMWF) and found that this sensitivity is small for values up to 50% different from the localisation radius they used. The localisation radius does not seem to vary strongly with depth in the range of observation depths (the upper ocean) and we have thus decided to only retain the variability with latitude, set as a bimodal Gaussian function (red line in Fig. 5).

5 Observation error variance

Prior estimation of observation error is another crucial setting for DA. Overestimation of the error may limit the efficiency of the observation assimilation while underestimation may lead to a collapse of the ensemble spread and a quick divergence in the performance of the system. The observation error in DA can be decomposed in two parts: the instrumental and the representativeness errors. The instrumental error is often given by data provider. The representativeness error is more complex to be estimated and it depends on the system. It accounts for the unresolved processes and scales (Janjić and Cohn, 2006; Oke and Sakov, 2008) or the misspecification of the observation operator (Liu and Rabier, 2002) and varies spatially. Oke and Sakov (2008) found that there is significant spatial variability in the representativeness error of T-S profiles. The representativeness error is often larger than the instrumental error, particularly in the strong mesoscale variability regions.

In this paper, we estimate the total observation error variance of T-S profiles using the ensemble-based technique of Karspeck (2016). The variance of innovations (observational values minus forecast ensemble means mapped to observational space) is essentially explained by the contributions of observation error variance and forecast error variance. A key assumption of Karspeck (2016) is that the ensemble is and should remain **reliable** at any stage of the assimilation cycle. An ensemble forecast system is reliable, if the truth and the ensemble members can be considered to be drawn from the same underlying probability distribution function (PDF) at any given time. However, it is impossible to verify such definition of reliability for all cases (Murphy and Winkler, 1987). In practise, an ensemble is defined to be reliable, if a given observed event which is forecasted with probability p occurs **on average over time** with the empirical probability p.

We consider a time series of N ensemble forecasts with an ensemble size of m for each grid cell. For the ensemble forecast at time j for a grid cell, the ensemble $\{x_{\mathrm{f}}^{i,j}\}_{i=1,m}$ is considered to be drawn from the forecast PDF $\mathcal{N}\left(\mu^j,\ (\sigma_{\mathrm{f}}^j)^2\right)$ and the observation y^j is considered to be drawn from the observation PDF $\mathcal{N}\left(x_{\mathrm{t}}^j,\ (\sigma_{\mathrm{o}}^j)^2\right)$. For the simplification of notation, $x_{\mathrm{f}}^{i,j}$ represents the model state mapped to observational space corresponding to the observation y^j and $\overline{x}_{\mathrm{f}}^j$ is the

ensemble mean of $\{x_{\mathrm{f}}^{i,j}\}_{i=1,m}$. The innovation can be written as

$$y^j - \overline{x}_{\mathrm{f}}^j = \left(y^j - x_{\mathrm{t}}^j\right) + \left(x_{\mathrm{t}}^j - \mu^j\right) + \left(\mu^j - \overline{x}_{\mathrm{f}}^j\right). \tag{10}$$

where the first term in the right side is the observation error with variance $(\sigma_{\mathrm{o}}^j)^2$, the second term in the right side is the forecast distribution error related to the truth with variance $(\sigma_{\mathrm{f}}^j)^2$ (considering the truth to be drawn from the forecast distribution) and the third term in the right hand side is the forecast sampling error with variance $(\sigma_{\mathrm{f}}^j)^2/m$. If we assume that the three terms (observation, distribution and sampling errors) in the right side are independent, the variance of innovation can be written as

$$\begin{aligned} \delta^2 &= \overline{\sigma_{\mathrm{o}}^2} + \overline{\sigma_{\mathrm{f}}^2} + \frac{1}{m}\overline{\sigma_{\mathrm{f}}^2}, \\ &= \overline{\sigma_{\mathrm{o}}^2} + \frac{m+1}{m}\overline{\sigma_{\mathrm{f}}^2}, \end{aligned} \tag{11}$$

where $\overline{\sigma_{\mathrm{o}}^2}$ is the expected value of observation error variances and $\overline{\sigma_{\mathrm{f}}^2}$ is the expected value of forecast error variances. We define an unbiased estimator of observation error variance Φ from Eq. (11) as follows:

$$\begin{aligned} \Phi = &\underbrace{\frac{1}{N-1}\sum_{j=1}^{N}\left(y^j - \overline{x}_{\mathrm{f}}^j\right)^2 - \frac{1}{(N-1)N}\left(\sum_{j=1}^{N}\left(y^j - \overline{x}_{\mathrm{f}}^j\right)\right)^2}_{\text{Estimator of innovation variance } \delta^2} \\ &- \underbrace{\frac{m+1}{(m-1)mN}\sum_{j=1}^{N}\sum_{i=1}^{m}\left(x_{\mathrm{f}}^{ij} - \overline{x}_{\mathrm{f}}^j\right)^2}_{\text{Estimator of forecast error variance } \overline{\sigma_{\mathrm{f}}^2}}. \end{aligned} \tag{12}$$

A similar ensemble-based technique was proposed for the reliability budget in Rodwell et al. (2016) and is used for the reliability validation of our system in section 6.

In this paper, instead of using the forecasts from the ensemble simulation with suboptimal DA (a first guess observation error variance), we use a free run (ensemble simulation without DA) from 1980 to 2010, since a suboptimal update violates the reliability criteria (Karspeck, 2016). Although the estimated observation error variance is not visualised in this paper, it has

significant spatial variability (Oke and Sakov, 2008) and is in good agreement with previous studies (Forget and Wunsch, 2007; Karspeck, 2016).

The estimated observation error consists of both the instrumental error and the representativeness error. The instrumental error is here provided by Levitus et al. (1994a,b), Stammer et al. (2002) and Xie and Zhu (2010). We can thus deduct the representativeness error by subtracting the instrumental error from the estimated observation error (and ensure that it is larger or equal to zero). In coarse resolution climate models, the dominating sources of the representativeness error are due to unresolved processes and scales (Janjić and Cohn, 2006; Oke and Sakov, 2008). This type of error is not expected to reduce when the number of observation increases in the same grid cell, as it relates to model limitation. We propose therefore that only the instrumental error is reduced with the number of observations superobed but the representativeness error is not.

6 Verification in a real framework

In this section, we test the behaviour of our system when assimilating T-S hydrographic profiles. This experiment is based on the settings that were found optimal in previous sections; namely we use the EnKF-z (section 3), the localisation radius varies with latitude (section 4) and the observation error includes the instrumental and representativeness errors tuned prior for our system (section 5). An analysis of NorCPM is carried out for 10 years from 2001 to 2010 with monthly assimilation of T-S profile observations (section 2.3). Our analysis starts in 2001 when the number of profile observations drastically increases because of the Argo program (Roemmich et al., 2009). The validation of the system is only performed against the assimilated profile data (assimilated dataset) and we ensure that the system shows accuracy with time and also that the stochastic system is reliable. If the system is well calibrated, it shows no degradation with time. The validation with unassimilated data will be performed in future work on long-term analyses.

6.1 Accuracy

We estimate the accuracy of the system based on the RMSE and bias (appendix A) of the monthly ensemble mean against the assimilated data. Figure 6 shows the time evolutions of bias

and RMSE for T and S over the analysis period. There are warm biases at 0 - 400 m depth, cold biases at 400-1000 m depth and warm biases below 1000 m. The salinity is slightly too fresh above 1000 m. Overall, NorCPM overestimates the heat content and underestimates the salt content above 2000 m (Fig. 3 in Bentsen et al., 2013). Because assimilated data are used for validation, it is not surprising that the accuracy improves with time. However, it is still encouraging that the error and bias decrease uniformly through the whole water column and that there is no sign of error reemergence which indicates an improperly calibrated system. The depth-averaged bias and RMSE in T and S (black lines in Fig. 6) have converged to stable values by 2006 (vertical dashed lines in Fig. 6). Overall, the system takes approximately 5 years to converge to a stable performance during the Argo period.

6.2 Reliability budget

Here, we follow the **reliability budget** from Rodwell et al. (2016). We decompose the variance of innovations into the forecast error and observation error variances as Eq. (12) in section 5 and obtain the reliability budget from Eq. (11)

$$\underbrace{\frac{1}{N-1}\sum_{j=1}^{N}\left(\overline{x}_{\mathrm{f}}^{j}-y^{j}\right)^{2}-\frac{1}{(N-1)N}\left(\sum_{j=1}^{N}\left(\overline{x}_{\mathrm{f}}^{j}-y^{j}\right)\right)^{2}}_{\text{Estimator of innovation variance } \delta^2}$$
$$=\underbrace{\frac{m+1}{(m-1)mN}\sum_{j=1}^{N}\sum_{i=1}^{m}\left(x_{\mathrm{f}}^{ij}-\overline{x}_{\mathrm{f}}^{j}\right)^{2}}_{\text{Estimator of } \overline{\sigma_{\mathrm{f}}^{2}}}+\underbrace{\frac{1}{N}\sum_{j=1}^{N}R^{j}}_{\text{Estimator of } \overline{\sigma_{\mathrm{o}}^{2}}}+r, \tag{13}$$

where R^j is the observation error variance related to the observation y^j and r is the residual. If the ensemble anomalies/perturbations represent correctly the forecast error variance (reliable ensemble) and if the observation error variance is correctly set, the residual r would converge to zero when $N \to \infty$.

In this paper, we verify the reliability for the period of 2006-2010, since the system needs about 5 years to converge to a stable performance (Fig. 6). The colours in Fig. 7 represent the residuals r for T and S at depth 200 m and 1000 m. We apply a statistical significance test on the residual r to verify the reliability of our system. The model bias is assumed to be constant and

we compute the time series of residual values $\{r^j\}_{j=1,N}$ (Rodwell et al., 2016). The Student's t-test (at the significance level of 5 %) is used to determine if r is statistically different from zero (null hypothesis H_0: $r = 0$). The black dots in Fig. 7 represent the regions where the hypothesis H_0 is rejected and our system is not found reliable (underdispersive). In most regions (more than 96% at 200 m and 90% at 1000 m) we cannot reject the hypothesis that our ensemble is reliable. Note that the size of the time series varies spatially with observation availability and that the significance test is not performed for samples less than 10 (white colour in Fig. 7).

7 Summary and conclusions

In this paper, we investigated several key implementation settings for the assimilation of hydrographic profiles into an isopycnal coordinate ocean model. This study was performed with a fully-coupled climate prediction system (NorCPM) that used to assimilate oceanic observations at monthly frequency with an EnKF. It is an initial verification and documentation of a system that we aim to use to perform long-term reanalyses of the ocean and seasonal-to-decadal climate predictions. Only the reanalysis capability of the system is investigated in the paper.

First we tested whether it is more optimal to construct the innovation vector in the observation native z-level coordinates or in the model isopycnal coordinates. While there are regional discrepancies in the efficiency of the two schemes, constructing the innovation in z-level coordinates is more accurate. It is challenging to directly compare our conclusions to Xie and Zhu (2010) as we modified the implementation of the two schemes. In the best scheme of Xie and Zhu (2010), profile observations interpolated to isopycnal coordinates were assimilated, S and T were separately updated with their corresponding observations, and below the mixed layer T was diagnosed from the equation of state of seawater. In our schemes, both T and S are simultaneously updated as recommended by Srinivasan et al. (2011). Furthermore, the improved formulation of the observation errors of T, S and DP in isopycnal coordinates may have influenced the results. We proposed that constructing the innovation in isopycnal coordinates degrades the efficiency of assimilation because the operator transforming z-level coordinate profiles (T and S) to isopycnal coordinate profiles (T, S and layer thickness) is strongly non-linear and non-unique and introduces uncertainties that deteriorate the efficiency of the assimilation.

Secondly, we identified horizontal localisation radii based on the spatial correlation length scale of T and S in our system. While the correlation length varies only marginally with depth, a pronounced variation is found with latitude. Such variation was previously assumed to be follow a cosine function of latitude in (Zhang et al., 2005), since the correlation length varies with the Rossby radius. However, we found that the correlation length is smaller near the Equator than at mid-latitudes. A bimodal Gaussian function was thus proposed to fit the empirical optimal localisation radius - being largest at mid-latitudes and smaller at high latitudes and at the Equator.

Thirdly, we estimated the observation error variance for T-S profiles using the ensemble technique of Karspeck (2016) so that the reliability of the system was preserved. This estimate accounts for both the instrumental and representativeness errors. The observation error variance estimated in NorCPM was in good agreement with previous studies (Forget and Wunsch, 2007; Oke and Sakov, 2008; Karspeck, 2016). It is higher than the observation error variance provided in the EN4 objective analysis because the error estimate of the EN4 objective analysis does not account for the representativeness error. Moreover, we proposed that the representativeness error should not be reduced during superobing because it relates to unresolved processes and scales in climate models.

Finally, a 10-year analysis (2001–2010) was carried out to verify whether our system with the above implementation settings performs consistently and with the expected gain in accuracy and reliability. The assimilation of T-S profiles improves the accuracy of the system (reducing the RMSE and bias). During the Argo observation period, the system requires a 5-year assimilation spin-up to converge to a stable level of performance. The reliability of the system was validated for the period of 2006–2010 using the reliability budget proposed in Rodwell et al. (2016). The reliability is rejected at a 5% significance level only in few places. This suggests that the assimilation of T-S profiles with the current implementation setting can be safely included in NorCPM to perform long-term reanalyses and seasonal-to-decadal predictions.

In this paper, we only verified the implementation of assimilation of T-S profiles for reanalysis purpose. For future works, we will use NorCPM to perform long-term reanalyses and seasonal-to-decadal climate predictions, which we will contribute to the Decadal Climate Prediction Project (DCPP) of the Coupled Climate Prediction Project (CMIP) phase 6 (Eyring et al., 2015).

Acknowledgement

This study was co-funded by the Center for Climate Dynamics at the Bjerknes Center and the Norwegian Research Council under the NORKLIMA research (EPOCASA; 229774/E10). This work received a grant for computer time from the Norwegian Program for supercomputer (NOTUR2, project number NN9039K) and a storage grant (NORSTORE, NS9039K). We thank anonymous reviewers for the comments, and L. Bertino, A. Carrassi, P. Raanes and J.-P. Xie for discussions.

Appendix A Statistical metrics

Let $\{x_{\mathrm{o}}^i\}_{i=1,N}$ and $\{\overline{x}_{\mathrm{f}}^i\}_{i=1,N}$ be the observations and forecast ensemble means. N is the number of available observations. The statistical metrics are defined as follows:

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N} w_i \left(\overline{x}_{\mathrm{f}}^i - x_{\mathrm{o}}^i\right)^2}, \tag{14}$$

$$\mathrm{bias} = \sum_{i=1}^{N} w_i \left(\overline{x}_{\mathrm{f}}^i - x_{\mathrm{o}}^i\right), \tag{15}$$

where w_i is the area-weight related to the area of grid cell a_i and is defined by

$$w_i = \frac{a_i}{\sum_{j=1}^{N} a_j}. \tag{16}$$

Bibliography

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, O., Drange, H., Roelandt, C., Seierstad, I. a., Hoose, C., and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1- Part 1: Description and basic evaluation of the physical climate, Geoscientific Model Development, 6, 687–720, doi:10.5194/gmd-6-687-2013, URL `http://www.geosci-model-dev.net/6/687/2013/`, 2013.

Bleck, R., Rooth, C., Hu, D., and Smith, L. T.: Salinity-driven Thermocline Transients in a Wind- and Thermohaline-forced Isopycnic Coordinate Model of the North Atlantic, Journal

546 of Physical Oceanography, 22, 1486–1505, doi:10.1175/1520-0485(1992)022⟨1486:SDTTIA⟩2.
547 0.CO;2, 1992.

548 Bocquet, M.: Localization and the iterative ensemble Kalman smoother, Quarterly Journal of
549 the Royal Meteorological Society, pp. 1075–1089, doi:10.1002/qj.2711, 2016.

550 Brune, S., Nerger, L., and Baehr, J.: Assimilation of oceanic observations in a global coupled
551 Earth system model with the SEIK filter, Ocean Modelling, 96, 254–264, doi:10.1016/j.ocemod.
552 2015.09.011, URL http://dx.doi.org/10.1016/j.ocemod.2015.09.011, 2015.

553 Counillon, F., Bethke, I., Keenlyside, N., Bentsen, M., Bertino, L., and Zheng, F.: Seasonal-
554 to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System
555 Model: a twin experiment, Tellus A, 66, 1–21, doi:10.3402/tellusa.v66.21074, URL http:
556 //www.tellusa.net/index.php/tellusa/article/view/21074, 2014.

557 Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., and Bentsen,
558 M.: Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the
559 Norwegian Climate Prediction Model, Tellus A, 68, 1–17, doi:10.3402/tellusa.v68.32437, URL
560 http://www.tellusa.net/index.php/tellusa/article/view/32437, 2016.

561 Craig, A. P., Vertenstein, M., and Jacob, R.: A new flexible coupler for earth system modeling
562 developed for CCSM4 and CESM1, Int. J. High Perform. Comput. Appl., 26, 31–42, doi:
563 10.1177/1094342011428141, 2012.

564 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U.,
565 Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L.,
566 Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L.,
567 Healy, S. B., Hersbach, H., H?lm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M.,
568 McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P.,
569 Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and
570 performance of the data assimilation system, Quarterly Journal of the Royal Meteorological
571 Society, 137, 553–597, doi:10.1002/qj.828, URL http://dx.doi.org/10.1002/qj.828, 2011.

572 Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E.,

Stouffer, R. J., Cooke, W., Dunne, K. A., Harrison, M. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Phillipps, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A. T., and Zadeh, N.: GFDL's ESM2 Global Coupled Climate-Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics, Journal of Climate, 25, 6646–6665, doi:10.1175/JCLI-D-11-00560.1, 2012.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynamics, 53, 343–367, doi:10.1007/s10236-003-0036-9, URL `http://link.springer.com/10.1007/s10236-003-0036-9`, 2003.

Eyring, V., Bony, S., Meehl, G. A., Senior, C., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation, Geoscientific Model Development Discussions, 8, 10 539–10 583, doi:10.5194/gmdd-8-10539-2015, URL `http://www.geosci-model-dev-discuss.net/8/10539/2015/`, 2015.

Forget, G. and Wunsch, C.: Estimated Global Hydrographic Variability, Journal of Physical Oceanography, 37, 1997–2008, doi:10.1175/JPO3072.1, 2007.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Q.J.R. Meteorol. Soc., 125, 723–757, doi:10.1002/qj.49712555417, 1999.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z. L., and Zhang, M.: The community climate system model version 4, Journal of Climate, 24, 4973–4991, doi:10.1175/2011JCLI4083.1, 2011.

Good, S. a., Martin, M. J., and Rayner, N. a.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, Journal of Geophysical Research: Oceans, 118, 6704–6716, doi:10.1002/2013JC009067, 2013.

Gouretski, V. and Reseghetti, F.: On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database, Deep-Sea

599 Research Part I: Oceanographic Research Papers, 57, 812–833, doi:10.1016/j.dsr.2010.03.011,
600 URL http://dx.doi.org/10.1016/j.dsr.2010.03.011, 2010.

601 Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-Dependent Filtering of Background
602 Error Covariance Estimates in an Ensemble Kalman Filter, Monthly Weather Review, 129,
603 2776–2790, doi:10.1175/1520-0493(2001)129⟨2776:DDFOBE⟩2.0.CO;2, 2001.

604 Hamrud, M., Bonavita, M., and Isaksen, L.: EnKF and Hybrid Gain Ensemble Data As-
605 similation. Part I: EnKF Implementation, Monthly Weather Review, 143, 4847–4864, doi:
606 10.1175/MWR-D-14-00333.1, 2015.

607 Holland, M. M., Bailey, D. A., Briegleb, B. P., Light, B., and Hunke, E.: Improved sea ice
608 shortwave radiation physics in CCSM4: the impact of melt ponds and aerosols on arctic sea
609 ice, J. Climate, 25, 1413–1430, doi:10.1175/JCLI-D-11-00078.1, 2012.

610 ICES, SCOR, and IAPSO: Tenth Report of the Join Panel on Oceanographic Tables and Stan-
611 dards (The Practical Salinity Scale 1978 and the International Equation of State of Seawater
612 1980), Tech. rep., UNESCO Technical Papers in Marine Science No. 36, 25 pp., 1981.

613 Janjić, T. and Cohn, S. E.: Treatment of Observation Error due to Unresolved Scales in Atmo-
614 spheric Data Assimilation, Monthly Weather Review, 134, 2900–2915, doi:10.1175/MWR3229.
615 1, 2006.

616 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha,
617 S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki,
618 W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph,
619 D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological
620 Society, 77, 437–471, doi:10.1175/1520-0477(1996)077⟨0437:TNYRP⟩2.0.CO;2, URL http://
621 dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

622 Karspeck, A. R.: An Ensemble Approach for the Estimation of Observational Error Illustrated
623 for a Nominal 1° Global Ocean Model, Monthly Weather Review, 144, 1713–1728, doi:10.1175/
624 MWR-D-14-00336.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-14-00336.
625 1, 2016.

626 Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., and Janssen, P.: A coupled data assim-
627 ilation system for climate reanalysis, Quarterly Journal of the Royal Meteorological Society,
628 142, 65–78, doi:10.1002/qj.2629, URL http://dx.doi.org/10.1002/qj.2629, 2016.

629 Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence,
630 P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.:
631 Parameterization improvements and functional and structural advances in version 4 of the
632 community land model, J. Adv. Model. Earth Syst., 3, M03001, doi:10.1029/2011MS000045,
633 2011.

634 Levitus, S., Burgett, R., and Boyer, T.: World Ocean Atlas 1994, vol. 3, Salinity, U.S. Dep. of
635 Commer., Washtiongton, DC, 1994a.

636 Levitus, S., Burgett, R., and Boyer, T.: World Ocean Atlas 1994, vol. 4, Temperature, U.S. Dep.
637 of Commer., Washtiongton, DC, 1994b.

638 Liu, Z. Q. and Rabier, F.: The interaction between model resolution, observation resolution and
639 observation density in data assimilation: A one-dimensional study, Quart. J. Roy. Meteor.
640 Soc., 128, 1367–1386, 2002.

641 McDougall, T. J. and Jackett, D. R.: An assessment of orthobaric density in the global ocean,
642 J. Phys. Oceanogr., 35, 2054–2075, doi:10.1175/JPO2796.1, 2005.

643 Meehl, G. a., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K.,
644 Giorgetta, M. a., Greene, A. M., Hawkins, E. D., Hegerl, G., Karoly, D., Keenlyside, N.,
645 Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., and Stockdale,
646 T.: Decadal prediction: Can it be skillful?, Bulletin of the American Meteorological Society,
647 90, 1467–1485, doi:10.1175/2009BAMS2778.1, 2009.

648 Mochizuki, T., Masuda, S., Ishikawa, Y., and Awaji, T.: Multiyear climate prediction with initial-
649 ization based on 4D-Var data assimilation, Geophysical Research Letters, 43, 3903–3910, doi:
650 10.1002/2016GL067895, URL http://dx.doi.org/10.1002/2016GL067895, 2016GL067895,
651 2016.

652 Murphy, A. H. and Winkler, R. L.: A general framework for forecast verification, Mon. Wea.
653 Rev., 115, 1330–1338, 1987.

654 Neale, R. B., Richter, J. H., Conley, A. J., Park, S., Lauritzen, P. H., Gettelman, A., Williamson,
655 D. L., Rasch, P. J., Vavrus, S. J., Taylor, M. A., Collins, W. D., Zhang, M., and jiann Lin, S.:
656 Description of the NCAR Community Atmosphere Model (CAM 4.0), 2010.

657 Oke, P. R. and Sakov, P.: Representation error of oceanic observations for data assimilation, Jour-
658 nal of Atmospheric and Oceanic Technology, 25, 1004–1017, doi:10.1175/2007JTECHO558.1,
659 2008.

660 Oke, P. R. and Schiller, A.: Impact of Argo, SST, and altimeter data on an eddy-resolving ocean
661 reanalysis, Geophysical Research Letters, 34, 1–7, doi:10.1029/2007GL031549, 2007.

662 Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J.,
663 Levis, S., Swenson, S. C., Thornton, P. E., Dai, A., Decker, M., Dickinson, R., Feddema, J.,
664 Heald, C. L., Hoffman, F., Lamarque, J.-F., Mahowald, N., Niu, G.-Y., Qian, T., Randerson,
665 J., Running, S., Sakaguchi, K., Slater, A., Stöckli, R., Wang, A., Yang, Z.-L., Zeng, X., and
666 Zeng, X.: Technical Description of version 4.0 of the Community Land Model (CLM), Tech.
667 Rep. NCAR/TN-478+STR, National Center for Atmospheric Research, Boulder, Colorado,
668 USA, 2010.

669 Rodwell, M. J., Lang, S. T. K., Ingleby, N. B., Bormann, N., Hólm, E., Rabier, F., Richardson,
670 D. S., and Yamaguchi, M.: Reliability in ensemble data assimilation, Quarterly Journal of the
671 Royal Meteorological Society, 142, 443–454, doi:10.1002/qj.2663, URL `http://doi.wiley.`
672 `com/10.1002/qj.2663`, 2016.

673 Roemmich, D., Johnson, G., Riser, S., Davis, R., Gilson, J., W.B., O., Garzoli, S., Schmid,
674 C., and Ignaszewski, M.: The Argo Program: Observing the global ocean with profiling
675 floats, Oceanography, 22(2), 34–43, doi:10.5670/oceanog.2009.36, URL `http://dx.doi.org/`
676 `10.5670/oceanog.2009.36`, 2009.

677 Sakov, P. and Bertino, L.: Relation between two common localisation methods for the EnKF,

678 Computational Geosciences, 15, 225–237, doi:10.1007/s10596-010-9202-6, URL http://link.
679 springer.com/10.1007/s10596-010-9202-6, 2010.

680 Sakov, P. and Oke, P. R.: A deterministic formulation of the ensemble Kalman filter: an al-
681 ternative to ensemble square root filters, Tellus A, 60, 361–371, doi:10.1111/j.1600-0870.2007.
682 00299.x, URL http://tellusa.net/index.php/tellusa/article/view/15252, 2008.

683 Sakov, P., Counillon, F., Bertino, L., Lisæ ter, K. A., Oke, P. R., and Korablev, A.: TOPAZ4:
684 an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Sci-
685 ence, 8, 633–656, doi:10.5194/os-8-633-2012, URL http://www.ocean-sci.net/8/633/2012/
686 os-8-633-2012.pdfhttp://www.ocean-sci.net/8/633/2012/, 2012.

687 Srinivasan, a., Chassignet, E. P., Bertino, L., Brankart, J. M., Brasseur, P., Chin, T. M., Counil-
688 lon, F., Cummings, J. a., Mariano, a. J., Smedstad, O. M., and Thacker, W. C.: A comparison
689 of sequential assimilation schemes for ocean prediction with the HYbrid Coordinate Ocean
690 Model (HYCOM): Twin experiments with stati c forecast error covariances, Ocean Mod-
691 elling, 37, 85–111, doi:10.1016/j.ocemod.2011.01.006, URL http://dx.doi.org/10.1016/j.
692 ocemod.2011.01.006, 2011.

693 Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, a.,
694 Hill, C., and Marshall, J.: The Global ocean circulation during 1992-1997, estimated from
695 ocean observations and a general circulation model, 107, doi:10.1029/2001JC000888, URL
696 http://eprints.soton.ac.uk/214/, 2002.

697 Steele, M., Morley, R., and Ermold, W.: PHC: A global ocean hydrography with a high quality
698 Arctic Ocean, Journal of Climate, 14 (9), 2079–2087, 2001.

699 Sun, S. and Bleck, R.: Multi-century simulations with the coupled GISS–HYCOM climate model:
700 control experiments, Climate Dynamics, 26, 407–428, doi:10.1007/s00382-005-0091-7, URL
701 http://dx.doi.org/10.1007/s00382-005-0091-7, 2006.

702 Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and
703 the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498,

704 doi:10.1175/BAMS-D-11-00094.1, URL http://journals.ametsoc.org/doi/abs/10.1175/
705 BAMS-D-11-00094.1, 2012.

706 Thacker, W. C. and Esenkov, O. E.: Assimilating XBT data into HYCOM, Journal of Atmo-
707 spheric and Oceanic Technology, 19, 709–724, doi:10.1175/1520-0426(2002)019⟨0709:AXDIH⟩
708 2.0.CO;2, 2002.

709 Thomson, R. E. and Fine, I. V.: Estimating mixed layer depth from oceanic profile data, Journal
710 of Atmospheric and Oceanic Technology, 20, 319–329, doi:10.1175/1520-0426(2003)020⟨0319:
711 EMLDFO⟩2.0.CO;2, 2003.

712 Vertenstein, M., Craig, T., Middleton, A., Feddema, D., and Fischer, C.: CESM1.0.3
713 User Guide, http://www.cesm.ucar.edu/models/cesm1.0/cesm/cesm_doc_1_0_4/ug.pdf,
714 accessed: 2015-01-23, 2012.

715 Wang, Y., Counillon, F., and Bertino, L.: Alleviating the bias induced by the linear analysis
716 update with an isopycnal ocean model, Quarterly Journal of the Royal Meteorological Society,
717 pp. 1064–1074, doi:10.1002/qj.2709, 2016.

718 Xie, J. and Zhu, J.: Ensemble optimal interpolation schemes for assimilating Argo profiles into
719 a hybrid coordinate ocean model, Ocean Modelling, 33, 283–298, doi:10.1016/j.ocemod.2010.
720 03.002, URL http://dx.doi.org/10.1016/j.ocemod.2010.03.002, 2010.

721 Zhang, S., Harrison, M. J., Wittenberg, a. T., Rosati, a., Anderson, J. L., and Balaji, V.:
722 Initialization of an ENSO Forecast System Using a Parallelized Ensemble Filter, Monthly
723 Weather Review, 133, 3176–3201, doi:10.1175/MWR3024.1, 2005.

724 Zhang, S., Rosati, A., and Harrison, M. J.: Detection of multidecadal oceanic variability by
725 ocean data assimilation in the context of a "perfect" coupled model, Journal of Geophysical
726 Research: Oceans, 114, C12 018, doi:10.1029/2008JC005261, URL http://doi.wiley.com/
727 10.1029/2008JC005261, 2009.

728 Zhang, S., Rosati, a., and Delworth, T.: The Adequacy of Observing Systems in Monitoring the
729 Atlantic Meridional Overturning Circulation and North Atlantic Climate, Journal of Climate,

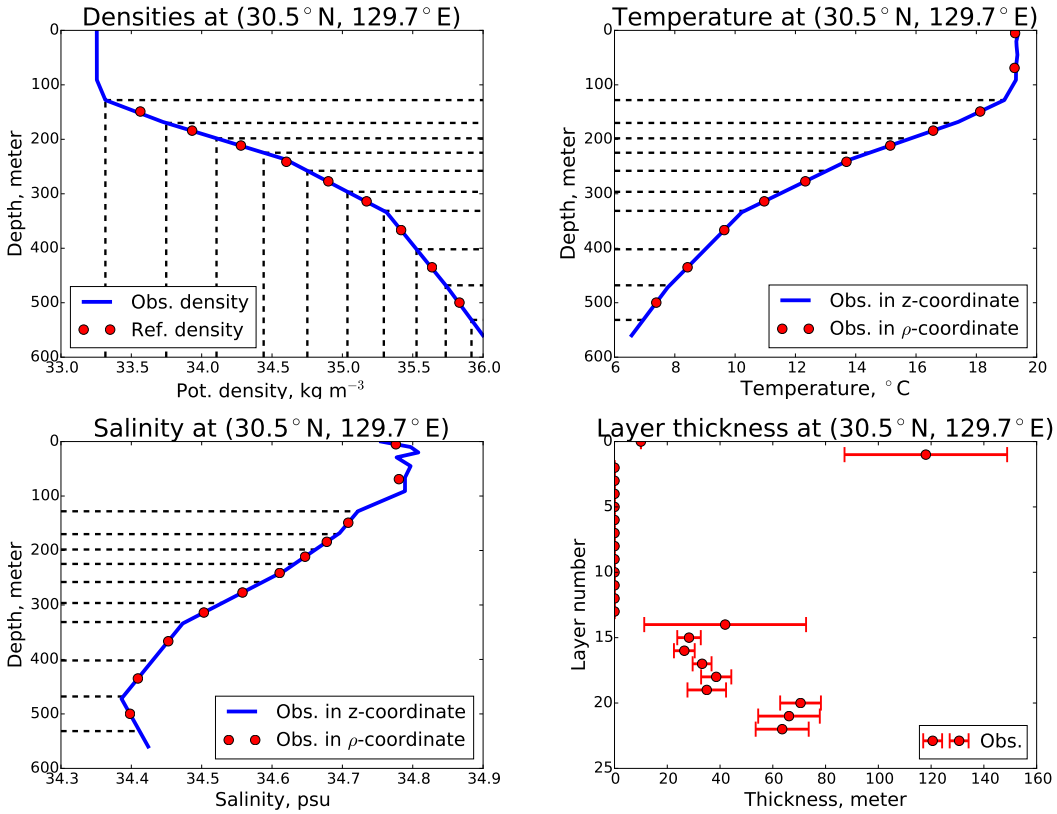

Figure 1: Observations in January 1980 from the EN4 dataset at (30.5° N, 129.7° E) in z-level (z-) and converted to isopycnal (ρ-) coordinates. Top left panel shows the potential density profile (blue line) and reference densities (red dots) in our implementation of MICOM. Vertical dashed lines present reference density interfaces. Horizontal dashed lines present the interfaces of isopycnal layers. Top right and bottom panels show the temperature and salinity in z-level coordinates (blue lines) and isopycnal coordinates (red dots). Bottom right panel shows observations (dots) and uncertainties (bars) of isopycnal layer thickness.

730 23, 5311–5324, doi:10.1175/2010JCLI3677.1, URL `http://journals.ametsoc.org/doi/abs/`

731 `10.1175/2010JCLI3677.1`, 2010.

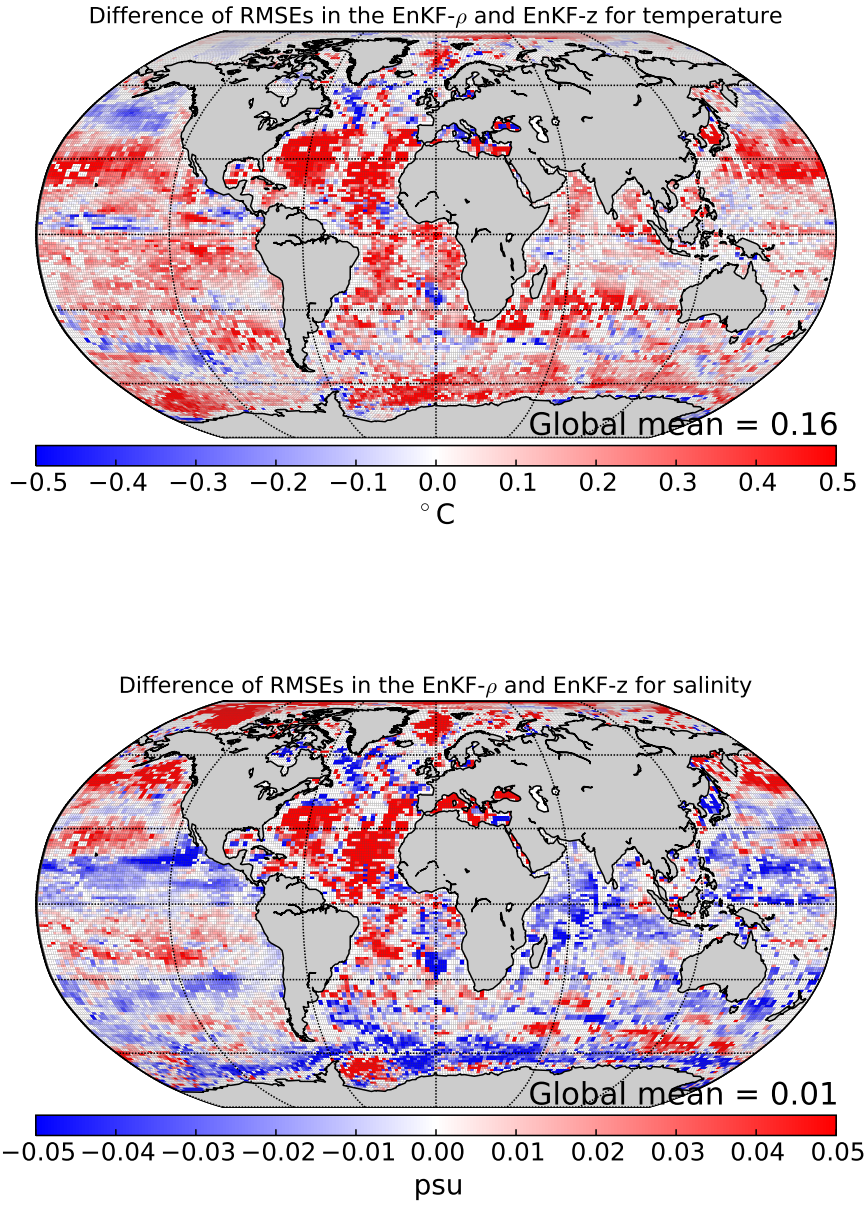

Figure 2: The difference in RMSE for EnKF-ρ and EnKF-z for temperature and salinity. Warm (cold) colours mean that EnKF-z performs better (worse) than EnKF-ρ. The global averages of the differences for temperature and salinity are respectively +0.16 °C and +0.01 psu.

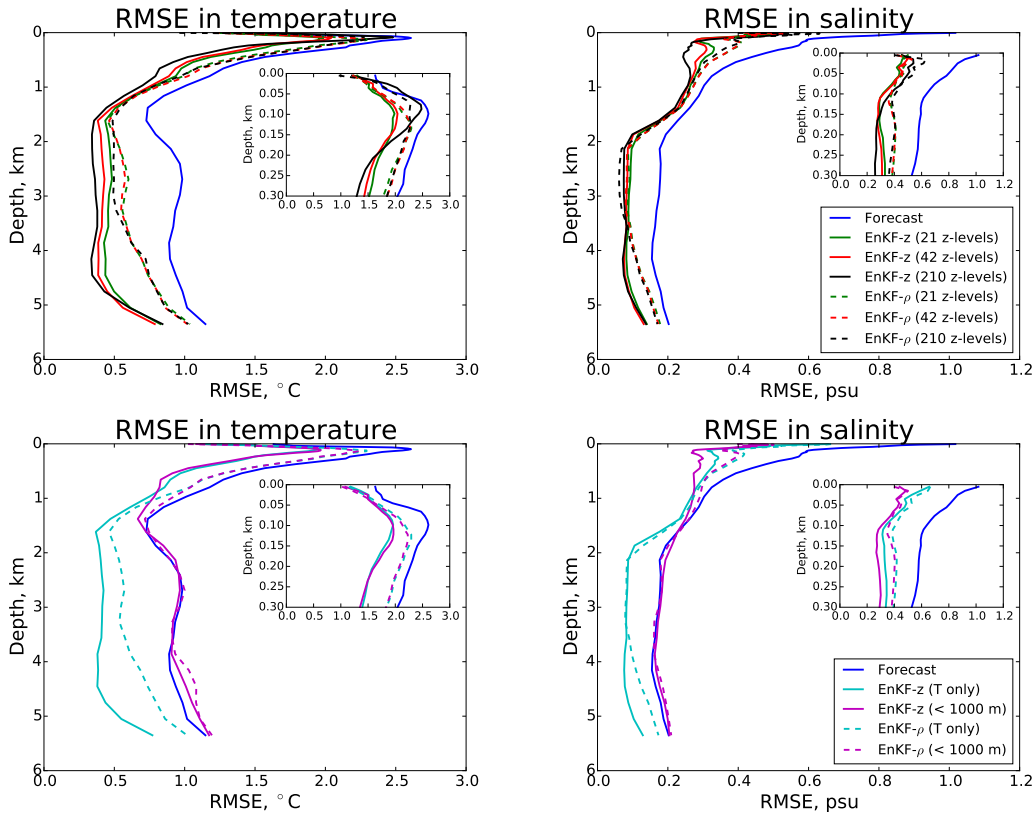

Figure 3: Horizontal RMSE for temperature and salinity by assimilating different datasets. Blue lines stand for the forecast. Green, red and black lines stand respectively for assimilation of observations in 21, 42 and 210 z-levels generated from the truth (42 z-levels). Cyan lines stand for assimilation of temperature observations and salinity climatology. Magenta lines stand for assimilation of observations in top 1000 meter. Solid (dashed) lines represent the RMSEs in the EnKF-z (EnKF-ρ). The inset in each panel shows the RMSEs zoomed in the top 300 m.

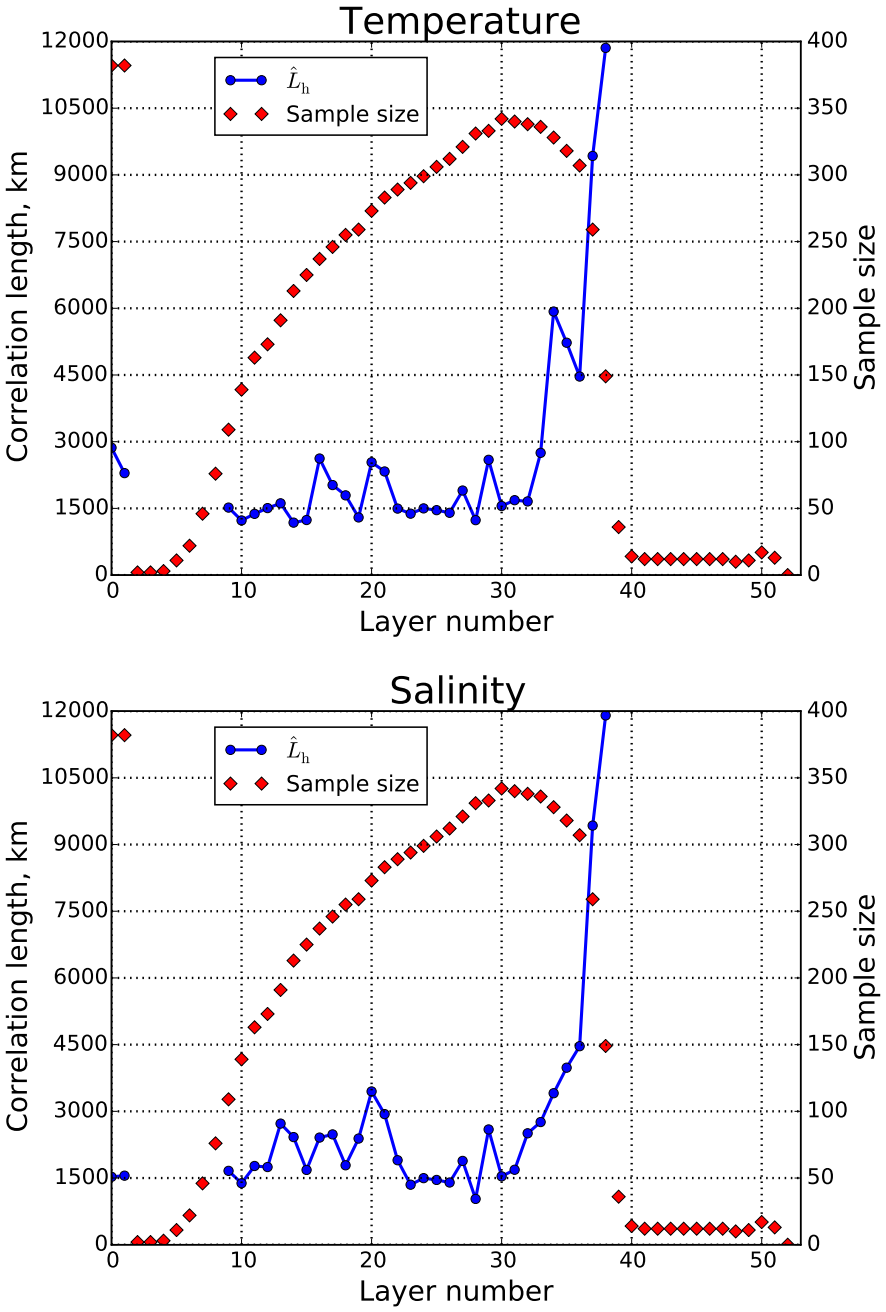

Figure 4: Optimal ensemble-based correlation length scale $\hat{L}_{\mathrm{h}}$ in different isopycnal layers for temperature and salinity. In each layer, the sample size of L_{h} used to estimate $\hat{L}_{\mathrm{h}}$ is shown by diamond (right vertical axis). There is no $\hat{L}_{\mathrm{h}}$ value in some layers because of a too small sample size (less than 100).

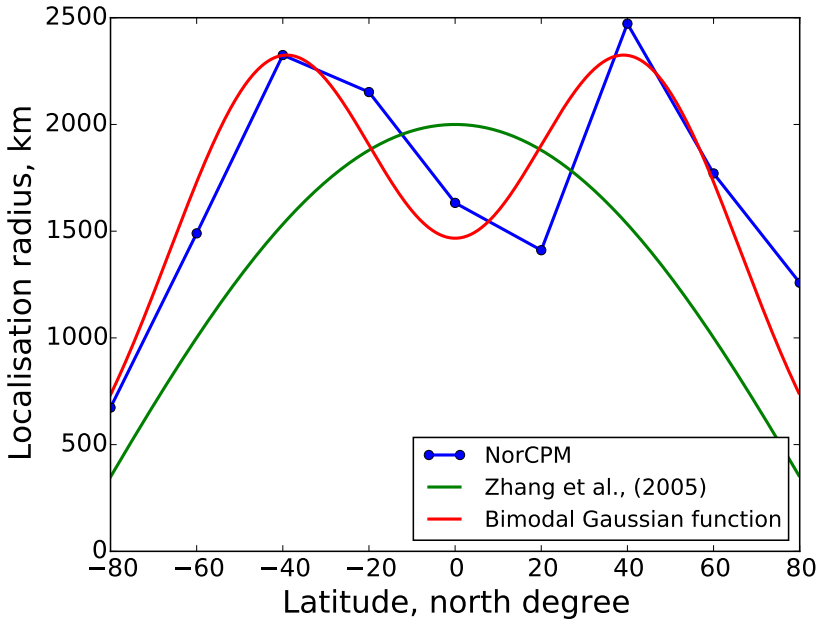

Figure 5: Horizontal localisation radius varying with latitude (20° band). The blue line represents the optimal ensemble-based correlation length scale used as proxy for the localisation radius in NorCPM. The green line represents the localisation radius used in Zhang et al. (2005). The red line represents a bimodal Gaussian function fitting the blue line.

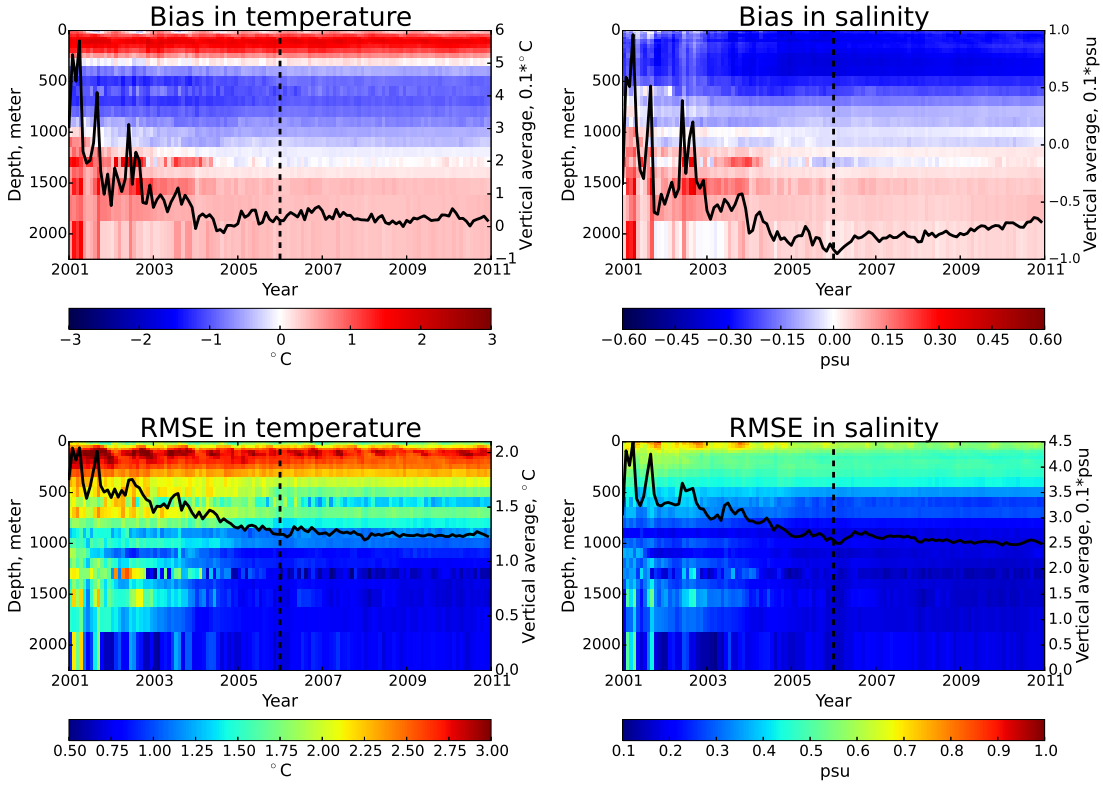

Figure 6: Bias (top) and RMSE (bottom) computed with assimilated profile data for temperature and salinity. Colours represent monthly statistics in depth (left vertical axis). Black lines represent the time evolutions of monthly depth-averaged statistics (right vertical axis). Vertical dashed lines show the moment when the system converges to a stable performance.

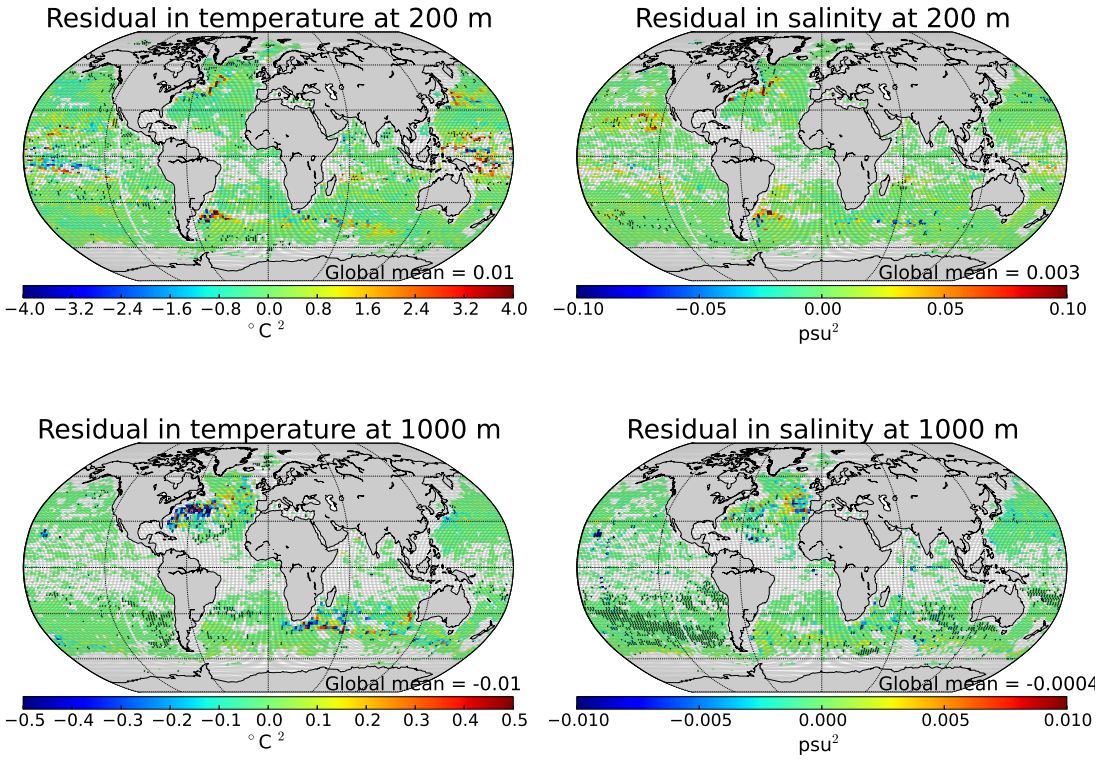

Figure 7: Residuals in the reliability budget for the period of 2006–2010 at 200 m (top) and 1000 m (bottom) for temperature (left) and salinity (right). The dot indicates grid cell where the residual is statistically different from zero at a 5% significance level in the Student's t-test. The white colour stands for grid cell with less than 10 observations over time where the statistical test was not performed.