RMetS
Royal Meteorological Society

Evaluation of the ECMWF ocean reanalysis system ORAS4

Magdalena Alonso Balmaseda,[a]* Kristian Mogensen[a] and Anthony T. Weaver[b]
[a]*European Centre for Medium-range Weather Forecasts, Reading, UK*
[b]*CERFACS, Toulouse, France*

*Correspondence to: M. A. Balmaseda, ECMWF, Shinfield Park, Reading RG2 9AX, UK.
E-mail: Magdalena.Balmaseda@ecmwf.int

A new operational ocean reanalysis system (ORAS4) has been implemented at ECMWF. It spans the period 1958 to the present. This article describes its main components and evaluates its quality. The adequacy of ORAS4 for the initialization of seasonal forecasts is discussed, along with the robustness of some prominent climate signals.

ORAS4 has been evaluated using different metrics, including comparison with observed ocean currents, RAPID-derived transports, sea-level gauges, and GRACE-derived bottom pressure. Compared to a control ocean model simulation, ORAS4 improves the fit to observations, the interannual variability, and seasonal forecast skill. Some problems have been identified, such as the underestimation of meridional overturning at 26°N, the magnitude of which is shown to be sensitive to the treatment of the coastal observations.

ORAS4 shows a clear and robust shallowing trend of the Pacific Equatorial thermocline. It also shows a clear and robust nonlinear trend in the 0–700 m ocean heat content, consistent with other observational estimates. Some aspects of these climate signals are sensitive to the choice of sea-surface temperature product and the specification of the observation-error variances. The global sea-level trend is consistent with the altimeter estimate, but the partition into volume and mass variations is more debatable, as inferred by discrepancies in the trend between ORAS4- and GRACE-derived bottom pressure.

Key Words: ocean reanalysis; climate variability; validation; quality metric; initialization

Received 15 June 2012; Revised 14 September 2012; Accepted 1 October 2012; Published online in Wiley Online Library 31 December 2012

Citation: Balmaseda MA, Mogensen K, Weaver AT. 2013. Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q. J. R. Meteorol. Soc.* **139**: 1132–1161. DOI:10.1002/qj.2063

1. Introduction

Ocean reanalyses are historical reconstructions of the ocean climate, based on the objective synthesis of the information provided by ocean models, atmospheric forcing fluxes and ocean observations, combined via data assimilation methods. They are also referred to as *ocean syntheses*. The designation *reanalysis* is not so intuitive, but mirrors the equivalent designation used for the more consolidated activities in the historical reconstruction of the atmosphere, and it will be used in the rest of the article.

Ocean reanalysis is now an established activity in several research and operational centres. Ocean reanalyses are revisited every so often, and new 'vintages' of reanalyses are produced at intervals of about five years, when improvements in ocean models, data assimilation methods, forcing fluxes or ocean observations are available. A review of the state of the art on ocean reanalysis in 2010 is given by Lee *et al.* (2010). A new vintage is now being generated (Saha *et al.*, 2010; Masina *et al.*, 2011; Yin *et al.*, 2011; Ferry *et al.*, 2012; Haines *et al.*, 2012), which has come about through the availability of new surface forcing fluxes (from new atmospheric reanalyses), and improved quality-controlled

ocean datasets, which include important corrections to observations (Lyman *et al.*, 2006; Wijffels *et al.*, 2009, among others).

Although new reanalysis vintages are produced infrequently, some of the reanalysis products are continuously brought to quasi-real time, with the model and data assimilation methodology kept frozen. This is the case of the ocean reanalyses produced in operational centres for the initialization of coupled (seasonal and monthly) forecasts, such as the European Centre for Medium-range Weather Forecasts (ECMWF). To initialize the coupled forecasts, ocean initial conditions are needed for the real time and for the historical record. The *a posteriori* calibration of model output requires an estimate of the model climatology, which is obtained by performing a series of coupled hindcasts during some historical period. These hindcasts are initialized from the ocean reanalysis. A historical record of hindcasts is also needed for skill assessment. The interannual variability represented by ocean reanalyses will have an impact on both the calibration and on the assessment of the skill (Balmaseda *et al.*, 2010). Often the impact of forecast skill can be used as a metric for the reanalysis quality. Ocean reanalyses are potentially a valuable resource for climate variability studies and have the advantage of being continuously brought up to real time, which allows monitoring of relevant climate variables (Xue *et al.*, 2010).

The production of a robust ocean reanalysis with uncertainty estimates is a major challenge. In addition to the three-dimensional estimation of the ocean state at a given time (the analysis problem), the estimation of the time evolution is also required in a reanalysis. The time evolution represented by an ocean reanalysis will be sensitive to the time variations of the observing system, to the errors of the ocean model, atmospheric fluxes and assimilation system, which are often flow-dependent, and not easy to estimate. Therefore, before the data of a reanalysis are used, the validation and intercomparisons of the reanalysis output with other independent products become essential. In this article we present a series of objective metrics that can be used to validate any reanalysis product, and we apply them to the new ECMWF ocean reanalysis.

The Ocean ReAnalysis System 4 (ORAS4) has recently been implemented operationally at ECMWF. It replaces the previous system ORAS3 (Balmaseda *et al.*, 2008). Both the ocean model and ocean data assimilation system have been changed. The Hamburg Ocean Primitive-Equation (HOPE) model (Wolff *et al.*, 1997) has been replaced with the Nucleus for European Modelling of the Ocean (NEMO) model (Madec, 2008). The Optimal Interpolation (OI) has been replaced by the newly developed variational assimilation system NEMOVAR (Mogensen *et al.*, 2012, hereafter MBW12). Aside from the use of NEMO and NEMOVAR, several other innovative features have been introduced in ORAS4 with respect to its predecessor ORAS3, such as the use of ERA-Interim forcing fluxes, revised quality-controlled datasets with corrections to the eXpendable BathyThermographs (XBTs), Argo data for the estimation of model bias, and a revised ensemble generation strategy that should sample better the uncertainty in the deeper ocean.

An early non-operational NEMOVAR-based ocean reanalysis, called COMBINE-NEMOVAR (hereafter COMB-NV; Balmaseda *et al.*, 2010), was produced as part of the European COMBINE (Comprehensive Modelling of the Earth System for Better Climate Prediction and Projection) project,* and has been used by some groups to initialize Coupled Model Intercomparison Project (CMIP5) decadal integrations (Hazeleger *et al.*, 2011). When compared with other reanalysis products, COMB-NV scores favourably, both in the representation of the tropical Atlantic variability (Zhu *et al.*, 2011) and in the initialization of seasonal forecasts (Zhu *et al.*, 2012). The COMB-NV did not assimilate altimeter-derived sea-level anomaly (SLA) data, and used different parameter choices in the data assimilation compared to ORAS4. This reanalysis will not be updated to real time, and it is superseded by ORAS4. The differences between COMB-NV and ORAS4 are discussed in this article.

The purpose of this article is to describe the different components of ORAS4, with special emphasis on the NEMOVAR data assimilation system, to validate the reanalysis outcome by means of a series of objective diagnostics, and to evaluate the robustness of some important climate signals. The article is organized as follows. Section 2 gives a general overview of NEMOVAR as implemented in ORAS4. Section 3 describes the datasets, forcing fields and ensemble generation procedure used in ORAS4. It also provides a summary of the differences with respect to the previous COMB-NV reanalysis, and discusses some interesting sensitivity experiments. Section 4 assesses the performance of NEMOVAR by evaluating the quality of the ORAS4 reanalysis according to a set of objective metrics, including the impact on the initialization of seasonal forecasts. Section 5 discusses the robustness of some relevant climate signals represented in ORAS4, such as trends in the Equatorial thermocline, ocean heat content in the upper 700 m, and sea level.

2. The NEMOVAR system

2.1. General formulation

NEMOVAR is a variational data assimilation system designed to produce an analysis of the global ocean. It is based on the OPAVAR data assimilation system (Weaver *et al.*, 2003, 2005; Daget *et al.*, 2009), but has been rewritten to be compatible with a newer version of the ocean model (NEMO), to run on parallel computer architectures, and to assimilate new datasets.

NEMOVAR is designed as an incremental four-dimensional variational assimilation (4D-Var) algorithm (Courtier *et al.*, 1994). Incremental three-dimensional variational assimilation (3D-Var) is also supported, using the First-Guess at Appropriate Time (FGAT) approach, as a simpler and cheaper alternative to 4D-Var. ORAS4 is based on 3D-Var FGAT. Let

$$\mathbf{y}^{\mathrm{o}} = \left\{ (\mathbf{y}_0^{\mathrm{o}})^{\mathrm{T}} \cdots (\mathbf{y}_i^{\mathrm{o}})^{\mathrm{T}} \cdots (\mathbf{y}_N^{\mathrm{o}})^{\mathrm{T}} \right\}^{\mathrm{T}}$$

be a 4D vector of observations where $\mathbf{y}_i^{\mathrm{o}}$ denotes the observation vector at time t_i and $t_0 \leq t_i \leq t_N$ defines the period of the assimilation cycle. The superscript $^{\mathrm{T}}$ denotes the transpose. On each cycle, the 3D-Var FGAT analysis is obtained by minimizing a quadratic cost function of the form

$$J[\delta\mathbf{w}] = \frac{1}{2}\delta\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{w} + \frac{1}{2}(\mathbf{G}\delta\mathbf{w}-\mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{G}\delta\mathbf{w}-\mathbf{d}),$$

*http://www.combine-project.eu

where $\delta\mathbf{w} = \mathbf{w} - \mathbf{w}^{\mathrm{b}}$ is an increment to the vector $\mathbf{w}$ of analysis control variables, $\mathbf{w}^{\mathrm{b}}$ is the background estimate of $\mathbf{w}$, $\mathbf{d} = \mathbf{y}^{\mathrm{o}} - G(\mathbf{w}^{\mathrm{b}})$ is a 4D vector of observation-minus-background state differences (innovations), and $\mathbf{B}$ and $\mathbf{R}$ are symmetric and positive-definite matrices that contain estimates of the background- and observation-error covariances, respectively. The nonlinear operator G is a generalized observation operator that maps the control vector into the space of the observation vector. It has the form

$$G(\mathbf{w}) = \begin{pmatrix} \vdots \\ G_i(\mathbf{w}) \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ H_i[M(t_i, t_0)\{K(\mathbf{w})\}] \\ \vdots \end{pmatrix}, \qquad (1)$$

where K is a balance operator that transforms $\mathbf{w}$, defined at t_0, into an initial state vector $\mathbf{x}(t_0)$ for the ocean model; $M(t_i, t_0) = M(t_i, t_{i-1}) \cdots M(t_1, t_0)$ is the nonlinear model operator that propagates $\mathbf{x}(t_0)$ to time t_i; and H_i is the observation operator that transforms $\mathbf{x}(t_i)$ into the measured quantities at the appropriate location. To compute $\mathbf{d}$, G is applied to the background field $\mathbf{w}^{\mathrm{b}} \equiv K^{-1}\{\mathbf{x}^{\mathrm{b}}(t_0)\}$.

In NEMO, the state vector $\mathbf{x}$ comprises the variables $(T, S, \eta, u, v)^{\mathrm{T}}$; i.e. potential temperature, salinity, sea-surface height (SSH) and the horizontal components of velocity.[†] The control variables comprising $\mathbf{w}$ are constructed so that their cross-covariances are much weaker than those of the state variables. In NEMOVAR, the control variables are taken to be $(T, S_{\mathrm{U}}, \eta_{\mathrm{U}}, u_{\mathrm{U}}, v_{\mathrm{U}})^{\mathrm{T}}$ where the variables with a subscript U correspond to an *unbalanced* component of that variable. The cross-covariances between these transformed variables are assumed negligible, so that $\mathbf{B}$ has a simplified, univariate structure involving block-diagonal components associated with each of the control variables. The balance operator is described in section 2.2 and Appendix A, while the univariate covariance model is described in section 2.3.

The linearized generalized observation operator

$$\mathbf{G} = \begin{pmatrix} \vdots \\ \mathbf{G}_i \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \mathbf{H}_i\, \mathbf{M}(t_i, t_0)\, \mathbf{K} \\ \vdots \end{pmatrix}, \qquad (2)$$

where $\mathbf{H}_i$, $\mathbf{M}(t_i, t_0)$ and $\mathbf{K}$ are linearizations of the observation operator, model propagator, and balance operator, respectively, is assumed to satisfy the approximation

$$G(\mathbf{w}^{\mathrm{b}} + \delta\mathbf{w}) \approx G(\mathbf{w}^{\mathrm{b}}) + \mathbf{G}\,\delta\mathbf{w}.$$

The major simplification in 3D-Var FGAT is to assume that $\mathbf{M}(t_i, t_0) = \mathbf{I}$, the identity operator. The same increment is thus used to compute the linear model equivalent of the innovations at all times in the assimilation window, but the background state is still evolved with the full nonlinear model to compute the innovations themselves.

The minimization of J is performed iteratively using a Lanzcos implementation of the conjugate gradient algorithm (Fisher, 1998), preconditioned by the $\mathbf{B}$ matrix (Weaver *et al.*, 2003). After the final iteration, the analysis increment in the space of the model initial conditions is approximated by applying the linearized balance operator to the iterative solution $\delta\mathbf{w}^{\mathrm{a}}$:

$$\delta\mathbf{x}^{\mathrm{a}} = K\left(\mathbf{w}^{\mathrm{b}} + \delta\mathbf{w}^{\mathrm{a}}\right) - K\left(\mathbf{w}^{\mathrm{b}}\right) \approx \mathbf{K}\,\delta\mathbf{w}^{\mathrm{a}},$$

where

$$\delta\mathbf{w}^{\mathrm{a}} \approx \mathbf{B}\,\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\,\mathbf{B}\,\mathbf{G}^{\mathrm{T}} + \mathbf{R}\right)^{-1}\mathbf{d}. \qquad (3)$$

Directly initializing the model with the background state corrected by an analysis increment from 3D-Var can lead to unphysical fast adjustment processes. The balance operator can alleviate this problem to some extent but is generally not sufficient. The Incremental Analysis Update (IAU; Bloom *et al.*, 1996) procedure is used in NEMOVAR to reduce the initialization shock. With IAU, the analysis increment is included as an additional forcing term in the prognostic equations. The corrected model trajectory through the assimilation cycle $t_0 \le t_i \le t_N$ can thus be expressed as

$$\mathbf{x}^{\mathrm{a}}(t_i) = M(t_i, t_{i-1})\left[\mathbf{x}^{\mathrm{a}}(t_{i-1}),\, F_i\,\delta\mathbf{x}^{\mathrm{a}}\right], \qquad (4)$$

where $\mathbf{x}^{\mathrm{a}}(t_0) = \mathbf{x}^{\mathrm{b}}(t_0)$ and F_i is a weighting function such that $\sum_{i=1}^{N} F_i = 1$. The forcing term acts as a low-pass time filter.

2.2. *Balance operator*

Central to the multivariate formulation of the 3D-Var analysis is the linearized balance operator $\mathbf{K}$ that transforms the increments $\delta\mathbf{w} = \left(\delta T, \delta S_{\mathrm{U}}, \delta\eta_{\mathrm{U}}, \delta u_{\mathrm{U}}, \delta v_{\mathrm{U}}\right)^{\mathrm{T}}$ for the control variables into increments $\delta\mathbf{x} = (\delta T, \delta S, \delta\eta, \delta u, \delta v)^{\mathrm{T}}$ for the initial state variables. The fundamental role of the adjoint of the balance operator, $\mathbf{K}^{\mathrm{T}}$, in determining the multivariate response of the 3D-Var analysis is evident from (3) (and (2)) which shows that the analysis increment is a linear combination of the columns of the matrix $\mathbf{B}\,\mathbf{K}^{\mathrm{T}}$.

The balance equations used in NEMOVAR can be summarized as

$$\left.\begin{array}{lll} \delta T &= \delta T & \\ \delta S &= \mathrm{K}^{\mathrm{b}}_{S,T}\,\delta T + \delta S_{\mathrm{U}} & = \delta S_{\mathrm{B}} + \delta S_{\mathrm{U}} \\ \delta\eta &= \mathrm{K}_{\eta,\rho}\,\delta\rho \;+ \delta\eta_{\mathrm{U}} & = \delta\eta_{\mathrm{B}} + \delta\eta_{\mathrm{U}} \\ \delta u &= \mathrm{K}_{u,p}\,\delta p \;+ \delta u_{\mathrm{U}} & = \delta u_{\mathrm{B}} + \delta u_{\mathrm{U}} \\ \delta v &= \mathrm{K}_{v,p}\,\delta p \;+ \delta v_{\mathrm{U}} & = \delta v_{\mathrm{B}} + \delta v_{\mathrm{U}} \end{array}\right\} \qquad (5)$$

where

$$\left.\begin{array}{lll} \delta\rho &= \mathrm{K}^{\mathrm{b}}_{\rho,T}\,\delta T \;+ \mathrm{K}^{\mathrm{b}}_{\rho,S}\,\delta S \\ \delta p &= \mathrm{K}_{p,\rho}\,\delta\rho \;+ \mathrm{K}_{p,\eta}\,\delta\eta \end{array}\right\}$$

are relations for the diagnostic variables of density and pressure, respectively. The balance formulation in ORAS4 follows Weaver *et al.* (2005), with some modifications to the density and SSH balance as outlined below.

The operator $\mathrm{K}^{\mathrm{b}}_{S,T}$, and the operator pair $(\mathrm{K}^{\mathrm{b}}_{\rho,T}, \mathrm{K}^{\mathrm{b}}_{\rho,S})$, are the linear balance transformations from temperature into salinity, and from temperature and salinity into density. The superscript $^{\mathrm{b}}$ indicates that these operators have been linearized about the background state. Variables with a subscript $_{\mathrm{B}}$ ($_{\mathrm{U}}$) denote the balanced (unbalanced)

[†]In this notation, the individual components of the state and control vectors are scalar functions of the spatial coordinates.

component of that variable. It is important to remark that the unbalanced component of a given variable can contribute to the balanced component of another variable following it in the balance sequence; it is only strictly independent of the variables *preceding* it in the balance sequence. For example, in (5), δS_{U} is independent of δT but is related to $\delta\eta$ through $\delta\rho$.

The balance relationships in NEMOVAR are derived from physical constraints, some of which are flow dependent. Following Troccoli and Haines (1999), the linearized transformation $\mathrm{K}^{\mathrm{b}}_{S,T}$ between increments of temperature and salinity is a local adjustment designed to preserve approximately the water-mass properties of the background state by making vertical displacements of salinity in response to changes in temperature. The SSH increment is partitioned into a *baroclinic* component, which is related to the local subsurface temperature and salinity increments, and a *barotropic* component, which is related to the barotropic velocity. The *baroclinic* formulation has similarities to the Cooper and Haines (1996) scheme, by which changes in the SSH correspond to vertical displacements of the water column. This is achieved by explicit use of the background estimate of the Brunt–Väisälä frequency $(N^{\mathrm{b}})^2$ in the density balance, which encourages density increments in strongly stratified regions when assimilating altimeter data and helps preserve static stability. The temperature, salinity and SSH increments influence the velocity fields by means of geostrophic balance, which near the Equator is imposed using a β-plane formulation. Near coastlines, where nonlinear and frictional terms are important, the geostrophic balance is relaxed. A more detailed description of the NEMOVAR balance operator is given in Appendix A and in MBW12.

2.3. Background-error covariances

The observations that are assimilated in ORAS4 are temperature and salinity profiles, and altimeter-derived sea-level anomalies. There is no information on velocity other than that provided by the balance operator, so the unbalanced velocity vector $\delta\mathbf{u}^{\mathrm{h}}_{\mathrm{U}}$ will remain unchanged from its initial zero value and thus can be ignored. The active control variables are thus δT, δS_{U} and $\delta\eta_{\mathrm{U}}$.

The background-error covariance matrix $\mathbf{B}$ is assumed to be block-diagonal with respect to these variables. Each block $\mathbf{B}_X$, where $X = T$, S_{U} or η_{U}, can be factored in standard form as

$$\mathbf{B}_X \;=\; \mathbf{D}_X^{1/2}\,\mathbf{C}_X\,\mathbf{D}_X^{1/2},$$

where $\mathbf{D}_X^{1/2}$ is a diagonal matrix of standard deviations, and $\mathbf{C}_X$ is a correlation matrix. The specification of the standard deviation matrices $\mathbf{D}_X^{1/2}$ is described in Appendix B.

The correlations are assumed to be approximately Gaussian and modelled in grid-point space using a normalized diffusion operator (Weaver and Courtier, 2001). For T and S_{U}, which are 3D fields, the diffusion operator is formulated as a product of a 2D horizontal diffusion operator acting along geopotential surfaces (z levels) and a 1D vertical diffusion operator acting perpendicular to these surfaces. For η_{U}, only the 2D horizontal diffusion operator is needed. The length-scales are controlled by the diffusion tensor, which is assumed to be spatially varying but diagonal with respect to the model coordinates. Near the Equator, where the horizontal model coordinates are approximately geographic, different horizontal length-scales have been used in the zonal and meridional directions to accommodate anisotropic features, as outlined in Appendix B.

2.4. Accounting for model bias

An algorithm to deal with model bias in sequential data assimilation has been implemented in NEMOVAR. It follows the one-step bias-estimation algorithm of Balmaseda *et al.* (2007a) where the bias variables are allowed to be different from the state vector variables. The algorithm includes a model for the time evolution of the bias. The model consists of two terms: a bias term $\overline{\mathbf{b}}$ estimated *a priori*, which can account for seasonal variations, and a bias term $\mathbf{b}'_c$ estimated online, which is updated each analysis cycle c. The total bias $\mathbf{b}_c$ is then given by

$$\mathbf{b}_c \;=\; \overline{\mathbf{b}} + \mathbf{b}'_c. \tag{6}$$

In NEMOVAR, the bias vector has four distinct components $\mathbf{b} = (b^{\mathrm{tr},T}, b^{\mathrm{tr},S}, b^{p,T}, b^{p,S})^{\mathrm{T}}$, where $b^{\mathrm{tr},T}$ and $b^{\mathrm{tr},S}$ are the bias correction terms used to adjust the tendencies in the tracer equations, and $b^{p,T}$ and $b^{p,S}$ are the bias corrections terms used to adjust the horizontal pressure gradient terms in the momentum equations.

The bias control vector on cycle c is estimated from the analysis increments as

$$\mathbf{b}'_c \;=\; \alpha\,\mathbf{b}'_{c-1} - \mathbf{A}\delta\mathbf{x}^{\mathrm{a}}_{c-1}, \tag{7}$$

where α is a memory factor that also determines a time-scale for the time evolution of the online-estimated bias term, $\mathbf{A}$ is a linear transformation from the state vector increment $\delta\mathbf{x}$ to the bias control vector, and $\mathbf{b}'_0 = \mathbf{0}$.

The *a priori* term $\overline{\mathbf{b}}$ can be seasonally varying. It has the potential to provide a smoother analysis by preventing abrupt changes in the analysis associated with the introduction of new observing systems.

The bias correction is used to modify the tendencies of the nonlinear model, so the time evolution of the background and corrected (analyzed) states can be expressed as

$$\begin{aligned}\mathbf{x}^{\mathrm{b}}_c(t_i) &= M(t_i, t_{i-1})\,[\,\mathbf{x}^{\mathrm{b}}_c(t_{i-1}),\,\mathbf{b}_{c-1}\,],\\ \mathbf{x}^{\mathrm{a}}_c(t_i) &= M(t_i, t_{i-1})\,[\,\mathbf{x}^{\mathrm{a}}_c(t_{i-1}),\,\mathbf{b}_{c-1},\,F_i\,\delta\mathbf{x}^{\mathrm{a}}_c\,],\end{aligned}$$

where the second equation is a generalization of (4).

Explicit details on the estimation of the parameters of the bias correction scheme are given MBW12. Here we summarize the main ideas. The term $\overline{\mathbf{b}}$ used in ORAS4 has been estimated using observational information from the Argo period (2000–2008), and then applied from the start of the reanalysis period in September 1957. In this way, the Argo information is extrapolated retrospectively. This assumes that the model bias is largely stationary (except for the seasonal dependence). Specifically, the offline bias term was estimated as the 2000–2008 monthly climatology of the total bias resulting from a data assimilation integration equivalent to ORAS4, but with the offline bias correction defined by the damping terms from a 1-year relaxation to climatology.

The online term $\mathbf{b}'$ is used to adjust the $\overline{\mathbf{b}}$ values and stabilize the solution. In ORAS4 the memory factor α is

equivalent to an e-folding time of 5 years. The elements of the matrix **A**, which specify the partition of the assimilation increments into the different bias control variables, are a function of latitude, to ensure that at low latitudes the dominant bias term is the pressure correction (Bell *et al.*, 2004), and that the terms for direct correction of temperature and salinity are weak at the Equator. There is no obvious methodology for the optimal estimation of the bias parameters. Instead, these have been determined by tuning. Several combinations of parameters were evaluated in long reanalysis experiments (about 15 years long). Criteria such as stability of the bias estimate and fit to the assimilated and independent data were used for evaluation of the results.

The temperature bias correction applied in ORAS4 can be seen in Figure 1. Figure 1(a) shows the spatial structure of the offline temperature bias $\overline{\mathbf{b}}$ for the depth range 300–700 m, exhibiting largest values by the western boundary currents. Figures 1(b, c) show vertical profiles of the temporal mean and standard deviation of the globally averaged temperature bias $\mathbf{b_c} = \overline{\mathbf{b}} + \mathbf{b_c}'$, as well as the temporal mean of the offline bias term $\overline{\mathbf{b}}$. The statistics have been computed for the period 1960–2009, and the profiles show the depth ranges 0–1000 m and 1000–5000 m. The mean of $\mathbf{b_c}$ is not too different from $\overline{\mathbf{b}}$, suggesting that the offline term is a good guess for the mean bias. The standard deviation of $\overline{\mathbf{b}}$ is small compared with the mean bias, indicating the temporal stability of the bias correction applied.

3. The ORAS4 reanalysis

ORAS4 uses version 3.0 of the NEMO ocean model (Madec, 2008) in the so-called ORCA1 horizontal discretization. ORCA is the generic name that refers to the tripolar grids used by the NEMO model; the ORCA1 configuration corresponds to a horizontal resolution of 1° in the Extratropics and refined meridional resolution in the Tropics with a minimum value of 0.3° directly at the Equator. The version of ORCA1 used in ORAS4 has been developed jointly by the National Oceanography Centre, Southampton (NOCS) and the Met Office. It has 42 vertical levels, 18 of which are in the upper 200 m. The first level has a 10 m thickness. The vertical discretization scheme uses partial steps to have better representation of the flow over steep topography. A weak (20-year time-scale) relaxation to temperature and salinity climatological values from the World Ocean Atlas 2005 (WOA05; Antonov *et al.*, 2006; Locarnini *et al.*, 2006) is applied throughout the water column.

The analysis cycle in ORAS4 is 10 days; every 10 days, the NEMO model is integrated forward forced by daily surface fluxes, relaxed to sea-surface temperature (SST) and bias corrected to produce the first guess and background trajectory. The model equivalent of each available observation is calculated to construct the innovation vector, and a quality control (QC) of the observations is performed. This information is the input for the 3D-Var minimization (section 2). In the final phase of the analysis cycle, the assimilation increment computed by 3D-Var is applied using IAU with constant weights ($F_i = 1/N$ in (4)) during a second model integration spanning the same time window as for the first guess.

NEMOVAR assimilates temperature and salinity profiles, and along-track altimeter-derived sea-level anomalies. In addition, information of SST and global mean sea-level variations is used to modify the heat and fresh-water budget, respectively. Figure 2 shows schematically the different datasets used for the production of ORAS4.

It is important to evaluate the impact of assimilation in ORAS4 by comparing it with a simulation that does not assimilate data. This simulation, called the control integration (CNTL), uses the same spin-up, forcing fields, SST/sea-ice relaxation and relaxation to climatology (with 20-year time-scale) as ORAS4.

The CNTL global mean sea level is also constrained by the altimeter-derived global mean values, following the same procedure as in ORAS4. No additive bias correction is applied, since the bias correction is part of NEMOVAR. The observation operators are applied in the same way as in ORAS4 in order to compare the CNTL fields with the observations that were assimilated in ORAS4.

3.1. *Forcing fields, SST and sea ice*

The ocean model is forced by atmospheric-derived daily surface fluxes, instead of being computed using a bulk formula within NEMO. Daily fluxes of solar radiation, total heat flux, evaporation-minus-precipitation and surface wind stress are taken from the ERA-40 reanalysis (Uppala *et al.*, 2005) from September 1957 to December 1989, ERA-Interim reanalysis (Dee *et al.*, 2011) from January 1989[‡] to December 2009, and the ECMWF operational archive from January 2010 onwards.

The heat fluxes from the atmospheric reanalysis are corrected using SST information. A strong relaxation to gridded SST products is applied as a flux correction. The relaxation coefficient is $200\,\mathrm{W\,m^{-2}{}^{\circ}C^{-1}}$, equivalent to about a 2–3-day time-scale over a depth of 10 m. Sea-ice concentration (SIC) data are also used to adjust the SST in polar regions: if the SIC is higher than a certain threshold (55%), the model SST is set to freezing point. Equally, if SIC is below the given threshold (i.e. there is no ice in the observations) but the model SST is below freezing point, the relaxation term to observed SST is increased. The use of strong nudging to assimilate SST is not optimal, and has the potential to degrade the mixed-layer depth, especially in those areas where the background temperature errors are not due to surface fluxes but to deficiencies in the ocean model. A better approach would be to assimilate the SST maps directly through the observation term in the cost function, although to do so effectively would require an adequate representation of the error correlations in the SST maps which is not straightforward. Alternatively, the along-track satellite SST products could be assimilated directly; this will be possible for future reanalyses as quality bias-corrected products become available.

The fresh-water flux is also adjusted using ocean observations (i) globally, by constraining the global model sea-level changes to the changes given by the altimeter data, and (ii) locally, via a relaxation to monthly climatology of surface salinity from WOA05, with a time-scale of about 1 year.

In the early period of the ocean reanalysis (from September 1957 to November 1981) the SST/SIC are taken from the ERA-40 archive. From December 1981 to

[‡]The ERA-interim extension back to 1979 did not exist when the production of ORAS4 started.

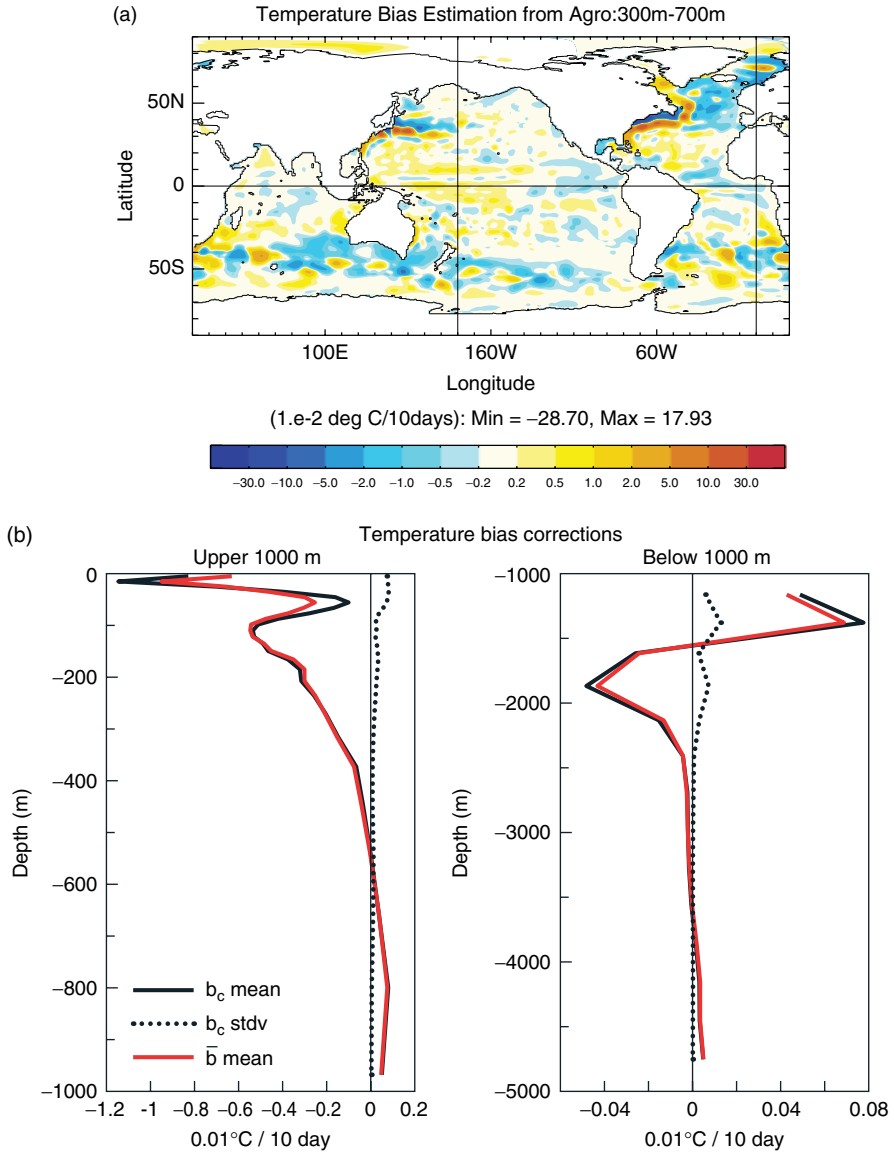

Figure 1. (a, b) show profiles of the mean temperature bias terms $\overline{\mathbf{b}}$ (red curves) and $\mathbf{b_c}$ (solid black curves), and the standard deviation of $\mathbf{b_c}$ (dashed curves), for the depth ranges (a) 0–1000 m and (b) 1000–5000 m. Units are 0.01 C/10 days.

December 2009, SST/SIC are taken from the NCEP OI_v2 weekly products (Reynolds *et al.*, 2002), and from January 2010 onwards from the OSTIA SST/SIC products (Stark *et al.*, 2007). There was some choice regarding the SST/SIC product to use, and several SST products were compared. Figure 3 shows time series of globally averaged SST from different products:

OIv2_1x1: NOAA global SST weekly product, 1° resolution.[§] Available from late 1981 onwards. This is the SST product currently used in the operational seasonal/monthly forecasting system at ECMWF. Black line in Figure 3.

OIv2_025_d1: NOAA global SST daily product, 0.25° resolution.* Available from late 1981 onwards. It uses only one instrument, AVHRR. Red line in Figure 3.

OIv2_025_d2: NOAA global SST daily product, 0.25° resolution.* Available from late 2002 onwards. It uses two instruments, AVHRR and AMSR. Blue line in Figure 3.

Ersstv3b: NOAA Extended Reconstruction of monthly SST, 2° resolution.[¶] Available since 1880. Green line in Figure 3.

CMIP5_proto: SST monthly maps recommended by the CMIP5 protocol. It is a blend of HadISST1 and the NOAA OIv2_1x1 weekly product.[‖] Purple line in Figure 3.

The most obvious feature in Figure 3 is the cold bias of the high-resolution SST products relative to the low-resolution products. The bias is especially pronounced in the case of OIv2_025_d1. This difference in SST has a significant impact in the evolution of the global ocean heat content, as will be discussed in section 3.5. The operational OSTIA product

[§]http://www.esrl.noaa.gov/psd/data/gridded/ data.noaa.oisst.v2.html

[¶]http://www.ncdc.noaa.gov/oa/climate/research/sst/ersstv3.php
[‖]http://pcmdi-cmip.llnl.gov/cmip5/forcing.html#amip
*http://www.ncdc.noaa.gov/oa/climate/research/sst/oi-daily.php

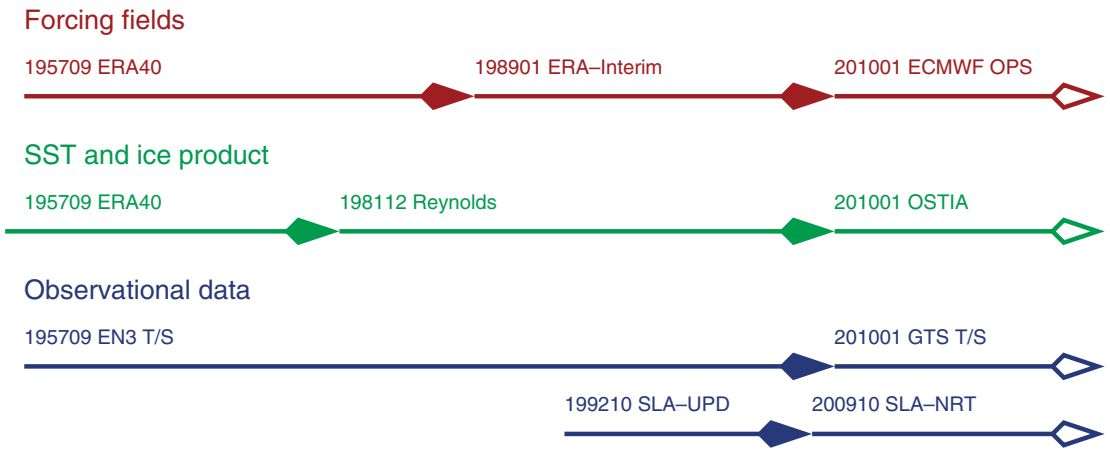

Figure 2. Timeline of changes to the reanalysis forcing and assimilation datasets for ORAS4.

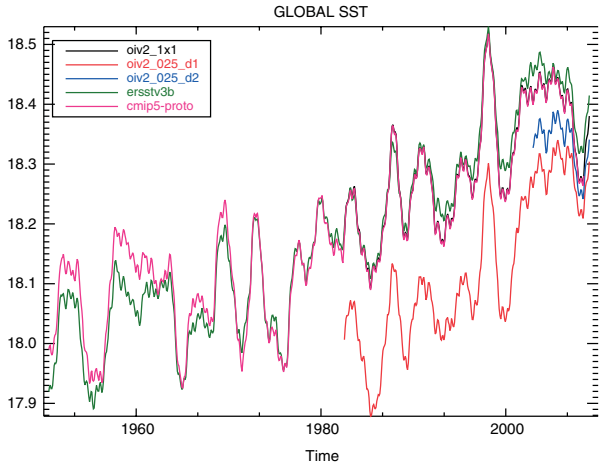

Figure 3. Time series of globally averaged SST (°C) from different products as described in the text. The NOAA OIv2_025_d1 is cold with respect to the other products. The OIv2_025_d2 is still colder than the low-resolution products, but the difference is not so pronounced. The curves for OIv2_1 × 1 and CMIP5_proto are virtually indistinguishable.

(not shown) has similar values to the high-resolution NOAA SST from 2008 onwards.

3.2. *Temperature and salinity profiles*

ORAS4 assimilates different types of temperature and salinity (T/S) profiles: XBTs (T only), Conductivity–Temperature–Depth sensors (CTDs, T/S), TAO/TRITON/PIRATA/RAMA** moorings (T/S), Argo profilers (T/S), and Autonomous Pinniped Bathythermograph (APBs or elephant seals, T/S). For the period September 1957 to January 2010, the profiles are from the quality-controlled EN3_v2a dataset,†† with XBT depth corrections from Wijffels *et al.* (2009).‡‡ In the following we will refer to this dataset as EN3_v2a_xbtc. From January 2010, operational data from the Global Telecommunications System (GTS) are used.

Both EN3 and GTS data are quality controlled by the NEMOVAR QC procedure. The main component of the QC procedure is described by Ingleby and Huddleston (2007), except that the background check here is performed against the model background interpolated T/S values rather than climatology. For the EN3 input data, the TAO/TRITON/PIRATA/RAMA moorings are already available as daily mean super-observations (superobs). For the GTS data, daily mean superobs are constructed for each platform in the same way as for the EN3 data. Some model-dependent QC checks and procedures have been implemented in addition to the main QC procedure, such as vertical thinning of the observations, and rejection of observations on the continental shelves (MBW12).

All observation errors are assumed to be uncorrelated in ORAS4, so that only the observation-error standard deviations (OESDs) need to be specified. In ORAS4, the OESDs $\sigma_X^{\mathrm{o}} = \sigma_X^{\mathrm{o}}(r, z)$, where X represents T or S, depend on the Cartesian distance r to the coast (Eq. (A2) in Appendix A) and depth:

$$\sigma_X^{\mathrm{o}} \;=\; W(r\,;\,\delta_{\mathrm{o}}, R_{\mathrm{o}})\,\widehat{\sigma}_X^{\mathrm{o}}(z),$$

where $\widehat{\sigma}_X^{\mathrm{o}}$ has been constructed to provide an approximate fit to the vertical profiles of globally averaged estimates from Ingleby and Huddleston (2007). For $\widehat{\sigma}_T^{\mathrm{o}}$, the maximum value is at 75 m depth where it is 1 °C compared to 0.75 °C at the surface and the minimum value of 0.07 °C in the deep ocean. For $\widehat{\sigma}_S^{\mathrm{o}}$, the values decrease exponentially with depth from 0.18 psu at the surface to 0.02 psu in the deep ocean (MBW12 gives more details). In ORAS4, the weighting coefficient W has been used with parameter values $\delta_{\mathrm{o}} = 6$ and $R_{\mathrm{o}} = 800$ km. Sensitivity experiments to other values of these parameters have been conducted, as discussed in section 3.5.

3.3. *Assimilation of altimeter-derived sea-level anomalies*

The sea-level data are provided by AVISO (Archiving, Validation and Interpretation of Satellite Oceanographic data). The sea-level data distributed by AVISO (η'^{o}) are anomalies relative to the 7-year period from 1993 to 1999. To enable comparison with the background field (η^{b}) to construct the SSH innovation d_η, a reference Mean Dynamic Topography (MDT; $\overline{\eta}^{\mathrm{o}}$) is required:

$$d_\eta = \left(\eta'^{\mathrm{o}} + \overline{\eta}^{\mathrm{o}}\right) - H(\eta^{\mathrm{b}}),$$

**Tropical Atmosphere Ocean; TRIangle Trans-Ocean buoy Network; Prediction and Research Moored Array in the Atlantic; Research Moored Array for African–Asian–Australian Monsoon Analysis and Prediction

††http://www.metoffice.gov.uk/hadobs/en3

‡‡Labelled on the website as EN3_v2a_NoCWT_WijffelsTable1XBTCorr

where the observation operator H represents horizontal interpolation to the observation point. In ORAS4, $\overline{\eta}^{\,o} = H(\overline{\eta}^{\,a})$ where $\overline{\eta}^{\,a}$ is the 1993–1999 mean sea level from an ocean reanalysis using ORAS4 but assimilating only temperature and salinity data (experiment E–TS). In what follows we refer to it as TS MDT. The possibility of using an alternative reference $\overline{\eta}^{\,o}$ derived from *in situ* observations and data from the Gravity Recovery and Climate Experiment (GRACE) mission has also been explored. In particular, attempts were made to use the MDT from Rio *et al.* (2011), which has large differences relative to TS MDT (not shown). As discussed by Vidard *et al.* (2009), without proper treatment of observation bias, this can introduce abrupt jumps in the analysis.

3.3.1. *Use of along-track SLA data from altimeters*

A superob scheme has been developed to assimilate the high spatial resolution along-track SLA data into the fairly low-resolution model set-up. In this scheme, a superob grid is constructed, with a resolution comparable to that of the model (typically 1° latitude/longitude). For a given point on this superob grid, the observations for the same day which have this grid point as the closest of all superob grid points are collected to form a statistical sample from which the superob is created. The sample mean of the SLA value and space/time positions become the superob SLA value and position. The superob error $\sigma_\eta^{\mathrm{sup}}$ is computed as the standard deviation of the sample. This standard deviation is a measure of the local variability of the SLA measurements at a scale similar to the model resolution. However, if the superob sample is small, this error might be underestimated. To compensate for this, the OESD is bounded below by a minimum value of 0.05 m. The full expression for the SLA OESD also includes the inflation factor near coastlines, as described in MBW12.

3.3.2. *Assimilation of sea-level trends*

There is clear evidence that the global sea level is rising, due to the combined effect of thermal expansion (steric) and mass changes over the ocean (Church and White, 2006). The steric component of the global mean sea level cannot be represented by the ocean model since, in common with most ocean models used for climate activities, the Boussinesq approximation is made, which means that the ocean model preserves volume. Therefore, if not treated correctly, the trend in sea level can be problematic when assimilating altimeter observations. To avoid inconsistencies, the spatial mean of the sea-level background field and of the input sea-level superobs is removed before assimilation.

The information about the global mean sea level is not neglected, however, as it is used to close the fresh-water budget, thus helping with the attribution of sea-level rise. Although the steric height η_s is not a prognostic variable of the ocean model, it can be diagnosed by vertically integrating the density field of the ocean analysis. By comparing trends in the global mean sea level from the altimeter data with the trends in steric height from the ocean analysis, it is possible to estimate the component of global mean sea level change due to mass variations. This approach, implemented in ORAS3 (Balmaseda *et al.*, 2008) is also followed in ORAS4. The information given by the altimeter data maps* about trends in the global mean sea level ($\Delta\overline{\eta}_o$) is compared every assimilation cycle with the trends in the ocean analysis steric height ($\Delta\overline{\eta}_s$). The trends are relative to the values of model SSH and altimeter data at the beginning of the inclusion of altimeter data (i.e. November 1992).

The residual $\Delta\overline{\eta}_m$, where

$$\Delta\overline{\eta}_m = \Delta\overline{\eta}_o - \Delta\overline{\eta}_s,$$

is applied as a spatially uniform fresh-water flux. The partition between volume change and mass change is quite valuable information since it can help to close the fresh-water budget over the oceans, which is currently a large source of uncertainty in the analysis of the ocean. However, uncertainty remains on the spatial distribution of the global fresh-water residual.

3.4. *Spin-up and ensemble generation*

As in the previous ECMWF ocean analysis system, the ORAS4 consists of five ensemble members. These have been generated by adding wind-stress perturbations, the same as those in ORAS3. These perturbations introduce uncertainty mainly in the upper ocean (Balmaseda *et al.*, 2008). The perturbations are added symmetrically to four of the five ensemble members (the perturbed members), leaving a control ensemble member unperturbed. The ensemble generation in ORAS4 has two new components which are designed to sample the uncertainty in both the initial conditions at the start of the reanalysis and the observation coverage.

To sample the uncertainty of the observation coverage, a scheme was developed which can reject a given observation with a certain random probability. For the perturbed members, the probability of rejection is 10% for Argo observations and 5% for other platforms measuring T/S. This procedure attempts to sample uncertainty in the QC decisions.

An aspect often overlooked is the description of the initial conditions at the start of the reanalysis. A common practice is to start from a climatological spin-up. However, the spin-up state is not representative of the state of the ocean at the given point in time. Therefore, there is an error in the initial state of the magnitude of the interannual/decadal variability of the ocean. One would expect the assimilation of ocean data to be efficient in bringing the state close to reality, but this will only be the case if enough observations are available. In the early 1960s, and in the deep ocean, a large uncertainty is likely to remain.

ORAS4 attempts to tackle this issue by the following ensemble-generation strategy. First, an 18-year climatological spin-up is conducted (INI0), where the NEMO model, starting from rest and the WOA05 T/S fields from January climatological conditions, is forced with fluxes from the ERA-Interim climatology. A strong relaxation (1-year time-scale) to WOA05 is employed. A second integration (INI1) from the final state at the end of INI0 is performed using time-varying fluxes and assimilating temperature and salinity, for the period 1958 to 1980. Five different ocean restarts from this second integration, sampled at 5-year

*http://www.aviso.oceanobs.com/en/data/products/sea-surface-height-products/global/index.html

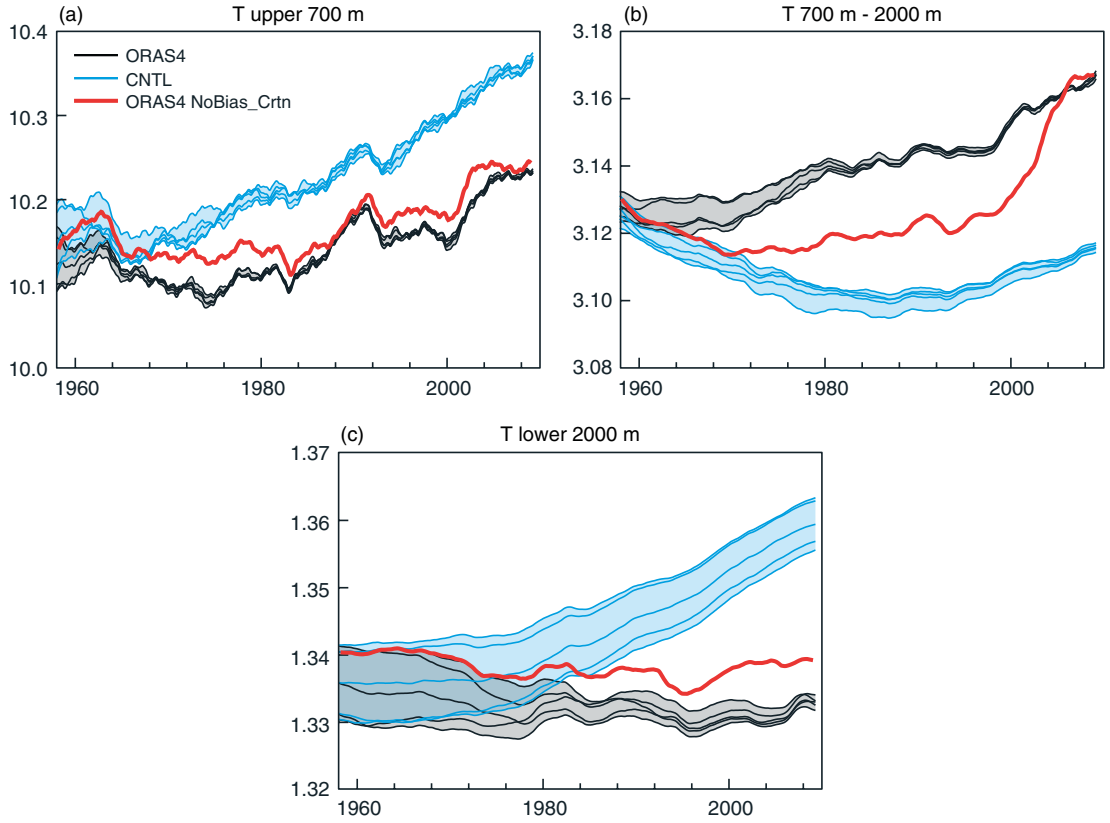

Figure 4. Time series of the globally averaged temperature (°C) for (a) 0–700 m, (b) 700–2000 m, and (c) below 2000 m, showing the five ensemble members of ORAS4, the five ensemble members of CNTL, and the sensitivity experiment ORAS4 NoBias_Crtn, which is equivalent to the unperturbed member of ORAS4 but without bias correction. The bias correction has a noticeable impact in the mean and variability.

intervals from September 1960, are used to initialize each of the ensemble members of ORAS4.

The impact of this second step in the ensemble generation is illustrated in Figure 4. This figure shows the evolution of the globally averaged temperature at different depth ranges. Compared with the interannual variability/trend, the ensemble spread increases with depth. The ensemble spread in ORAS4 is smaller than in CNTL at all depths, but especially in the upper ocean, and can be taken as an indication of the effect of assimilation in reducing uncertainty. In the upper ocean, the increased spread associated with the spin-up uncertainty lasts for approximately 5 years, in both CNTL and ORAS4. At mid-depth range (700 to 2000 m), the memory of the initial perturbations lasts longer. In CNTL the spread is reduced slowly, with the ensemble members tending to converge only near the end of the period. In ORAS4 the spread is reduced faster than in CNTL, as it is affected by the assimilation of observations. The spread is clearly reduced around the 1980s, and it is further reduced with the advent of Argo (after 2000). Below 2000 m, the memory of the initial conditions is long and, in the case of CNTL, the spread remains large (relative to the signal) during the whole period. All CNTL ensemble members drift significantly towards higher temperatures. In contrast, the temperatures of the ORAS4 ensemble members remain relatively stationary, which is possible because of the bias correction. The memory of the initial conditions is shorter with than without assimilation, since the spread is reduced at a faster rate in ORAS4 than in CNTL. As before, there is a sudden reduction of the spread around the 1980s and a gradual reduction thereafter. It is not possible to say if this is due to the assimilation of (sparse) observations at this depth or as a consequence of constraining the ocean above. This figure also reveals different mean states in CNTL and ORAS4.

The effect of the bias correction is also illustrated in Figure 4, where the red line shows results from an experiment equivalent to ORAS4 but without bias correction (ORAS4 NoBias_Crtn). The bias correction affects both the mean and variability, and this effect is not sampled by the ensemble generation. In the upper 700 m, CNTL has a warm bias with respect to ORAS4, and shows a stronger linear trend. A warm bias with respect to ORAS4 is also visible in the experiment without bias correction, albeit with a smaller amplitude than in CNTL. (CNTL and NoBias_Crtn also have a warm bias with respect to the observations, as seen in Figure 6.) The two data assimilation experiments exhibit similar but not identical variability, with nonlinear warming trends characterized by a pronounced warming during the 2000s. In the experiment without bias correction, this warming is sharper and saturates earlier; the saturation values are similar in NoBias_Crtn and ORAS4, as if the observing system at this stage were finally able to constrain the solution. This convergence in the later years leads to an underestimation of the warming trend in the experiment NoBias_Crtn, since it is biased warm in the preceeding years. The impact of the bias correction on the mean and variability is more noticeable in the depth range 700–2000 m. Here both CNTL and NoBias_Crtn have a cold bias with respect to ORAS4. CNTL does not show any warming trend, ORAS4 exhibits a smooth upward trend, and NoBias_Crtn exhibits a pronounced warming only during the Argo period, as it is only at this stage that there are enough observations available to correct the cold bias in the ocean at this depth range.

Table 1. Summary of the sensitivity experiments.

Experiment name	SST	(δ_o, R_o)	$W_B(\phi)$	Altimeter
ORAS4	OIv2_1x1	(6, 800 km)	$\exp\{-(\phi/\phi_o)^2\}$, $\phi_o = 10^\circ$	Yes
NoBias_Crnt	OIv2_1x1	(6, 800 km)	No bias correction	Yes
SST_Coast1	OIv2_025_d1	(20, 800 km)	$\exp\{-(\phi/\phi_o)^2\}$, $\phi_o = 10^\circ$	No
SST_Coast2	OIv2_025_d1	(20, 300 km)	$\exp\{-(\phi/\phi_o)^2\}$, $\phi_o = 10^\circ$	No
SST_Coast3	OIv2_025_d1	(2, 300 km)	$\exp\{-(\phi/\phi_o)^2\}$, $\phi_o = 20^\circ$	No
COMB-NV	OIv2_1x1	(2, 300 km)	$1 - \sin^2(\pi\phi/180^\circ)$	No

Although the values of ORAS4 and NoBias_Crtn converge at the end of the period, the magnitude and timing of the trend is very different in both experiments. Below 2000 m, the three sets of experiments have again very different mean values. This time NoBias_Crtn and ORAS4 do not converge at the end of the record, as could be expected by the lack of observations at this depth range.

As shown in Figure 1, the temporal variablity of the bias term is quite small, suggesting that the impact of the bias correction on the temporal variability of the reanalysis is due mainly to the correction of the mean state.

3.5. *Sensitivity experiments*

Several sensitivity experiments were conducted in the process of choosing the final configuration for ORAS4. These include sensitivity to the SST products, to the standard deviation and length-scales of the background error, the function controlling the weight to the observations near the coast, the bias correction parameters, the use of ERA-40 versus ERA-Interim, and various other specifications. Some of the found sensitivities are worth mentioning, since they have a dramatic impact on some climate indices, and have not been considered in the ensemble generation of ORAS4.

The sensitivity experiments are described in Table 1, and comprise sensitivity to the SST product, the OESD parametrization near the coast, and the choice of latitudinal dependency of the bias correction partition (BCP) into pressure gradient and T/S components (section 2.4). The pressure gradient bias correction term is weighted by the function $W_B(\phi)$ in Table 1, while $\{1 - W_B(\phi)\}$ is the approximate weighting function of the T/S bias correction term (MBW12 provides details). One of these experiments is the COMB-NV reanalysis, which has been used by several groups to initialize decadal forecasts. Table 1 provides a summary of the relevant changes between ORAS4 and COMB-NV. COMB-NV used a different BCP (weaker in T/S; stronger in pressure), reduced OESDs near coasts, and does not assimilate altimeter data (although this particular aspect does not influence the climate indices presented here). All the other sensitivity experiments, with names starting with SST, use the OIv2_025_d1 SST (NOAA global SST daily product, 0.25° resolution) and different specifications of the coastal OESDs. Differences between ORAS4, SST_Coast1 and SST_Coast2 arise from different OESD parametrizations and different SST products as they use the same BCP. Differences between COMB-NV and SST_Coast3 are due to the SST product and BCP. The COMB-NV and SST_Coast3 use different BCP from the other three experiments. Before 1982, all the experiments use the same SST.

Figure 5 shows the time evolution of two climate indices as represented by the five ensemble members of ORAS4 (black lines) and by a selection of sensitivity experiments.

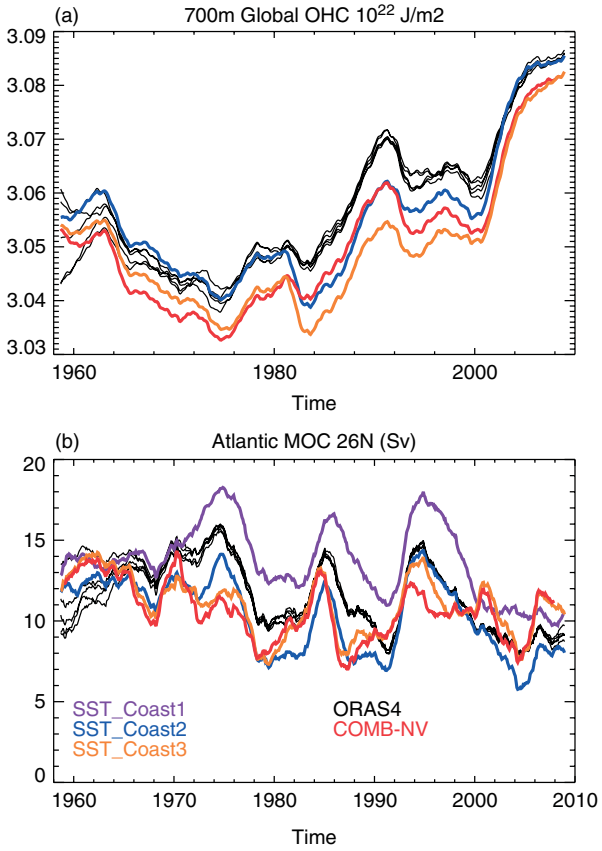

Figure 5. Time series of (a) the upper 700 m global ocean heat content (normalized by the global area; 10^{22} J m^{-2}) and (b) the Atlantic MOC (Sv) at 26°N. The black curve shows the five ensemble members from ORAS4. The red curve shows the COMB-NV reanalysis. The other curves show the sensitivity experiments specified in Table 1. See text for details.

The selected indices are the upper 700 m ocean heat content (OHC700), normalized by the global area, and the Atlantic Meridional Overturning Circulation (MOC) at 26°N (AMOC26N). The OHC700 exhibits a robust warming trend, but its interannual/decadal variability seems sensitive to the choice of SST and, as expected from Figure 4, to the BCP. The impact of BCP in OHC700 is noticeable from the beginning of the experiments, by comparing SST_COAST3 and COMB-NV, which both use the same SST product (from ERA-40) before December 1981. The impact of the different SST products is very clear from 1982 onwards, where the experiments using the OIv2_025_d1 exhibit a decay in the ocean heat content. The differences are maintained for the following two decades, and it is only at the beginning of 2000 that the impact of different SST products in OHC700 becomes small.

The time evolution of the AMOC26N and its sensitivities is interesting. All the experiments exhibit a profound

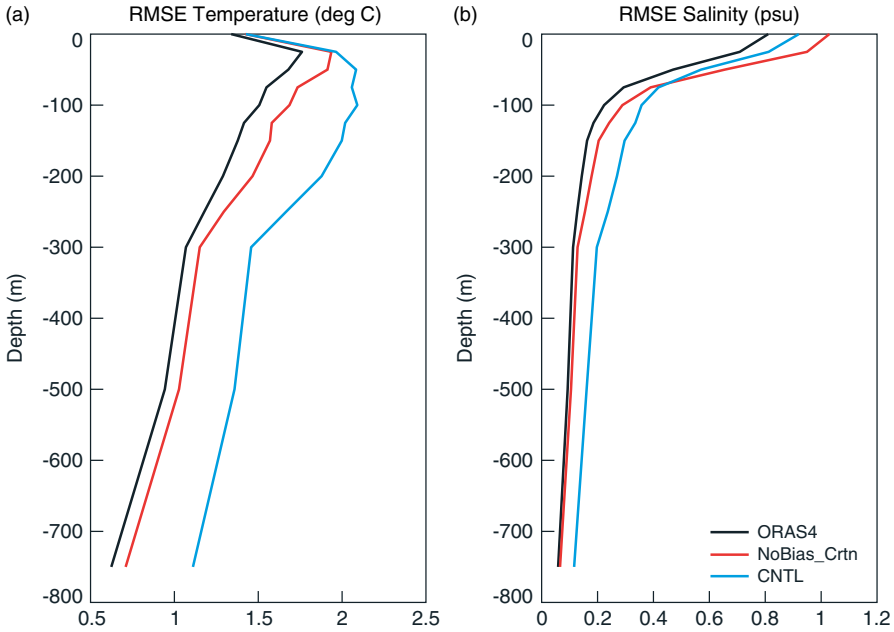

Figure 6. Fit to (a) the temperature (°C) and (b) salinity (psu) observations, as measured by the RMSE. The globally averaged statistics are shown for CNTL (blue), ORAS4 (black) and NoBias_Crtn experiment (red; after the increment has been applied via IAU, as in (4)).

decrease of the AMOC26N after 1995, which remains low for the decade following 2000. Comparison with RAPID data (see next section) indicates that the AMOC26N is underestimated in all the experiments. All experiments also show a marked decadal variability, which is largely (but not completely) in phase. The largest sensitivity comes from the OESD parametrization near coastal boundaries. When the coastal OESDs are large ($\delta_o = 20$) and inflated within 800 km of the coast (SST_Coast1), the AMOC26N amplitude is relatively stable, oscillating around 15 Sv. If the OESDs are equally large but are inflated only within 300 km of the coast (SST_Coast2), the AMOC26N amplitude is substantially weaker. ORAS4 is in between these two experiments. The lowest amplitude of the AMOC26N is for experiments COMB-NV and SST_Coast3, which use the smallest OESDs near the coast.

4. Assessment of ORAS4

The diagnostics presented in this section focus on evaluating the unperturbed ensemble member of ORAS4.

4.1. Comparison with the assimilated observations

4.1.1. Comparison with temperature and salinity data

The quality-controlled EN3 dataset has been used to evaluate the fit of ORAS4 to the observations. Figures 6(a, b) show vertical profiles of the root mean square error (RMSE) of temperature and salinity, respectively. The strong relaxation to SST constrains the temperature at the top of the ocean so the fit is quite similar in all simulations. Largest errors are in the thermocline. The fit to the observations (in terms of bias (not shown) as well as RMSE) is improved in ORAS4 compared to CNTL, the improvement being most visible in the mixed layer and thermocline. The inclusion of bias correction helps to reduce the RMSE in ORAS4.

The largest salinity errors occur at the surface, since the surface salinity is poorly constrained and the impact of erroneous fresh-water fluxes and ocean mixing manifest themselves in errors in the salinity field. As for temperature, data assimilation improves the fit to the salinity observations, but in the upper 100 m this is only achieved by means of the bias correction. The improvement in salinity by data assimilation is more robust below the mixed layer. This improvement is not only due to the direct assimilation of salinity observations, but also to the assimilation of temperature observations, through the balance relations between temperature and salinity.

The results presented here are the cumulative effect of many assimilation cycles during a long integration, and indicate that the assimilation reduces the RMSE in CNTL by about 30%. MBW12 also show statistics of the error growth within the assimilation cycle. They conclude that on the 10-day assimilation window, the assimilation, after the IAU, reduces the RMSE by about 10% with respect to the background.

Time series of the global values of analysis-minus-observation monthly statistics are displayed in Figure 7. (Statistics for other regions are given in MBW12). The time variation in the statistics is due to a mixture of model performance and changes to the observation network, but the relative comparison between CNTL and ORAS4 for any given time should filter the effect of the changes in the observation coverage since the same observations are used. During the first years of the record, all the experiments show similar values of the error statistics, since there are few observations, and CNTL has not drifted significantly from the state used to initialize the simulation.

From around 1970 onwards, the temperature error in CNTL becomes quite substantial, resulting in an increase in both RMSE and bias. In comparison, the RMSE and bias of ORAS4 are kept relatively stable at lower values than those of CNTL, since they are constrained by the assimilation and by the bias correction. ORAS4 errors are also smaller than the NoBias_Crtn experiment. The RMSE in both assimilation experiments has a decreasing trend, showing the cumulative

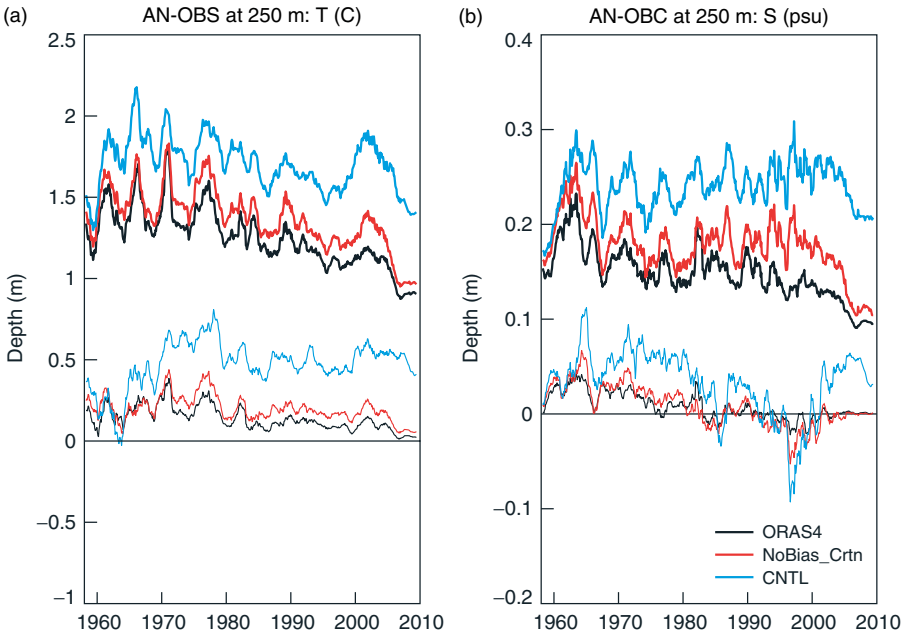

Figure 7. Time series of monthly fit to (a) temperature and (b) salinity observations at depths between 225 and 275 m for the global domain for ORAS4 (black), CNTL (blue) and NoBias_Crtn (red). The RSME (bias) values are shown as thick (thin) lines.

effect of the data in improving the estimate. A similar narrative applies to the errors in salinity (Figure 7(b)).

There is a pronounced decrease in the temperature and salinity error during the Argo period (post 2004), which is also apparent in CNTL, suggesting that the error reduction cannot be attributed only to the increase in the number of observations. These decreasing errors suggest that factors other than data assimilation are important, such as improvement of the forcing fluxes. The reduced error can also result from a more homogeneous spatial distribution of the observations. (With Argo, there are more observations in the open ocean, where the coarse-resolution models have smaller errors.)

4.1.2. *Comparison with altimeter data*

A good fit to the data does not guarantee a good representation of the time variability of the ocean state. The time variability can be gauged by the temporal correlation of the SLA analysis with the altimeter-derived SLA maps provided by AVISO. Figure 8 shows the correlation of monthly means from the period 1993–2008 for three different experiments: CNTL, the experiment E-TS (described in section 3.3) used to derive the MDT (which is equivalent to ORAS4 but without altimeter data assimilation), and ORAS4. It can be seen that the assimilation of T and S profiles improves the correlation with the altimeter data in most of the tropical regions, including the Equatorial Atlantic Ocean. However, there is some degradation in the Northeastern Atlantic by the Iberian Peninsula, for reasons that are not fully understood, although they seem related to the representation of the water masses from the Mediterranean Outflow. As expected, the inclusion of altimeter data in the assimilation further increases the correlation with the AVISO data. The correlation is not improved in the coastal areas (for instance around Hawaii), where the prescribed OESDs are large ((A2) and Table 1), and in high latitudes where the stratification is weak.

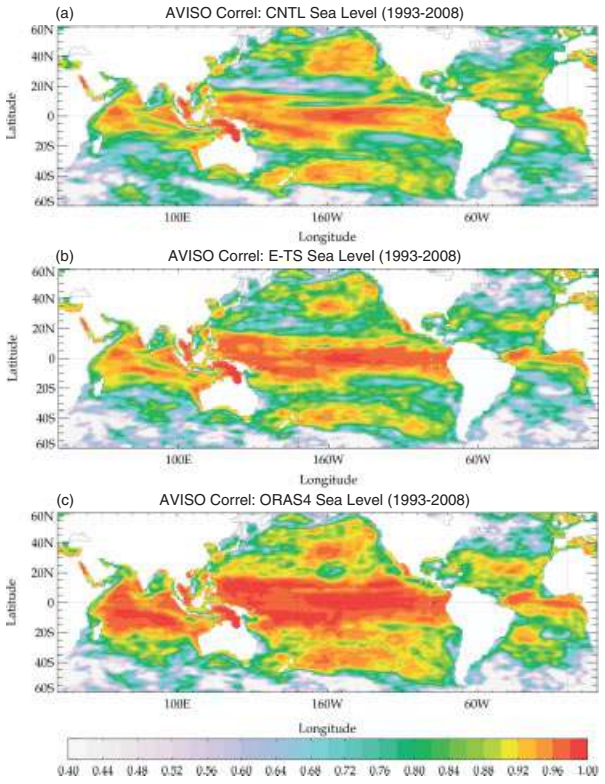

Figure 8. Temporal correlation between analysis and AVISO sea level for (a) CNTL, (b) E-TS, which assimilates T and S but not altimeter data, and (c) ORAS4. The statistics have been computed with monthly means for the period 1993–2008 and values only above 0.4 are shown.

4.2. *Comparisons with independent observations: ocean currents*

Velocity observations are not assimilated in ORAS4, and therefore can be treated as independent data for validation of ORAS4. The assimilation process influences the currents of the model both directly via the analysis increments

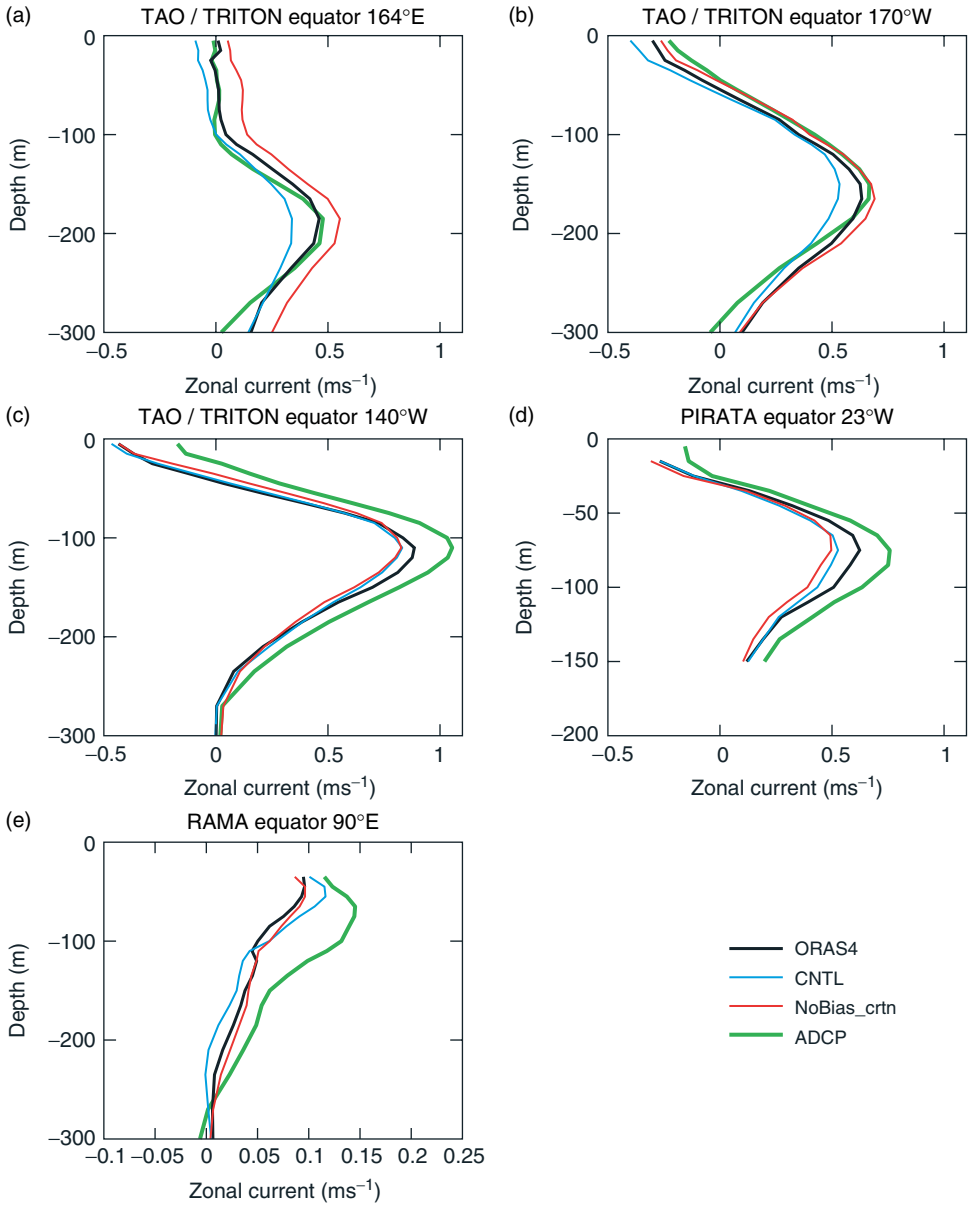

Figure 9. Temporal mean of the zonal current from five moorings in the TAO/PIRATA/RAMA array (green) compared to ORAS4 (black), CNTL (blue) and NoBias_Crtn (red).

(which through the balance relations between temperature, salinity and SSH (Appendix A) includes increments for the horizontal velocity components), and indirectly via the model's response to the IAU (which includes linear and nonlinear adjustments).

4.2.1. *Validation against current-meter data*

Some of the moorings in the Pacific (TAO/TRITON), Atlantic (PIRATA) and Indian Ocean (RAMA) possess current meters and Acoustic Doppler Current Profilers (ADCPs), providing vertical profiles of the currents. These data are available from the RAMA/TAO Project Office of NOAA/PMEL's website[†] and are used here for validation.

Figure 9 shows the temporal mean profile of the zonal currents from ADCP data for three TAO moorings at positions (0°N, 165°E), (0°N, 170°E) and (0°N, 140°W), a PIRATA mooring at (0°N, 23°E) and a RAMA mooring at (0°N, 90°E). Also shown are the corresponding profiles for ORAS4, CNTL and NoBias_Crtn. The figure shows that the assimilation increases the mean value of the Equatorial undercurrent in the Pacific and Atlantic Oceans, improving the fit to the data there. The strength of the Equatorial undercurrent is noticeably improved by the data assimilation in the Western and Central Pacific; the undercurrent is more confined in the vertical in ORAS4 than in CNTL, in better agreement with the observations, and consistent with the data assimilation correcting for an overly diffuse thermocline. In the Eastern Pacific (0°N, 170°E), even though the undercurrent is slightly stronger with assimilation, it is still too weak compared with the observations. In the Central and Eastern Pacific, the surface westward currents are too strong in both CNTL and ORAS4 compared to the observations. The story is similar for the PIRATA mooring, where again the assimilation strengthens the undercurrent although it is still too weak compared to the observations. The RAMA mooring does

[†]http://www.pmel.noaa.gov/tao/

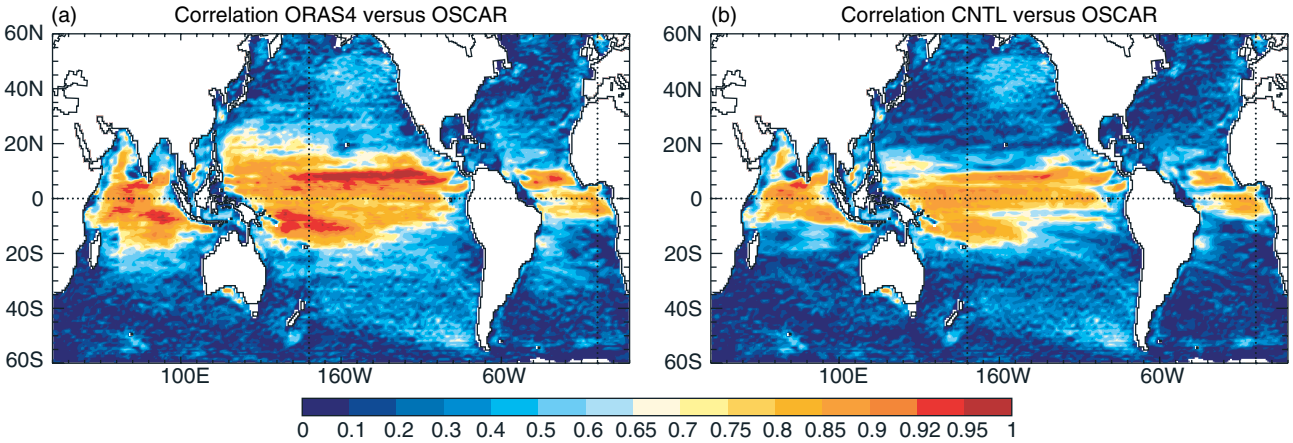

Figure 10. Time correlation with OSCAR surface zonal currents from (a) ORAS4 and (b) CNTL, for the period 1993–2009. Note the nonlinear colour scale.

not show any obvious undercurrent, and here there is very little difference in the mean current profile with and without assimilation.

The impact of the bias correction is also apparent in the Equatorial currents. Without bias correction, the strength of the undercurrent in the Eastern Pacific (140°W) and Atlantic (23°W), is not better than the CNTL. In the Western Pacific (165°E) the bias correction acts to reduce the strength of the zonal current, which would be too strong otherwise. In the Western Pacific (170°W) and Indian Ocean (90°E), the impact of the bias on the currents is not so noticeable.

The time evolution of the zonal current is described in more detail in MBW12. They show that the features of the current variability associated with the El Niño-Southern Oscillation (ENSO) are better captured by data assimilation, such as the weak/strong vertical shear during warm/cold ENSO events in the Western Pacific. The positive impact of assimilation is also visible in the interannual migrations of the Atlantic undercurrent.

4.2.2. *Validation against OSCAR current analyses*

The Ocean Surface Current Analysis –Real-time (OSCAR) project provides analyses of oceanic surface currents derived from satellite altimeter and scatterometer data (Bonjean and Lagerloaf, 2002). They are available from the end of 1992 up to near-real time and now cover the whole ocean from 60°S to 60°N. The OSCAR currents are not completely independent from sea-level data since altimeter data are used in their production. However, we do not use SLA data directly to estimate the currents and therefore there is no guarantee that they should lead to improved velocity analyses. Therefore, it is a good metric to assess the quality of our analysis.

Figure 10 shows the correlation between the zonal component of the surface velocities from OSCAR monthly means and the ORAS4 and CNTL analyses, over the period 1993–2009. The colour scale is nonlinear, and different from that in Figure 8, since the correlation for currents is lower than for sea level. ORAS4 has higher correlation values than CNTL, except at the Equator and along the coasts. Comparison with experiment E-TS (which does not assimilate altimeter data) indicates that, within 10° of the Equator, the increased positive correlation is mainly due to the assimilation of *in situ* data, while the assimilation of SLA improves the currents almost everywhere.

4.3. *Comparison with independent observations: RAPID-derived transports*

The RAPID observations can be used as independent data for validation, allowing comparison of the strength, variability and vertical structure of the Atlantic MOC with observations at 26°N. Processed datasets are available at NOCS.[‡]

Figure 11 shows the Atlantic MOC at 26°N from ORAS4 (black), CNTL (blue) and RAPID (dark green). The total MOC in ORAS4 and CNTL has been estimated following the RAPID method and by directly integrating the meridional velocity. The small difference between the two estimates confirms the validity of the RAPID methodology for MOC estimation, which derives the transport from temperature and salinity profiles at the RAPID locations (Cunningham *et al.*, 2007).

The MOC in ORAS4 and CNTL is substantially lower than in RAPID, and ORAS4 is slightly lower than CNTL. The time series are not long enough to make strong statements about the time variability of the MOC, which is dominated by the seasonal cycle. Therefore, the time series appear relatively consistent in ORAS4, CNTL and RAPID, except for the 2004 peak in RAPID, which is absent in ORAS4 and CNTL. The Ekman components (Figure 11(b)) are very similar, since they are determined by the zonal wind stress, although the Ekman components in ORAS4 and CNTL correspond to the ageostrophic residual rather than being directly calculated from wind stress. ORAS4 and CNTL capture the low MOC values at the end of 2009 and early 2010, and importantly the total MOC–Ekman difference (Figure 11(c)) indicates that this minimum has a geostrophic component as well as an Ekman component.

The vertical profile of the MOC in CNTL and ORAS4, averaged for the period 2004–2009, is shown in Figure 12. Compared to RAPID, the maximum of the MOC is about 100 m and 50 m shallower in CNTL and ORAS4, respectively. However, in both CNTL and ORAS4, the MOC is too weak. The MOC vertical profiles also show that the deep Antarctic Bottom Water (AABW) cell is shallower in ORAS4 and

[‡]http://www.noc.soton.ac.uk/rapidmoc

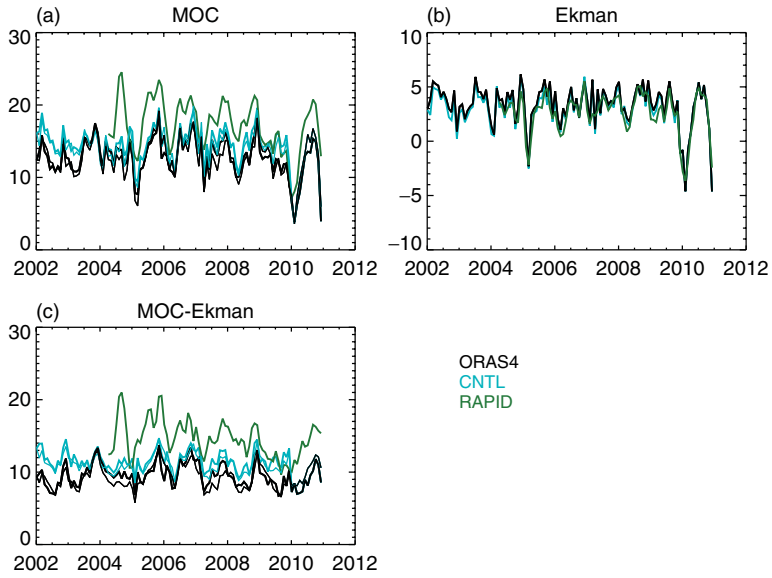

Figure 11. Timeseries of the Atlantic (a) MOC, (b) Ekman component, and (c) MOC minus Ekman (all Sv) from ORAS4 (black), CNTL (blue) and RAPID (green). The total MOC in ORAS4 and CNTL have been estimated following the RAPID method (thick line) and by directly integrating the meridional velocity (thin line).

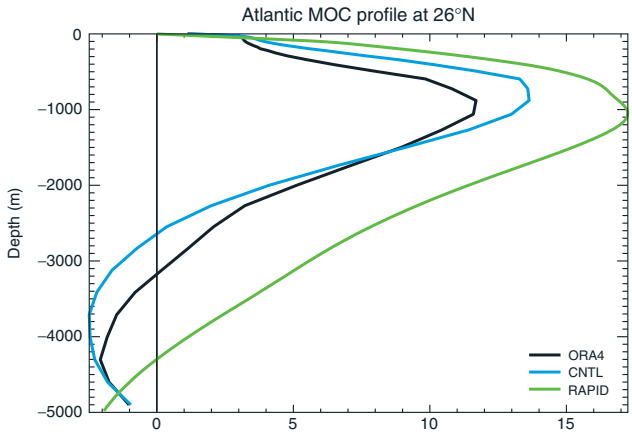

Figure 12. Mean vertical profile of the Atlantic MOC (Sv) at 26°N (Sv) from RAPID (green), ORAS4 (black) and CNTL (blue), averaged over 2005–2009.

CNTL than in RAPID, covering a larger depth range. The shallow and strong AABW cell seems to be model-related and not caused by the data assimilation. Compared with CNTL, ORAS4 has weaker MOC maximum, but it does not overestimate the AABW cell as much as CNTL.

The different MOC components (Ekman or ageostrophic, Gulf Stream or Florida Straits Transport (FST), and Mid-Ocean transport) are shown in Figure 13. Also shown are the FST cable estimates.[§] The time series span the 50-year record (1960–2009), and have been smoothed with a 12-month running mean. The long time series show a large degree of low-frequency variability, highlighting that the current record of RAPID is not yet sufficient to evaluate the decadal variability of the MOC.

The figure shows that underestimation of the MOC comes mainly from the FST component, which is too low in the model-based estimates compared to observations. CNTL appears to overestimate the FST for the first 25 years, but tends toward lower than observed values in the second part of the record. The assimilation in ORAS4 seems to accelerate this tendency. Both ORAS4 and CNTL decline after the mid-1980s. The assimilation values remain slightly lower than those of CNTL. This result seems to contrast with results from the previous ECMWF data assimilation system (ORAS3) with HOPE/OI (Balmaseda *et al.*, 2007b), in which data assimilation was shown to maintain a stronger Atlantic MOC. This is also the case in ORAS4 poleward of 30°N, where the assimilation substantially strengthens the MOC (at 40°N, the average MOC in ORAS4 is 20 Sv, about 8 Sv stronger than in the CNTL; not shown).

Both CNTL and ORAS4 show a decreasing trend in the MOC (stronger in CNTL) which appears to be associated with the FST. The trend in the FST observations is not so obvious. Sensitivity experiments with different initial conditions indicate that the strong decreasing trend in CNTL in the 1960s and 1970s is mainly the artifact of model drift, but the decreasing trend after the 1980s seems quite robust in all the integrations performed with NEMO and NEMOVAR. The figures also show an obvious, albeit small, increasing trend in the Ekman component of the MOC.

The weaker than observed MOC and FST in ORAS4 and CNTL is likely due to the low resolution of the ocean model. It is expected that the underestimation of the FST would be alleviated by increasing the horizontal resolution of the ocean model. The weaker MOC and FST in ORAS4 compared to those in CNTL are related to limitations in the assimilation system, as discussed in section 3.5, where results suggest that the assimilation of observations near coasts needs improvement.

4.4. Comparison with independent observations: sea-level gauges

Tide gauges provide one of the few independent ocean datasets for which a long record is available. These data are thus relevant for validating long ocean reanalyses and potentially important for validating the atmospheric fluxes

[§]www.aoml.noaa.gov/phod/floridacurrent

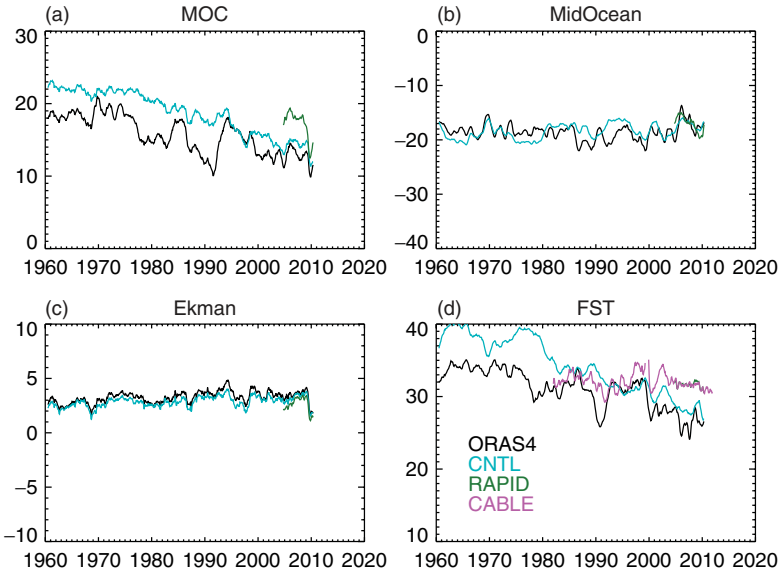

Figure 13. Time series at 26°N of (a) the Atlantic MOC, and its components (b) Mid-Ocean transport, (c) Ekman transport, and (d) Florida Straits transport (all Sv), from ORAS4 (black), CNTL (blue) and RAPID (green). In (d), the FST transport cable estimates are shown in pink.

from long atmospheric reanalyses. In this study, tide-gauge data are used to assess the quality of ORAS4 and CNTL.

A set of 91 tide-gauge records selected from the Permanent Service for Mean Sea Level (PSMSL) database[¶] (Woodworth and Player, 2003) has been used. The sea-level variability in the selected tide gauges is correlated with large-scale sea-level variations, and has been used for historical reconstruction of sea-level changes by Meyssignac *et al.* (2012). The data are yearly means at selected stations. For each station, the sea-level values at model points averaged in a $2° \times 2°$ box centred at that station are chosen for comparison. The time correlation for the period 1960–2009 has been calculated for ORAS4 and CNTL. Figure 14 shows the values of this correlation for ORAS4, and the difference in the correlation values of ORAS4 and CNTL.

The correlations are often large, since they are dominated by the trends. In some areas, like off the west coast of Europe, there are several stations that are clustered together, but have different correlations. This may point to structures in the observations not captured by the coarse resolution of the model, or it could also point to non-representative outliers in the tide-gauge selection. The lowest correlation values appear on the western boundaries. The largest values appear in the interior of the Pacific, where ENSO variability is well captured by the reanalyses. Even if the CNTL global sea-level trend is also constrained by the altimeter observations, the correlation values of ORAS4 are generally superior to those of CNTL. There are a few exceptions, such as the Gulf Stream and Kuroshio current regions, and off the southern tip of India and the coast of Chile.

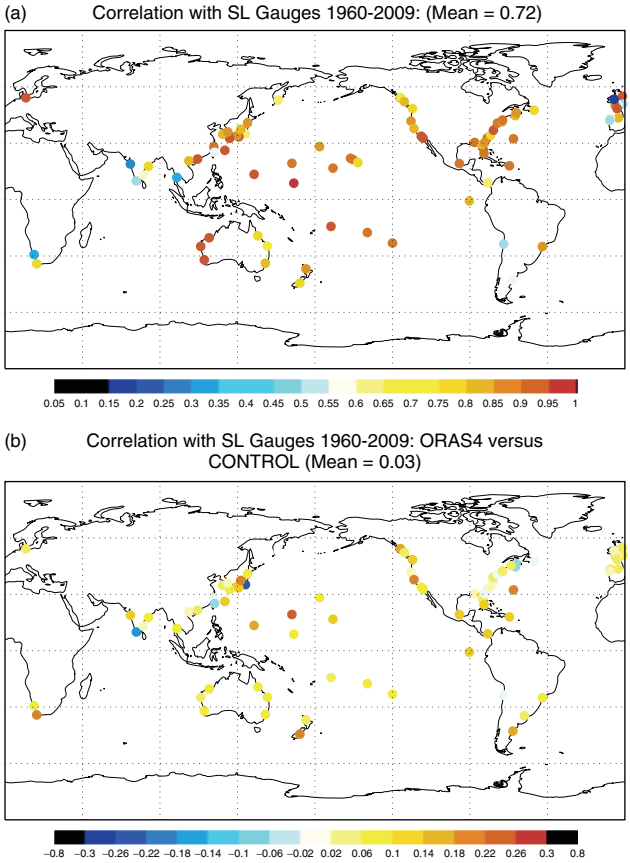

Figure 14. (a) Correlation of ORAS4 with tide-gauge sea-level data for the period 1960–2009 (mean 0.72). (b) shows the correlation values of ORAS4 minus the correlation values of CNTL (mean 0.03).

4.5. Comparison with independent observations: GRACE-derived equivalent bottom pressure

A short record of monthly mean bottom-pressure maps derived from gravity data from the GRACE mission is now publicly available.[‖] Chambers (2006) describes the procedure involved in the conversion of the GRACE gravity signal into mass variations in the ocean, specifically on the equivalent bottom pressure (EBP), i.e. variations in sea level

[¶]http://www.psmsl.org/

[‖]http://grace.jpl.nasa.gov/data/GRACEMONTHLYMASSGRIDSOCEAN

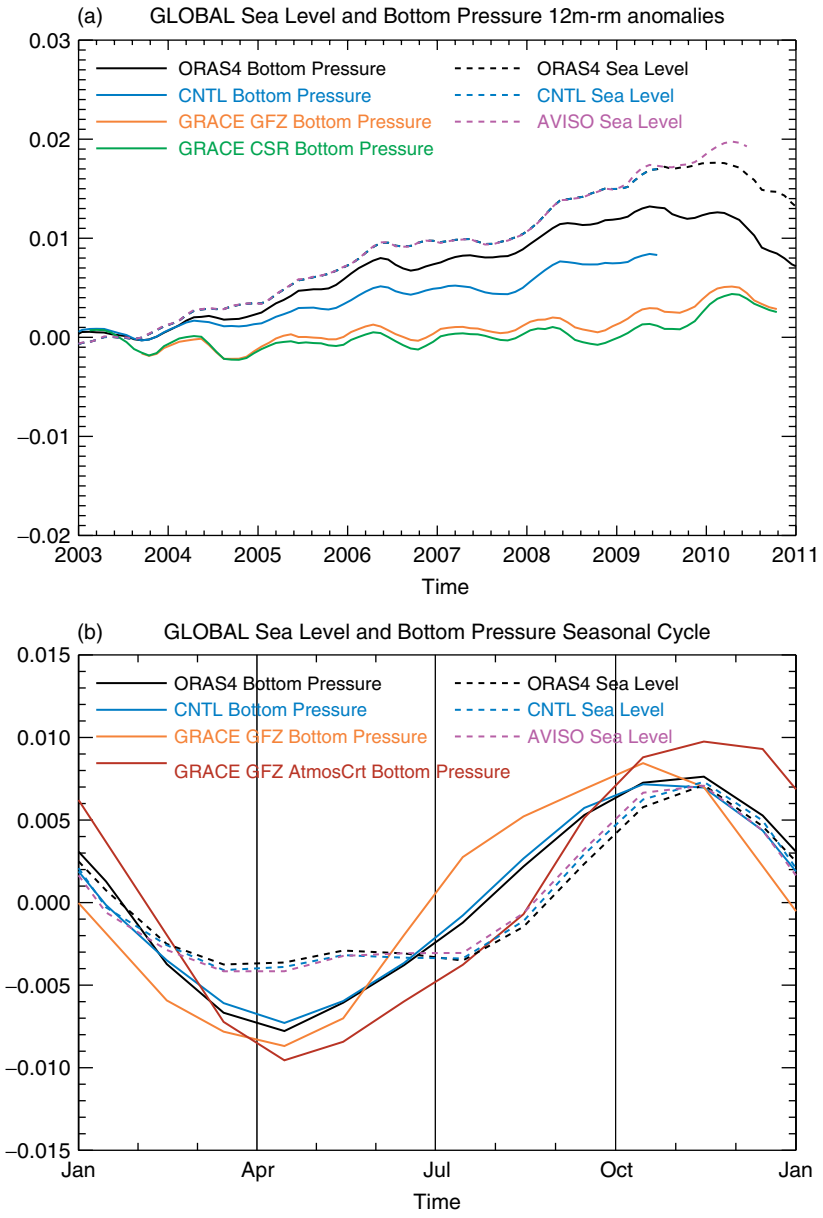

Figure 15. (a) Anomaly and (b) seasonal cycle of equivalent bottom pressure (solid lines) and sea level (dashed lines). The anomalies are smoothed with a 12-month running mean (12m-rm). Shown are the estimates from ORAS4 (black), CNTL (blue), AVISO SL (pink), and the two GRACE products GFZ (orange) and CSR (green). In (b), the seasonal cycle for GRACE GFZ with atmospheric correction is shown in brown. Units are metres.

due to changes in mass over the ocean basins, which also include variations of atmospheric mass. The contribution from atmospheric mass can be removed by subtracting the atmospheric sea-level pressure (SLP) over the oceans from the GRACE estimates (GRACE AtmosCrt in what follows). Here we have used the two versions of GRACE-derived data (specifically GFZ and CSR versions, realease RL04 with a 750 km Gaussian filter), and the ERA-Interim SLP, but the results shown below are similar if estimates from other atmospheric reanalysis products are used (not shown).

In ORAS4/CNTL, the global EBP is computed as the residual between the global sea-level (which should be very similar to the altimeter values) and the global steric height, computed as the vertical integral of the density. We refer to this method as *OceanResidual* in what follows. In this way, the sea-level variations are divided into volume changes (steric changes) and mass changes (visible in the EBP).

Figure 15 shows the time series of the EBP for ORAS4 and CNTL, as well as the two estimates from GRACE. The dashed lines are the time series of the global sea level estimated from the AVISO reference product, ORAS4, and CNTL. The 12-month running mean anomalies (Figure 15(a)) are dominated by a near-linear trend, and the seasonal cycle (Figure 15(b)) which is very similar in both GRACE products; only the GRACE GFZ is shown, together with the estimate GRACE GFZ AtmosCrt. The removal of SLP has little impact on the GRACE ocean mass trends, and it is not shown.

The linear trend in the altimeter sea level for the period 2003–2010 is about 27 mm y^{-1}. There is a large discrepancy between the trends in EBP from the different products. ORAS4 exhibits the largest trend in EBP (17 mm y^{-1}), which would indicate that approximately two thirds of the global sea-level trends in ORAS4 is due to mass changes, and about one third is due to steric changes. In CNTL, the EBP trend is smaller (13 mm y^{-1}). The lowest EBP trend is in the GRACE-derived EBP, which for the period 2003–2010 amounts to 0.43 mm y^{-1} in the GFZ product and 0.57 mm y^{-1} in the

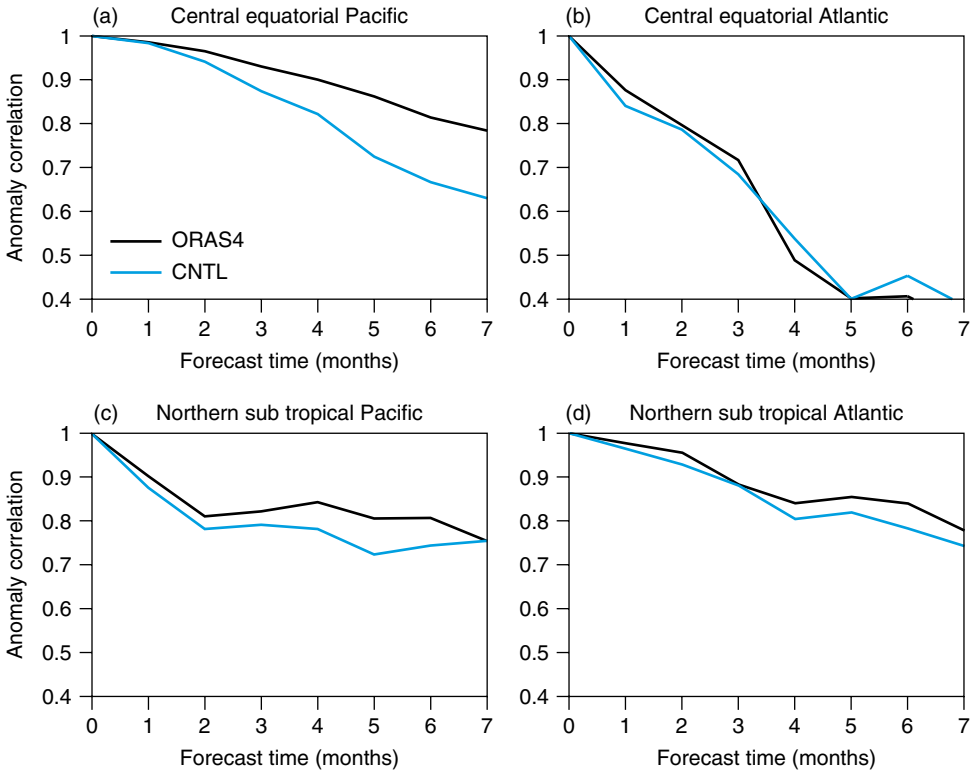

Figure 16. The impact of initialization (ORAS4 versus CNTL) in seasonal forecast skill. Anomaly correlations as a function of forecast lead time for (a) central Equatorial Pacific, (b) central Equatorial Atlantic, (c) northern Subtropical Pacific, and (d) northern Subtropical Atlantic. The regions are defined in Table 2.

CSR product. This would imply that according to GRACE, most of the changes in sea level for the period 2003–2007 are steric changes (volume expansion). Discrepancies in trends between Argo-derived steric height, altimeter sea level and GRACE bottom pressure have also been noticed by Willis *et al.* (2008).

Removing the atmospheric SLP from GRACE bottom pressure increases the amplitude of the seasonal cycle by approximately 1 mm, and also affects its phase: GRACE peaks in mid-October while GRACE AtmosCrt peaks around mid-November. Both the ORAS4 and CNTL EBP seasonal cycle agree well with GRACE, although the amplitude ($\sim$ 8 mm) is slightly lower than GRACE ($\sim$ 9 mm) and GRACE AtmosCrt ($\sim$ 10 mm). There is also a good agreement in the extremes of the seasonal cycle, with a clear consensus for a minimum in mid-April, and slightly broader consensus for the maximum, which occurs between mid-October/November. ORAS4 peaks at the same time as GRACE AtmosCrt, around November. The spread is larger during the transition periods, the model-based estimates lying between GRACE and GRACE AtmosCrt. Ponte *et al.* (2007) also found good agreement in the EBP seasonal cycle of Estimating the Circulation and Climate of the Ocean–Global Ocean Data Assimilation Experiment (ECCO–GODAE; Wunsch and Heimbach, 2007) and GRACE. The consistency between these different products indicates robust knowledge of the global mean seasonal cycle of mass fluctuations over the oceans. The agreement of model-derived EBP with GRACE also validates the seasonal cycle of the ORAS4/CNTL steric height, since the sea level in ORAS4/CNTL is strongly constrained by altimeter data.

In this section we have only used the global values of EBP. In principle, spatial variations of EBP could be assimilated in conjuction with the sea level from altimeter data to better constrain the barotropic circulation of the ocean.

4.6. *Impact on coupled forecasts*

The ocean analysis system at ECMWF is primarily used for initialization of coupled-model forecasts. It is therefore natural to benchmark the performance of the assimilation system by using the output of ORAS4 as ocean initial conditions for the seasonal forecasting system and comparing the results to those using output from CNTL as ocean initial conditions.

Two seasonal forecasting experiments, initialized by ORAS4 and CNTL, were conducted with a prototype of the new S4 seasonal forecasting system (Molteni *et al.*, 2011), but at a resolution of T159 with 62 levels in the vertical. Each experiment consists of 40 start dates, three months apart, over the period 1989–2008. For each date, an ensemble of five members is integrated for seven months. Figure 16 shows the skill of seasonal forecasts of SST in terms of anomaly correlation as a function of forecast time for different regions. The SST used for verification are from the NOAA OIv2_1x1 dataset (Reynolds *et al.*, 2002) described in section 3.1.

The experiment initialized by ORAS4 has better skill than the experiment initialized by CNTL. The improvement is most noticeable in the Central Pacific, a region important for ENSO prediction. The improvement is consistent in most regions for all lead times and in all regions for short lead times (less than four months). The positive impact of the assimilation in the Atlantic regions, although modest, is worth highlighting, since this is the first time in the history of seasonal forecasts at ECMWF that the ocean data assimilation is able to improve the forecasts in

Table 2. Regions shown in Figure 16.

Region	Longit. range	Latit. range
Central Eq. Pacific	190° to 230°	−5° to 5°
Central Eq. Atlantic	340° to 360°	−3° to 3°
N. Subtrop Pacific	105° to 270°	10° to 30°
N. Subtrop Atlantic	280° to 20°	5° to 28°

the Equatorial Atlantic region. The adequacy of an ocean reanalysis to initialize seasonal forecasts ultimately depends on the quality of the coupled model. Zhu *et al.* (2012), in a comparative study using several ocean reanalysis products to initialize a single coupled model (different from the one used here), find that the previous NEMOVAR ocean reanalysis COMB-NV also produced improved seasonal skill scores relative to other ocean reanalyses.

5. Robustness of some climate signals

In this section we compare the evolution of some climate signals in ORAS4 with those in the previous ECMWF ocean reanalysis ORAS3, and the equivalent control experiments for ORAS3 and ORAS4, called here ORAS3-NoObs and ORAS4-NoObs. (Note that ORAS4-NoObs is the same as CNTL in the previous sections). Differences between ORAS3 and ORAS4 are summarized in Table 3. Differences include the ocean model, data assimilation system, forcing fluxes, the version of the *in situ* data, the source and use of SLA data, and the ensemble generation strategy. The treatment of the mean state is also slightly different: in ORAS3 there is no latitudinal dependence of the pressure and density bias terms, the relaxation to WOA05 is 10 years, and the offline bias was estimated from the long-term climatology instead of from the Argo period.

In order to investigate the impact of the XBT corrections on the climate variability, an experiment similar to ORAS3, but with the corrected XBTs was conducted. We refer to this experiment as ORA-XBTc in what follows. The impact of assimilation can be measured by comparing results with ORAS3-NoObs and ORAS4-NoObs. The -NoObs experiments differ not only in the ocean model, but also in the spin-up, the control of the mean state and forcing fluxes. For instance, ORAS3 and ORAS3-NoObs use E4/NWP fluxes (this is to say, ERA-40 until 2002 and operational NWP fluxes thereafter), while ORAS4 and ORAS4-NoObs use E4/EI/OPS (ERA-40 until 1989, ERA-Interim from 1989-2010 and NWP thereafter).

5.1. Equatorial Pacific thermocline

Figure 17 shows the time evolution of the Equatorial Pacific depth of the 20°C isotherm (D20), as represented by the various experiments in Table 3. This variable is a proxy for the depth of the thermocline. The shaded areas are for the ensemble of reanalyses in ORAS4 and ORAS3. ORAS3-XBTcrt (dark blue line) closely follows ORAS3, except for the period 1975–1985, where the XBT-corrected version produces a shallower thermocline than ORAS3. The model-only simulations ORAS3-NoObs and ORAS4-NoObs are the outliers, and their difference with respect to the other reanalyses is indicative of errors in the ocean models: the thermocline is too shallow in HOPE and slightly too deep in NEMO. The data assimilation, in both ORAS4 and ORAS3, constrains the solution by reducing the uncertainty coming from the models and forcing fields. The assimilation changes the mean state of the model-only runs: it deepens the thermocline in the HOPE model (ORAS3 deeper than ORAS3-NoObs) and produces a shallower thermocline in the NEMO model (ORAS4 shallower than ORAS4-NoObs). In the case of ORAS4, it is also clear that the assimilation reduces the spread, which comes mostly from the wind stress perturbations.

The uncertainty in the ORAS3 and ORAS4 reanalyses decreases with time, as the number of ocean observations increases. In the later years, not only is the ensemble spread clearly reduced, but it is also commensurate with the uncertainty in the analyses. This is not the case in the earlier years where, in spite of both reanalyses showing larger spread, it is not enough to cover the differences between reanalyses.

All the experiments exhibit a pronounced shallowing trend ($\sim$ 30 m in the last 50 years). This shallowing of the Equatorial thermocline has been discussed by Balmaseda *et al.* (2007d), and is caused by changes in the wind stress, which induce an enhanced heat export by intensification of the Equatorial meridional circulation. Corre *et al.* (2010) identify the shallowing trend of the Equatorial thermocline in several ocean reanalysis products as a fingerprint of global warming.

Equatorial Hovmöller diagrams of D20, SST and wind stress can be seen on the ORAS4 web pages** which we reproduce in Appendix C to facilitate the visual inspection. The figures display the 12-month running mean anomalies for the different viariables, spanning the period 1958–2011. The anomalies have been computed with respect to the 1981–2009 climate. (The climatological period is the same used for the calibration of the S4 seasonal forecasts.)

The shallowing of D20 in the Equatorial Pacific is clearly seen in Figure C1 . The changes are more prominent in the Western Pacific. Long-term changes have occurred also in the Equatorial Atlantic, where the thermocline in the eastern part has become shallower. The changes in D20 are likely related to changes in ENSO variability, favouring the occurrence of the so-called Modoki ENSO (Ashok and Yamagata, 2009), characterized by the appearance of SST anomalies in the central Pacific (Figure C2). The Equatorial SST also shows a long-term trend, which manifests itself in the Indian Ocean and Pacific warm pool, before appearing in the Atlantic Ocean at a later stage. No obvious trend is apparent in the Eastern Pacific.

Dramatic changes are also visible in the wind stress (Figures C3 and C4). Before the 1980s there are stronger easterlies in the Indian Ocean and eastern Pacific. The latter may be responsible for the deeper thermocline in the Western Pacific. The easterlies are weaker in the Atlantic, also consistent with the flatter thermocline there. Noticeable are the changes in the Equatorial Pacific meridional wind, where the anomalies are consistently positive prior to the 1980s. The changes in the wind stress can be explained by changes in the SST, and can cause the changes in the thermocline. The changes also coincide with the beginning of the satellite period. These results would suggest that good knowledge of the SST is essential to pin down changes in the tropical climate, especially in the early stages of

**http://www.ecmwf.int/products/forecasts/d/charts/oras4/ reanalysis/

Table 3. Experiments involving the ORAS3 and ORAS4 comparison.

Name	Model/Assim	Forcing	Observations	Comments
ORAS3	HOPE/OI	E4 (−2002) NWP (2002−)	EN2(−2004) GTS (2004−) SLA maps (1993−)	XBTs not corrected; all Argo SOLO/FSI sensors included; five ensemble members
ORAS3-XBTcrtd	HOPE/OI	E4 (−2002) NWP (2002−)	EN2_xbtc (−2007) GTS (2007−) SLA maps (1993−)	XBT corrected; some Argo SOLO/FSI sensors blacklisted
ORAS3-NoObs	HOPE	E4 (−2002) NWP (2002−)	None	
ORAS4-NoObs	NEMO	E4 (−1989) EI (1989−2010) NWP (2010−)	None	Five ensemble members
ORAS4	NEMO/3D-Var	E4 (−1989) EI (1989−2010) NWP (2010−)	EN3_v2a_xbtc Along track (1993−)	XBTs corrected; Argo blacklisted; Argo corrections; five ensemble members

the atmospheric and ocean reanalysis, when the observing system may not be sufficient to constrain the atmospheric and ocean circulation.

5.2. *Upper 300 m global ocean heat content*

Figure 18 shows the time evolution of the global ocean heat content in the upper 300 m (OHC300), according to experiments based on ORAS3 and ORAS4. The set of experiments includes the model-only (ORAS3-NoObs and ORAS4-NoObs) and data assimilation (ORAS3, ORAS3-XBTcrtd and ORAS4) experiments. The observation-only reconstruction of Domingues *et al.* (2008) (D08), which uses corrected XBTs as in Table 1 of Wijffels *et al.* (2009), is shown in dark grey. A 2-year running mean has been applied to all the curves, including that of D08, which is based on yearly values. All the experiments show a clear warming trend, but the details of the time evolution differs substantially. The anomalies are computed with respect to the 1980−2005 climatology, and that explains why the largest differences are seen at the beginning and end of the record.

The OHC300 is clearly affected by the XBT correction, as can be seen by comparing the OHC300 from experiments ORAS3 and ORAS3-XBTcrtd (Figure 18). In ORAS3 the XBTs were not corrected, which caused the spurious signals in the 1970s and sharp warming post-1995. These signals are not in ORAS3-XBTcrtd, D08, ORAS4 and the ocean-only simulations. The ORAS3 variability is also contaminated by the faulty SOLO/FSI sensors in Argo floats (Lyman *et al.*, 2006) and lack of manual blacklisting in Argo. In ORAS3, the combination of XBT errors and SOLO-FSI errors created a heat content peak in 2003 and a large drop post-2003. In ORAS3-XBTcrtd the XBTs have been corrected, and SOLO-FSI sensors blacklisted, but still a drop in the last part of the record is observed, coinciding with the switch from the quality-controlled EN3 dataset to the GTS data.

ORAS4 is not affected by the SOLO-FSI or XBT problems. Therefore, it does not show the peaks in the 1970s or in 2003, and it seems to be in good agreement with D08. The impact of data assimilation, visible as the difference between ORAS4 and ORAS4-NoObs, is quite pronounced

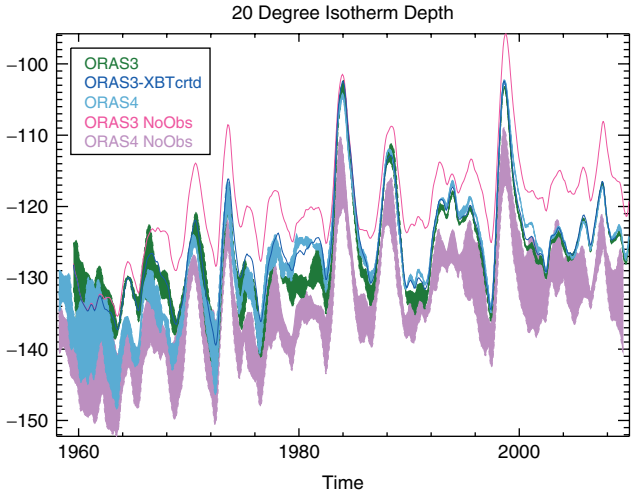

Figure 17. Time series (12-month running mean) of the depth (m) of the 20°C isotherm (D20) in the Equatorial Pacific (130°E−80°W, 5°N−5°S): ORAS4 (light blue), ORAS4-NoObs (violet), ORAS3 (green), ORAS3-NoObs (pink), and ORAS3-XBTcrtd (dark blue). The shaded areas encompass the spread of the different ensemble members in ORAS3, ORAS4 and ORAS4-NoObs. Only one ensemble member was used in ORAS3-NoObs and ORAS3-XBTcrtd. The shallowing trends in D20 are present in all the experiments.

in the magnitude of the trend, which is stronger in ORAS4-NoObs. Additional experiments indicate that the trend in ORAS4-NoObs is contaminated by drift in the ocean model; the model is too diffusive and tends to become too warm. The assimilation is correcting for model error. The impact of assimilation is also noticeable in the variability in the longer part of the record; in NEMO-NoObs the heat content continues to increase nearly monotonically after 1995, while in ORAS4 it declines after 1998, with a local minimum around 2000, and the warming trend seems to stabilize after 2003. This local minimum is also in D08.

All the model-derived estimates of OHC300 produce a local maximum in the early 1990s which is not so obvious in D08. However, all the estimates have a local minimum or plateau around 1995. There are large discrepancies at the

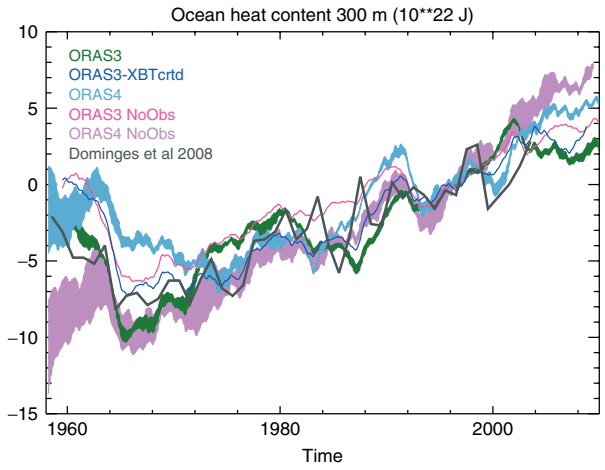

Figure 18. Time series of the upper 300 m global ocean heat content anomalies (10^{22} J) from ORAS3 and ORAS4, their respective NoObs integrations, and the ORAS3-XBT corrected reanalysis. The values from the observational product of Domingues *et al.* (2008) are also plotted. A 2-year running mean has been applied. The anomalies are computed with respect to the 1993–2000 climatology.

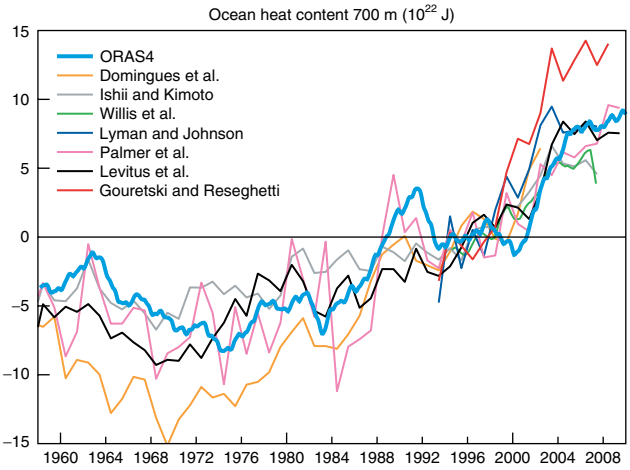

Figure 19. Time evolution of the upper 700 m ocean heat content from ORAS4 and different observational estimates (as in legend). The magnitude of the warming in ORAS4 is consistent with the average. The values are referred to the 1993–2000 mean.

beginning of the record, but all the estimates (assimilation, model-only and observation-only) show a decrease in the mid-1960s.

All the estimates show a non-monotonic trend in OHC300, which exhibits a step-like behaviour, with occasional cooling events, when it either stabilizes or decreases. If we eliminate ORAS3 from the estimates, which is badly affected by the errors in the XBT fall rate, there is some agreement on the timing of three of these cooling events: mid-1960s, mid-1990s and early 2000. There is a cooling event in the 1980s, but the agreement is not so good. Understanding the reasons for these cooling events is obviously important.

5.3. *Upper 700 m global ocean heat content*

Figure 19 shows the estimates of the upper 700 m global ocean heat content from ORAS4, and the different observational estimates reported in Lyman *et al.* (2010) used for evaluating the state of the climate. The ORAS4 values are consistent with the ensemble of existing OHC700 estimates. The most visible difference is that ORAS4 produces a smoother estimate since it is based on monthly mean rather than annual values. The trends in ORAS4 OHC700 are similar to Palmer *et al.* (2007), Levitus *et al.* (2009), and Ishii and Kimoto (2009) estimates.

There is also some degree of agreement in the timing of the variability: local maxima in the first half of the 1960s, 1978–1981 and early 1990s; a steep rise after 2000; and stabilization after 2005. The cooling events in the early 1980s and 1990s may be related to major volcanic eruptions such as El Chichon and Mount Pinatubo. There are also obvious differences concerning the exact timing and duration of the maxima and minima. Note that Palmer *et al.* (2007) and ORAS4 are based on the EN3 dataset, and therefore their differences can be interpreted as the added information from the model and forcing fluxes.

5.4. *Sea-level changes*

Figure 20 shows the attribution of global sea-level changes according to ORAS4 and ORAS4-NoObs. By construction, the global sea level in both products follows closely the AVISO sea-level product. The changes in sea level are dominated by the increasing trend, and by occasional interannual variability (like the 1998 ENSO event). According to ORAS4 and ORAS4-NoObs, the 1998 increase in sea level is mainly due to an increase in mass. This attribution is the same according to ORAS3 and ORAS3-NoObs (not shown).

The partition of the trends into steric and mass changes is less robust. According to ORAS4, the trends in sea level were equally distributed into steric height and mass trends up to 2005, but were dominated by trends of mass from 2005 to 2010. The comparison with GRACE bottom pressure in Figure 15 would suggest that ORAS4 overestimates the trends in bottom pressure (mass) and underestimates the trends in steric height. In contrast, the ORAS4-NoObs trends in steric height are quite linear, and similar in magnitude to the trends in bottom pressure.

The steric height changes can be decomposed into thermo-steric (changes caused by temperature only, keeping salinity constant), and halo-steric (changes caused by salinity only, keeping temperature constant). Figure 21 shows the partition of the ORAS4 steric height into thermo-steric and halo-steric. The sum of the thermo- and halo-contributions is also displayed, and follows closely the direct (nonlinear) estimate. The time series is extended back to the beginning of the record. As expected, the variability in the thermal component has the same signature as the heat content (Figure 19). In comparison, the variability in the halo-steric component is weaker. After 2002, there is a decrease in the halo-steric component, which is directly linked to the assimilation of Argo data. This is confirmed by an additional experiment where Argo data have been removed (not shown), and consistent with the results presented in Balmaseda *et al.* (2007c) which show Argo substantially modifying the global salinity field. This decrease in the halo-steric height may be spurious (if the global salinity is not well constrained in ORAS4 before Argo), resulting in an underestimation of the total steric height by ORAS4, which could explain part of the disagreement in bottom pressure trends between ORAS4 and GRACE.

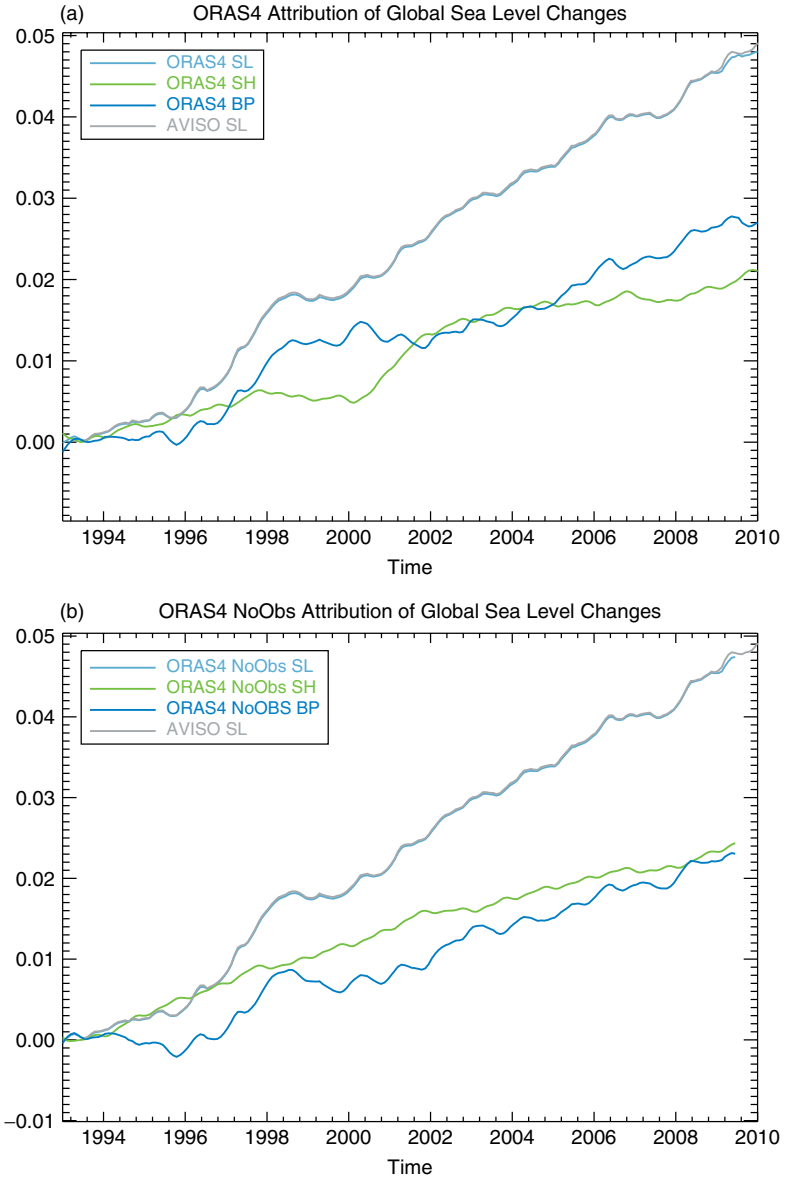

Figure 20. Attribution of global sea level trends (m) by (a) ORAS4 and (b) ORAS4-NoObs, showing the sea level (light blue), steric height (green), and bottom pressure (darker blue). The global sea level from AVISO is also shown (grey). All the values are based on a 12-month running mean, with the mean from 1993 removed.

6. Summary and conclusions

A new operational ocean reanalysis system (ORAS4) has been implemented at ECMWF. It replaces the previous ORAS3 system. Most of the components of the reanalysis system have been changed in the transition between ORAS3 and ORAS4: the ocean model, data assimilation method, observation datasets, forcing fluxes, ensemble-generation procedure and model bias-correction scheme.

ORAS4 consists of five ensemble members, starting in September 1957 and is continuously updated to the present. ORAS4 uses a 10-day assimilation window, and it is updated every 10 days with a 6-day delay. Details about the various components of the ORAS4 system have been provided.

ORAS4 is based on the NEMO ocean model and the NEMOVAR ocean data assimilation system. ORAS4 uses NEMOVAR in its 3D-Var FGAT configuration to assimilate data from temperature, salinity and SLA observations. The multivariate balance relations that are used in NEMOVAR are similar to those described by Weaver *et al.* (2005), except for the expression for linearized density which has been modified to account explicitly for information about the background vertical stratification.

The forcing fluxes in ORAS4 from 1989 to 2010 are from ERA-Interim. The impact of using ERA-Interim versus ERA-40 has not been discussed here, but it is documented in Balmaseda and Mogensen (2010).

The ensemble-generation procedure in ORAS4 uses wind stress perturbations as in ORAS3, but in addition attempts to account for uncertainty in the deep ocean and in the observation coverage. The model bias-correction scheme in ORAS4 should be an improvement with respect to that used in ORAS3: a latitudinal dependence in the partition of the bias has been introduced (it was not in ORAS3), and the prescribed bias term has been estimated from the Argo period, instead of just from climatology. This procedure is based on the assumption that errors in the model are essentially stationary, and provides a

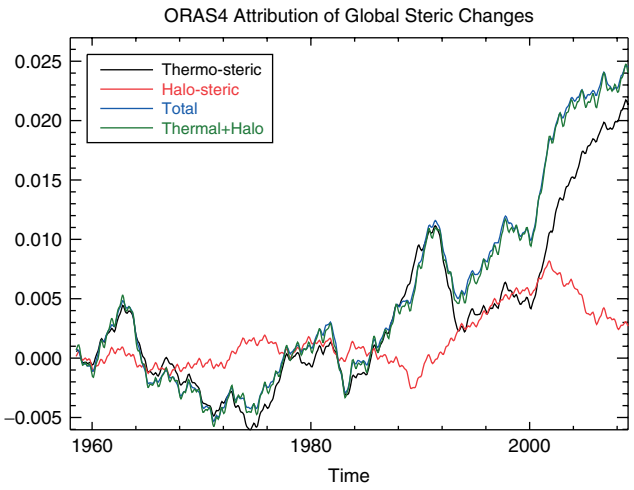

Figure 21. Time evolution of the steric height (m) components in ORAS4: thermo-steric (black), halo-steric (red) and total (blue). The sum of thermo- and halo-components appears in green, and agrees quite well with the total. Note the decrease/increase in the halo/thermal components after the introduction of Argo.

way of extrapolating Argo information into the past. It has been shown that the bias correction improves the Equatorial currents and the fit to the temperature and salinity profiles. Results also indicate that the presence of model bias can be detrimental in the estimation of climate trends.

An ocean model simulation, similar to ORAS4 but without assimilating data, has been used as a control experiment to assess the quality of ORAS4. Assimilation statistics indicate that ORAS4 consistently improves the fit to observations of temperature and salinity in terms of both bias and RMS. Compared to the control experiment, the bias in ORAS4 is relatively small and, more importantly, quite stationary in time. Time series of the RMS fit to observations exhibit an interesting decreasing tendency. Several possible reasons to explain this tendency have been discussed. Interannual variability of the ocean estimate also seems to be improved in ORAS4, as suggested by the improved temporal correlation with sea-level data and surface currents. Observations of ocean currents and transports, which have not been assimilated, have been used to provide an independent measure for evaluating the performance of ORAS4. The conclusion is that ORAS4 improves the ocean estimate for most of the parameters investigated. An exception is the Atlantic MOC, where ORAS4 tends to underestimate its mean strength.

Initial conditions from ORAS4 have been used to initialize seasonal forecasts, with positive impact on the skill of forecasting SST. The positive impact includes the Atlantic, a region where previous ocean reanalysis systems showed poor performance, at least at ECMWF.

It has been shown that important large-scale climate indices are sensitive to the choice of SST product, the model bias-correction scheme and the treatment of observations near the coast. The global ocean heat content is sensitive to the SST product, and the amplitude of the Atlantic MOC is quite sensitive to the prescription of the OESDs near the coast. These sensitivities are not covered by the ensemble-generation procedure in ORAS4.

There appears to be a robust shallowing trend in the depth of the Equatorial Pacific thermocline (as represented by D20), which is present in both model-only and data assimilation experiments, and is consistent with changes in the wind stress. The robustness of this signal should be investigated further by examining changes in wind stress from atmospheric reanalyses, as well as changes in the SST (since the tropical wind stress in the early part of the record may be largely determined by SST forcing).

The global heat content in ORAS4 differs with respect to ORAS3, which was badly affected by the XBT fall-rate errors and faulty Argo sensors. The ORAS4 heat content is consistent with other observational estimates used to evaluate the state of the climate. Although there are differences regarding decadal variability, it is probably more surprising that there is some degree of coherence.

The partition of sea-level changes into changes in mass and volume have been evaluated, and compared with estimates from AVISO sea level and GRACE bottom pressure. The seasonal cycle of ocean mass changes in ORAS4 is in good agreement with the GRACE bottom pressure, in both phase and amplitude. This agreement supports the robustness of current estimates of the seasonal cycle of global ocean mass and steric height. In contrast, there is discrepancy in the representation of the long-term trends, with ORAS4 assigning two thirds of the trend to bottom pressure (during the period 1993–2010) and one third to steric changes. The trends in steric height are dominated by changes in temperature, but changes in salinity are not negligible. In the past few years, coinciding with the advent of Argo, a declining halo-steric component may contribute to the underestimation of the steric height trends in ORAS4.

Appendices

A. Balance operator in NEMOVAR

A1. Salinity balance

The linearized transformation between increments of temperature and salinity is a local adjustment that approximately preserves the water-mass properties of the background state by making vertical displacements (δz) of salinity in response to changes in temperature. The expression for the balanced part of the salinity increment is given by (Ricci *et al.*, 2005)

$$\delta S_{\mathrm{B}} = \gamma^{\,\mathrm{b}} \left(\frac{\partial S}{\partial z}\right)^{\mathrm{b}} \delta z \qquad \text{(A1)}$$

where

$$\delta z \;=\; \left(\frac{\partial z}{\partial T}\right)^{\mathrm{b}} \delta T. \qquad \text{(A2)}$$

The coefficient $\gamma^{\,\mathrm{b}}$ is a switch that has values of either zero or one, depending on various conditions in the background state. It is set to one if the multivariate T/S relationship is applicable. It is set to zero in areas where temperature and salinity are expected to be weakly correlated, such as in the mixed layer, the barrier layer, or areas of weak vertical temperature stratification where $(\partial T/\partial z)^{\mathrm{b}}$ is very small.

When $\gamma^{\mathrm{b}} = 0$, the salinity increment is described entirely by its unbalanced component (MBW12 give more details).

A2. SSH balance

The SSH increment is partitioned into a *baroclinic* component $\delta\eta_{\mathrm{B}}$ that is balanced with δT, δS and a *depth-dependent* geostrophic velocity increment $\delta\widehat{\mathbf{u}}^{\mathrm{h}}_{\mathrm{B}} = (\delta\widehat{u}_{\mathrm{B}}, \delta\widehat{v}_{\mathrm{B}})^{\mathrm{T}}$, and a *barotropic* component $\delta\eta_{\mathrm{U}}$ that is balanced with a *depth-independent* geostrophic velocity increment $\delta\overline{\mathbf{u}}^{\mathrm{h}}_{\mathrm{B}} = (\delta\overline{u}_{\mathrm{B}}, \delta\overline{v}_{\mathrm{B}})^{\mathrm{T}}$ where $\delta\mathbf{u}^{\mathrm{h}}_{\mathrm{B}} = \delta\widehat{\mathbf{u}}^{\mathrm{h}}_{\mathrm{B}} + \delta\overline{\mathbf{u}}^{\mathrm{h}}_{\mathrm{B}}$ (section A3). The balanced SSH component is estimated by computing the dynamic height at the surface $z = 0$ relative to the (spatially varying) ocean bottom $z = -H$:

$$\delta\eta_{\mathrm{B}} = -\int_{-H}^{0} (\delta\rho/\rho_0)\ \mathrm{d}z, \tag{A3}$$

where $\rho_0 = 1020\,\mathrm{kg\,m^{-3}}$ is a reference density, and

$$\frac{1}{\rho_0}\delta\rho = -\alpha^{\mathrm{b}}\,\delta T \ + \ \beta^{\mathrm{b}}\,\delta S \tag{A4}$$

is a linearized equation of state.

The coefficients α^{b} and β^{b} in (A4) are the thermal expansion and saline contraction coefficients, respectively, and are functions of the background potential temperature and salinity fields, and depth. Two formulations have been tested for the estimation of the coefficients α^{b} and β^{b} in (A4). The first formulation derives these coefficients through a direct tangent-linearization of the equation of state of Jackett and McDougall (1995):

$$\alpha^{\mathrm{b}} \ = \ -\frac{1}{\rho_0}\left(\frac{\partial\rho}{\partial T}\right)^{\mathrm{b}} \quad \text{and} \quad \beta^{\mathrm{b}} \ = \ \frac{1}{\rho_0}\left(\frac{\partial\rho}{\partial S}\right)^{\mathrm{b}}.$$

This formulation was proposed by Weaver *et al.* (2005). A similar formulation is used by Storto *et al.* (2011) in their variational scheme to assimilate altimeter data.

When assimilating altimeter data, it was found that this formulation, in combination with the SSH balance (A3), produced undesirable density increments in the deep ocean where grid cells are thicker. It had a detrimental impact on the steric height and also in the upper ocean salinity field. The vertical stratification is not considered explicitly in this formulation, although some provision can be made for it indirectly by specifying depth-dependent background-error standard deviations (discussion in Appendix B).

An alternative formulation has been implemented which takes into account the vertical stratification of the background density field. This is done using the background value of the Brunt–Väisälä frequency (Gill, 1982),

$$\left(N^{\mathrm{b}}\right)^2 = g\beta^{\mathrm{b}}\left\{\frac{\alpha^{\mathrm{b}}}{\beta^{\mathrm{b}}}\left(\frac{\partial T}{\partial z}\right)^{\mathrm{b}} - \left(\frac{\partial S}{\partial z}\right)^{\mathrm{b}}\right\}, \tag{A5}$$

where β^{b} and the ratio $\alpha^{\mathrm{b}}/\beta^{\mathrm{b}}$ are defined from the polynomial expressions of McDougall (1987), evaluated using the background temperature and salinity fields. Rearranging (A5) gives

$$\alpha^{\mathrm{b}} \ = \ \beta^{\mathrm{b}}\left\{\frac{\left(N^{\mathrm{b}}\right)^2}{g\beta^{\mathrm{b}}} + \left(\frac{\partial S}{\partial z}\right)^{\mathrm{b}}\right\}\left(\frac{\partial z}{\partial T}\right)^{\mathrm{b}}. \tag{A6}$$

Substituting (A6) into (A4), and writing $\delta S = \delta S_{\mathrm{B}} + \delta S_{\mathrm{U}}$ with δS_{B} given by (A1)–(A2) and γ^{b} assumed to be equal to 1, yields

$$\frac{1}{\rho_0}\delta\rho = -\frac{\left(N^{\mathrm{b}}\right)^2}{g}\left(\frac{\partial z}{\partial T}\right)^{\mathrm{b}}\delta T + \beta^{\mathrm{b}}\,\delta S_{\mathrm{U}}.$$

The first term on the right-hand side is the contribution from the vertical displacement of the water column (which implicitly accounts for the balanced salinity increment). The second term involves the unbalanced salinity increment, and has been neglected in the current formulation. The normalized density increment is then approximated by

$$\frac{1}{\rho_0}\delta\rho \quad \approx \quad -\widetilde{\gamma}^{\mathrm{b}}\,\frac{\left(N^{\mathrm{b}}\right)^2}{g}\,\delta z,$$

where δz is given by (A2) and $\widetilde{\gamma}^{\mathrm{b}}$ is a coefficient analogous to γ^{b} in (A1), which has been introduced to account for the case when $(\partial T/\partial z)^{\mathrm{b}}$ is very small. The above expression illustrates the similarity with the Cooper and Haines (1996) scheme, as discussed in section 2.2.

A3. Velocity balance

Away from the Equator, the balanced part $\delta\mathbf{u}^{\mathrm{h}}_{\mathrm{B}}$ of the horizontal velocity increment is assumed to be geostrophic, i.e. proportional to the horizontal gradient of pressure divided by the Coriolis parameter f. The unbalanced part $\delta\mathbf{u}^{\mathrm{h}}_{\mathrm{U}} = (\delta u_{\mathrm{U}}, \delta v_{\mathrm{U}})^{\mathrm{T}}$ corresponds to the ageostrophic component of the velocity increment. The pressure increment at a given depth z can be computed by integrating the hydrostatic equation from z to the free surface $\delta\eta$. After substituting $\delta\eta = \delta\eta_{\mathrm{B}} + \delta\eta_{\mathrm{U}}$, where $\delta\eta_{\mathrm{B}}$ is given by (A3), this yields

$$\delta p(z) \quad = \quad -\int_{-H}^{z} \delta\rho(s)\, g\, \mathrm{d}s + \rho_0\, g\, \delta\eta_{\mathrm{U}}. \tag{A1}$$

The horizontal gradient of the first term in (A1) is associated with the baroclinic geostrophic velocity increment, while the horizontal gradient of the second term is associated with the barotropic geostrophic velocity. Special care is required near the Equator where the f-plane geostrophic relation becomes singular. Following Burgers *et al.* (2002), the zonal component δu_{B} is taken to be geostrophically balanced at the Equator while the meridional component is reduced to zero. Geostrophic balance for δu_{B} is imposed near the Equator using a β-plane approximation combined with a modified pressure increment $\delta\widetilde{p} = \delta p + \delta p_{\mathrm{c}}$ to enforce a symmetric pressure increment about the Equator (Picaut and Tournier, 1991). The correction term δp_{c} is independent of longitude λ (leaving δv_{B} unaffected), is negligible far from the Equator, and leaves the β-plane approximation unaltered (i.e. directly at the Equator, the second derivatives of the unmodified and modified pressure increments, with respect to latitude ϕ, are equal).

The transition in the geostrophic balance between the Equatorial β-plane and the f-plane approximations is handled by introducing weighting functions $W_\beta = \exp\left(-\phi^2/L_\beta^2\right)$ and $W_f = 1 - W_\beta$, where L_β is an e-folding scale, set to 2.2°. At the Equator, $W_\beta = 1$ and

Table B1. Parameter values in the ORAS4 background-error covariance model.

X	$\left(L_X^\lambda, L_X^\phi\right)$	L_X^z	σ_X^b
T	$\delta_\lambda = 2$ $\delta_\phi = 0.5$ $\overline{L} = 2^\circ$ $\phi_L = 15^\circ$	$\alpha = 1$	$\sigma_T^{max} = 1.5\,^\circ$C $\sigma_T^{ml} = 0.5\,^\circ$C $\sigma_T^{do} = 0.07\,^\circ$C $\delta z = 10$ m
S_U	Same as for T	Same as for T	$\sigma_{S_U}^{max} = 0.25$ psu $\sigma_{S_U}^{do} = 0.01$ psu
η_U	Same as for T	n/a	$\phi_{ex} = \pm 20^\circ$ $\sigma_{\eta_U}^{ex} = 0.01$ m $\delta_{eq} = 0$

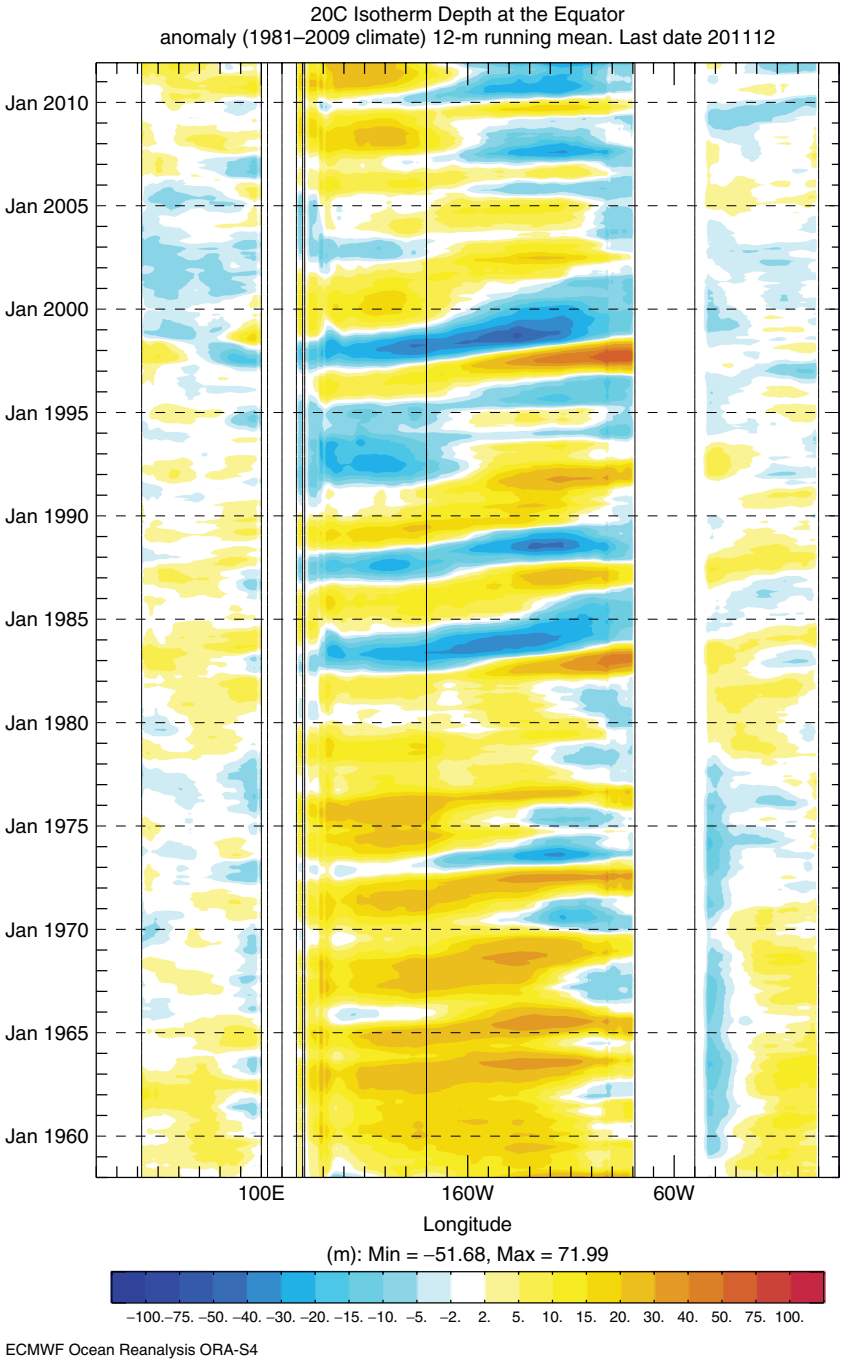

Figure C1. Equatorial Hovmöller diagrams of ORAS4 D20, displaying 12-month running mean anomalies (m) with respect to the 1981–2009 climate. Note the nonlinear colour scale.

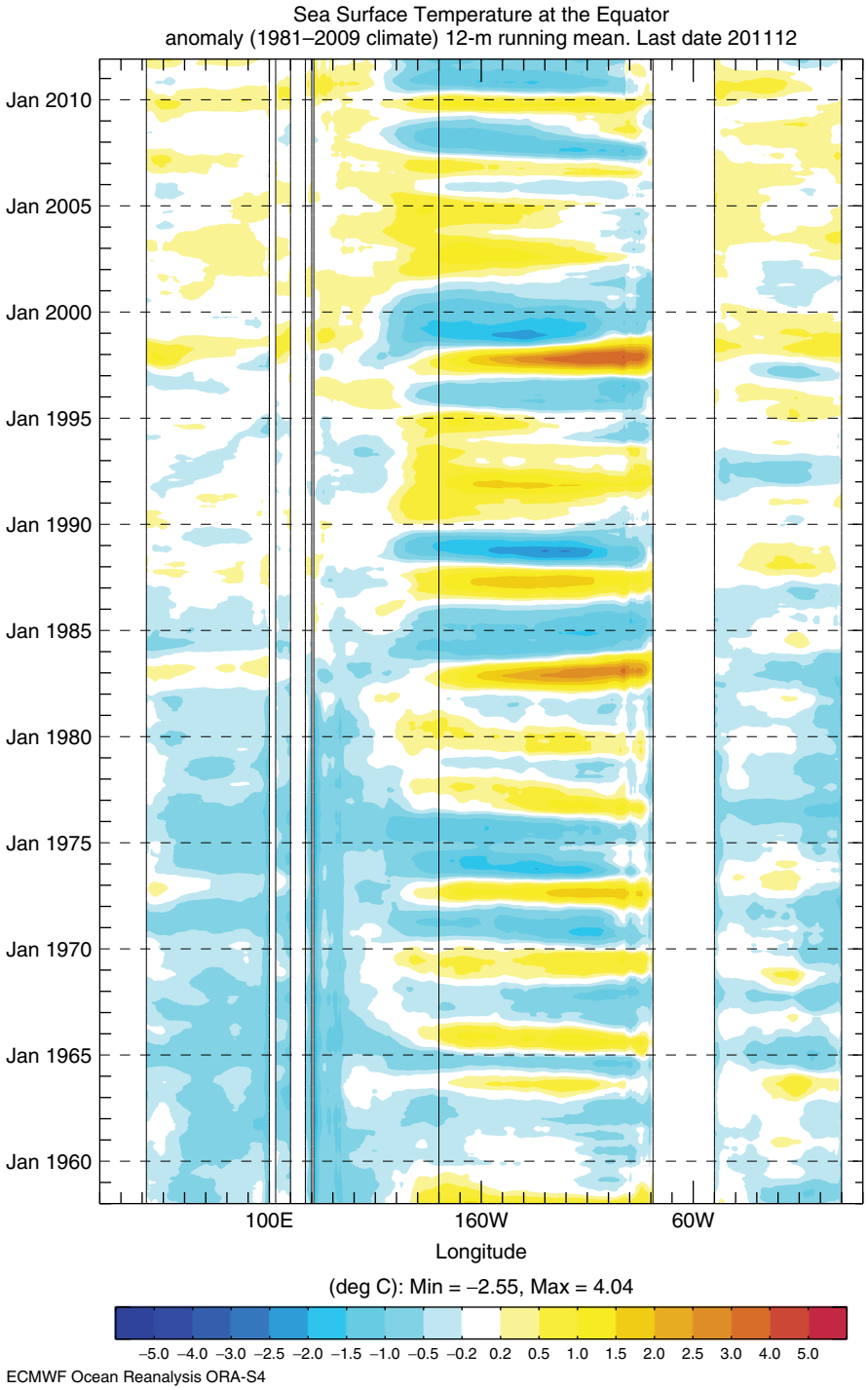

Figure C2. Equatorial Hovmöller diagrams of SST, displaying 12-month running mean anomalies (°C) with respect to the 1981–2009 climate. Note the nonlinear colour scale.

$W_f = 0$, while far away from the Equator, $W_\beta \approx 0$ and $W_f \approx 1$.

The geostrophic balance is reduced near the boundaries by multiplying $\delta \mathbf{u}_B^h$ with a weighting function $W = W(r\,;\,\delta, R)$ of the form

$$W = \begin{cases} 1 + \dfrac{1}{2}(\delta - 1)\left\{1 + \cos\left(\dfrac{\pi\, r}{R}\right)\right\} & \text{if } |r| \le R, \\ 1 & \text{if } |r| > R. \end{cases} \quad \text{(A2)}$$

Equation (A2) is a smooth function of distance r and depends on the parameters $\delta = W(0)$ and the cut-off distance R beyond which the weight is set to one. For the velocity balance, r is the Cartesian distance between the model grid-point and the nearest model coastline point, and the parameters R and δ are set to 200 km and 0, respectively. Distance-dependent weighting functions of the same basic form are also used to modulate parameters in the covariance model (Appendix B).

B. Specification of the background-error covariance parameters in ORAS4

The background-error covariance model described in section 2.3 requires specifying (for $X = T$, S_U and η_U)

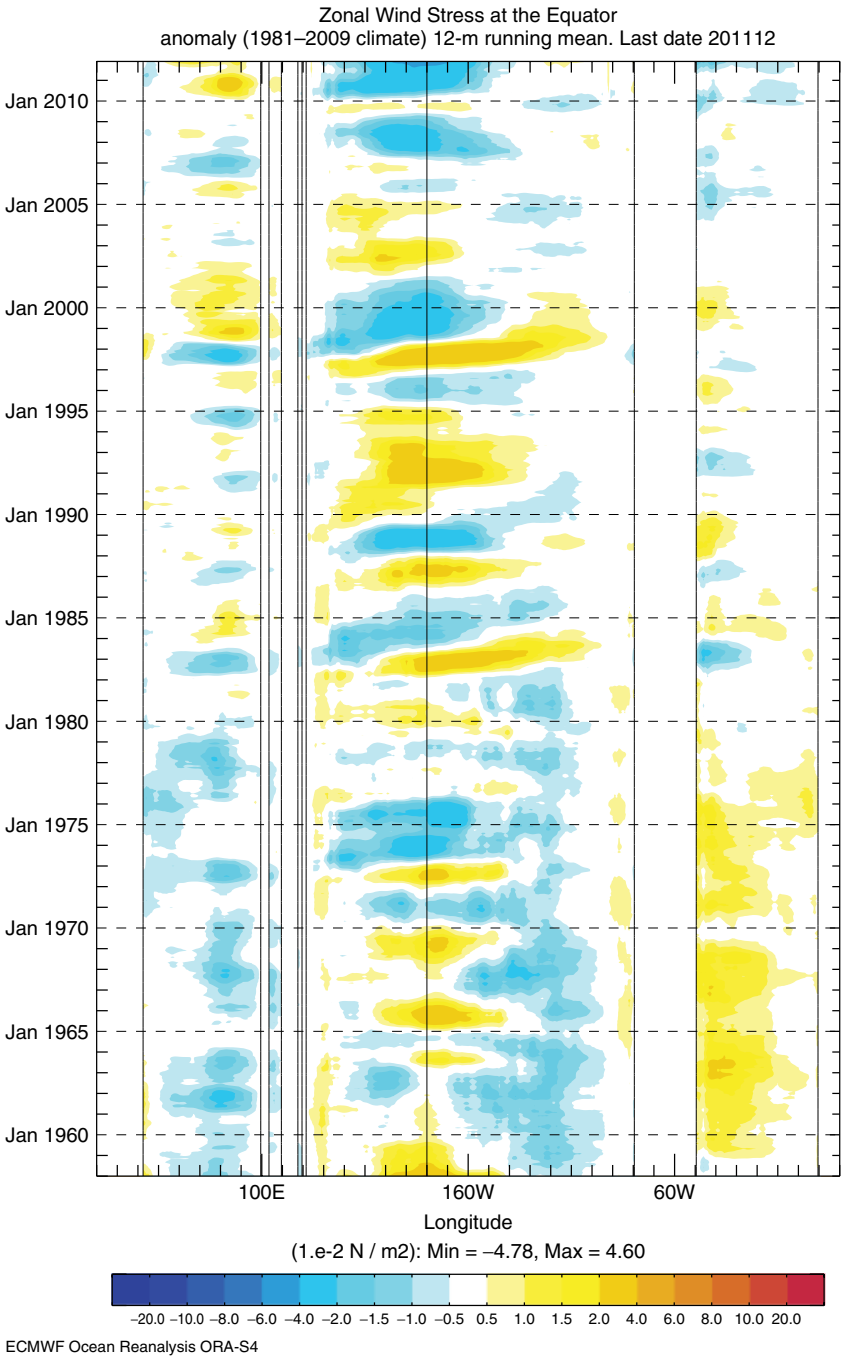

Figure C3. Equatorial Hovmöller diagrams of ORAS4 zonal wind stress, displaying the 12-month running mean anomalies ($0.01\,\mathrm{N\,m^{-2}}$) with respect to the 1981–2009 climate. Note the nonlinear colour scale.

the standard deviations σ_X^{b} (the diagonal elements of $\mathbf{D}_X^{1/2}$) and the directional length-scales L_X^{λ}, L_X^{ϕ} and L_X^{z} of the quasi-Gaussian correlation functions represented by the horizontal and vertical diffusion operators. These parameters have been defined somewhat heuristically in ORAS4, using the values in ORAS3 (Balmaseda *et al.*, 2008) and OPAVAR (Daget *et al.*, 2009) as a guideline. Ideally these parameters should be estimated from background-error statistics.

The standard deviations $\sigma_T^{\mathrm{b}} = \sigma_T^{\mathrm{b}}(\lambda, \phi, z)$ are parametrized in terms of the vertical gradient of the background temperature field, $(\partial T/\partial z)^{\mathrm{b}}$. Weaver *et al.* (2003) showed that such a parametrization could capture flow-dependent aspects of variance propagation, which are implicit in 4D-Var, since the model dynamics tend to move the level of maximum σ_T^{b} to the level of the local thermocline where $(\partial T/\partial z)^{\mathrm{b}}$ is large. This effect was shown to be particularly pronounced near the Equator. It is then reasonable to assume that $\sigma_T^{\mathrm{b}} \sim |(\partial T/\partial z)^{\mathrm{b}}\,\delta z|$ where δz corresponds to a vertical displacement error in the background temperature profile. To avoid unrealistically small values in the mixed layer (ml) and deep ocean (do), where $(\partial T/\partial z)^{\mathrm{b}}$ is small, the standard deviations are bounded by minimum values σ_T^{ml} and σ_T^{do}. The maximum value is also bounded by σ_T^{max} (Table B1).

The standard deviations $\sigma_{S_{\mathrm{U}}}^{\mathrm{b}} = \sigma_{S_{\mathrm{U}}}^{\mathrm{b}}(\lambda, \phi, z)$ are also flow dependent. They are bounded by a maximum value $\sigma_{S_{\mathrm{U}}}^{\mathrm{max}}$ in the upper ocean and minimum value $\sigma_{S_{\mathrm{U}}}^{\mathrm{do}}$ in the deep ocean. The transition between these two values depends on depth and $\left|(\partial S/\partial T)^{\mathrm{b}}\right|$. The parametrization defines the largest

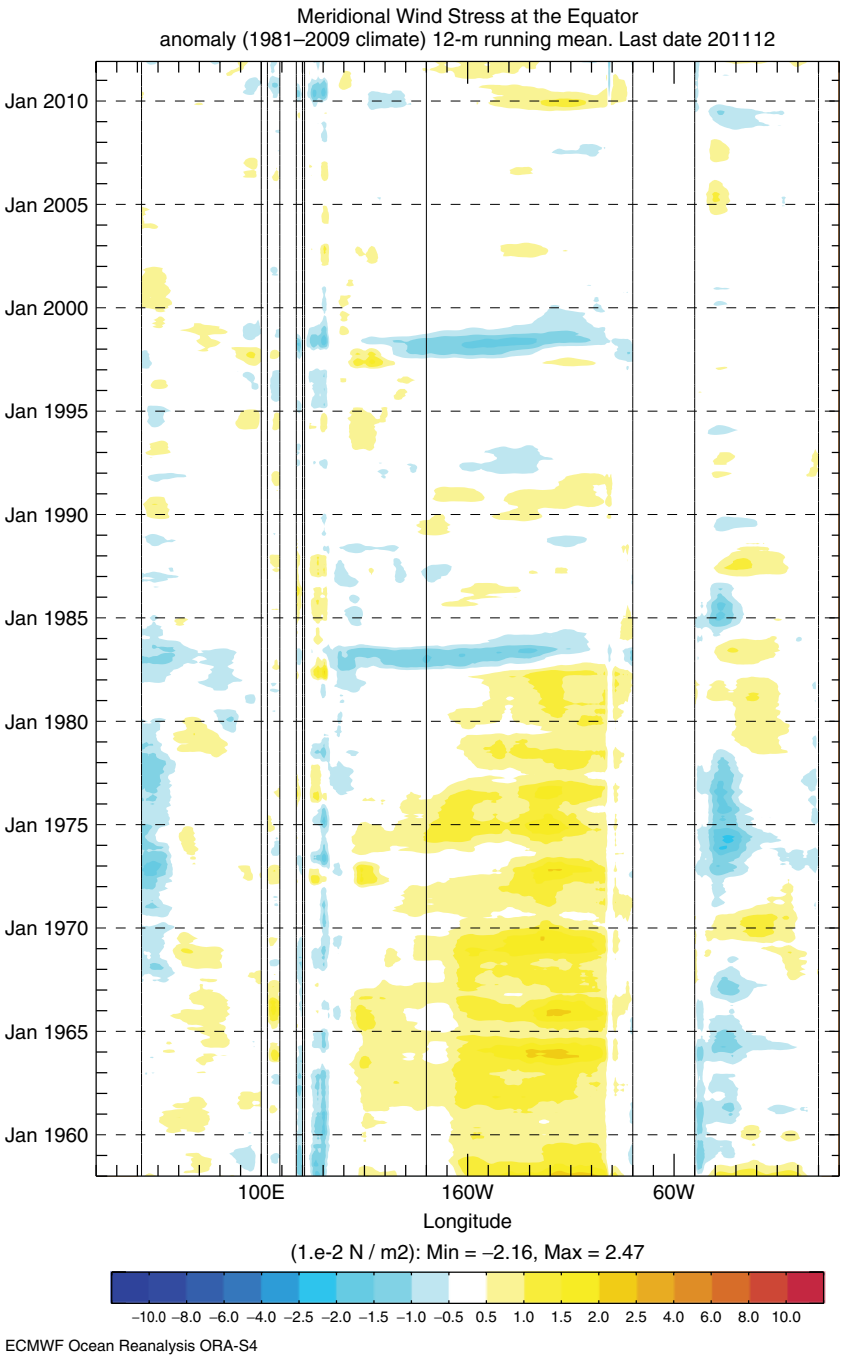

Figure C4. Equatorial Hovmöller diagrams of ORAS4 meridional wind stress, displaying the 12-month running mean anomalies ($0.01\,\mathrm{N\,m^{-2}}$) with respect to the 1981–2009 climate. Note the nonlinear colour scale.

$\sigma^{\mathrm{b}}_{S_{\mathrm{U}}}$ between the surface and the level of maximum $S(T)$ gradients, and decreases $\sigma^{\mathrm{b}}_{S_{\mathrm{U}}}$ monotonically below this level (Table B1). Large values in the mixed layer are especially important since there salinity is described entirely by its unbalanced component (Ricci *et al.*, 2005). Since σ^{b}_{T} and $\sigma^{\mathrm{b}}_{S_{\mathrm{U}}}$ depend on the state, they will evolve from one assimilation cycle to the next and thus will contain information about the impact of data assimilation from previous cycles.

The standard deviations $\sigma^{\mathrm{b}}_{\eta_{\mathrm{U}}} = \sigma^{\mathrm{b}}_{\eta_{\mathrm{U}}}(\phi)$ depend on latitude through the weighting function (A2) (with r replaced by ϕ): $\sigma^{\mathrm{b}}_{\eta_{\mathrm{U}}} = W(\phi\,;\,\delta_{\mathrm{eq}},\phi_{\mathrm{ex}})\,\sigma^{\mathrm{ex}}_{\eta_{\mathrm{U}}}$ where $\sigma^{\mathrm{ex}}_{\eta_{\mathrm{U}}}$ is constant (Table B1). This parametrization is designed to account for the greater importance of the barotropic component (η_{U}) in extratropical regions.

The horizontal background-error correlations for $X = T$, S_{U} and η_{U} are assumed to be isotropic poleward of a given latitude ϕ_L, with an identical length-scale $L^{\lambda}_{X} = L^{\phi}_{X} = \overline{L}$ used for all variables and at all depths. Equatorward of ϕ_L, the length-scales L^{λ}_{X} and L^{ϕ}_{X} are modulated by (A2):

$$\left.\begin{aligned} L^{\lambda}_{X}(\phi) &= W(\phi\,;\,\delta_{\lambda},\phi_L)\,\overline{L}, \\ L^{\phi}_{X}(\phi) &= W(\phi\,;\,\delta_{\phi},\phi_L)\,\overline{L}. \end{aligned}\right\}$$

This feature is included to allow stretching/shrinking of the zonal/meridional length-scales in the Equatorial wave-guide (Table B1). The vertical correlation scales for T and S_{U} are specified as a scalar multiple α of the local vertical grid

resolution dz,

$$L_X^z(z) = \alpha \, \mathrm{d}z,$$

in order to ensure adequate smoothing between vertical levels. In ORAS4, a minimum smoothing is employed by setting $\alpha = 1$.

C. Figures from the ORAS4 web pages

Figures C1–C4 show Equatorial Hovmöller diagrams of ORAS4 from the ORAS4 web pages, displaying the anomalies (12-month running mean) for the period 1958–2011, computed with respect to the 1981–2009 climate. (The climatological period is the same used for the calibration of the S4 seasonal forecasts.) The variables displayed are D20, SST, and zonal and meridional wind stress respectively.

Acknowledgements

The development of NEMOVAR is collaborative effort with contributions from numerous people from different institutions (CERFACS, ECMWF, Met Office, INRIA). We thank Andrew Coward from NOCS for his help in setting up the ORCA1 configuration at ECMWF, and Simon Good from the Met Office for his help in implementing the observational QC system in NEMOVAR. The Hadley Centre provided the EN3 XBT-corrected dataset. The RAMA/TAO Project Office of NOAA/PMEL provided the TAO/TRITON/RAMA currents used for validation of ORAS4. The Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it (http://www.argo.ucsd.edu; http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System. GRACE ocean data were processed by Don P. Chambers, supported by the NASA MEASURES Program, and are available at http://grace.jpl.nasa.gov. The tide-gauge sea-level data have been kindly provided by A. Cazenave and B. Meyssignac, from the LEGOS laboratory in Toulouse, France. The Florida Current cable and section data are made freely available on the Atlantic Oceanographic and Meteorological Laboratory web page (http://www.aoml.noaa.gov/phod/floridacurrent/) and are funded by the NOAA Office of Climate Observations. We acknowledge the RAPID project for providing the MOC transport data. The MOC diagnostics have been done under the VALOR project. ATW would like to acknowledge financial support received from the French LEFE-ASSIM and GMMC programmes which enabled him to make several visits to ECMWF, and from the French National Research Agency (ANR) COSINUS programme (VODA project). The COMB-NV product and the collaboration with CERFACS have been funded by the EU FP7 project COMBINE. We also thank Paul Berrisford for providing the ERA-Interim SLP over the oceans. Two anonymous reviewers provided many helpful suggestions for improving the article.

References

Antonov JI, Locarnini RA, Boyer TP, Mishonov AV, Garcia HE. 2006. Salinity. In *World Ocean Atlas 2005, vol 2*. Levitus S. (ed.) *NOAA Atlas NESDIS* 62. US Government Printing Office: Washington DC.

Ashok K, Yamagata T. 2009. Climate change: The El Niño with a difference. *Nature* **461**: 481–484.

Bell MJ, Martin MJ, Nichols NK. 2004. Assimilation of data into an ocean model with systematic errors near the Equator. *Q. J. R. Meteorol. Soc.* **130**: 873–893.

Balmaseda MA, Mogensen K. 2010. 'Evaluation of ERA-Interim forcing fluxes from an ocean perspective'. ERA Report Series No. 6. ECMWF: Reading, UK.

Balmaseda MA, Dee D, Vidard A, Anderson DLT. 2007a. A multivariate treatment of bias for sequential data assimilation: application to the tropical oceans. *Q. J. R. Meteorol. Soc.* **133**: 167–179.

Balmaseda MA, Smith GC, Haines GC, Anderson DLT, Palmer TN, Vidard A. 2007b. Historical reconstruction of the Atlantic meridional overturning circulation from ECMWF operational ocean reanalysis. *Geophys. Res. Lett.* **34**: L23615, DOI: 10.1029/2007GL031645

Balmaseda MA, Anderson DLT, Vidard A. 2007c. Impact of Argo on analyses of the global ocean. *Geophys. Res. Lett.* **34**: L16605, DOI: 10.1029/2007GL030452

Balmaseda MA, Anderson DLT, Molteni F. 2007d. 'Climate variability from the new ECMWF System 3 ocean reanalysis'. *ECMWF Newsletter* **113**: 8–16.

Balmaseda MA, Vidard A, Anderson DLT. 2008. The ECMWF Ocean Analysis System: ORA-S3. *Mon. Weather Rev.* **136**: 3018–3034.

Balmaseda MA, Mogensen K, Molteni F, Weaver AT. 2010. 'The NEMOVAR-COMBINE ocean re-analysis'. COMBINE Technical Report No. 1. http://www.combine-project.eu

Bloom SC, Takacs LL, Da Silva AM, Ledvina D. 1996. Data assimilation using incremental analysis updates. *Mon. Weather Rev.* **124**: 1256–1271.

Bonjean F, Lagerloef GSE. 2002. Diagnostic model and analysis of the surface currents in the tropical Pacific Ocean. *J. Phys. Oceanogr.* **32**: 2938–2954.

Burgers G, Balmaseda MA, Vossepoel F, van Oldenborgh GJ, van Leeuwen PJ. 2002. Balanced ocean-data assimilation near the Equator. *J. Phys. Oceanogr.* **32**: 2509–2519.

Chambers DP. 2006. Evaluation of new GRACE time-variable gravity data over the ocean. *Geophys. Res. Lett.* **33**: LI7603.

Church J, White NJ. 2006. A 20th century acceleration in global sea-level rise. *Geophys. Res. Lett.* **33**: L01602, DOI: 10.1029/2005GL024826

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

Corre L, Terray L, Balmaseda MA, Ribes A, Weaver AT. 2010. Can oceanic reanalyses be used to assess recent anthropogenic changes and low-frequency internal variability of upper ocean temperature? *Clim. Dyn.* **38**: 877–896.

Courtier P, Thépaut J-N, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**: 1367–1388.

Cunningham SA, Kanzow T, Rayner D, Baringer MO, Johns WE, Marotzke J, Longworth HR, Grant EM, Hirschi JJ-M, Beal LM, Meinen CS, Bryden HL. 2007. Temporal variability of the Atlantic Meridional Overturning Circulation at 26°N. *Science* **317**: 935–938. DOI: 10.1126/science.1141304

Daget N, Weaver AT, Balmaseda MA. 2009. Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Q. J. R. Meteorol. Soc.* **135**: 1071–1094.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597.

Domingues CM, Church JA, White NJ, Glecker PJ, Wijffels SE, Barker PM, Dunn JR. 2008. Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature* **453**: 1090–1093.

Ferry N, Parent L, Garric G, Bricaud C, Testut CE, Le Galloudec O, Lellouche JM, Drevillon M, Greiner E, Barnier B, Molines JM, Jourdain NC, Guinehut S, Cabanes C, Zawadzki L. 2012. GLORYS2V1 global ocean reanalysis of the altimetric era (1992–2009) at meso scale. *Mercator Quarterly Newsletter* **44**: 29–39. http://www.mercator-ocean.fr/eng/

Fisher M. 1998. Minimization algorithms for variational data assimilation. In *Recent Developments in Numerical Methods for Atmospheric Modelling*. 364–385. ECMWF: Reading, UK.

Gill AE. 1982. *Atmospheric-Ocean Dynamics*. Academic Press: London.

Haines K, Valdivieso M, Zuo H, Stepanov VN. 2012. Transports and budgets in a 1/4^o global ocean reanalysis 1989–2010. *Ocean Sci.* **8**: 333–344.

Hazeleger W, Wang X, Severijns C, Stefanescu S, Bintanja R, Sterl A, Wyser K, Semmler T, Yang S, van den Hurk B, van Noije T, van der Linden E, van der Wieland K. 2011. EC-Earth V2: description and validation of a new seamless Earth system prediction model. *Clim. Dyn.* in press. DOI: 10.1007/s00382-011-1228-5

Ingleby B, Huddleston M. 2007. Quality control of ocean temperature and salinity profiles – historical and real-time data. *J. Marine Sys.* **65**: 158–175.

Ishii M, Kimoto M. 2009. Re-evaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *J. Oceanogr.* **65**: 287–299.

Jackett DR, McDougall TJ. 1995. Minimal adjustment of hydrographic data to achieve static stability. *J. Atmos. Ocean. Technol.* **12**: 381–389.

Lee T, Stammer D, Awaji T, Balmaseda MA, Behringer D, Carton J, Ferry N, Fischer A, Fukumori I, Giese B, Haines K, Harrison E, Heimbach P, Kamachi M, Keppenne C, Köhl A, Masina S, Menemenlis D, Ponte R, Remy E, Rienecker M, Rosati A, Schroeter J, Smith D, Weaver AT, Wunsch C, Xue Y. 2010. Ocean state estimation for climate research. In *OceanObs'09: Conference on Sustained Ocean Observations and Information for Society* (vol. 2), Venice, 21–25 September 2009. Hall J, Harrison DE, Stammer D (eds). ESA Publication WPP-306. DOI: 10.5270/OceanObs09

Levitus S, Antonov JI, Boyer TP, Locarnini RA, Garcia HE, Mishonov AV. 2009. Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys. Res. Lett.* **36**: L07608, DOI: 10.1029/2008GL037155

Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia HE. 2006. Temperature. In *World Ocean Atlas 2005, volume 1*. Levitus S. (ed.) *NOAA Atlas NESDIS* 61. US Government Printing Office: Washington DC.

Lyman JM, Willis JK, Johnson GC. 2006. Recent cooling of the upper ocean. *Geophys. Res. Lett.* **33**: L18604, DOI: 10.1029/2006GL027033

Lyman JM, Good SA, Gouretski VV, Ishii M, Johnson GC, Palmer MD, Smith DM, Willis JK. 2010. Robust warming of the global upper ocean. *Nature* **465**: 334–337.

Madec G. 2008. 'NEMO reference manual, ocean dynamics component: NEMO-OPA. Preliminary version'. Note du Pôle de modélisation 27. Institut Pierre-Simon Laplace: Paris.

Masina S, Di Pietro P, Storto A, Navarra A. 2011. Global ocean re-analyses for climate applications. *Dyn. Atmos. Oceans* **52**: 341–366.

McDougall TJ. 1987. Neutral surfaces. *J. Phys. Oceanogr.* **17**: 1950–1964.

Meyssignac B, Becker M, Llovel W, Cazenave A. 2012. An assessment of two-dimensional past sea-level reconstructions over 1950–2009 based on tide-gauge data and different input sea-level grids. *Surv. Geophys.* **33**: 945–972. DOI: 10.1007/s10712-011-9171-x

Mogensen K, Balmaseda MA, Weaver AT. 2012. 'The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4'. Tech. Memo. 668. ECMWF: Reading, UK.

Molteni F, Stockdale T, Balmaseda MA, Balsamo G, Buizza R, Ferranti L, Magnusson L, Mogensen K, Palmer TN, Vitart F. 2011. 'The new ECMWF seasonal forecast system (System 4)'. Tech. Memo. 656. ECMWF: Reading, UK.

Palmer MD, Haines K, Tett SFB, Ansell TJ. 2007. Isolating the signal of ocean global warming. *Geophys. Res. Lett.* **34**: L23610, DOI: 10.1029/2007GL031712

Picaut J, Tournier R. 1991. Monitoring the 1979-1985 equatorial Pacific current transports with expendable bathythermograph data. *J. Geophys. Res.* **96**: 3263–3277.

Ponte RM, Quinn KJ, Wunsch C, Heimbach P. 2007. A comparison of model and GRACE estimates of the large-scale seasonal cycle in ocean bottom pressure. *Geophys. Res. Lett.* **34**: L09603, DOI: 10.1029/2007GL029599

Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W. 2002. An improved *in situ* and satellite SST analysis for climate. *J. Climate.* **15**: 1609–1625.

Ricci S, Weaver AT, Vialard J, Rogel P. 2005. Incorporating temperature-salinity constraints in the background-error covariance of variational ocean data assimilation. *Mon. Wea. Rev.* **133**: 317–338.

Rio MH, Guinehut S, Larnicol G. 2011. New CNES-CLS09 global mean dynamic topography computed from the combination of GRACE data, altimetry, and *in situ* measurements. *J. Geophys. Res.* **116**: C07018, DOI: 10.1029/2010JC006505

Saha S, Moorthi S, Pan H-L, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang J, Hou Y-T, Chuang H-Y, Juang H-MH, Sela J, Iredell M, Treadon R, Kleist D, Van Delst P, Keyser D, Derber J, Ek M, Meng J, Wei H, Yang R, Lord S, Van Den Dool H, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm J-K, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou C-Z, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds RW, Rutledge G, Goldberg M. 2010. The NCEP climate forecast system reanalysis. *Bull. Am. Meteorol. Soc.* **91**: 1015–1057.

Stark JD, Donlon CJ, Martin MJ, McCulloch ME. 2007. OSTIA: An operational, high-resolution, real time, global sea surface temperature analysis system. In *Oceans 2007 – Europe*. 18–21 June 2007, Aberdeen, UK. IEEE.

Storto A, Dobricic S, Masina S, Pietro P. 2011. Assimilating along-track altimetric observations through local hydrostatic adjustment in a global ocean variational assimilation system. *Mon. Weather Rev.* **139**: 738–754.

Troccoli A, Haines K. 1999. Use of Temperature–Salinity relation in a data assimilation context. *J. Atmos. Oceanic Technol.* **16**: 2011–2025.

Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Van De Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**: 2961–3012.

Vidard A, Balmaseda MA, Anderson DLT. 2009. Assimilation of altimeter data in the ECMWF Ocean Analysis System 3. *Mon. Wea. Rev.* **137**: 1393–1408.

Wahr J, Swenson S, Zlotnicki V, Velicogna I. 2004. Time-variable gravity from GRACE: First results. *Geophys. Res. Lett.* **31**: L11501, DOI: 10.1029/2004GL019779

Wahr J, Swenson S, Velicogna I. 2006. Accuracy of GRACE mass estimates. *Geophys. Res. Lett.* **33**: L06401, DOI: 10.1029/2005GL025305

Weaver AT, Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.* **127**: 1815–1846.

Weaver AT, Vialard J, Anderson DLT. 2003. Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks. *Mon. Weather Rev.* **131**: 1360–1378.

Weaver AT, Deltel C, Machu E, Ricci S, Daget N. 2005. A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.* **131**: 3605–3625.

Wijffels S, Willis J, Domingues CM, Barker P, White NJ, Gronell A, Ridgway K, Church JA. 2009. Changing expendable bathythermograph fall rates and their impact on estimates of thermosteric sea level rise. *J. Climate* **21**: 5657–5672.

Wolff J, Maier-Reimer E, Legutke S. 1997. 'The Hamburg ocean primitive-equation model'. Tech. Note 13, Deutsches Klimarechenzentrum: Hamburg, Germany.

Woodworth PL, Player R. 2003. The Permanent Service for Mean Sea Level. *J. Coastal Res.* **19**: 287–295.

Willis JK, Chambers DP, Nerem RS. 2008. Assessing the globally averaged sea level budget on seasonal to interannual timescales. *J. Geophys. Res.* **123**: C06015, DOI: 10.1029/2007JC004517

Wunsch C, Heimbach P. 2007. Practical global oceanic state estimation. *Pysica D* **230**: 197–208.

Xue Y, Alves O, Balmaseda MA, Ferry N, Good S, Ishikawa I, Lee T, McPhaden MJ, Peterson KA, Rienecker M. 2010. Ocean state estimation for global ocean monitoring: ENSO and beyond ENSO. In *OceanObs'09: Conference on Sustained Ocean Observations and Information for Society*, vol. 2, Venice, 21–25 September 2009. Hall J, Harrison DE, Stammer D. (eds). ESA publication WPP-306, DOI: 10.5270/OceanObs09

Yin Y, Alves O, Oke PR. 2011. An ensemble ocean data assimilation system for seasonal prediction. *Mon. Weather Rev.* **139**: 786–808.

Zhu J, Huang B, Balmaseda MA. 2011. An ensemble estimation of the variability of upper-ocean heat content over the tropical Atlantic Ocean with multi-ocean reanalysis products. *Clim. Dyn.* in press. DOI: 10.1007/s00382–011–1189–8

Zhu J, Huang B, Marx L, Kinter III JL, Balmaseda MA, Zhang R-H, Hu Z-Z. 2012. Ensemble ENSO hindcasts initialized from multiple ocean analyses. *Geophys. Res. Lett.* **39**: L09602, DOI: 10.1029/2012GL051503

