



Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts

E. W. Blockley, M. J. Martin, A. J. McLaren, A. G. Ryan, J. Waters, D. J. Lea, I. Mirouze, K. A. Peterson, A. Sellar, and D. Storkey

Met Office, FitzRoy Road, Exeter, EX13PB, UK

Correspondence to: E. W. Blockley (ed.blockley@metoffice.gov.uk)

Received: 1 November 2013 – Published in Geosci. Model Dev. Discuss.: 29 November 2013

Revised: 10 September 2014 – Accepted: 15 September 2014 – Published: 10 November 2014

Abstract. The Forecast Ocean Assimilation Model (FOAM) is an operational ocean analysis and forecast system run daily at the Met Office. FOAM provides modelling capability in both deep ocean and coastal shelf sea regimes using the NEMO (Nucleus for European Modelling of the Ocean) ocean model as its dynamical core. The FOAM Deep Ocean suite produces analyses and 7-day forecasts of ocean tracers, currents and sea ice for the global ocean at $1/4^\circ$ resolution. Satellite and in situ observations of temperature, salinity, sea level anomaly and sea ice concentration are assimilated by FOAM each day over a 48 h observation window. The FOAM Deep Ocean configurations have recently undergone a major upgrade which has involved the implementation of a new variational, first guess at appropriate time (FGAT) 3D-Var, assimilation scheme (NEMOVAR); coupling to a different, multi-thickness-category, sea ice model (CICE); the use of coordinated ocean-ice reference experiment (CORE) bulk formulae to specify the surface boundary condition; and an increased vertical resolution for the global model.

In this paper the new FOAM Deep Ocean system is introduced and details of the recent changes are provided. Results are presented from 2-year reanalysis integrations of the Global FOAM configuration including an assessment of short-range ocean forecast accuracy. Comparisons are made with both the previous FOAM system and a non-assimilative FOAM system. Assessments reveal considerable improvements in the new system to the near-surface ocean and sea ice fields. However there is some degradation to sub-surface tracer fields and in equatorial regions which highlights specific areas upon which to focus future improvements.

1 Introduction

The Forecast Ocean Assimilation Model (FOAM) system is an operational ocean forecasting system run daily at the Met Office which provides modelling capability in both deep ocean and shelf sea regimes. FOAM has been producing global analyses and forecasts for the deep ocean operationally since 1997 (Bell et al., 2000). The FOAM Deep Ocean system was radically overhauled at the end of the last decade, when it was upgraded to use the Nucleus for European Modelling of the Ocean (NEMO; Madec, 2008) community model as its dynamical core. As part of this change, termed FOAM version 10 (FOAM v10), the deep ocean configurations were rationalised to comprise a $1/4^\circ$ global model with three one-way-nested $1/12^\circ$ regional models in the North Atlantic, Indian Ocean and Mediterranean Sea (Storkey et al., 2010).

Forecasts are primarily produced for use by the Royal Navy, but there is also an increasing requirement for FOAM within the commercial, ecological and government sectors for applications involving safety at sea and shipping; monitoring of oil spills and pollutants as well as offshore commercial operations (Davidson et al., 2009; Brushett et al., 2011; Jacobs et al., 2009). Additionally ocean and sea ice analyses from the Global FOAM configuration are used as initial conditions for the Met Office's GloSea5 coupled ocean–atmosphere seasonal and medium-range forecasting systems (MacLachlan et al., 2014). This coupled forecasting system provides short-range $1/4^\circ$ global ocean forecasts as part of the MyOcean2 project (www.myocean.eu); with previous versions of FOAM having provided global analyses

and forecasts as part of the original MyOcean project. FOAM was also one of the systems contributing to the Global Ocean Data Assimilation Experiment (GODAE; Bell et al., 2009; Dombrowsky et al., 2009) and is participating in the GODAE OceanView follow-on project (Le Traon et al., 2010).

January 2013 saw the operational implementation of a major upgrade to the FOAM Deep Ocean system, denoted FOAM version 12 (FOAM v12). The new system retains the NEMO ocean model which is coupled to the Los Alamos sea ice model (CICE) of Hunke and Lipscomb (2010) in place of NEMO's native Louvain-la-Neuve Sea Ice Model 2 (LIM2; Fichefet and Maqueda, 1997; Bouillon et al., 2009). This change from the LIM2 model to CICE was driven by the need to be consistent with the Met Office seasonal forecasting (GloSea; MacLachlan et al., 2014; Arribas et al., 2011) and climate modelling (HadGEM; Hewitt et al., 2011; Johns et al., 2006) systems to support the Met Office's aim of producing seamless forecasts across all timescales (Brown et al., 2012). In particular, as the FOAM analyses are used as initial condition for the GloSea5 seasonal forecasting system – which uses the CICE sea ice model with five thickness categories – it is important that the two systems are consistent so as to minimise coupled initialisation shock. The ocean surface boundary condition (SBC) has been upgraded from direct forcing, with fluxes derived by the atmospheric model, to use the CORE bulk formulation of Large and Yeager (2004). This change means that the bulk formulae calculations are now performed in the ocean model using an evolving ocean surface to provide a more realistic representation of atmosphere interactions at the ocean and ice surface. The analysis correction (AC) assimilation scheme described in Storkey et al. (2010) and Martin et al. (2007) has been replaced with a newly developed variational (3D-Var) assimilation scheme called NEMOVAR (Mogensen et al., 2012; Balmaseda et al., 2013; Mogensen et al., 2009). NEMOVAR has been specifically developed for use with NEMO and has been further tuned for the $1/4^\circ$ global model by Waters et al. (2013, 2014). Initial comparisons between NEMOVAR and AC show considerable improvements to ocean surface fields, particularly in areas of high variability, as well as the Atlantic meridional overturning circulation (AMOC) at 26.5° N (Waters et al., 2014; Roberts et al., 2013). Improvements to the initial ocean conditions can play an important role in the improvement of coupled seasonal forecasts (Barnston et al., 2012), whilst the potential importance of the AMOC for controlling sub-surface temperature anomalies in the subtropical Atlantic has recently been shown by Cunningham et al. (2013).

This paper documents the developments that were made to the Global FOAM configuration and provides an assessment of the new global analyses and forecasts made relative to the previous FOAM v11 system. The paper is structured as follows: in Sect. 2 the FOAM v12 system is described and the evolution of the system is detailed from FOAM v10 through to FOAM v12. Details of Global FOAM reanalyses

and forecast experiments are documented in Sect. 3, and results from these integrations are presented in Sect. 4. The paper concludes with a summary in Sect. 5.

2 System description

2.1 Physical model

The Global FOAM configuration is based on the ORCA025 setup developed by Mercator Océan (Dréville et al., 2008). This tripolar grid is effectively a regular Mercator grid over the majority of the globe with a $1/4^\circ$ (28 km) horizontal grid spacing at the Equator reducing to 7 km at high southern latitudes in the Weddell and Ross seas. To avoid singularities associated with the convergence of meridians at the North Pole, a stretched grid is used in northern latitudes with two poles in the Arctic (on the North American and Eurasian landmasses respectively) as described by Madec (2008). Using this irregular grid gives a typical grid spacing of approximately 10 km in the Arctic Ocean basin.

The vertical coordinate system is based on geopotential levels using the DRAKKAR 75 level set. These levels are prescribed using a double-tanh function distribution to give an increased concentration of levels in the near-surface without compromising the resolution in deeper waters. The model has a 1 m top box in order to better resolve shallow mixed layers and potentially capture diurnal variability (Bernie et al., 2005). Partial cell thickness is used at the sea floor (Adcroft et al., 1997; Pacanowski and Gnanadesikan, 1998) to better resolve the bottom topography. The model bathymetry is the DRAKKAR G70 bathymetry, which is based on the ETOPO2v2 data set and created using methods described in Barnier et al. (2006).

The modelling component of the FOAM v12 system is version 3.2 of the NEMO ocean model (Madec, 2008) – a primitive equation model with variables distributed on a three-dimensional Arakawa C grid. The model uses a linear filtered free surface (Roullet and Madec, 2000) and free-slip lateral momentum boundary condition. A vector invariant formulation of the momentum equations is used, with the total vorticity term discretised using an energy- and enstrophy-conserving scheme adapted from Arakawa and Lamb (1981). Barnier et al. (2006) show that this combined use of partial cells, the energy- and enstrophy-conserving momentum advection scheme and the free-slip lateral boundary condition gives an improved representation of the mesoscale circulation in the DRAKKAR NEMO ORCA025 configuration and, in particular, western boundary currents such as the Gulf Stream, Kuroshio and Agulhas.

Horizontal momentum diffusion is performed using a bi-Laplacian operator along geopotential levels with diffusion coefficient $-1.5 \times 10^{11}\text{ m}^4\text{ s}^{-1}$. Meanwhile tracer diffusion is Laplacian and along isopycnals using diffusion coefficient $300\text{ m}^2\text{ s}^{-1}$. These diffusion values are valid at the Equator,

where the grid spacing is a maximum and the coefficients are reduced with decreasing grid spacing to prevent numerical instabilities and unrealistically high diffusion in areas of increased horizontal resolution (such as the Weddell Sea). The Laplacian coefficient scales linearly with the grid spacing, and the bi-Laplacian coefficient scales with the cube of the grid spacing. The tracer equations use a total-variation-diminishing (TVD) advection scheme (Zalesak, 1979) to avoid the problem of overshooting where sharp gradients exist in the tracer fields (Lévy et al., 2001).

Vertical mixing is parametrised using the turbulent kinetic energy (TKE) scheme of Gaspar et al. (1990) (embedded into NEMO by Blanke and Delecluse, 1993). This scheme includes a prognostic equation for the TKE and a diagnostic equation for the turbulent mixing length based on the local stability profile. Convection is parametrised using an enhanced vertical diffusion, and the mixing effect of Langmuir circulations is prescribed using the simple parametrisation proposed by Axell (2002). The scheme uses background vertical eddy viscosity and diffusivity coefficients of $1.0 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$ and $1.0 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$ respectively and buoyancy mixing length scale minimum values of 0.01 m at the surface and 0.001 m in the interior – consistent with the values used within DRAKKAR. The TKE scheme within NEMO was updated at version 3.2 to ensure dynamical consistency in the space–time discretisations (Burchard, 2002).

A quadratic bottom friction boundary condition is applied together with an advective and diffusive bottom boundary layer for temperature and salinity tracers (Beckmann and Döscher, 1997). There is a geographical variation of parameters to provide enhanced mixing in the Indonesian Throughflow (ITF), Denmark Strait and Bab el-Mandeb. Bottom intensified tidal mixing is parametrised following the formulation proposed by St. Laurent et al. (2002) using K1 and M2 mixing climatologies provided by the DRAKKAR project. The Indonesian Throughflow area is treated as a special case, and the parametrisations of Koch-Larrouy et al. (2007) (adapted from those of St. Laurent et al., 2002) are employed to better reproduce the effects of the strong internal tides that exist in this highly dynamic region.

The model is forced at the surface using the CORE bulk formulae scheme of Large and Yeager (2004) using fields provided by the Met Office Unified Model (UM) global Numerical Weather Prediction (NWP) system (Davies et al., 2005) – currently running at a horizontal resolution of approximately 25 km. These forcing fields consist of 3-hourly radiative fluxes, 3-hourly 10 m temperature and humidity fields and 1-hourly 10 m wind speeds. An RGB (red, green, blue) scheme is used for the penetration of solar radiation (Lengaigne et al., 2007) with a uniform chlorophyll value of 0.05 g L^{-1} . A Haney flux correction (Haney, 1971) is applied to the sea surface salinity (SSS) based on the difference between the model and climatology. River outflow is input to the model as a surface freshwater flux with an enhanced vertical diffusion at river mouths – with mixing

coefficient $2.0 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ over the top 10 m – to mix the fresh water to depth. The climatological river run-off fields for ORCA025 were derived by Bourdalle-Badie and Treguier (2006) based on estimates given in Dai and Trenberth (2002).

The long-time evolution of sub-surface tracer fields is controlled by way of 3-D Newtonian damping using temperature and salinity climatologies with a 360-day timescale. The temperature and salinity climatologies used for this damping – and also for the Haney flux salinity correction – were created by averaging the EN3v2a analysis (updated from Ingleby and Huddleston, 2007) over the years 2004–2008. However, as there were problems with the ingestion of data in the Black Sea into EN3v2a during this period, the temperature and salinity climatologies in this region were taken from the WOA2001 $1/4^\circ$ analysis of Boyer et al. (2005).

The sea ice model used is version 4.1 of the Los Alamos CICE model of Hunke and Lipscomb (2010) based on the HadGEM3 implementation of Hewitt et al. (2011). The CICE model determines the spatial and temporal evolution of the ice thickness distribution (ITD) due to advection, thermodynamic growth and melt, and mechanical redistribution/ridging (Thorndike et al., 1975). At each model grid point the ice pack is divided into five thickness categories (lower bounds: 0, 0.6, 1.4, 2.4 and 3.6 m) to model the sub-grid-scale ITD, with an additional ice-free category for open water areas.

The thermodynamic growth and melt of the sea ice is calculated using the zero-layer thermodynamic model of Semtner (1976), with a single layer of ice and a single layer of snow. Although the standard CICE configuration uses multi-layer thermodynamics, this scheme is not currently compatible with the coupling used in HadGEM3 or GloSea5, and so the zero-layer scheme is used for consistency. The calculated growth or melt rates are used to transport ice between thickness categories using the linear remapping scheme of Lipscomb (2001). Ice dynamics are calculated using the elastic–viscous–plastic (EVP) scheme of Hunke and Dukowicz (2002), with ice strength determined using the formulation of Rothrock (1975). Sea ice ridging is modelled using a scheme based on work by Thorndike et al. (1975), Hibler (1980), Flato and Hibler (1995) and Rothrock (1975). The ridging participation function proposed by Lipscomb et al. (2007) is used, with the ridged ice being distributed between thickness categories assuming an exponential ITD.

The CICE model runs on the same ORCA025 tripolar grid as the NEMO ocean model with NEMO–CICE coupling as detailed in the HadGEM3 documentation (Hewitt et al., 2011). Unlike HadGEM3 however, the freezing temperature in the FOAM system is dependent on salinity to provide a more realistic representation of ice melting and freezing mechanisms and to give better consistency when assimilating both sea surface temperature (SST) and sea ice concentration. The CICE model uses its own CORE bulk formulation to specify surface boundary conditions, which is based on the CICE standard values.

2.2 Data assimilation

The data assimilation component of the FOAM v12 system is NEMOVAR (Mogensen et al., 2012). NEMOVAR is a multivariate, incremental 3D-Var, first guess at appropriate time (FGAT) data assimilation scheme that has been developed specifically for NEMO in collaboration with CERFACS (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique), ECMWF (European Centre for Medium-range Weather Forecasting) and INRIA-LJK (French Institute for Research in Computer Science and Automation–Jean Kuntzmann Laboratory). The state vector in NEMOVAR consists of temperature, salinity, surface elevation, sea ice concentration and horizontal velocities. Key features of NEMOVAR are the multivariate relationships which are specified through a linearised balance operator (Weaver et al., 2005) and the use of an implicit diffusion operator to model background error correlations (Mirouze and Weaver, 2010).

The NEMOVAR system has been tuned at the Met Office for the ORCA025 configuration (Waters et al., 2014). In this implementation the state vector was extended to include sea ice concentration, which is treated as an unbalanced variable in the linearised balance relationships. The background error variances for temperature and salinity are specified as a combination of statistical errors and vertical parametrisations. This allows for flow-dependent errors while incorporating climatological information. The background error variances for sea surface height (SSH) and sea ice concentration are statistical errors. The statistical error variances were calculated using the NMC method (developed at the National Meteorology Center; Parrish and Derber, 1992) on 2 years' worth of 24 and 48 h forecast fields and were then scaled using background error variances calculated from the Hollingsworth and Lonnberg (1986) method. In a similar way, the observation variances are calculated from the NMC method scaled by observation error variances calculated from the Hollingsworth and Lonnberg method. Martin et al. (2007) provides more details on the method used to calculate these statistical error variances. The horizontal background error correlations for temperature, salinity and sea ice concentration are prescribed based on the Rossby radius (Cummings, 2005), while the barotropic SSH correlation length scales are set at 4° . The vertical background error correlations are flow-dependent and parametrised based on the mixed layer depth (Waters et al., 2014).

The NEMOVAR system includes bias correction schemes for SST and altimeter data, and their implementations are detailed in Waters et al. (2013). The SST bias correction scheme aims to remove bias in SST data due to errors in the non-constant atmospheric constituents used in the retrieval algorithms by correcting data to a reference data set of assumed unbiased SST observations (Martin et al., 2007; Donlon et al., 2012). The SST biases are determined using a 2-D version of NEMOVAR which calculates a large-scale

analysis of the match-ups between the SST observations and the reference data set. An altimeter bias correction scheme is used to correct biases in the mean dynamic topography (MDT) which is added to the sea level anomaly (SLA) altimeter observations prior to assimilation. The bias correction is applied in a similar way to Lea et al. (2008), by adding an additional altimeter bias field to the data assimilation control vector and including extra terms in the 3D-Var cost function. The mean dynamic topography used is the CNES09 (Centre National d'Etudes Spatiales) MDT of Rio et al. (2011). Systematic errors in the wind forcing near the Equator are counteracted by the addition of a correction term to the subsurface pressure gradients in the tropics to improve the retention of temperature and salinity increments by the model (Bell et al., 2004).

Observations are read into NEMO and model fields are mapped into observation space using the NEMO observation operator to create model counterparts using bilinear interpolation in the horizontal and cubic splines in the vertical directions. These FGAT model–observation comparisons, called the innovations, are subsequently used as inputs to the NEMOVAR assimilation system. NEMOVAR assimilates satellite and in situ observations of SST, in situ observations of sub-surface temperature and salinity, altimeter observations of SSH and satellite observations of sea ice concentration. Velocity data are not assimilated into NEMOVAR, but balanced velocities are determined through the multivariate balance relationships.

Observations are assimilated using a 24 h assimilation window, and increments are applied to the model using a 24 h incremental analysis update (IAU) step (Bloom et al., 1996) with constant increments. Analysis updates are made to the state variables in the NEMO model with the exception of sea ice concentration updates which are made in the CICE model, taking into account the distribution of ice concentration between the different ice thickness categories (Peterson et al., 2014). Updates increasing ice concentration are always made to the thinnest (0–0.6 m) category ice at a thickness of 0.5 m, whilst updates decreasing ice concentration are made to the thinnest ice thickness category available in that grid cell.

2.2.1 Observations assimilated

The satellite SST data assimilated include sub-sampled level 2 Advanced Very High Resolution Radiometer (AVHRR) data from NOAA and MetOp satellites supplied by the Global High-Resolution Sea Surface Temperature (GHRSSST) project. In situ SST from moored buoys, drifting buoys and ships are obtained from the Global Telecommunications System (GTS). This in situ data set is considered unbiased and is used as the reference for the satellite SST bias correction scheme. Sea level anomaly observations from Jason-2, CryoSat2 and Jason-1 satellite altimeters are provided by CLS (Collecte Localisation Satellites) in near-real

time through the MyOcean project. Sub-surface temperature and salinity profiles are obtained from the GTS and include measurements taken by Argo profiling floats, underwater gliders, moored buoys and marine mammals as well as manual profiling methods such as expendable bathythermograph (XBT) and conductivity–temperature–depth (CTD). The sea ice concentration observations are Special Sensor Microwave Imager/Sounder (SSMIS) data provided by the EUMETSAT Ocean Sea Ice Satellite Application Facility (OSI-SAF). These OSI-SAF sea ice data are derived using data from several different SSMIS satellites and provided as a daily gridded product on a 10 km polar stereographic projection (OSI-SAF, 2012).

2.3 Operational implementation and daily running

The FOAM Deep Ocean system is run daily in the Met Office operational suite in an early morning slot. Starting from $T-48$ h each day, the system performs two 24 h data assimilation cycles before running a 7-day forecast. Performing data assimilation over a 48 h observation window in this manner allows the FOAM system to assimilate considerably more observations than would be possible with a single 24 h window owing to the inclusion of late-arriving observations. A detailed breakdown of the daily operational running is as follows:

1. Observations (as detailed in Sect. 2.2 above) are obtained from the Met Office’s observations database separately for the ($T-48$ h, $T-24$ h] and ($T-24$ h, $T+00$ h] time periods and are quality-controlled using the methods described in Storkey et al. (2010) and Ingleby and Lorenc (1993). The satellite SST bias correction is then performed using the reference data sets (at present only in situ SST) to correct for biases in the satellite SST data.
2. Surface boundary conditions are processed from Met Office UM Global NWP system output (Davies et al., 2005), using analysis fields from $T-48$ h up to $T+00$ h followed by forecast fields out to $T+168$ h. The resulting SBCs are then translated onto the FOAM model grids using bilinear interpolation.
3. A 24 h NEMO model forecast is then run for the period $T-48$ h to $T-24$ h using the observation operator described in Sect. 2.2 to create FGAT model–observation differences (innovations) valid at the observation locations/times.
4. The FGAT innovations output by the observation operator are then used by the NEMOVAR assimilation scheme to generate fields of daily increments as detailed in Sect. 2.2 and Waters et al. (2013, 2014).
5. The model is then rerun for the period $T-48$ h to $T-24$ h, and these increments are applied evenly over

the 24 h period using an incremental analysis update (IAU) method (Bloom et al., 1996). At the end of this first IAU step the $T-24$ h NEMO and CICE “best estimate” analyses are saved for the initialisation of the $T-48$ h observation operator step on the following day.

6. The daily data assimilation cycle described in items 3–5 above is then repeated for the period $T-24$ h to $T+00$ h, and the model is run out to $T+168$ h to produce a 7-day forecast. The $T+00$ h NEMO and CICE analyses are saved for initialisation of the GloSea5 (MacLachlan et al., 2014) coupled seasonal and medium-range forecasts. Owing to the variation in observation arrival times, this ($T-24$ h, $T+00$ h] “update run” will have been performed using fewer observations than the ($T-48$ h, $T-24$ h] “best estimate” analysis. Typically it will only have used 65 % of the sub-surface profiles and may not have had access to CryoSat2 SLA or OSI-SAF sea ice concentration data.
7. The forecasts are then post-processed to produce specific forecasts for various users as well as boundary conditions for the FOAM $1/12^\circ$ regional configurations (Storkey et al., 2010) and FOAM Shelf Seas configurations (O’Dea et al., 2012; Hyder et al., 2013) for the next day. Products are delivered to the Royal Navy via a dedicated communications link and to other customers via FTP.

The Met Office operational suite benefits from round-the-clock operator technical support, with additional out-of-hours support being provided by ocean forecasting scientists where required. This helps to make the operational delivery of FOAM products robust – keeping failures and instances of late delivery to a minimum.

The new v12 FOAM Deep Ocean operational system was initialised in autumn 2012 from pre-operational trials (detailed in Sect. 3). It was implemented operationally on 17 January 2013 after a successful period of trial running in the Met Office’s parallel suite.

2.4 Evolution of the global FOAM configuration from v11 to v12

In this section differences are highlighted between the new v12 FOAM global configuration described in Sect. 2.1 and the previous v11 version. Details of the FOAM v11 upgrade relative to the Storkey et al. (2010) FOAM v10 system can be found in Appendix A, whilst a summary of the differences between the global model configurations for FOAM v10, v11 and v12 can be found in Table 1.

The main differences between the new FOAM v12 system and the v11 system are as follows: data assimilation change from the analysis correction scheme to NEMOVAR 3D-Var FGAT scheme; sea ice model change from LIM2 to CICE; and SBC change from direct forcing to CORE bulk formulae.

There have additionally been a number of changes made to the input files and parameters used by the NEMO ocean model (Table 1) as well as an upgrade to the vertical resolution from 50 levels to 75. Motivation for the sea ice model, SBC and assimilation changes was provided in Sect. 1, and the remaining NEMO changes were made to align the FOAM system with the Met Office's climate modelling (HadGEM) and seasonal forecasting (GloSea) systems as part of the Met Office's seamless forecasting agenda (Brown et al., 2012). This was accomplished by using a shared standard UK NEMO Global Ocean configuration which was developed by the NERC–Met Office Joint Ocean Modelling Programme (JOMP) and is based on the DRAKKAR configuration of Barnier et al. (2006).

3 Experiment setup

In order to investigate the quality of the new FOAM v12 system, a series of reanalysis and hindcast trials have been performed using three separate FOAM configurations: the full FOAM v12 system; the full FOAM v11 system; and a free-running FOAM v12 system with no data assimilation (hereafter the “v12”, “v11” and “free” trials). The main purpose of these trials is twofold: first to show the difference between the new FOAM system and the existing system (i.e. v12 versus v11) and second to assess the impact that the data assimilation has on the accuracy of FOAM predictions (v12 versus free). The assessment period for these experiments is the 2-year period from 1 December 2010 until 30 November 2012. The v12 and v11 reanalyses were performed using a single 24 h data assimilation cycle only, rather than the 2 days performed operationally because, as they are run in delayed time rather than near-real time, there would be no benefit in running a longer observation window to capture late-arriving observations.

To assess the model forecast skill, a series of forecast experiments were performed by spawning off 5-day hindcasts from the FOAM v11 and v12 reanalysis trials every day during the middle month of each season (January, April, July and October for both 2011 and 2012). These hindcasts were performed using SBCs generated from forecast, as opposed to analysis, NWP fields to reflect the true manner in which forecasts are run operationally. As April only has 30 days, a 5-day hindcast was also spawned off on 1 May each year to ensure that an equal number of hindcasts were performed per season. The surface forcing for all three trials was derived using output from the same UM Global NWP system which was run at a horizontal resolution of approximately 25 km for the entire duration of the trials.

3.1 Initial conditions

The FOAM v11 experiment was initialised from operational FOAM fields from 1 November 2010 and spun up for 30 days

with full assimilation. Initialisation of the FOAM v12 experiments (v12 and free) was more complicated owing to a change in vertical resolution, the change to use of the multi-category CICE sea ice model and the updated bathymetry. Initial conditions for the CICE model were obtained from a climatology derived from the HadGEM1 coupled climate system of Johns et al. (2006). Sea ice concentration, sea ice thickness and snow thickness fields were taken from a 20-year mean (1986–2005) of a HadGEM1 integration performed with time-varying anthropogenic and natural forcing (Jones et al., 2011; Stott et al., 2006). All other fields required for the CICE model, including ice velocities, were initialised to zero, and these fields were then spun up in a fully assimilative FOAM system (Waters et al., 2013) for a further 3.5 years until 10 June 2010. The ocean temperature and salinity initial conditions for the trials were taken from archived operational FOAM v10 initial conditions on 10 June 2010 and interpolated vertically to the new FOAM v12 grid. Owing to a known problem with Black Sea subsurface salinity in the v11 Global FOAM system, temperature and salinity fields throughout this region were replaced using the climatology developed by the World Ocean Atlas 2001 1/4° analysis (Boyer et al., 2005). All other fields required for the NEMO model, including ocean velocities, were set to zero. The resulting NEMO and CICE initial conditions were then integrated for 21 days without data assimilation to allow the currents to spin up naturally, before commencing a fully assimilative 5-month spin-up from 1 July 2010. After this spin-up both the v12 and free runs were started from the same conditions on 1 December 2010.

3.2 Observations assimilated

Owing to the changing availability of satellite observations during the reanalysis period, the observations used for the trials differ slightly from those used operationally. In particular SST data from the Advanced Microwave Scanning Radiometer–Earth Observing System (AMSRE) and Advanced Along-Track Scanning Radiometer (AATSR) instruments as well as SLA data from the ENVISAT (ENVironmental SATellite) altimeter are available at the start of the reanalyses for a limited period. The CryoSat2 SLA data meanwhile is only available towards the end of the period. The availability of satellite SST and SLA observations for the trial period is detailed in Table 2. The in situ SST and sea ice concentration observations are the same as used operationally, coming from the GTS and OSI-SAF respectively. However the temperature and salinity profiles used for the reanalyses are quality-controlled data provided by the EN3v2a analysis (updated from Ingleby and Huddleston, 2007). Owing to the high accuracy of the ENVISAT AATSR instrument (Donlon et al., 2012, Sect. 2), AATSR data are used alongside the in situ SST data, where present, as reference for the satellite SST bias correction scheme.

Table 1. Differences between the Global FOAM configuration in the new v12 system, the previous v11 system and the v10 system of Storkey et al. (2010).

	FOAM v10	FOAM v11	FOAM v12
Ocean model	NEMO vn3.0	NEMO vn3.2	NEMO vn3.2
Ice model	LIM2: 1 thickness category	LIM2: 1 thickness category	CICE vn4.1: 5 thickness categories
Data assimilation	Analysis Correction (AC)	Analysis Correction (AC)	NEMOVAR (3D-Var FGAT)
Observation window	24 h	48 h	48 h
Mean dynamic topography	Rio05	CNES09	CNES09
Error (co)variances	time invariant	seasonally varying	interpolated daily from seasonally varying estimates
Bathymetry	Mercator Océan ORCA025	Mercator Océan ORCA025	DRAKKAR G70 ORCA025
Vertical levels	50	50	75
Surface forcing	direct fluxes	direct fluxes	CORE bulk formulae
Penetrating solar radiation scheme	2-band	2-band	RGB
Haney retroaction	SST	SST	SSS
Vertical mixing length scale min	0.4 m everywhere	0.4 m everywhere	surface: 0.01 m; interior: 0.001 m
Langmuir cell parametrisation	none	none	yes
Horizontal momentum advection	Laplacian	mixed Laplacian–bi-Laplacian	bi-Laplacian
Lateral momentum BCs	free slip	partial (half) slip	free slip
Enhanced mixing at river mouths	$1.5 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ over top 25 m	$1.5 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ over top 25 m	$2.0 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ over top 10 m
Tidal mixing parametrisations	none	none	DRAKKAR M2 and K1 climatologies
3-D Newtonian damping	none	none	temperature and salinity 1-year timescale
Bottom boundary layer	none	none	advective and diffusive BBL for temperature and salinity
Enhanced bottom friction mixing	none	none	Indonesian Throughflow, Denmark Strait and Bab el Mandeb

Table 2. Availability of satellite SST (upper half) and satellite altimeter (lower half) observations used within the operational implementation of FOAM v12 and the trials described in Sect. 3. If an instrument was operational before the start of the trials on 10 June 2010 or is still operational at the time of writing, “–” is used.

Data source	Start	End
AATSR	–	8 Apr 2012
AMSRE	–	4 Oct 2011
NOAA AVHRR	–	–
MetOp AVHRR	–	–
ENVISAT	–	8 Apr 2012
Jason-1	–	21 Jun 2013
Jason-2	–	–
CryoSat2	4 May 2012	–

4 Assessments

Assessment of the Global FOAM trials described above is split into three parts. Section 4.1 details validation of the analysis fields for all three of the FOAM trials and is concerned with documenting the differences between the new and the old FOAM systems (i.e. v12 versus v11) as well as the impact that the NEMOVAR data assimilation has on the

new v12 model (i.e. v12 versus free). Section 4.2 contains an assessment of the 5-day hindcasts performed during the assimilative trials and describes the difference in forecast skill between the v12 and v11 systems. Section 4.3 describes a qualitative assessment of FOAM model fields performed by comparing SST, SSH and surface velocity fields with gridded observational products.

4.1 Reanalysis validation

Throughout the duration of the reanalyses, FGAT model–observation differences (innovations) are output each day from the NEMO observation-operator step. As well as being used by the data assimilation scheme, these innovations can be used to assess the quality of the FOAM fields during this initial 24 h forecast. Although these observations have not yet been assimilated, data from the same instrument may have been assimilated in previous cycles – 1 day before in most cases and 10 (5) days before for Argo (MedArgo) profiles. Therefore these observations are not strictly independent, but they still provide a very useful assessment and, owing to the sparsity of independent observations, it is common practice to validate assimilative models in this manner (Lellouche et al., 2013; Balmaseda et al., 2013; Storkey et al., 2010). The reanalysis innovations are filtered to ensure that a common subset of observations is used to assess each trial because, owing to differences in the model bathymetry,

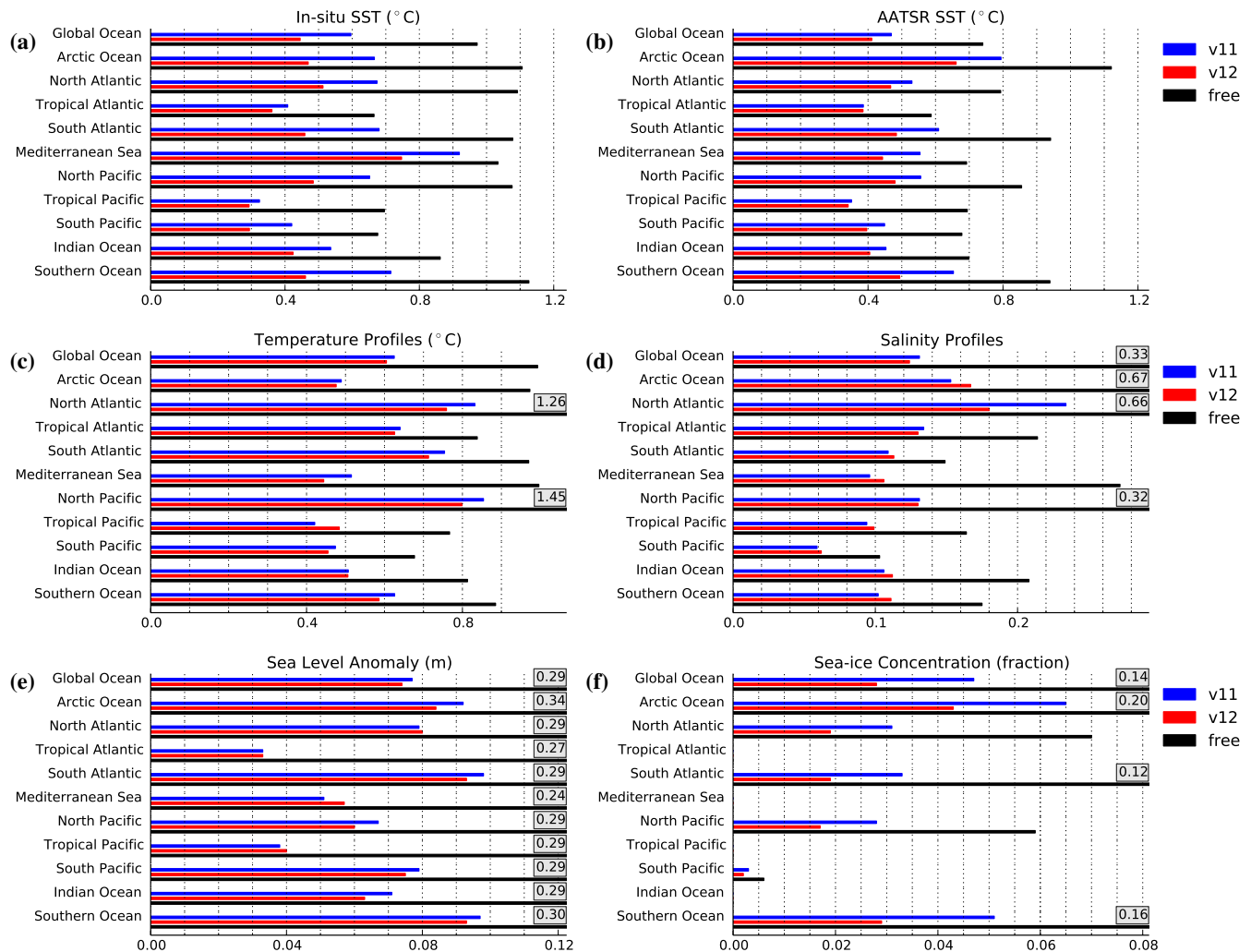


Figure 1. Root-mean-square (rms) errors against observations of (a) in situ surface temperature ($^{\circ}\text{C}$), (b) AATSR satellite surface temperature ($^{\circ}\text{C}$), (c) sub-surface temperature profiles ($^{\circ}\text{C}$), (d) sub-surface salinity profiles (measured on the practical salinity scale), (e) sea level anomaly (m) and (f) sea ice concentration (fraction) for the v12 (red), v11 (blue) and free (black) trials. All statistics are compiled as averages over the full 2-year assessment period save for comparisons with AATSR data, which are only available until 8 April 2012. Where the rms errors for the free run are considerably higher than those for the assimilative runs, the x axis has been truncated in order to allow the reader to see the finer detail for the v12 and v11 runs. In these situations the rms value has been added as an annotation above the corresponding bar.

different numbers of observations were ingested into the v11 and v12 system trials.

Root-mean-square (rms) errors calculated using the re-analysis innovations for SST, SSH, sea ice concentration and sub-surface temperature and salinity profiles can be found in Fig. 1. Meanwhile mean errors (for temperature and salinity fields only) can be found in Fig. 2. SST assessment is made relative to the unbiased data sets that are used for the satellite SST bias correction scheme. These are displayed separately in Figs. 1 and 2 for in situ and AATSR observations (the latter only for the reduced period 1 December 2010–8 April 2012). Profile errors are calculated over all depth levels, so the mean errors displayed in Fig. 2 are actually

depth-averaged biases. These plots are included to provide details of how sub-surface biases, in particular for the free run, are distributed geographically. A better understanding of how the biases change with depth can be obtained from Fig. 3 which shows temperature and salinity profile errors both globally and for the North Atlantic and tropical Pacific regions.

4.1.1 Sea surface temperature (SST)

SST statistics show a clear improvement in the FOAM system at v12 compared to v11, with a reduction in global rms error of over 25% – from 0.60 to 0.45 $^{\circ}\text{C}$ – against in situ

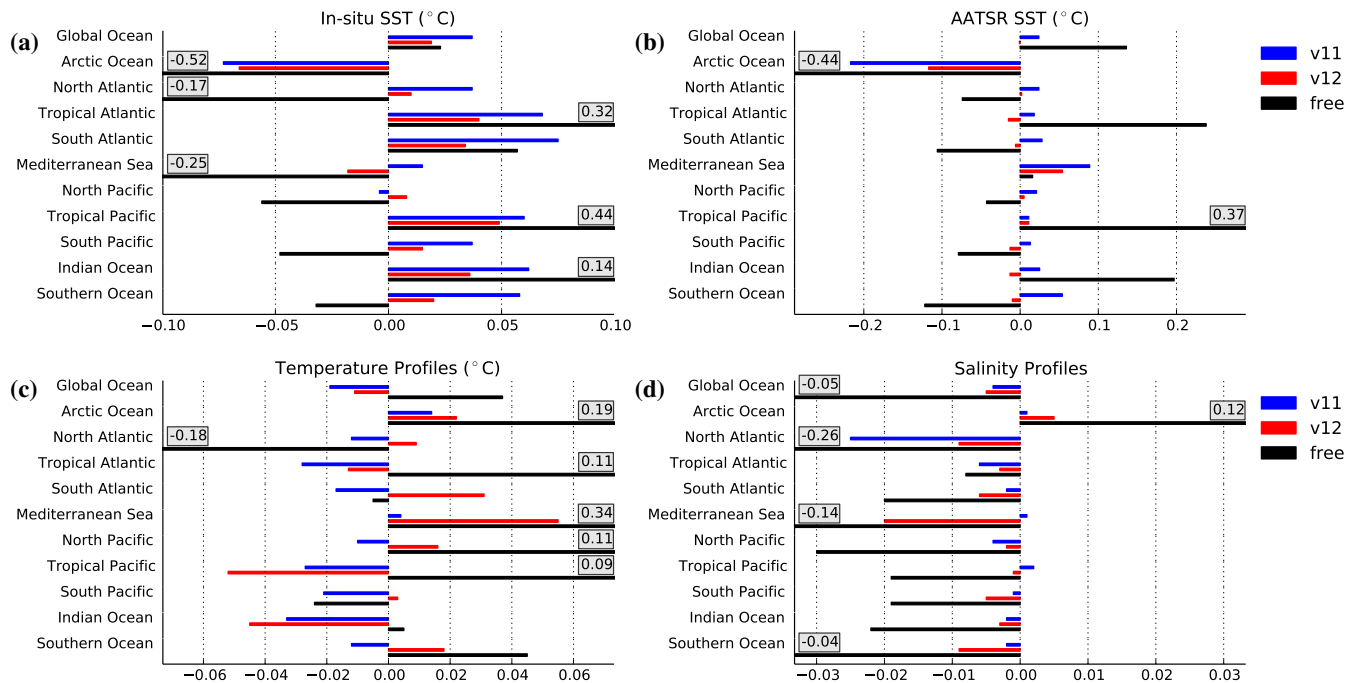


Figure 2. Mean errors against observations of (a) in situ surface temperature ($^{\circ}\text{C}$), (b) AATSR satellite surface temperature ($^{\circ}\text{C}$), (c) sub-surface temperature profiles ($^{\circ}\text{C}$) and (d) sub-surface salinity profiles (measured on the practical salinity scale) for the v12 (red), v11 (blue) and free (black) trials. Mean errors are plotted as modelled–observed, meaning that positive temperature (salinity) values indicate that the model is too warm (salty). All statistics are compiled as averages over the full 2-year assessment period save for comparisons with AATSR data, which are only available until 8 April 2012. Where the mean errors for the free run are considerably higher than those for the assimilative runs, the x axis has been truncated in order to allow the reader to see the finer detail for the v12 and v11 runs. In these situations the mean error value has been added as an annotation above the corresponding bar.

SST observations (Fig. 1a). This decrease is mainly due to lower errors in extratropical areas, with the largest improvements at high latitudes (a reduction in rms error of over 35% in the Southern Ocean and almost 30% in the Arctic). These large SST improvements at high latitudes can be mainly attributed to the NEMOVAR data assimilation scheme fitting smaller-scale features better than the old AC scheme – which is particularly noticeable in high latitudes where the Rossby radius is smaller. Additionally, in ice-covered areas such as the Arctic, improvements are also caused by a more consistent representation of ice–ocean–atmosphere interactions resulting from the CICE and CORE bulk formulae changes.

The free trial performed considerably worse against SST observations than either of the assimilative trials. In particular there are fairly large biases in the free-running model fields in the tropics, where the model is too warm. Fig. 2a shows that mean errors against in situ SST are 0.44°C in the tropical Pacific and 0.32°C in the tropical Atlantic. Root-mean-square errors meanwhile are relatively low in the tropics (see Fig. 1a), which suggests that the majority of the tropical errors in the free run are prescribed by these biases. There is also a significant bias in the Arctic Ocean which is even larger than the tropical biases and is of opposite sign (-0.52°C against in situ SST), showing that the model is too

cold there. This Arctic bias is caused mainly by observations in the boreal summer months, which is consistent with the decreased May–July Arctic sea ice melting detailed later in this section and shown in Fig. 4 (below).

4.1.2 Temperature profiles

Globally the full-depth temperature profile rms errors are lower for the v12 trial (0.61°C) than for the v11 trial (0.63°C). Areas of particular improvement are the North Atlantic, North Pacific and Mediterranean Sea regions (see Fig. 1c). However rms errors are larger in the tropical Pacific and mean errors are worse in the tropical Pacific and Indian Ocean amongst other regions. Log-depth profile plots show that globally v12 temperature errors are considerably lower than for v11 in the top 80 m or so and in particular around 50 m depth, where the v11 system has a cold bias (Fig. 3a). However at 100 m there is a warm bias in the new v12 system that is not seen in the v11 system. Below 100 m depth the rms and mean errors are very similar for the two assimilative systems although they are marginally better for the v11 system. This improvement to the upper 80 m is present over most of the world ocean as illustrated for the North Atlantic in Fig. 3b. One notable exception however is the tropical Pacific

(Fig. 3c), where errors are slightly worse for v12 in the surface layers, with a clear increase in rms centred around 100 m depth.

The free run has worse errors than the v12 assimilative run, with a global rms error of 0.99°C and rms errors exceeding this in the North Atlantic and Pacific. There are also substantial mean errors in the North Atlantic, Arctic Ocean and Mediterranean Sea. Temperature profile errors are considerably worse for the free run through all depths as shown by the black line in Fig. 3a–c. In particular the mean profile errors show that the free-running model has the same warm bias centred at 100 m as can be seen in the v12 run – albeit much more pronounced. This suggests that the degraded temperature fields at 100 m are caused by the new NEMOVAR assimilation system failing to fully constrain a persistent model bias there.

4.1.3 Salinity profiles

The global full-depth salinity profile rms errors are also lower for the v12 trial (0.12) than the v11 trial (0.13). However this improvement seems to be almost exclusively restricted to the North Atlantic, where the rms errors are lower by 23%. There are marginal improvements in the North Pacific and tropical Atlantic, but all other regions are slightly worse for v12 (see Fig. 1d). The somewhat large (23%) reduction in salinity errors seen in the North Atlantic is associated with improvements to the near-surface salinity in coastal locations caused by the upgrade to bulk formulae SBCs (see Fig. 3e). This improvement appears to be limited to the North Atlantic region only because a large proportion of these shallow coastal observations are situated along the east coast of North America. If observations in shallow water areas ($< 100\text{ m}$) are ignored, then the rms errors for v11 and v12 are of similar magnitude.

Log-depth profile plots (Fig. 3d) show that near-surface ($< 70\text{ m}$) salinity is better in the v12 system, with the most notable improvement occurring at around 20 m, where the v11 system has a fresh bias. Error statistics for v12 and v11 are roughly comparable through the rest of the water column. Fig. 2d shows the v11 system to have a significant fresh bias in the North Atlantic region, which is reduced for v12 (Fig. 3e), although mean errors are more pronounced in the Southern Ocean and Mediterranean Sea.

Although the shorter horizontal correlation length scales employed by NEMOVAR allow for tighter matching of small-scale features for dense observation sets such as SST, they also make it harder for the assimilation to constrain the tracer fields when observations are sparse (Waters et al., 2014). This is thought to be responsible for the degradation of salinity in the ocean interior.

The free-running model has particularly bad salinity profile errors with a fairly substantial fresh bias above 120 m depth and rms errors in excess of 0.5 in the top 50 m. The regional distribution of the depth-averaged profile mean errors

(Fig. 2d) shows that the free-running model is too fresh everywhere save for in the Arctic Ocean. These fresh biases are particularly large in the North Atlantic (0.26) and Mediterranean Sea (0.14) regions and are believed to be an artefact of the increased number of coastal observations in these areas. Further investigation into the Arctic salty bias shows that it is probably not a fair reflection of conditions throughout the whole Arctic Ocean owing to the lower number of profile observations (ca. eight per day for the assessment period) and their somewhat restrictive spatial and temporal distribution.

4.1.4 Sea surface height (SSH)

Comparisons against SLA observations are better for v12 than v11, with rms errors reduced by approximately 4% from 7.7 to 7.4 cm (see Fig. 1e). Again the majority of the improvement can be seen in mid–high latitudes (South Atlantic, North Pacific and Southern Ocean). Statistics are better in the Indian Ocean whilst comparable in the tropical Atlantic and worse in the tropical Pacific and the Mediterranean Sea. The fact that v12 statistics are better in the Indian Ocean suggests that, consistent with the findings of Waters et al. (2014), the new system is doing a better job recreating the fronts and mesoscale eddies in highly dynamic regions of this sort.

In order to test this hypothesis, and quantify the relative improvement to the mesoscale eddy fields at v12, the extratropical ocean (between 23° and 66° latitude) was partitioned into high- and low-variability regimes dependent on the spatial distribution of the variability of SLA observations for the full 2-year assessment period. This partitioning was performed using a threshold standard deviation of $\sigma = 0.11\text{ m}$, which was shown to provide the most sensible split between high- and low-variability areas. Root-mean-square errors were calculated separately for both regimes, and the relative improvements can be found in Table 3, which shows the percentage reduction in rms error for v12 relative to v11. This process was performed for the SSH fields using the reanalysis innovations but also for near-surface velocities using drifter-derived current observations as detailed in Sect. 4.1.6 below. Results show that, although the v12 SSH fields are improved over most of the midlatitude areas, the improvement is considerably more (by a factor of 10) in areas of high mesoscale activity, which confirms our hypothesis.

There is clearly a large bias in the free run which causes the statistics to be significantly worse than for the assimilative runs. Time series plots reveal that this is caused by a long-term drift in the model surface height with an approximate increase of 28 cm globally over the course of the 2-year trial period (not shown). This SSH drift appears to be the result of a mismatch between the precipitation and riverine freshwater inputs and is most likely the result of a precipitation bias in the NWP forcing fields. This ties in with the aforementioned surface salinity drifts seen in Fig. 3 and Fig. 2d. It is worth noting that a 2-year drift of 28 cm corresponds to a daily drift of approximately 0.4 mm, which

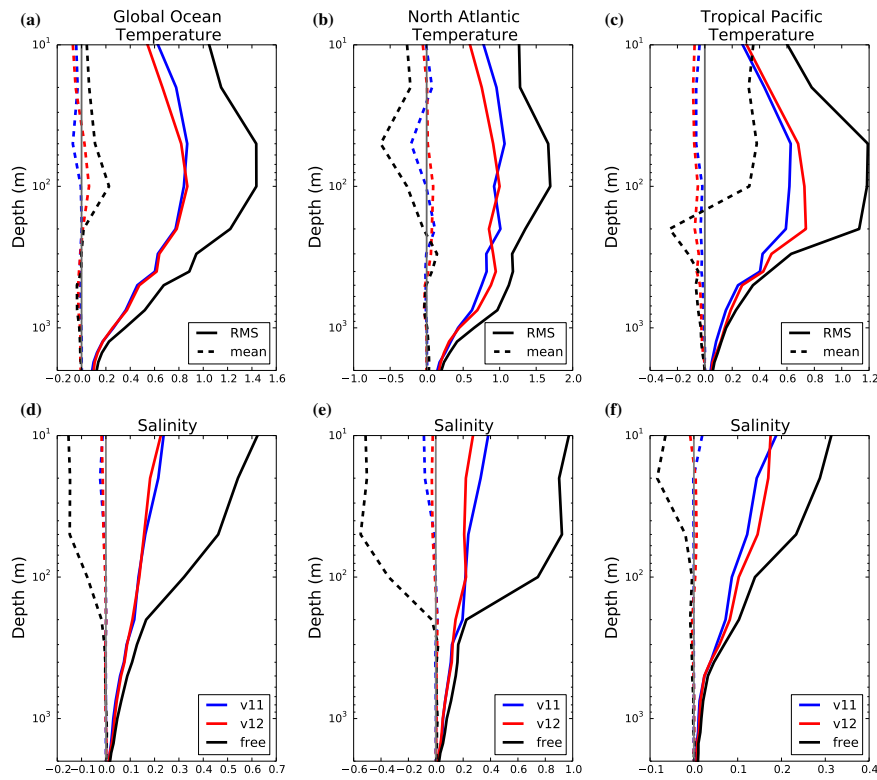


Figure 3. Mean error profiles against EN3 data for temperature (top row; °C) and salinity (bottom row; measured on the practical salinity scale) plotted against model depth (m) on a log scale for the global ocean (left), North Atlantic (centre) and tropical Pacific (right) regions. Solid lines denote rms errors and dashed lines denote mean errors for the v12 (red), v11 (blue) and free (black) trials. Mean errors are plotted as modelled–observed, meaning that positive temperature (salinity) values indicate that the model is too warm (salty).

Table 3. Percentage reduction in rms error for the v12 trial relative to the v11 trial calculated separately for areas of high and low mesoscale variability in the extratropics from 23 to 66° latitude. The variability threshold used is based on the standard deviation of SLA observations with $\sigma = 0.11$ m. The proportion of the extratropical ocean surface classified as either high or low variability using this threshold can be seen in the bottom row of the table.

	High variability (SLA $\sigma \geq 0.11$ m)	Low variability (SLA $\sigma < 0.11$ m)
Sea level anomaly	4.2 %	0.43 %
Zonal velocity	5.8 %	3.0 %
Meridional velocity	5.7 %	3.3 %
Percentage of extra-tropical ocean surface	13 %	87 %

is unlikely to have an adverse effect on the quality of the short-range FOAM forecasts. Mean errors for the assimilative models are typically less than 5 mm and, globally, are less than 5 % of the rms error. Meanwhile for the free run the SSH drift causes mean errors of around 20–25 cm, which are approximately 80% of the rms error. For these reasons SSH mean errors are not included in Fig. 2.

4.1.5 Sea ice concentration and thickness

Sea ice concentration statistics are significantly improved in the v12 system compared to the v11 system, with an approximate reduction of 40 % in global rms error. The reduction in rms error appears to be of similar magnitude both in the Arctic and the Antarctic regions (Fig. 1f).

This improvement comes in part from the sea ice upgrade to the multi-category CICE model; the SBC upgrade to CORE bulk formulae; and the change to NEMOVAR – which has been shown to better resolve smaller-scale features when used with dense observation sets such as the OSI-SAF gridded data (Waters et al., 2014). Initial testing of the component parts of the v12 upgrade (not shown) suggests that roughly half of this improvement is down to the NEMOVAR assimilation upgrade while the remaining half is split evenly between the CICE sea ice model and the CORE bulk formulae SBC upgrades. There is clearly a large difference in sea ice concentration rms errors between the free-running model and the assimilative models in all areas (Fig. 1f). This is also apparent in the mean errors (not shown) and suggests that there are considerable biases in the free-running model over the polar regions.

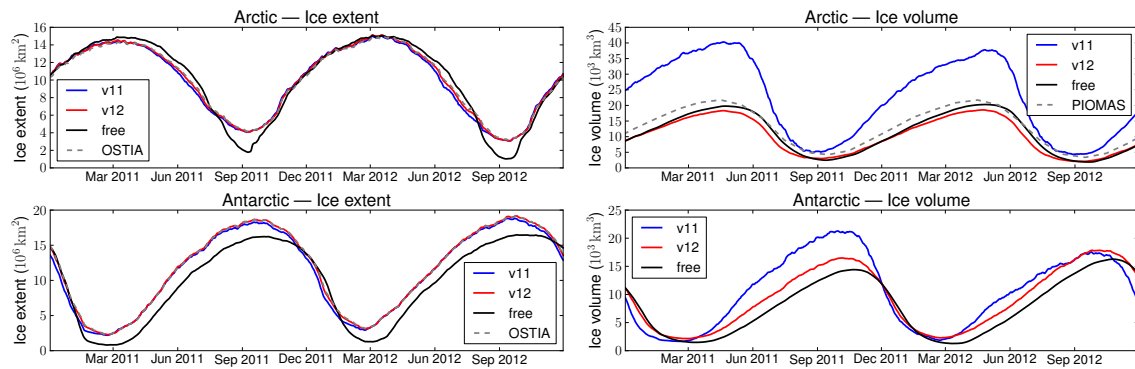


Figure 4. Time series of Arctic (upper) and Antarctic (lower) sea ice extent (left; 10^6 km^2) and volume (right; 10^3 km^3) derived from the v12 (red), v11 (blue) and free (black) trials. Daily OSTIA (Donlon et al., 2012) sea ice extent derived from OSI-SAF ice concentrations and monthly PIOMAS (Schweiger et al., 2011) sea ice volume (Northern Hemisphere only) are plotted as grey dashed lines.

At this stage it should be noted that the sea ice statistics shown in Fig. 1f are obtained from all of the OSI-SAF gridded data over the entire 2-year assessment period. As the OSI-SAF grid (as detailed in OSI-SAF, 2012) is designed to cover all areas of the globe where sea ice may be present at any point during the year, this means that these data include many areas where both model and observations have zero concentration values. This is particularly true during summer months. These statistics therefore will be diluted by the large number of observations taken away from the ice pack where the ocean is ice-free and will not truly represent the changes at the, highly variable, ice edge where the majority of ice concentration differences would be expected to occur.

It is therefore more interesting to consider errors in sea ice extent – i.e. the area of all grid cells which contain ice concentration of 15 % or more – rather than ice concentration. Fig. 4 shows time series of ice extent (left-hand plots) derived from the v12, v11 and free trials. Also plotted is sea ice extent derived from $1/20^\circ$ OSTIA (Donlon et al., 2012) ice concentration fields. These OSTIA ice fields are interpolated each day from the 10 km OSI-SAF observations after performing filling to account for differences in the land–sea masks and the fact that OSI-SAF observations do not extend right to the North Pole (Donlon et al., 2012). Ice extents are calculated from the OSTIA analysis in the same manner as the FOAM extents after first being re-gridded onto the coarser ORCA025 model grid.

The v12 and v11 systems are very similar to each other and to the OSTIA system although the v12 extents follow the OSTIA analyses slightly closer than do the v11 extents. In fact the v12 and OSTIA extents (red and grey lines respectively) are indistinguishable from one another for most of the trial period save for during the Arctic melt season (mid-May to July), where the v12 extent is slightly lower than for OSTIA. Closer inspection of the time series shows that the v12 extents are consistently higher than the v11 extents during the melt periods but are slightly lower than those derived from

the OSTIA analyses (which can be seen by considering the dashed lines in Fig. 8 below).

Ice extent in the free run is significantly different than the (v12/v11) assimilative runs and the OSTIA observations. Ice initially melts slower in the Arctic (March to July), leading to too high of an extent, but then starts to melt excessively from mid-July/August, leading to an exaggerated sea ice minimum in September. In the Antarctic meanwhile the free run consistently underestimates the ice extent, save for a small period during the melt season. It seems also that there is a phase lag between the free run and the assimilative runs and OSTIA analysis, with the free run growing (and melting) ice slightly behind the analyses.

As in situ observations of sea ice thickness are very sparse and satellite observations are not available during the melt season, direct model–observation comparisons of sea ice thickness have not been performed. In order to assess the quality of the FOAM ice thickness distributions, sea ice volumes are instead compared with the reanalysis volume estimates of the Pan-Arctic Ice–Ocean Modelling and Assimilation System (PIOMAS) of Schweiger et al. (2011). These PIOMAS data are considered to be the best available year-round estimates of Arctic ice volume and compare well against the available ice thickness observations (Laxon et al., 2013; Schweiger et al., 2011). Comparisons with PIOMAS data show that Arctic sea ice volume in the v12 system is much better than in the v11 system, which has a significant bias most pronounced in the boreal winter. Given that the ice extent and concentration are very similar in the v11 and v12 systems (Fig. 4), this excessive volume can be interpreted as a too-thick bias in the LIM2 model, which is consistent with the findings of Massonnet et al. (2011). Although much better than in the v11 system, ice volume in v12 is consistently lower than the PIOMAS data, suggesting that the v12 CICE ice fields are a little too thin.

Curiously the Arctic ice volume in the free model run is comparable to that in the assimilative v12 run even during September, when the ice extent is very different. Given that the summer ice extent is much lower, it follows that the ice is thicker in the free-running model than in the v12 assimilative run. This suggests that the assimilation of ice concentration data is thinning the ice within the Arctic ice pack, where there is a larger proportion of older, and hence thicker, multi-year ice. Ice thickness comparisons (not shown) support this hypothesis and reveal that the ice is on average 5 % thicker over the central Arctic in the free run than the v12 assimilative run (this figure rises to between 10 and 20 % thicker during June and July). This is in keeping with the findings of Lindsay and Zhang (2006), who show that assimilating concentration observations within the ice pack with as much weight as at the ice edge can have detrimental effects on the ice thickness distribution.

In the Antarctic however the free run has lower ice volume than the v12 run, which is presumably caused by the considerable reduction in ice extent and the low proportion of multi-year ice in the region. Again the free-run Antarctic ice fields show evidence of a phase lag relative to the assimilative model, with ice volume minima and maxima occurring approximately 1 month after the v12 run.

4.1.6 Near-surface velocities

As well as analysing the FGAT model–observation match-ups output from the NEMO observation operator step, the positions of drifting buoys are also used to give an independent assessment of the quality of the FOAM near-surface velocity fields. Using the methods of Blockley et al. (2012) daily-mean velocities are derived from the daily displacement of Global Drifter Program (GDP) buoys obtained via the GTS. These drifters have a drogue centred at 15 m depth to ensure that the drifter follows the 15 m currents with a wind slip of less than 0.1 % of the wind speed. All drifters known to have lost their drogues are blacklisted, and velocities derived from the remaining buoys are compared with FOAM 15 m modelled velocities for the entire 2-year assessment period – providing an average of approximately 725 model–observation match-ups per day. It should be emphasised here that this verification is based on independent data as velocities are not assimilated by the FOAM system.

Results show that globally the v12 system is better than the old v11 system, with zonal correlation increasing from 0.57 to 0.59 and the corresponding rms error reducing by 2 % to under 21 cm s^{-1} . The most notable improvements are in the Southern Ocean and extratropical regions such as the North Atlantic. Although it is better in the Indian Ocean, the v12 system is worse elsewhere in the tropics – in particular in the tropical Pacific. Further comparisons with currents measured by the TAO/TRITON (Tropical Atmosphere Ocean project/Triangle Trans-Ocean Buoy Network; McPhaden et al., 1998) and PIRATA (Prediction and

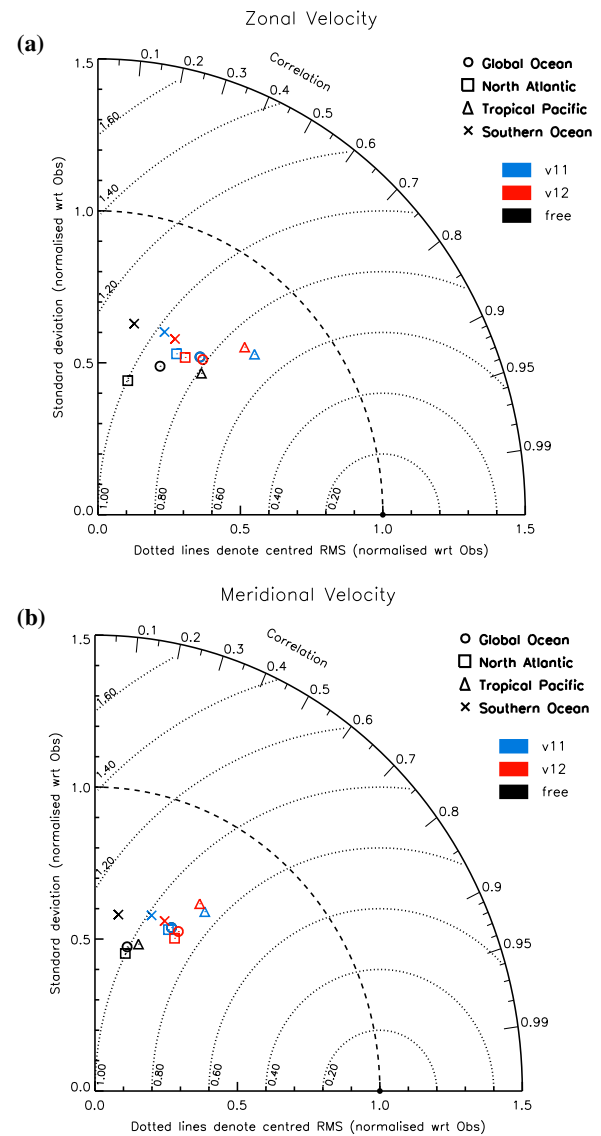


Figure 5. Taylor plots showing comparisons between model near-surface currents and velocities derived from drifter locations for the v12 (red), v11 (blue) and free (black) trials. Results are shown for the global ocean (circles), North Atlantic (squares), tropical Pacific (triangles) and Southern Ocean (crosses) regions for (a) zonal velocity and (b) meridional velocity.

Research Moored Array in the Atlantic; Servain et al., 1998) tropical moorings (not shown) confirm the findings of the drifter regional results that the skill of current predictions is reduced in the tropical Pacific and tropical Atlantic. Taylor plots (Taylor, 2001) of these results for the v12, v11 and free trials can be found in Fig. 5 for the global ocean, North Atlantic, tropical Pacific and Southern Ocean regions. Results in Table 3 (see Sect. 4.1.4 above) show a twofold reduction in rms error for near-surface velocities in areas of high variability compared to low-variability regions, which suggests

that the v12 system is providing a better representation of mesoscale eddies.

Comparison of drifter-velocity statistics for the v12 and free trials shows that, in keeping with the findings of Blockley et al. (2012), the data assimilation is generally having a positive impact on the near-surface currents even though velocity data are not assimilated. Interestingly however, the situation is not so clear-cut in the tropics, where data assimilation only has a notable improvement on the meridional velocity, with much less impact on zonal velocity. Figure 5a shows that the free run has actually a very good representation of zonal velocity in the tropical Pacific region, with a correlation of 0.62. Data assimilation results in an increase in correlation of 11 % to 0.68, which, although a considerable increase, is significantly smaller than the corresponding 70 % increase in meridional correlation in this region, or the 120 % increase in zonal correlation seen in the North Atlantic. The main effect however seems to be to increase the variability of the near-surface currents in the region, which, although not shown in Fig. 5, is also true for the tropical Atlantic. This result may be indicative of the data assimilation artificially increasing the variability in the tropics, which could be caused by the tracer increments initialising waves that travel zonally along the equatorial waveguide (similar to the findings of Moore, 1989). This theory would also be supported by the degradation to the SSH and sub-surface tracer fields in the tropical Pacific.

4.2 Forecast validation

To analyse the performance of the 5-day forecasts for the two assimilative FOAM trials, comparisons are made between model daily-mean fields and a common observation set. The observations used are in situ SST drifters courtesy of US-GODAE and sub-surface profiles of temperature and salinity from the EN3 data set of Ingleby and Huddleston (2007).

The analysis is performed using an off-line version of the NEMO observation operator (as described in Sect. 2) which has been modified to read in forecast (and analysis) fields and create model counterparts mapped to observation space for each data set. The reason for performing the analysis in this way is to mimic the FOAM operational verification systems which use this method to produce model–observation differences for the GODAE intercomparison project and the MyOcean verification systems.

In addition to calculating model counterparts for the forecast and analysis fields at the correct time, match-ups are also produced using temporally interpolated monthly climatologies (using linear interpolation) and analyses persisted from previous days. It should be noted here that, unlike for NWP systems, skill versus persistence is not a user-driven metric for ocean forecasting as users do not generally know the ocean state on a given day to make their own persistence forecasts. Persistence however is useful from a scientific perspective and is used here to highlight the impact of the NEMO

model and to identify any potential problems. The equivalent “naive” forecast for the average ocean user would be climatology rather than persistence. Climatological comparisons are made here using the modified EN3 climatology detailed in Sect. 2.

4.2.1 Sea surface temperature (SST)

Results for the SST comparisons can be found in Fig. 6, which shows rms and mean errors against forecast lead time averaged globally as well as separately for the tropical Pacific, North Pacific and Southern Ocean regions. The rms errors show that the v12 forecasts are better than the v11 forecasts throughout the 5-day forecast. In particular the $T + 60$ h (day 3) forecast error for the v12 system is comparable to the v11 $T + 12$ h (day 1) forecast error (see Fig. 6a). Forecasts are also much better than climatology for both the v12 and v11 systems. This is most pronounced in the tropics, where rms errors are less than 0.4°C for the v12 system throughout the entirety of the forecast (Fig. 6b).

However the dotted rms lines in Fig. 6 show that globally v12 SST forecasts are not better than persistence, albeit only marginally, which is not the case for the v11 system. This problem appears to be much worse in the Southern Ocean, where persistence is considerably better over the latter parts of the forecast (see Fig. 6d). This situation is believed to be caused by a mixing bias in the ORCA025 model, which has been highlighted by the change in SBCs from direct forcing to CORE bulk formulae. The SBC upgrade inadvertently removed an error in the NEMO code that was preventing wind-induced mixing from being included in the TKE vertical mixing scheme – an error that seems to have been compensating for a general over-specification of vertical mixing in the system. Furthermore an additional error has been found in the TKE scheme at NEMO vn3.2, caused by the enhanced vertical diffusion used to parametrise convection being fed back into the TKE equations. This error has been shown to increase mixing in the system particularly in the winter and can lead to a threefold increase in winter mixed layer depths at mid–high latitudes (D. Calvert, personal communication, 2013). Forecast versus analysis comparisons (not shown) indicate a cold bias in the system during summer months (July for Northern Hemisphere and January for Southern Hemisphere), which, along with the cold bias visible in the North Pacific in Fig. 6b, strengthens this over-mixing argument. The fact that the v12 analysis surface temperature fields are better than v11 suggests that the NEMOVAR assimilation scheme is doing a better job of correcting this mixing bias in the surface layers.

It should be stressed here however that although the v12 forecasts are worse than persistence, they are still much better than the v11 forecasts even in the Southern Ocean. In particular the rms error of the $T + 84$ h (day 4) Southern Ocean forecasts for the v12 system are comparable to the rms error of the v11 $T + 12$ h (day 1) forecasts.

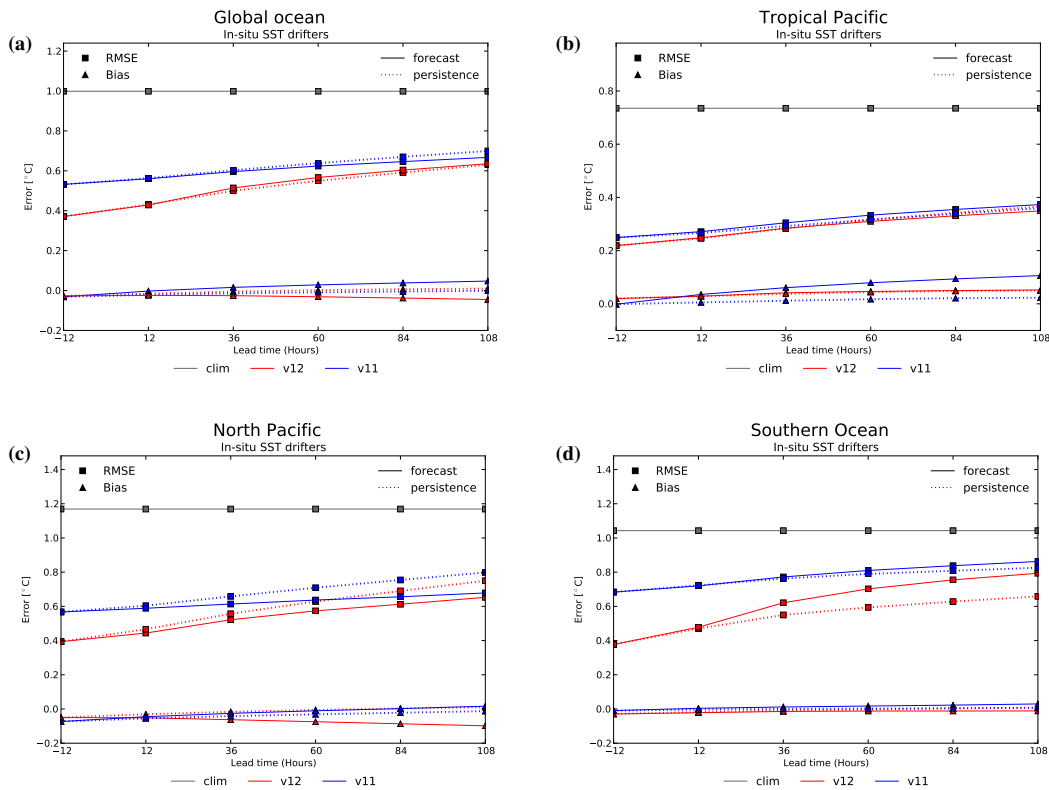


Figure 6. Forecast lead-time plots showing rms errors (squares) and mean errors (triangles) against surface temperature measurements ($^{\circ}\text{C}$) taken by in situ drifting buoys for the (a) global, (b) tropical Pacific, (c) North Pacific and (d) Southern Ocean regions. Statistics are shown for model forecasts (solid lines) and persistence (dotted lines) averaged over all forecasts performed during the trials for the v12 (red) and v11 (blue) systems and the EN3 climatology (grey). The x axis represents forecast lead time (in hours) ranging from the analysis fields at $T - 12$ h up to the 5-day forecasts at $T + 108$ h.

4.2.2 Temperature profiles

Results for the comparisons with sub-surface temperature profiles can be found in Fig. 7, which shows rms errors and mean errors averaged globally against (a) forecast lead time and (b) depth. The plots show that, in keeping with the analysis results in Section 4.1 above, the v12 forecasts are initially better than v11 globally. However at forecast day 2 ($T + 48$ h) the two converge and rms errors are higher for v12 by the end of the 5-day forecast (Fig. 7a). A regional breakdown of the results shows that v12 sub-surface temperature forecasts are generally better in the extratropics, and the Southern Ocean in particular, but worse in the tropics. Additionally the v12 system shows a marked improvement against temperature profiles in waters less than 200 m deep (not shown). This is most likely caused by the fact that the NEMOVAR scheme is better at resolving smaller-scale features and, in particular, SST, which will have a strong impact in well-mixed shelf regions.

Once again the v12 forecasts do not beat persistence globally throughout the whole forecast, which, as was the case for SST, is worse in the Southern Ocean. This issue is also thought to be caused by the over-specification of vertical

mixing in the system in exactly the same way as described for SST above. Error profiles in Fig. 7b show that forecasts are slightly cold-biased over the top 50 m and warm-biased below this (as far down as 500 m in the Southern Ocean), which further supports this over-mixing hypothesis. The tropical Pacific forecasts are more skilful than persistence (not shown), which was also the case for SST.

Perhaps the most noticeable feature in the sub-surface lead-time plots (Fig. 7a and b) is the increase in error between the $T - 12$ h analysis and the $T + 12$ h forecast for both the v12 and v11 systems – behaviour not seen in the SST forecast results in Fig. 6. This feature may be caused by the data assimilation over-fitting the sub-surface profile data but may also be caused by differences in the abundance and independence of the sub-surface profile and SST data sets.

The sub-surface profile observations are rather sparsely distributed in both space and time with Argo profiles, which make up the majority of these observations, reporting only every 10 days. This means that observations received from any particular float will most likely be compared with ocean forecasts in areas outside the radius of influence of prior observations from the same instrument. Therefore the sub-surface profiles can be considered nearly independent,

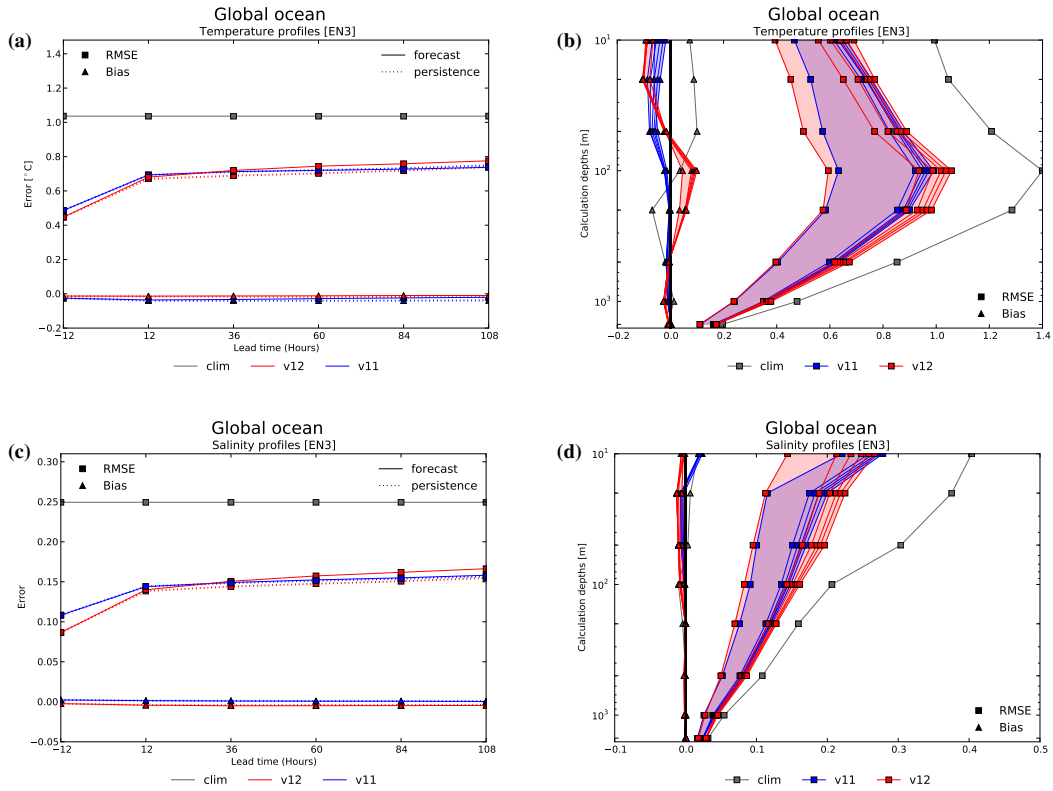


Figure 7. Global rms errors (squares) and mean errors (triangles) against sub-surface profiles of temperature (upper; °C) and salinity (lower; measured on the practical salinity scale) from the EN3 data set averaged over all forecasts performed during the trials in waters deeper than 200 m. Plots show results for the v12 (red) and v11 (blue) trials as well as the EN3 climatology (grey). The left-hand plots, (a) and (c), show model forecast errors (solid lines) and persistence errors (dotted lines) against forecast lead time (in hours) ranging from the analysis fields at $T - 12$ h up to the 5-day forecasts at $T + 108$ h. The right-hand plots, (b) and (d), show forecast profile errors against model depth (m) on a log scale for the analysis and each of the 5 forecast days ($T + 12$ h– $T + 108$ h) to show the evolution of error profiles with forecast lead time. The area between the analysis ($T - 12$ h) and forecast day 5 ($T + 108$ h) is shaded red for v12 or blue for v11.

particularly in the upper ocean. This means that the apparent deterioration in profile error suggested by Fig. 7 will most likely be exaggerated by the lack of independence of the observations at $T - 12$, where comparisons are made using daily-mean analysis fields into which the observations have already been partially assimilated.

In contrast, the SST observations are considerably more abundant in both space and time, with the majority of drifters reporting SST hourly. The abundance of these SST data, in conjunction with the large number of satellite observations available, allows the data assimilation to provide a better initialisation for the forecasts each day, and so a smaller jump in error between the analysis and forecast is to be expected. Additionally the assimilation spreads the information from each observation into surrounding areas of ocean, and so drifter observations in the early part of the forecast may still be within the radius of influence of observations from the same instrument that were assimilated during the analysis. Furthermore, the drifters have a drogue centred at 15 m depth and so will tend to propagate with the same ocean water masses into which they have previously been assimilated, meaning that

the errors will be correlated. Therefore the drifter observations should be considered less independent than the subsurface profiles.

4.2.3 Salinity profiles

Results for the comparisons with sub-surface salinity profiles can be found in Fig. 7c and d, which show rms errors and mean errors averaged globally against forecast lead time and depth respectively. As with temperature, the global v12 forecasts are initially better than v11, but the errors grow at a greater rate through the forecast so that errors are higher in the v12 system after forecast day 2. This improvement in the analysis and subsequent degradation at longer lead times appears to be driven by a freshening of the upper ocean fields (roughly above 110 m depth) which is most pronounced at around 20 m (Fig. 7d). This is in keeping with the precipitation bias discussed in Sect. 4.1 above in relation to salinity and SSH drifts in the free-running system.

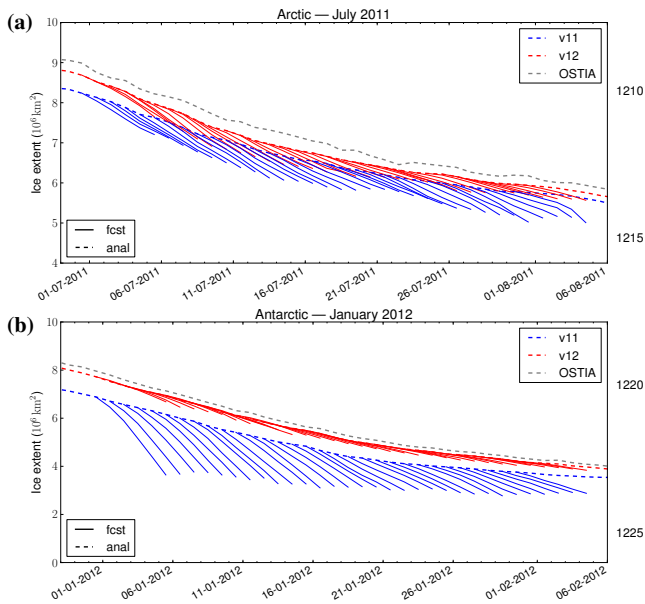


Figure 8. Time series of (a) Arctic sea ice extent (10^6 km^2) for the forecasts performed in July 2011 and (b) Antarctic sea ice extent (10^6 km^2) for the forecasts performed in January 2012 from the v12 (red), v11 (blue) and OSTIA (grey) systems. Dashed lines show extents calculated from analysis ice concentration fields, redrawn from Fig. 4, whilst solid lines show the evolution of the ice extent over each of the 5-day hindcasts performed during the 31-day periods.

In contrast to the FGAT results in the previous section the v11 error profiles show a considerable salty bias in the near-surface 10 m salinity fields. This error appears to be caused by comparisons with a few isolated moorings in the tropics, mostly located in the Caribbean Sea, that are not in the filtered FGAT analysis and where the v11 system is not so good.

As with the sub-surface temperature forecasts, there is a marked increase in error between the analysis and day 1 forecast in both the v12 and v11 systems. This is not unexpected, in light of the discussions in Sect. 4.2.2 above, given that sub-surface salinity observations are generally even more sparse than sub-surface temperature observations. The v12 forecasts do not beat persistence throughout the whole forecast, which again is most pronounced in the Southern Ocean. Although the global salinity profiles in Fig. 7d do not show evidence of excessive mixing, the mixing bias is apparent in mid-latitude regions such as the North Atlantic and North Pacific (not shown).

4.2.4 Sea ice concentration

For reasons discussed in Sect. 4.1 above, the quality of the ice forecasts is assessed by considering sea ice extent (i.e. the total area of all ocean grid points with ice concentration of at least 15%). Results show that the evolution of forecast

ice extent is generally in keeping with the behaviour of the free run shown in Fig. 4. The model tends to somewhat exaggerate Arctic (Antarctic) ice melt for the forecasts performed during the July (January) melting periods and over-predict the growth of Arctic ice during the January forecasts – albeit only slightly – consistent with the ice being a little too thin in the marginal ice zones. Forecasts performed during the April and October months however show good agreement with the analyses. Some examples of this over-melting can be seen in Fig. 8, which shows the model forecasts and analyses for the July 2011 Arctic melt period and the January 2012 Antarctic melt period. The v12 forecast ice extents are much closer to the OSTIA analysis values than the v11 ones, and this is particularly true in the Antarctic (Fig. 8b). As an example the sea ice extent predicted by the v11 5-day forecast for 5 January 2012 ($3.64 \times 10^6 \text{ km}^2$) is 41% below the corresponding analysis for that day – which in turn is 14% lower than the ($7.14 \times 10^6 \text{ km}^2$) extent derived from the OSTIA analysis for this day. The v12 5-day forecast meanwhile predicts an ice extent of ($6.93 \times 10^6 \text{ km}^2$) for 5 January 2012, which is much closer to the OSTIA observational product as well as the corresponding v12 analysis.

As well as diagnosing forecast errors the dashed lines in Fig. 8 can be used, as a zoom of Fig. 4, to see the finer detail of the analysis ice extents. These dashed lines show how much closer the v12 analysis ice extents compare to the OSTIA extents, particularly in the Antarctic. Differences between the ice extents in the Arctic could arise from the way coastal filling is used to augment the OSI-SAF observations as part of the OSTIA interpolations. So it is therefore not realistic to expect the FOAM analyses, which only assimilate the raw OSI-SAF observations, to match OSTIA exactly – particularly in the Arctic where the land–sea mask is considerably more complicated.

4.2.5 Near-surface velocities

The drifter-current analysis performed as part of Sect. 4.1 is extended here to assess the daily-mean forecast fields generated during January, April, July and October each year. Drifter-derived velocity observations are compared to model analysis and forecast fields, persisted analyses and climatology, as was done for SST and sub-surface profiles above.

Results from this analysis can be found in Fig. 9, which shows rms errors and correlations against forecast lead time separately for zonal and meridional velocity forecasts. These results show that globally the v12 velocities are better than the v11 velocities throughout the 5-day forecast. This is particularly true for meridional velocity and is consistent with the reanalysis results in Sect. 4.1. Forecasts beat persistence and climatology across the board, with only a marginal decrease in correlation with forecast lead time. The climatology used here is derived from drifter locations (Lumpkin and Garraffo, 2005), and so beating it shows a good level of skill. Both models show a considerable benefit to using the forecast

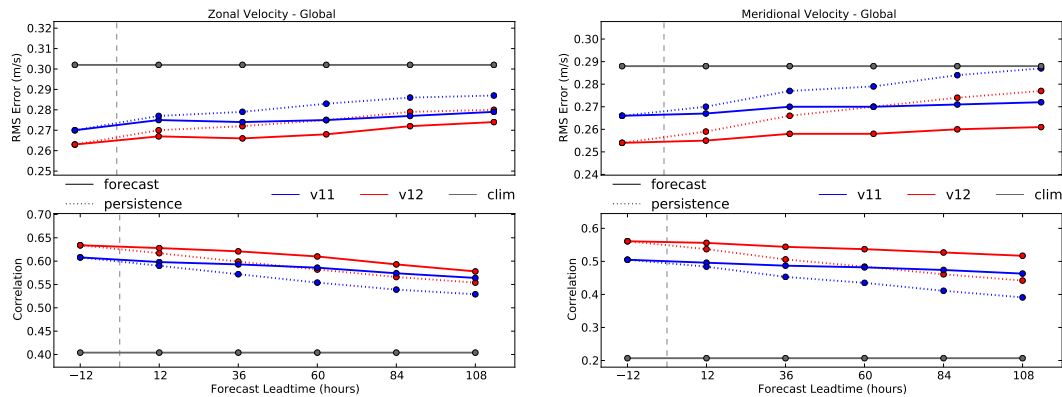


Figure 9. Forecast lead-time plots showing rms errors (upper) and correlation coefficients (lower) against zonal (left) and meridional (right) velocity observations (m s^{-1}) derived from drifter locations. Lines plotted are forecasts (solid lines) and persistence (dotted lines) from the v12 (red) and v11 (blue) trials. Also shown are the corresponding results for climatological velocities (grey solid lines) from the GDP drifter climatology of Lumpkin and Garraffo (2005). The x axis represents forecast lead-time (in hours) ranging from the (daily-mean) analysis fields valid at $T - 12$ h up to the 5-day forecasts at $T + 108$ h. The grey dashed line indicates the location of $T + 00$ h.

rather than persistence for meridional velocity, particularly in the tropics.

Global correlation coefficients ranging from almost 0.65 down to 0.6 for zonal velocity and over 0.55 down to 0.52 for meridional velocity show a good level of skill in agreement with the reanalysis assessments in Sect. 4.1 and Fig. 5. Regional statistics and comparisons (not shown) show that velocities are better for the v12 forecasts everywhere apart from the tropical Pacific (both zonal and meridional) and the tropical Atlantic (zonal only). This is consistent with the drifter results for the full reanalysis period as shown in Fig. 5. Although the v12 system has lower correlations than v11 in the tropics, the zonal correlations are still well above 0.6 (and over 0.75 in the tropical Pacific). Meridional correlations are also good for v12, being above 0.5 for the duration of the 5-day forecast in the Indian Ocean for v12 (up to 0.7 against tropical moorings).

4.3 Comparisons with gridded observations

To augment the quantitative, statistical assessments detailed above, a qualitative assessment of the FOAM analyses has also been performed by comparing 2-D spatial maps of modelled SSH, SST and surface velocity against gridded observational products. Modelled SSH fields were compared with $1/4^\circ$ AVISO gridded absolute dynamic height altimeter products; modelled SST fields were compared with $1/20^\circ$ OSTIA SST analyses; and model surface velocities, integrated over the top 15 m, were compared with $1/3^\circ$ OSCAR (Ocean Surface Current Analyses – Real time: Bonjean and Lagerloef, 2002) ocean surface currents derived from satellite altimeter and scatterometer winds. In addition to performing a visual comparison of these fields, anomaly correlations were calculated between model and observational

fields to provide further insight into the quality of the FOAM data.

Comparisons using monthly-mean analysis fields show good agreement between the v12 and v11 assimilative FOAM systems and the observational products. In general the v12 fields agree better with the observations than do the v11 fields as they seem to be better at resolving the smaller-scale features and mesoscale eddies – reinforcing the results of Table 3 (see Sects. 4.1.4 and 4.1.6). This is also consistent with the findings of Waters et al. (2014), who show that NEMOVAR produces better SST and SSH fields in frontal regions.

Details of one particular case study can be found in Fig. 10 which shows an example of such comparisons over the Agulhas retroflection region using September 2012 monthly-mean fields. This period and location were chosen for illustration because a pair of rather interesting cyclonic, cold-core eddies had traversed the frontal zone of the Agulhas retroflection and made their way northwards into the warmer waters that flow southwards from the Mozambique Channel. The eddies persisted for a considerable period, moving relatively slowly, which made them easily detectable in the September 2012 monthly-mean AVISO SSH (Fig. 10: left, row 2) and OSCAR velocity (Fig. 10: right, row 2) fields. The larger of these eddies can be seen located at approximately (26° E, 37.5° S) with a smaller eddy at (30° E, 36.5° S). These eddies are also visible in the OSTIA SST fields (Fig. 10: centre, row 2) albeit not so pronounced.

The v12 system does a very good job at reproducing these eddies, which can be seen in the SSH, SST and velocity plots (Fig. 10: row 1). However the v11 system does not capture these so well (see Fig. 10: row 3). Although there is a suggestion of lower SSH in the correct locations, the surface circulation is somewhat different in the v11 model and the eddies do not feature in the current fields (Fig. 10: right, row 1).

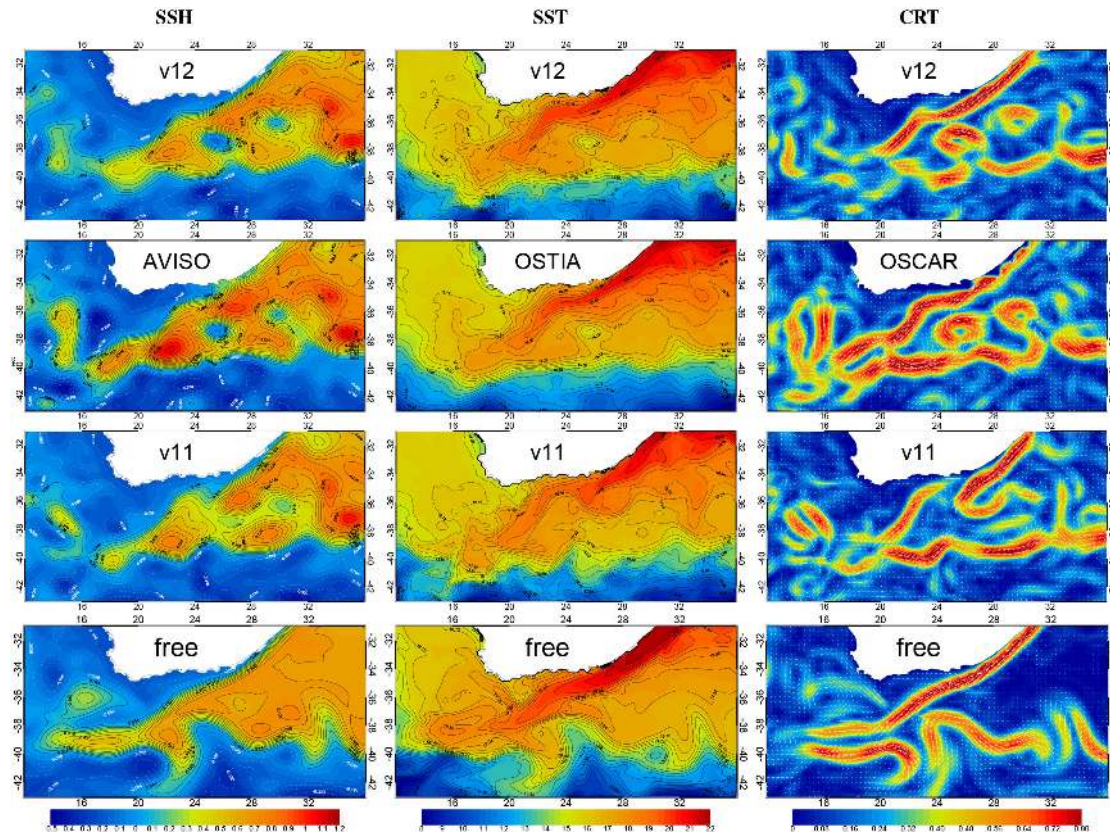


Figure 10. An array of monthly-mean gridded contour plots over the Agulhas retroflexion region (longitude: 12–36° E; latitude: 31–43° S) for September 2012. Sea surface height (left column; m) and temperature (centre column; °C) are plotted as coloured contours and overlaid with black contour lines. For the SSH plots solid black lines denote positive contour values and broken white lines are used for negative values. Surface currents (right column; m s^{-1}) are displayed as coloured contours of current intensity (speed), with white arrows overlaid to show direction. Output from the v12, v11 and free trials are plotted in the first, third and fourth rows respectively, whilst the second row plots show the gridded observational products: AVISO SSH, OSTIA SST and OSCAR near-surface currents. Model currents shown (i.e. v12, v11 and free) are total integrated velocity over the top 15 m. Anomaly correlations for each of the modelled fields against the corresponding gridded observations can be found in Table 4.

Aside from the position of the two cyclonic eddies, the v12 fields look more like the observational products throughout the majority of the rest of the domain. This is particularly true for the SST, which agrees very well with the OSTIA SST analysis throughout the whole of the domain plotted in Fig. 10. There is a suggestion that the model is resolving smaller-scale features than the OSTIA product, which is in keeping with the fact that, by design, OSTIA produces an analysis that is smoother than the true surface temperature, particularly in areas of sharp fronts (Donlon et al., 2012). The v12 SLA also compares well with the AVISO product but does not quite capture the high intensity of the anticyclonic features at (22° E, 39.5° S) and (27.5° E, 36° S). Additionally the cyclonic structure at (17° E, 36–38° S) is underestimated in both the v12 and v11 systems, as is the northwards projection to the west of the retroflexion at 15° E.

The free model does not do a bad job here and, to a certain extent, does represent the large-scale flow quite well. It does not manage to capture the finer-scale features seen in the observations and assimilative runs though, which is not surprising given that ORCA025 is only an eddy-permitting, rather than a fully eddy-resolving, model.

Anomaly correlations against the relevant observational data for each of the model fields in Fig. 10 can be found in Table 4. These reinforce the outcomes of the qualitative assessment showing that there is a better agreement between the FOAM v12 surface fields and the gridded observational products, which is particularly true for SST. Although velocities are not assimilated in any of the systems, the near-surface velocity fields in the (v12/v11) assimilative runs are considerably closer to the OSCAR product than are those of the free run. This will have been caused by the SLA assimilation successfully constraining the circulation.

Table 4. Anomaly correlations for modelled SSH, SST and the magnitude of near-surface velocity fields (speed) against the corresponding gridded observational products (AVISO, OSTIA and OSCAR) for all the 2-D spatial maps shown in Fig. 10. Correlations are calculated over all ocean points, and anomalies are calculated relative to the WOA2001 1/4° analysis (Boyer et al., 2005) for SST, the CNES09 MDT (Rio et al., 2011) for SSH and the GDP drifter climatology (Lumpkin and Garraffo, 2005) for the speed.

	Sea surface height (SSH)	Sea surface temperature (SST)	Near-surface speed
v12	0.68	0.89	0.63
v11	0.66	0.73	0.58
Free	0.22	0.18	0.35

5 Summary and future plans

In this paper recent developments to the Met Office FOAM system have been introduced, the new FOAM v12 system has been described and changes relative to the previous v11 FOAM system have been highlighted. Results have been presented from three 2-year FOAM experiments, and the performance of the new v12 system has been compared to the old v11 system and a free-running, non-assimilative v12 system to investigate the respective impacts of the v12 upgrade and the data assimilation. Assessments have focused on the analysis of FGAT innovations throughout the reanalysis period as well as daily-mean model–observation match-ups derived from a series of 5-day forecasts spun off the assimilative trials for 8 months during the assessment period (Jan, Apr, Jul and Oct each year). An additional qualitative assessment of the reanalysis surface fields has been performed by comparing 2-D spatial maps of SSH, SST and surface currents from all three FOAM trials against AVISO, OSTIA, and OSCAR gridded observational products.

Results show that improvements are mixed with some considerable advantages where the observation density is high but with some deterioration where observations are sparse.

Surface fields, and in particular surface temperature, are generally improved in the new v12 system, with global SST and SSH rms errors of 0.45 °C and 7.4 cm respectively. Comparisons with gridded observational products suggest that the v12 system provides a better representation of mesoscale features in the extratropics – an improvement that will have been caused primarily by the shorter horizontal correlation length scales used within NEMOVAR (Waters et al., 2014). Data assimilation is shown to have a positive effect on the surface fields, with a reduction in surface temperature biases and correction of a long-term drift in surface height. Comparisons with gridded data sets show a considerable improvement for the assimilative runs and an increased spatial structure to the surface fields.

The quality of near-surface (< 80 m) temperature and salinity fields is also improved in the new v12 system. The increased accuracy of near-surface temperatures is caused by

the move to NEMOVAR and the associated improvements to SST. However the salinity improvement is in contrast to the results of Waters et al. (2014) and is driven by the surface boundary condition upgrade to use CORE bulk formulae.

Temperature at 100 m is slightly degraded in the v12 system, and this seems to be a result of the present version of the NEMOVAR assimilation scheme not being able to constrain a persistent model bias quite as well as the old AC scheme did. Although sub-surface salinity is better globally and in the North Atlantic, there is a slight degradation in most other regions. In particular, salinity is worse in the Southern Ocean throughout most of the water column.

Although the shorter horizontal correlation length scales employed by NEMOVAR allow for tighter matching of mesoscale features (Table 3), they also make it harder for the assimilation to constrain the tracer fields at depth owing to the sparsity of sub-surface observations (Waters et al., 2014). This is thought to be responsible for the degradation of temperature and salinity at depths below 80 m. Further research is required here, but it is hoped that the extension of NEMOVAR to include multiple horizontal length scales (as used in AC) will better constrain the tracer fields at depth.

Assessment of the forecast fields shows that the v12 SST fields remain better than the v11 system and considerably better than climatology throughout the 5-day forecasts. However the v12 forecasts do not beat analysis persistence for SST or near-surface temperature and salinity profiles, which is particularly true in the Southern Ocean. It is believed that this result is caused by excessive mixing in the NEMO model, which seems to have been made worse at v12 by reinstating wind-induced mixing that was erroneously being ignored at v11 – an error that was seemingly compensating for the excessive mixing. The NEMOVAR assimilation scheme is doing a good job correcting for these mixing biases, and the v12 analyses are considerably improved compared to the v11 analyses and, in particular, the free-running model forecasts. However this relative improvement in analysis quality, coupled with the mixing bias, causes the propagation of errors through the forecasts to be higher in the new system for SST and near-surface temperature and salinity fields. There has been a lot of work carried out in the UK, under the framework of the NERC–Met Office Joint Ocean Modelling Programme, to better understand the cause of these vertical mixing errors within the Global NEMO model configurations (Calvert and Siddorn, 2013), and an improved set of NEMO TKE scheme parameter values has been developed for the latest release of the JOMP Global Ocean configuration (GO5.0: Megann et al., 2013). The FOAM system will be upgraded to use GO5.0 in 2014, and it is hoped that this will considerably reduce these forecast errors in the future. This change will also include the NEMO vn3.4 TKE convective bug fix, which should help reduce the evolution of erroneously deep winter mixed layers.

For both the v12 and v11 systems there is a substantial jump in errors between the analyses and the start of the fore-

cast when comparing against sub-surface temperature and salinity observations. This apparent jump could be the result of the data assimilation schemes over-fitting the relatively sparse sub-surface profiles but is most likely caused by the lack of independence of the observations when comparing with the analysis fields. This jump is not seen in the SST results owing to differences in the level of abundance and independence of the respective data sets as discussed in Sect. 4.2.2 above. It is hoped that recalculating error variances as part of the implementation of dual horizontal correlation length scales will reduce this problem in future versions of FOAM.

Sea ice fields are considerably improved in the v12 system, with a significant reduction in concentration errors revealed by the innovation statistics. Comparisons of ice extent against gridded OSTIA observations confirm this ice concentration improvement, showing that the v12 fields are closer to the SSMIS observations. The smaller horizontal correlation length scales used within the NEMOVAR assimilation scheme account for a significant portion of this improvement (Waters et al., 2014), with the bulk formulae surface boundary condition and CICE multi-category sea ice model upgrades accounting for the rest. The impact of the SBC and CICE changes can be seen by the improvement in sea ice extent evolution during the model forecasts (Fig. 8). Ice volume is also improved for v12 and compares much better with the Arctic PIOMAS volumes of Schweiger et al. (2011) than does the v11 system, which overestimates the volume of Arctic winter sea ice considerably. However there seems to be an underestimation of ice volume in the v12 CICE system albeit considerably less extreme than the overestimation in the v11 LIM2 system. Assimilation of sea ice concentration data has a significant impact on the ice edge, particularly during the summer months, where the free-running model tends to melt the ice too aggressively, leading to an underestimation of the ice extent minima. However the Arctic ice is thinner in the v12 system compared to the free run. This is thought to be caused by the assimilation of ice concentration in regions of thick multi-year ice (Lindsay and Zhang, 2006), and work is currently underway to investigate whether changing the way ice concentration is assimilated will reduce these detrimental effects. Additionally there are plans to investigate the ocean–ice–atmosphere interactions within CICE with the aim of improving sea ice fields in the free-running model.

Near-surface velocity statistics are generally better in the new v12 system, with lower rms errors and higher correlations, save for in the tropical Atlantic and tropical Pacific. The same is true for the forecast experiments, with v12 velocities outperforming v11 velocities throughout the forecast as well as beating both persistence and climatology. Comparisons with independent velocities derived from drifter positions suggest a good level of skill in the zonal velocity fields with a correlation of 0.59 globally and correlations above 0.6 in the tropical Atlantic, tropical Pacific and Indian Ocean regions. Data assimilation has a positive effect on the near-

surface velocity fields, particularly for meridional velocities, even though the velocities themselves are not assimilated.

In general there is a degradation of model skill in the tropics at v12, which is particularly pronounced in the tropical Pacific. One hypothesis is that assimilation of data in the tropics causes spurious variability in the system, which in turn is responsible for degrading the quality of model fields here. Mean and standard deviations of assimilation tracer increments (not shown) reveal that, in general, NEMOVAR is doing a lot more work than AC and at smaller length scales. This is particularly true in the tropics, which would exaggerate this issue and could be responsible for the degradation seen in the v12 assessments. This hypothesis is partially supported by the drifter-velocity results that show that the assimilation increases the zonal velocity variability in the tropics with comparatively little increase in model skill. In an attempt to improve the situation in the tropics, a number of modifications to the NEMOVAR scheme are being tested, including the use of a second-order velocity balance in the tropics and adjusting the IAU window to apply increments over both shorter and longer time periods.

As well as the previously mentioned development of dual horizontal correlation length scales, the upgrade to GO5.0 and the proposed modifications to the assimilation of sea ice concentration, there are a number of other changes planned to the FOAM system. As part of a continual upgrade to the FOAM observing system to use new data sources, Jason-1 SLA data will soon be replaced with AltiKA/SARAL data and the satellite SST observations will be extended to include microwave data from the Advanced Microwave Scanning Radiometer 2 (AMSR2) instrument onboard the GCOM-W1 (Global Change Observation Mission – Water) satellite. Since the loss of the AATSR instrument, the reference data set used for the satellite SST bias correction scheme has consisted of only in situ SST observations. There are plans to increase this reference data set by inclusion of an accurate subset of night-time MetOp-AVHRR data, defined based on low satellite zenith angle, as has already been implemented in the OSTIA system. Another planned change is the extension of the FOAM system to produce estimates of diurnal skin temperature using the parametrisations described in Sykes et al. (2014). There are also substantial upgrades planned to the Met Office global NWP model in summer 2014, including a resolution increase from 25 to 17 km, which will hopefully have a positive effect on the precipitation biases described in Sect. 4.

In the medium term, over the next year, FOAM forecasts will start to be produced by a coupled ocean–ice–atmosphere short-range forecasting system initialised from the FOAM and NWP analyses each day. In the longer term, there are also plans to extend the FOAM and NWP assimilation schemes to produce an analysis within the coupled framework. This move to a fully coupled system would mean that the ocean surface fields become more important for effective ocean–ice–atmosphere interactions.

Appendix A: FOAM v11 upgrade

Details of the FOAM v11 upgrade are described here to reference the changes made at v11 relative to the Storkey et al. (2010) FOAM v10 system. It is important to provide details of the v11 system because assessment of the new v12 system in Sect. 4 is made relative to the v11 system, which has not been specifically documented in the literature. For brevity the details and justifications for the v11 changes are not described in depth but are merely highlighted to allow the reader to get a better background picture of the evolution of the FOAM system since the initial FOAM-NEMO implementation described in Storkey et al. (2010). A summary of the differences between the global model configurations for FOAM v10, v11 and v12 can be found in Table 1.

The following changes were made as part of the FOAM v11 system upgrade: the use of CNES09 MDT (Rio et al., 2011) in place of the Rio2007 MDT (Rio et al., 2007b); implementation of newly calculated and seasonally varying error covariances for the data assimilation scheme; changing from free-slip to partial-slip lateral boundary conditions; and the implementation of a mixed Laplacian–bi-Laplacian horizontal momentum diffusion scheme. Although the v10 documentation of Storkey et al. (2010) describes the use of a mixed Laplacian–bi-Laplacian scheme for horizontal momentum diffusion, there was found to be an error in the scheme and only the Laplacian part was being applied. Correct implementation of the bi-Laplacian component reduced grid-scale noise in the velocity fields and improved mesoscale variability in the system. The partial-slip change was made to prevent the generation of spurious currents around islands in regions of steep topography, which were caused by the SLA assimilation within the analysis correction scheme. This issue has been solved by the move to NEMOVAR, and so free-slip lateral boundary conditions are used at v12 once again.

The upgrade to FOAM v11 also saw the extension of the operational FOAM system to include an additional 24 h data assimilation cycle to allow the assimilation of data over a 48 h window (as briefly outlined in Sect. 2.3 above). The addition of this retrospective assimilation cycle allowed the FOAM v11 system to assimilate considerably more sub-surface profiles than was possible with a single 24 h cycle because more late-arriving observations could be included (i.e. observations arriving more than 24 h behind time, but less than 48 h would now be assimilated, which was not the case at v10). This was particularly true for Argo (Roemmich et al., 2009) and marine mammal observations, which saw an average increase of over 50 % from approximately 220 to 340 profiles per day. The effect of assimilating these extra profiles was a major reduction in rms error of between 5 and 6 % globally against sub-surface temperature and salinity observations.

The v11 changes are further described in Blockley et al. (2012) and Storkey (2011), who also provide assessments of the impacts of the v11 upgrade on near-surface currents and temperature and salinity biases respectively. Readers should note that in these publications the FOAM v10 and FOAM v11 systems are referred to as “FOAMV0” and “FOAM V1” respectively – the reason being that, as these particular FOAM configurations were implemented as part of the MyOcean project, MyOcean version numbers were used to reference the configurations.

The Supplement related to this article is available online at doi:10.5194/gmd-7-2613-2014-supplement.

Acknowledgements. The data used for this study were obtained from numerous sources, all of which are gratefully acknowledged by the authors: drifter status and drogoue information were obtained from the Global Drifter Program (GDP, <http://www.aoml.noaa.gov/phod/dac/>); OSCAR surface current data was obtained from the OSCAR Project Office via the OSCAR website (<http://www.oscar.noaa.gov/>); gridded altimeter products were produced by SSALTO/DUACS and distributed by AVISO, with support from CNES (<http://www.aviso.oceanobs.com/duacs/>); and OSTIA gridded SST and sea ice analyses were produced by the Met Office and are distributed under the MyOcean project. Additionally the FOAM v12 model bathymetry and tidal mixing climatologies were obtained from the DRAKKAR project and the ORCA025 grid, and river run-off climatology were provided by Mercator Océan through the MyOcean project.

The authors would like to thank Catherine Guiavarc'h and James While for help and advice regarding the implementation of the CORE bulk formulae and the satellite SST bias correction scheme. Thanks also go to Daley Calvert for some useful discussions on vertical mixing and to Chris Harris for his advice concerning NEMO-CICE coupling.

Funding support from the UK Ministry of Defence and the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 283367 (MyOcean2) is gratefully acknowledged.

The authors would like to thank Gary Brassington and two other anonymous referees for providing useful suggestions for ways in which the original submission of this paper could be improved.

Edited by: O. Marti

References

- Adcroft, A., Hill, C., and Marshall, J.: Representation of topography by shaved cells in a height coordinate ocean model, *Mon. Weathee Rev.*, 125, 2293–2315, 1997.
- Arakawa, A. and Lamb, V. R.: A potential enstrophy and energy conserving scheme for the shallow water equations, *Mon. Weather Rev.*, 109, 18–36, 1980.
- Arribas, A., Glover, M., Maidens, A., Peterson, K. A., Gordon, M., MacLachlan, C., Graham, R., Fereday, D., Camp, J., Scaife, A. A., Xavier, P., McLean, P., Colman, A., and Cusack, S.: The GloSea4 ensemble prediction system for seasonal forecasting, *Mon. Weather Rev.*, 139, 1891–1910, doi:10.1175/2010MWR3615.1, 2011.
- Axell, L. B.: Wind-driven internal waves and Langmuir circulations in a numerical ocean model of the southern Baltic Sea, *J. Geophys. Res.*, 107, 3204, doi:10.1029/2001JC000922, 2002.
- Balmaseda, M. A., Mogensen, K., and Weaver, A. T.: Evaluation of the ECMWF ocean reanalysis system ORAS4, *Q. J. Roy. Meteor. Soc.*, 139, 1132–1161, doi:10.1002/qj.2063, 2012.
- Barnier, B., Madec, G., Penduff, T., Molines, J.-M., Treguier, A.-M., Le Sommer, J., Beckmann, A., Biastoch, A., Böning, C., Dengg, J., Derval, C., Durand, E., Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., McClean, J., De Cuevas, B.: Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution, *Ocean Dynam.*, 56, 543–567, doi:10.1007/s10236-006-0082-1, 2006.
- Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., and DeWitt, D. G.: Skill of Real-Time Seasonal ENSO Model Predictions during 2002–11: Is Our Capability Increasing? *Bull. Amer. Meteor. Soc.*, 93, 631–651, doi:10.1175/BAMS-D-11-00111.1, 2012.
- Beckmann, A. and Döscher, R.: A method for improved representation of dense water spreading over topography in geopotential-coordinate models, *J. Phys. Oceanogr.*, 27, 581–591, 1997.
- Bell, M. J., Forbes, R. M., and Hines, A.: Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting, *J. Marine Sys.*, 25, 1–22, 2000.
- Bell, M. J., Martin, M. J., and Nichols, N. K.: Assimilation of data into an ocean model with systematic errors near the equator, *Q. J. Roy. Meteor. Soc.*, 130, 873–893, 2004.
- Bell, M. J., Le Traon, P.-Y., Smith, N., Lefebvre, M., and Wilmer-Becker, K.: The Global Ocean Data Assimilation Experiment (GODAE), *Oceanography*, 22, 14–2, 2009.
- Bernie, D. J., Woolnough, S. J., Slingo, J. M., Guilyardi, E.: Modeling Diurnal and Intraseasonal Variability of the Ocean Mixed Layer, *J. Climate*, 18, 1190–1202, doi:10.1175/JCLI3319.1, 2005.
- Blanke, B. and Delecluse, P.: Variability of the tropical Atlantic Ocean simulated by a general circulation model with two different mixed-layer physics, *J. Phys. Oceanogr.*, 23, 1363–1388, 1993.
- Blockley, E. W., Martin, M. J., and Hyder, P.: Validation of FOAM near-surface ocean current forecasts using Lagrangian drifting buoys, *Ocean Sci.*, 8, 551–565, doi:10.5194/os-8-551-2012, 2012.
- Bloom, S. C., Takacs, L. L., da Silva, A. M., and Ledvina, D.: Data assimilation using incremental analysis updates, *Mon. Weather Rev.*, 124, 1256–1271, 1996.
- Bonjean, F. and Lagerloef, G. S. E.: Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean, *J. Phys. Oceanogr.*, 32, 2938–2954, 2002.
- Bouillon, S., Maqueda, M. A. M., Legat, V., and Fichet, T.: An elastic-viscous-plastic sea ice model formulated on Arakawa B and C grids, *Ocean Model.*, 27, 174–184, doi:10.1016/j.ocemod.2009.01.004, 2009.
- Bourdalle-Badie, R., and Treguier, A. M.: A climatology of runoff for the global ocean-ice model ORCA025, Report, Mercator-Ocean. Reference: MOO-RP-425-365-MER, 2006.
- Bourles, B., Lumpkin, R., McPhaden, M. J., Hernandez, F., Nobre, P., Campos, E., Yu, L., Planton, S., Busalacchi, A., Moura, A. D., Servain, J., and Trotte, J.: The PIRATA Program: History, Accomplishments, and Future Directions, *Bull. Amer. Meteor. Soc.*, 89, 1111–1125, 2008.
- Boyer, T., Levitus, S., Garcia, H., Locarnini, R. A., Stephens, C., and Antonov, J.: Objective analyses of annual, seasonal, and monthly temperature and salinity for the World Ocean on a 0.25° grid, *Int. J. Climatol.*, 25, 931–945, doi:10.1002/joc.1173, 2005.
- Brown, A., Milton, S., Cullen, M., Golding, B., Mitchell, J., and Shelly, A.: Unified Modeling and Prediction of Weather and Cli-

- mate: A 25-Year Journey. *Bull. Amer. Meteor. Soc.*, 93, 1865–1877, doi:10.1175/BAMS-D-12-00018.1, 2012.
- Brushett, B. A., King, B. A., and Lemckert, C. J.: Evaluation of met-ocean forecast data effectiveness for tracking drifters deployed during operational oil spill response in Australian waters, *J. Coastal Res.*, SI64, 991–994, 2011.
- Burchard, H.: Energy-conserving discretisation of turbulent shear and buoyancy production, *Ocean Model.*, 4, 347–361, 2002.
- Calvert, D. and Siddorn, J. R.: Revised vertical mixing parameters for the UK community standard configuration of the global NEMO ocean model, Met Office Hadley Centre Technical Note, 95, 2013.
- Cummings, J. A.: Operational multivariate ocean data assimilation, *Q. J. Roy. Meteor. Soc.*, 133, 3583–3604, 2005.
- Cunningham, S. A., Roberts, C. D., Frajka-Williams, E., Johns, W. E., Hobbs, W., Palmer, M. D., Rayner, D., Smeed, D. A., and McCarthy, G.: Atlantic Meridional Overturning Circulation slowdown cooled the subtropical ocean, *Geophys. Res. Lett.*, 40, 6202–6207, doi:10.1002/2013GL058464, 2013.
- Dai, A. and Trenberth, K. E.: Estimates of Freshwater Discharge from Continents: Latitudinal and Seasonal Variations, *J. Hydrometeorol.*, 3, 660–687, 2002.
- Davidson, F. J. M., Allen, A., Brassington, G. B., Breivik, Ø., Daniel, P., Kamachi, M., Sato, S., King, B., Lefevre, F., Sutton, M., and Kaneko, H.: Applications of GODAE ocean current forecasts to search and rescue and ship routing, *Oceanography*, 22, 176–181, 2009.
- Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N.: A new dynamical core for the Met Office's global and regional modelling of the atmosphere, *Q. J. Roy. Meteor. Soc.*, 131, 1759–1782, doi:10.1256/qj.04.101, 2005.
- Dombrowsky, E., Bertino, L., Brassington, G. B., Chassignet, E. P., Davidson, F., Hurlburt, H. E., Kamachi, M., Lee, T., Martin, M. J., Mei, S., and Tonani, M.: GODAE systems in operation, *Oceanography*, 22, 80–95, 2009.
- Donlon, C. J., Martin, M., Stark, J. D., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice analysis (OSTIA), *Remote Sens. Environ.*, 116, 140–158, doi:10.1016/j.rse.2010.10.017, 2012.
- Drévilion, M., Bourdallé-Badie, R., Derval, C., Drillet, Y., Lelouche, J. M., Rémy, E., Tranchant, B., Benkiran, M., Greiner, E., Guinehut, S., Verbrugge, N., Garric, G., Testut, C. E., Laborie, M., Nouel, L., Baharel, P., Bricaud, C., Crosnier, L., Dombrowsky, E., Durand, E., Ferry, N., Hernandez, F., Le Galloudec, O., Messal, F., and Parent, L.: The GODAE/Mercator-Océan global ocean forecasting system: results, applications and prospects, *J. Oper. Oceanogr.*, 1, 51–57, 2008.
- ETOPO2v2 2-minute Gridded Global Relief Data, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Geophysical Data Center, available at: <http://www.ngdc.noaa.gov/mgg/fliers/06mgg01.html> (last access: 5 November 2014), 2006.
- Flato, G. M. and Hibler, W. D.: Ridging and strength in modeling the thickness distribution of Arctic sea ice, *J. Geophys. Res.*, 100, 18611–18626, doi:10.1029/95JC02091, 1995.
- Fichefet, T. and Maqueda, M. A. M.: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, *J. Geophys. Res.*, 102, 12609–12646, doi:10.1029/97JC00480, 1997.
- Gaspar, P., Gregoris, Y., and Lefevre, J. M.: A simple eddy kinetic energy model for simulations of the ocean vertical mixing: Tests at station Papa and long-term upper ocean study site, *J. Geophys. Res.*, 95, 16179–16193, doi:10.1029/JC095iC09p16179, 1990.
- Haney, R. L.: Surface thermal boundary condition for ocean circulation models, *J. Phys. Ocean.*, 1, 241–248, 1971.
- Hayes, S. P., Mangum, L. J., PiCaut, J., Sumi, A., and Takeuchi, K.: TOGA-TAO: A moored array for real-time measurements in the Tropical Pacific Ocean, *B. Am. Meteorol. Soc.*, 72, 339–347, 1991.
- Hewitt, H. T., Copsey, D., Culverwell, I. D., Harris, C. M., Hill, R. S. R., Keen, A. B., McLaren, A. J., and Hunke, E. C.: Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system, *Geosci. Model Dev.*, 4, 223–253, doi:10.5194/gmd-4-223-2011, 2011.
- Hibler, W. D.: Modeling a variable thickness sea ice cover, *Mon. Wea. Rev.*, 108, 1943–1973, 1980.
- Hyder, P., While, J., Arnold, A., O'Dea, E., Furner, R., Siddorn, J., Martin, M., and Sykes, P.: Evaluating a new NEMO-based Persian/Arabian Gulf tidal operational model, *J. Oper. Oceanogr.*, 6, 3–16, 2013.
- Hollingsworth, A. and Lonnberg, P.: The statistical structure of short-range forecast errors as determined from radiosonde data. part i: The wind field, *Tellus*, 38A, 111–136, doi:10.1111/j.1600-0870.1986.tb00460.x, 1986.
- Hunke, E. C. and Dukowicz, J. K.: The Elastic-Viscous-Plastic Sea Ice Dynamics Model in General Orthogonal Curvilinear Coordinates on a Sphere – Incorporation of Metric Terms, *Mon. Weather Rev.*, 130, 1848–1865, 2002.
- Hunke, E. C. and Lipscomb, W. H.: CICE: the Los Alamos sea ice model. Documentation and software users manual, Version 4.1 (LA-CC-06-012), T-3 Fluid Dynamics Group, Los Alamos National Laboratory, Los Alamos, 2010.
- Ingleby, B. and Huddleston, M.: Quality control of ocean temperature and salinity profiles – historical and real-time data, *J. Mar. Syst.*, 65, 158–175, doi:10.1016/j.jmarsys.2005.11.019, 2007.
- Ingleby, N. B. and Lorenc, A. C.: Bayesian quality control using multivariate normal distributions, *Q. J. Roy. Meteor. Soc.*, 119, 1195–1225, 1993.
- Jacobs, G. A., Woodham, R., Jourdan, D., and Braithwaite, J.: GODAE applications useful to navies throughout the world, *Oceanography*, 22, 182–189, 2009.
- Johns, T. C., Durman, C. F., Banks, H. T., Roberts, M. J., McLaren, A. J., Ridley, J. K., Senior, C. A., Williams, K. D., Jones, A., Rickard, G. J., Cusack, S., Ingram, W. J., Crucifix, M., Sexton, D. M. H., Joshi, M. M., Dong, B. W., Spencer, H., Hill, R. S. R., Gregory, J. M., Keen, A. B., Pardaens, A. K., Lowe, J. A., Bodas-Salcedo, A., Stark, S., and Searl, Y.: The new Hadley Centre Climate Model (HadGEM1): Evaluation of coupled simulations, *J. Climate*, 19, 1327–1353, 2006.
- Jones, G. S., Christidis, N., and Stott, P. A.: Detecting the influence of fossil fuel and bio-fuel black carbon aerosols on near surface temperature changes, *Atmos. Chem. Phys.*, 11, 799–816, doi:10.5194/acp-11-799-2011, 2011.
- Kara, A. B., Rochford, P. A., and Hurlburt, H. E.: An optimal definition for ocean mixed layer depth, *J. Geophys. Res.*, 105, 16803–16822, 2000.

- Koch-Larrouy, A., Madec, G., Bouruet-Aubertot, P., Gerkema, T., Bessieres, L., and Molcard, R.: On the transformation of Pacific Water into Indonesian Throughflow Water by internal tidal mixing, *Geophys. Res. Lett.*, 34, L04604, doi:10.1029/2006GL028405, 2007.
- Large, W. G. and Yeager, S. G.: Diurnal to decadal global forcing for ocean and sea ice models: The data sets and flux climatologies, Technical Report TN-460+STR, NCAR, 10 5pp., 2004.
- Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume, *Geophys. Res. Lett.*, 40, 732–737, doi:10.1002/grl.50193, 2013.
- Lea, D. J., Drecourt, J. P., Haines, K., and Martin, M. J.: Ocean altimeter assimilation with observational- and model-bias correction, *Q. J. Roy. Meteor. Soc.*, 134, 1761–1774, doi:10.1002/qj.320, 2008.
- Lellouche, J. -M., Le Galloudec, O., Drévilion, M., Régnier, C., Greiner, E., Garric, G., Ferry, N., Desportes, C., Testut, C.-E., Bricaud, C., Bourdallé-Badie, R., Tranchant, B., Benkiran, M., Drillet, Y., Daudin, A., and De Nicola, C.: Evaluation of global monitoring and forecasting systems at Mercator Océan, *Ocean Sci.*, 9, 57–81, doi:10.5194/os-9-57-2013, 2013.
- Lengaigne, M., Menkes, C., Aumont, O., Gorgues, T., Bopp, L., André, J.-M., and Madec, G.: Influence of the oceanic biology on the tropical Pacific climate in a coupled general circulation model, *Clim. Dynam.*, 28, 502–516, doi:10.1007/s00382-006-0200-2, 2007.
- Le Traon, P., Bell, M., Dombrowsky, E., Schiller, A., and Wilmer-Becker, K.: GODAE OceanView: from an experiment towards a long-term ocean analysis and forecasting international program, in: Proceedings of the OceanObs09 Conference: Sustained Ocean Observations and Information for Society, Venice, Italy, 21–25 Sep 2009, vol. 2, ESA Publication WPP-306, 2010.
- Lévy, M., Estublier, A., and Madec, G.: Choice of an advection scheme for biogeochemical models, *Geophys. Res. Lett.*, 28, 3725–3728, doi:10.1029/2001GL012947, 2001.
- Lindsay, R. W. and Zhang, J.: Assimilation of Ice Concentration in an Ice-Ocean Model, *J. Atmos. Ocean. Techn.*, 23, 742–749, doi:10.1175/JTECH1871.1, 2006.
- Lipscomb, W. H.: Remapping the thickness distribution in sea ice models, *J. Geophys. Res.*, 106, 13989–14000, doi:10.1029/2000JC000518, 2001.
- Lipscomb, W. H., Hunke, E. C., Maslowski, W., and Jakacki, J.: Improving ridging schemes for high-resolution sea ice models, *J. Geophys. Res.*, 112, C03S91, doi:10.1029/2005JC003355, 2007.
- Lorenc, A. C., Ballard, S. P., Bell, R. S., Ingleby, N. B., Andrews, P. L. F., Barker, D. M., Bray, J. R., Clayton, A. M., Dalby, T., Li, D., Payne, T. J., and Saunders, F. W.: The Met. Office global three-dimensional variational data assimilation scheme, *Q. J. Roy. Meteor. Soc.*, 126, 2991–3012, doi:10.1002/qj.49712657002, 2000.
- Lumpkin, R. and Garraffo, Z.: Evaluating the decomposition of Tropical Atlantic drifter observations, *J. Atmos. Ocean. Techn.*, 22, 1403–1415, 2005.
- MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec, G.: Global Seasonal Forecast System version 5 (GloSea5): a high resolution seasonal forecast system, *Q. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2396, 2014.
- Madec, G.: NEMO ocean engine. Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27 ISSN No, 1288–1619, 2008.
- Martin, M. J., Hines, A., and Bell, M. J.: Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact, *Q. J. Roy. Meteor. Soc.*, 133, 59–89, 2007.
- Massonnet, F., Fichet, T., Goosse, H., Vancoppenolle, M., Mathiot, P., and König Beatty, C.: On the influence of model physics on simulations of Arctic and Antarctic sea ice, *The Cryosphere*, 5, 687–699, doi:10.5194/tc-5-687-2011, 2011.
- McPhaden, M. J., Busalacchi, A. J., Cheney, R., Donguy, J. R., Gage, K. S., Halpern, D., Ji, M., Julian, P., Meyers, G., Mitchum, G. T., Niiler, P. P., Picaut, J., Reynolds, R. W., Smith, N., and Takeuchi, K.: The Tropical Ocean-Global Atmosphere observing system: A decade of progress, *J. Geophys. Res.*, 103, 14169–14240, doi:10.1029/97JC02906, 1998.
- Megann, A., Storkey, D., Aksenov, Y., Alderson, S., Calvert, D., Graham, T., Hyder, P., Siddorn, J. R., and Sinha, B.: GO5.0: The joint NERC-Met Office NEMO global ocean model for use in coupled and forced applications, Submitted to *Geosci. Model Dev.*, 2013.
- Mirouze, I. and Weaver, A. T.: Representation of correlation functions in variational assimilation using an implicit diffusion operator, *Q. J. Roy. Meteor. Soc.*, 136, 1421–1443, 2010.
- Mogensen, K. S., Balmaseda, M. A., Weaver, A., Martin, M. J., and Vidard, A.: NEMOVAR: A variational data assimilation system for the NEMO ocean model, ECMWF newsletter, summer 2009.
- Mogensen, K. S., Balmaseda, M. A., and Weaver, A.: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4, ECMWF Tech. Memo. 668., 2012.
- Moore, A. M.: Aspects of geostrophic adjustment during tropical ocean data assimilation, *J. Phys. Oceanogr.*, 19, 435–461, 1989.
- O’Dea, E. J., Arnold, A. K., Edwards, K. P., Furner, R., Hyder, P., Martin, M. J., Siddorn, J. R., Storkey, D., While, J., Holt, J. T., and Liu, H.: An operational ocean forecast system incorporating NEMO and SST data assimilation for the tidally driven European North-West shelf, *J. Oper. Oceanogr.*, 5, 3–17, 2012.
- OSI-SAF: Sea ice product manual version 3.8, SAF/OSI/met.no/TEC/MA/125, available at: <http://www.osi-saf.org> (last access: 5 November 2014), 2012.
- Pacanowski, R. C. and Gnanadesikan, A.: Transient Response in a Z-Level Ocean Model That Resolves Topography with Partial Cells, *Mon. Weather Rev.*, 126, 3248–3270, 1998.
- Parrish, D. and Derber, J. C.: The national meteorological center’s spectral statistical interpolation analysis system, *Mon. Weather Rev.*, 120, 1747–1763, 1992.
- Penduff, T., Le Sommer, J., Barnier, B., Treguier, A.-M., Moline, J.-M., and Madec, G.: Influence of numerical schemes on current-topography interactions in 1/4° global ocean simulations, *Ocean Sci.*, 3, 509–524, doi:10.5194/os-3-509-2007, 2007.
- Peterson, K. A., Arribas, A., Hewitt, H. T., Keen, A. B., Lea, D. J., and McLaren, A. J.: Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system, *Clim. Dynam.*, 1–16, doi:10.1007/s00382-014-2190-9, 2014.

- Rio, M.-H., Poulain, P.-M., Pascual, A., Mauri, E., Larnicol, G., Santoleri, R.: A Mean Dynamic Topography of the Mediterranean Sea computed from altimetric data, in situ measurements and a general circulation model, *J. Marine Sys.*, 65, 484–508, 2007a.
- Rio, M.-H., Schaeffer, P., Hernandez, F., and Lemoine, J.-M.: The estimation of the ocean Mean Dynamic Topography through the combination of altimetric data, in-situ measurements and GRACE geoid: from global to regional studies, Proceedings of the GOCINA international workshop, Luxembourg, 2007b.
- Rio, M.-H., Guinehut, S., and Larnicol, G.: 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, 2011.
- Roberts, C. D., Waters, J., Peterson, K. A., Palmer, M. D., McCarthy, G. D., Frajka-Williams, E., Haines, K., Lea, D. J., Martin, M. J., Storkey, D., Blockley, E. W., and Zuo, H.: Atmosphere drives recent interannual variability of the Atlantic meridional overturning circulation at 26.5° N, *Geophys. Res. Lett.*, 40, 5164–5170 doi:10.1002/grl.50930, 2013.
- Roemmich, D., Belbéoc, M., Freeland, H., Garzoli, S., Gould, J., Grant, F., Ignaszewski, M., King, B., Klein, B., Le Traon, P.-Y., Mork, K. A., Owens, W. B., Pouliquen, S., Ravichandran, M., Riser, S., Sterl, A., Suga, T., Suk, M.-S., Sutton, P., and Thierry, V.: Argo: The Challenge of Continuing 10 Years of Progress, *Oceanography*, 22, 46–55, 2009.
- Rothrock, D. A.: The energetics of the plastic deformation of pack ice by ridging, *J. Geophys. Res.*, 80, 4514–4519, doi:10.1029/JC080i033p04514, 1975.
- Roullet, G. and Madec, G.: Salt conservation, free surface and varying levels: a new formulation for ocean general circulation models, *J. Geophys. Res.*, 105, 23927–23942, 2000.
- St. Laurent, L., Simmons, H., and Jayne, S.: Estimating tidally driven mixing in the deep ocean, *Geophys. Res. Lett.*, 29, 2106, doi:10.1029/2002GL015633, 2002.
- Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R.: Uncertainty in modeled Arctic sea ice volume, *J. Geophys. Res.*, 116, C00D06, doi:10.1029/2011JC007084, 2011.
- Semtner, A. J.: A model for the thermodynamic growth of sea ice in numerical investigations of climate, *J. Phys. Oceanogr.*, 6, 379–389, 1976.
- Servain, J., Busalacchi, A. J., McPhaden, M. J., Moura, A. D., Reverdin, G., Vianna, M., and Zebiak, S. E.: A pilot research moored array in the Tropical Atlantic (PIRATA), *B. Am. Meteorol. Soc.*, 79, 2019–2032, 1998.
- Storkey, D.: Summary of large-scale biases in the global FOAM ocean forecasting system, Met Office Forecasting Research Technical Report 554, 2011.
- Storkey, D., Blockley, E. W., Furner, R., Guiavarc'h, C., Lea, D., Martin, M. J., Barciela, R. M., Hines, A., Hyder, P., and Siddorn, J. R.: Forecasting the ocean state using NEMO: The new FOAM system, *J. Oper. Oceanogr.*, 3, 3–15, 2010.
- Stott, P. A., Jones, G. S., Lowe, J. A., Thorne, P., Durman, C., Johns, T. C., and Thelen, J.-C.: Transient Climate Simulations with the HadGEM1 Climate Model: Causes of Past Warming and Future Climate Change, *J. Climate*, 19, 2763–2782, doi:10.1175/JCLI3731.1, 2006.
- Sykes, P. A., While, J., Martin, M. J., Sellar, A., and McLaren, A. J.: Assessing and modelling skin SST, *J. Geophys. Res.*, in preparation, 2014.
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, 106, 7183–7192, 2001.
- Thorndike, A., Rothrock, D., Maykut, G., and Colony, R.: The thickness distribution of sea ice, *J. Geophys. Res.*, 80, 4501–4513, 1975.
- Waters, J., Lea, D. J., Martin, M. J., Storkey, D., and While, J.: Describing the development of the new foam-nemovar system in the global 1/4 degree configuration, Technical Report 578, Met Office, 2013.
- Waters, J., Lea, D. J., Martin, M. J., Mirouze, I., Weaver, A. T., and While, J.: Implementing a variational data assimilation system in an operational 1/4 degree global ocean model, *Q. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2388, 2014.
- Weaver, A. T., Deltel, C., Machu, E., Ricci, S., and Daget, N.: A multivariate balance operator for variational ocean data assimilation, *Q. J. Roy. Meteor. Soc.*, 131, 3605–3625, 2005.
- Zalesak, S. T.: Fully multidimensional flux corrected transport algorithms for fluids, *J. Comput. Phys.*, 31, 335–362, 1979.