Received: 25 August 2023 | Revised: 2 May 2024 | Accepted: 3 May 2024

DOI: 10.1002/qj.4798

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

# The Met Office Forecast Ocean Assimilation Model (FOAM) using a 1/12-degree grid for global forecasts

**Ana Barbosa Aguiar**[1] | **Michael J. Bell**[1] | **Edward Blockley**[1] | **Daley Calvert**[1] | **Richard Crocker**[1] | **Gordon Inverarity**[1] | **Robert King**[1] | **Daniel J. Lea**[1] | **Jan Maksymczuk**[1] | **Matthew J. Martin**[1] | **Martin R. Price**[1] | **John Siddorn**[1,2] | **Kerry Smout-Day**[1] | **Jennifer Waters**[1] | **James While**[1]

[1]Weather Science, Met Office, Exeter, UK

[2]National Oceanography Centre, Southampton, UK

**Correspondence**
Ana Barbosa Aguiar, Weather Science, Met Office, FitzRoy Road, Exeter, EX1 3PB, UK.
Email: ana.aguiar@metoffice.gov.uk

**Funding information**
-European Union's Horizon 2020 research and innovation programme, Grant/Award Number: 821926 (IMMERSE)

## Abstract

The Met Office Forecast Ocean Assimilation Model (FOAM) ocean–sea-ice analysis and forecasting operational system has been using an ORCA tripolar grid with 1/4° horizontal grid spacing since December 2008. Surface boundary forcing is provided by numerical weather prediction fields from the operational global atmosphere Met Office Unified Model. We present results from a 2-year simulation using a 1/12° global ocean–sea-ice model configuration while keeping a 1/4° data assimilation (DA) set-up. We also describe recent operational data assimilation enhancements that are included in our 1/4° control and 1/12° simulations: a new bias-correction term for sea-level anomaly assimilation and a revised pressure correction algorithm. The primary effect of the first is to decrease the mean and variability of sea-level anomaly increments at high latitudes, whereas the second significantly reduces the vertical velocity standard deviation in the tropical Pacific. The level of improvement achieved with the higher resolution configuration is moderate but consistently satisfactory when measured using neighbourhood verification metrics that provide fairer quantitative comparisons between gridded model fields at different spatial resolutions than traditional root-mean-square metrics. A comparison of the eddy kinetic energy from each configuration and an observation-based product highlights the regions where further system developments are most needed.

**KEYWORDS**

1/12 degree grid, data assimilation, eddy kinetic energy, forecast, global, neighbourhood verification, ocean

# 1 | INTRODUCTION

The Met Office global ocean and sea-ice Forecast Ocean Assimilation Model (FOAM) analysis and forecasting system has been operational since August 1997, using a tripolar ORCA grid with 1/4° horizontal grid spacing since December 2008 (Bell *et al.*, 2000; Blockley *et al.*, 2014; Storkey *et al.*, 2010). FOAM runs the hydrodynamic model Nucleus for European Modelling of the Ocean (Madec *et al.*, 2016) for the ocean forecast component, the Community Ice Code (Hunke *et al.*, 2015) for the sea-ice forecast component, and a multivariate incremental variational data assimilation (DA) system to incorporate observational updates (Waters *et al.*, 2015). The system assimilates observations from diverse sources: in-situ temperature and salinity profiles, in-situ and satellite sea-surface temperature (SST), satellite altimeter sea-level anomaly (SLA), and sea-ice concentration (SIC).

In September 2018, global FOAM was updated to the Global Ocean configuration version 6 (GO6) coupled to the Global Sea Ice configuration version 8.1 (GSI8.1) (Ridley *et al.*, 2018; Storkey *et al.*, 2018), retaining the 1/4° horizontal grid. This configuration has recently been incorporated into the Met Office's operational global atmosphere numerical weather prediction (NWP) system, which became a coupled atmosphere–land–ocean–sea-ice model in May 2022 (Guiavarc'h *et al.*, 2019). The results reported here are from coupled ocean–sea-ice simulations with prescribed Met Office Unified Model atmospheric forcing (Walters *et al.*, 2017).

The present study is motivated by the growing interest in higher resolution ocean models over recent years. Progressing from eddy-permitting to eddy-resolving configurations is important for climate projections and for operational short-range forecasting systems (Hewitt *et al.*, 2017). Moreover, using comparable resolutions in the different components of the global coupled NWP system should yield better performance overall. Other operational centres have been running high-resolution global ocean forecasting systems for several years. For example, Mercator Ocean International runs a NEMO system with a 1/12° horizontal grid (Lellouche *et al.*, 2018a) and produces a global ocean reanalysis GLORYS12 (Lellouche *et al.*, 2021), and the US Naval Research Laboratory runs a Hybrid Coordinate Ocean Model system at 1/25° (Barton *et al.*, 2021).

The performance of a forced 1/12° ocean–sea-ice configuration (ORCA12) is evaluated here and compared with the forced version of the current 1/4° configuration (ORCA025), both using the same GO6 and GSI8.1 model configurations but with specific parameters calibrated for each resolution. ORCA025 is eddy permitting, whereas ORCA12 enables eddies to be resolved over a wider range of latitudes. The oceanic eddy length scale is largely determined by the first baroclinic Rossby radius of deformation, which is highly dependent on the Coriolis parameter and depth, varying by an order of magnitude between the Equator (about 200 km) and high latitudes (up to 10 km). As a result, even the higher resolution configuration described here can barely resolve any eddies poleward of 50° latitude (Hallberg, 2013) and will only resolve eddies in 8% of regions shallower than 500 m (Holt *et al.*, 2017).

Traditional root-mean-square (RMS) metrics are not ideal when verifying model fields on grids with different resolutions against observations. If forecast features are offset in space and time compared with corresponding observations then there is a double-penalty effect (Rossa *et al.*, 2008), which is more apparent in the higher resolution model when compared with the smoother features of the lower resolution model. To provide a fairer comparison and validation between ORCA025 and ORCA12 experiments, in addition to traditional metrics we use the High-Resolution Assessment (HiRA) framework, which applies neighbourhood verification methods to compare model results with observations at common spatial length scales (Crocker *et al.*, 2020).

This article documents how the system has evolved since Blockley *et al.* (2014). Subsequent key DA developments have been described by Mirouze *et al.* (2016) and While and Martin (2019). In Section 2 we describe the NEMO and CICE numerical models used in the latest FOAM configuration, and Section 3 describes the NEMOVAR DA system, including some recent developments in its operational implementation in FOAM not documented elsewhere. Section 4 then describes the observations assimilated in the experiments that are described in Section 5. An overview of the results for the global ocean is presented in Section 6, first using traditional metrics and then HiRA. Conclusions and plans for future work are finally given in Section 7.

# 2 | OCEAN AND SEA-ICE MODELS

## 2.1 | NEMO/GO6

The ORCA grid used in NEMO is based on a global isotropic Mercator grid in which the meridional grid spacing is reduced polewards to match the zonal spacing (Madec & Imbard, 1996). To avoid the singularity at the North Pole, the Northern Hemisphere grid is quasi-isotropic, with two poles over land in Russia and Canada. FOAM uses 75 vertical levels, with thicknesses increasing from 1 m near the surface to 200 m at depth, which has been shown to be the minimum required to resolve the second baroclinic mode (Stewart *et al.*, 2017).

A partial-cell formulation is used at the lower boundary (Barnier *et al.*, 2006), where the bathymetry has been derived from Earth Topography 1′ for ORCA025 (Amante & Eakins, 2009), and General Bathymetric Chart of the Oceans 2014 for ORCA12 (Weatherall *et al.*, 2015). The prognostic model state variables are three-dimensional potential temperature, practical salinity, horizontal velocities, and the two-dimensional nonlinear sea-surface height (SSH).

The GO6 global ocean configuration, based on NEMO v3.6, is described by Storkey *et al.* (2018) and uses the 1980 International Equation of State for seawater densities UNESCO (1981). A nonlinear free-surface boundary condition is used, allowing for time-varying cell thicknesses throughout the water column and a better representation of surface freshwater flux. Bulk formulae from Large and Yeager (2009) are called at every time step and account for turbulent flux transfer of momentum, sensible heat, and evaporation. Lateral diffusion of momentum uses bi-Laplacian viscosity with coefficients of $-1.5 \times 10^{11}\,\mathrm{m^4 \cdot s^{-1}}$ for ORCA025 and $-1.25 \times 10^{10}\,\mathrm{m^4 \cdot s^{-1}}$ for ORCA12, both reducing polewards with the cube of the grid spacing. Lateral diffusion of tracers is along isoneutral surfaces using Laplacian mixing with coefficients of $150\,\mathrm{m^2 \cdot s^{-1}}$ for ORCA025 and $125\,\mathrm{m^2 \cdot s^{-1}}$ for ORCA12. Fresh water runoff from land is specified as climatological monthly estimates and augmented by point-like Lagrangian icebergs, which were implemented in NEMO by Marsh *et al.* (2015). The grid extends southwards to include closed (for computational efficiency, at present) ice-shelf cavities around Antarctica with a prescribed climatological freshwater input along their vertical boundary, mimicking the effect of ice shelf basal melt on the wider circulation, a parametrisation described in Mathiot *et al.* (2017). To avoid instabilities associated with artificial cliffs at the edge of the ice shelves (due to closed cavities), the momentum lateral boundary condition (free slip elsewhere) changes around the coastline of Antarctica to a partial slip condition in ORCA025 and a no-slip condition in ORCA12. In contrast to Storkey *et al.* (2018), the model time steps used here are shorter: 900 s for ORCA025 and 180 s for ORCA12. This is needed for model robustness since the system here has more energy due to (a) the DA, and (b) the higher resolution NWP forcing (see Section 5). Haney (1971)'s retroaction is used to control sea-surface salinities, adding a flux correction based on the difference between the model and climatology, with a restoring coefficient of $-33.33\,\mathrm{mm \cdot day^{-1}} \cdot \mathrm{psu^{-1}}$. A three-dimensional Newtonian damping is applied, with a 360-day time-scale, to compensate for the poor spatio-temporal observational coverage and consequent limited ability of DA to constrain values at depth. The purpose is to control the long-term evolution of subsurface tracer fields, nudging model values to temperature and salinity climatologies created by averaging EN.4.1.1 analyses (Good *et al.*, 2013) over the years 2011–2015.

## 2.2 | CICE/GSI8.1

GSI8.1 is based on version 5.1.2 of CICE (Ridley *et al.*, 2018) and uses an ORCA grid like NEMO, and for our experiments the CICE time step matches that of NEMO at both grid resolutions. This configuration has a prognostic ice thickness distribution so that, in each grid cell, there are five ice thickness categories (plus open water). The model uses a multilayer thermodynamics scheme with four vertical layers of ice plus one of snow, defined independently in each thickness category with an associated prognostic temperature profile and a fixed salinity profile.

Several other dynamics and thermodynamics prognostic variables are incorporated. The dynamics prognostics are the ice horizontal velocity, plus three principal ice stress components (in the form of a tensor with units of $\mathrm{kg \cdot s^{-2}}$) evaluated at the same grid points as velocity; that is, at each corner of a cell in the sea-ice model (Storkey *et al.*, 2018). Three thermodynamics prognostics are computed for each ice category (ice concentration, ice thickness, and snow thickness), and enthalpy per unit volume (not temperature) is computed for each of the four layers of ice plus the snow layer.

The model includes prognostic melt ponds and an explicit calculation of their impact on the sea-ice albedo, using the topographic melt pond formulation described by Flocco *et al.* (2010, 2012) and Hunke *et al.* (2015). Melt ponds play a critical role in the radiative balance since they change the sea-ice albedo.

# 3 | DATA ASSIMILATION

In this section we provide a consolidated description of the DA approaches used in FOAM. Many aspects of the system have been detailed elsewhere, but we recapitulate these and provide a brief description of three recent developments that have not yet been described in the literature. These developments are used in the experiments presented here.

The NEMOVAR multivariate incremental variational DA system has been used operationally in FOAM at the Met Office since January 2013. It minimises a cost function measuring the least-squares departures of a model state vector[1] from both the previous forecast and recent observations, while taking into account bias vectors. NEMOVAR generates analysis increments (corrections) to the previous forecast and bias increments to the estimated biases. On

each forecast/assimilation cycle, NEMOVAR is run separately for ocean and sea-ice observations. In this section, we describe the main developments since the implementation described by Mirouze *et al.* (2016) and some specific aspects of running NEMOVAR on an ORCA025 grid, while both NEMO and CICE use an ORCA12 grid.

Our NEMOVAR implementation uses a three-dimensional variational first guess at appropriate time (3DVar-FGAT Huang *et al.*, 2002; Lawless, 2010) algorithm, assimilating observations valid within a 1-day assimilation window that runs between $T-24$ and $T+0$, where the notation $T+n$ denotes a validity time $n$ hr after (positive $n$) or before (negative $n$) the nominal analysis time $T+0$ (0000 UTC). NEMOVAR produces three-dimensional increments of potential temperature, practical salinity, and horizontal velocity, together with two-dimensional increments of SSH and SIC. Additionally, using the methods of While and Martin (2019) and Lea *et al.* (2008), two-dimensional increments are produced to correct biases associated with satellite observations of SST and SLA (related to SSH; Section 3.1.2). NEMOVAR also assimilates SIC observations using a separate NEMOVAR minimisation to the ocean variables. This is because we have not yet developed multivariate relationships in NEMOVAR between the ocean and sea-ice variables and performing the minimisation separately for both improves convergence in the cost function.

For SST observations, biases are assumed to arise from atmospheric effects on the SST retrievals and are expected to differ between the different SST products. As described in Section 3.1.2, biases in SLA are assumed to come from inaccuracies in the mean dynamic topography (MDT) and dynamic atmosphere correction (DAC); these biases are treated separately but are assumed to affect all SLA observations equally. For the purposes of bias correction we split the observations into a number of different groups. First, all observations that are not bias corrected (temperature and salinity profiles, sea ice, in-situ SST, and reference satellite SST) are put into the observation vector $\mathbf{y}$. SLA data, which are all considered biased, are in the vector $\mathbf{y}^{\mathbf{SLA}}$. Finally, SST observations that need bias correction (most satellite SST) are put into separate vectors for each type of sensor $\mathbf{y}_i^{\mathrm{SST}}$, with $i=1,\ldots,p$ and $p$ being the number of sensors. The length of each vector is equal to the number of observations in each group.

Each day, a model forecast is run from $T-24$ to $T+0$ (the assimilation window) by applying the nonlinear model (NEMO/CICE) to the initial model state valid at $T-24$ to output innovations spanning the assimilation window. Innovations are the differences between observations (ideally unbiased) and model equivalent values, obtained by applying an observation operator to the model state vector at the appropriate observation time. Denoting the combined state vectors throughout the assimilation window by the single trajectory vector $\mathbf{x}^{\mathrm{f}}$, our 3DVar-FGAT innovations can be written as

$$\mathbf{d} = \mathbf{y} - H(\mathbf{x}^{\mathrm{f}}), \tag{1}$$

$$\mathbf{d}_i^{\mathrm{SST}} = \mathbf{y}_i^{\mathrm{SST}} - H_i^{\mathrm{SST}}(\mathbf{x}^{\mathrm{f}}) - \mathbf{H}_i^{\mathrm{SST}}\mathbf{b}_i^{\mathrm{f,SST}}, \quad \text{for } i=1,\ldots,p, \tag{2}$$

$$\mathbf{d}^{\mathrm{SLA}} = \mathbf{y}^{\mathrm{SLA}} - H^{\mathrm{SLA}}(\mathbf{x}^{\mathrm{f}}) - \mathbf{H}^{\mathrm{SLA}}(\mathbf{b}_{\mathrm{MDT}}^{\mathrm{f,SLA}} + \mathbf{b}_{\mathrm{DAC}}^{\mathrm{f,SLA}}), \tag{3}$$

where, as with the observation vectors, $\mathbf{d}^{\mathrm{SLA}}$, $\mathbf{d}_i^{\mathrm{SST}}$, and $\mathbf{d}$ are the innovations of the bias-corrected SLAs, bias-corrected SSTs, and other observations respectively. Likewise, $\mathbf{b}_i^{\mathrm{f,SST}}$ are the forecast vectors of the two-dimensional SST bias fields projected onto the top level of the model grid, with the vectors for the SLA MDT and DAC biases given by $\mathbf{b}_{\mathrm{MDT}}^{\mathrm{f,SLA}}$ and $\mathbf{b}_{\mathrm{DAC}}^{\mathrm{f,SLA}}$ respectively treated similarly. The observation operators ($H$, $H_i^{\mathrm{SST}}$, and $H^{\mathrm{SLA}}$) include both a time and space interpolation of the model forecast. The linear operators applied to the bias vectors ($\mathbf{H}_i^{\mathrm{SST}}$ and $\mathbf{H}^{\mathrm{SLA}}$) only include a bilinear interpolation in space and so are written as matrices. For convenience, we have omitted any time indexing from the observation vectors (which represent a collection of observations spanning the assimilation window), prior forecast trajectory $\mathbf{x}^{\mathrm{f}}$, and observation operators.

The bias-correction scheme of While and Martin (2019) also includes what are referred to as "observations-of-bias", which are the differences between biased observations and co-located, assumed unbiased, reference observations (as described in Section 3.1.1). In our system we only use observations-of-bias, $\mathbf{z}_i^{\mathrm{SSTb}}$, for the bias correction of SST. As implied by the subscript $i$, observations-of-bias are calculated separately for each SST satellite sensor, with details of the calculation given in Section 3.1.1. The observations-of-bias vectors have their own associated innovation vector $\mathbf{d}_i^{\mathrm{SSTb}}$, which is defined as

$$\mathbf{d}_i^{\mathrm{SSTb}} = \mathbf{z}_i^{\mathrm{SSTb}} - \mathbf{H}_i^{\mathrm{SSTb}}\mathbf{b}_i^{\mathrm{f,SST}}, \quad \text{for } i=1,\ldots,p, \tag{4}$$

where $\mathbf{H}_i^{\mathrm{SSTb}}$ is the observation operator for $\mathbf{z}_i^{\mathrm{SSTb}}$ and represents a bilinear interpolation in space to the position of the observations-of-bias. For SLA data, there is no independent reference dataset available with global coverage, so no observations-of-bias can be used (but note that the SLA data preprocessing already includes inter-satellite calibration).

Using the aforementioned definitions, NEMOVAR finds the best-fitting increment $\delta\mathbf{x}$ to the prior forecast as well as a set of bias increments $\delta\mathbf{b}_i^{\mathrm{SST}}$, $\delta\mathbf{b}_{\mathrm{MDT}}^{\mathrm{SLA}}$, and $\delta\mathbf{b}_{\mathrm{DAC}}^{\mathrm{SLA}}$ to

the various prior bias estimates. This is done by minimising

$$
\begin{aligned}
&2J(\delta\mathbf{x}, \delta\mathbf{b}_1^{\mathrm{SST}}, \ldots, \delta\mathbf{b}_p^{\mathrm{SST}}, \delta\mathbf{b}_{\mathrm{MDT}}^{\mathrm{SLA}}, \delta\mathbf{b}_{\mathrm{DAC}}^{\mathrm{SLA}}) \\
&= \|\delta\mathbf{x}\|_{\mathbf{B}}^2 + \|\mathbf{d} - \mathbf{H}\delta\mathbf{x}\|_{\mathbf{R}}^2 + \left\|\delta\mathbf{b}_{\mathrm{MDT}}^{\mathrm{SLA}}\right\|_{\mathbf{O}_{\mathrm{MDT}}^{\mathrm{SLA}}}^2 + \left\|\delta\mathbf{b}_{\mathrm{DAC}}^{\mathrm{SLA}}\right\|_{\mathbf{O}_{\mathrm{DAC}}^{\mathrm{SLA}}}^2 \\
&\quad + \left\|\mathbf{d}^{\mathrm{SLA}} - \mathbf{H}^{\mathrm{SLA}}\left(\mathbf{S}^{\mathrm{SLA}}\delta\mathbf{x} + \delta\mathbf{b}_{\mathrm{MDT}}^{\mathrm{SLA}} + \delta\mathbf{b}_{\mathrm{DAC}}^{\mathrm{SLA}}\right)\right\|_{\mathbf{R}^{\mathrm{SLA}}}^2 \\
&\quad + \sum_{i=1}^{p}\left[\left\|\delta\mathbf{b}_i^{\mathrm{SST}}\right\|_{\mathbf{O}_i^{\mathrm{SST}}}^2 + \left\|\mathbf{d}_i^{\mathrm{SST}} - \mathbf{H}_i^{\mathrm{SST}}\left(\mathbf{S}^{\mathrm{SST}}\delta\mathbf{x} + \delta\mathbf{b}_i^{\mathrm{SST}}\right)\right\|_{\mathbf{R}_i^{\mathrm{SST}}}^2\right. \\
&\quad \left. + \left\|\mathbf{d}_i^{\mathrm{SSTb}} - \mathbf{H}_i^{\mathrm{SSTb}}\delta\mathbf{b}_i^{\mathrm{SST}}\right\|_{\mathbf{L}_i^{\mathrm{SSTb}}}^2\right], \qquad (5)
\end{aligned}
$$

where $J$ is the cost function and $\|\mathbf{v}\|_{\mathbf{A}} = \sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{v}}$ is the Mahalanobis norm for column vector $\mathbf{v}$ and weighting matrix $\mathbf{A}$, using superscript T to denote transposition from a column to a row vector. The error covariance matrices $\mathbf{B}$, $\mathbf{R}$, $\mathbf{R}^{\mathrm{SLA}}$, $\mathbf{R}_i^{\mathrm{SST}}$, $\mathbf{O}_{\mathrm{MDT}}^{\mathrm{SLA}}$, $\mathbf{O}_{\mathrm{DAC}}^{\mathrm{SLA}}$, $\mathbf{O}_i^{\mathrm{SST}}$, and $\mathbf{L}_i^{\mathrm{SSTb}}$ correspond respectively to the errors in the prior (background) forecast, observations assimilated without bias correction, biased SLA observations, biased SST observations, background bias vectors for SLA (MDT and DAC components) and SST, and observations-of-bias for SST. More detail on the meaning and specification of these error covariance matrices is provided in the following subsections. The matrices $\mathbf{S}^{\mathrm{SST}}$ and $\mathbf{S}^{\mathrm{SLA}}$ extract the surface temperature and SLA from the state vector. The linear observation operator $\mathbf{H}$ is a set of bilinear (horizontal)/cubic spline (vertical) interpolations and variable transforms that match the model state increments to the innovations.[2] The other observation operators are the same as in Equations 2–4.

The minimisation of Equation (5) provides the analyses of the model and bias increments $\delta\mathbf{x}^{\mathrm{a}}$, $\delta\mathbf{b}_i^{\mathrm{a,SST}}$, $\delta\mathbf{b}_{\mathrm{MDT}}^{\mathrm{a,SLA}}$, and $\delta\mathbf{b}_{\mathrm{DAC}}^{\mathrm{a,SLA}}$, where the "a" superscript indicates an analysis (updated) vector. The bias increments are used to update the bias states for the next assimilation cycle as

$$\mathbf{b}_{\mathrm{yyy}}^{\mathrm{a,xxx}} = \mathbf{b}_{\mathrm{yyy}}^{\mathrm{f,xxx}} + \delta\mathbf{b}_{\mathrm{yyy}}^{\mathrm{a,xxx}}, \qquad (6)$$

where "xxx" is a substitute for "SST" or "SLA", and "yyy" represents the different bias subscripts "MDT", "DAC", or SST sensor index $i$. The analysis increments $\delta\mathbf{x}^{\mathrm{a}}$, on the other hand, are added into another reforecast of the same 1-day assimilation window, this time using an incremental analysis update (IAU) method, in which a fraction of the increments is added to the state variables on each model time step as a time-tendency forcing (Bloom *et al.*, 1996). The full benefit of the DA is consequently obtained by the end of the time window at $T + 0$.

For SIC, the assimilation increments are still applied using an IAU, but CICE adds positive increments with an assumed thickness of 0.5 m and progressively subtracts negative increments from the different ice categories, starting with the thinnest. This and other aspects of the SIC assimilation have not changed since the description by Blockley *et al.* (2014).

## 3.1 | Bias correction

Equation (5) includes penalty terms for the bias increments $\delta\mathbf{b}_i^{\mathrm{SST}}$, $\delta\mathbf{b}_{\mathrm{MDT}}^{\mathrm{SLA}}$, and $\delta\mathbf{b}_{\mathrm{DAC}}^{\mathrm{SLA}}$ used to correct biases arising from remotely sensed SST and SLA observations. The associated covariance matrices $\mathbf{O}_i^{\mathrm{SST}}$, $\mathbf{O}_{\mathrm{MDT}}^{\mathrm{SLA}}$, and $\mathbf{O}_{\mathrm{DAC}}^{\mathrm{SST}}$ use 10 iterations of an implicit diffusion operator to model spatial correlations (Mirouze *et al.*, 2016), replacing the second-order auto-regressive function of Lea *et al.* (2008).

Following cost function minimisation, each analysis bias vector is further relaxed (denoted by an "r" superscript) through multiplication by a factor $\alpha_{\mathrm{yyy}}^{\mathrm{xxx}} \leq 1$ (detailed in Sections 3.1.1 and 3.1.2), which has the effect of relaxing the bias towards zero if there are no observations:

$$\mathbf{b}_{\mathrm{yyy}}^{\mathrm{r,xxx}} = \alpha_{\mathrm{yyy}}^{\mathrm{xxx}}\mathbf{b}_{\mathrm{yyy}}^{\mathrm{a,xxx}}. \qquad (7)$$

The relaxed bias vector then becomes the forecast bias vector $\mathbf{b}_{\mathrm{yyy}}^{\mathrm{f,xxx}}$ in the next forecast/assimilation cycle. The relaxation e-folding time is given by $-T/\log\alpha_{\mathrm{yyy}}^{\mathrm{xxx}}$, where $T$ is the forecast/assimilation cycling period (24 hr).

### 3.1.1 | SST

Unlike SLA data, remotely sensed SST data use observations-of-bias (Equations 4 and 5) when calculating bias corrections. Observations-of-bias are formed by taking the difference between co-located pairs of biased observations and, assumed unbiased, reference observations. In our system we use in-situ drifting buoys and nocturnal observations from a high-quality satellite source (SNPP-VIIRS; Section 4) as reference observations. A significant number of these (up to 20% of in-situ and up to 5% of nocturnal SNPP-VIIRS observations) are withheld from direct assimilation and used instead to generate the observations-of-bias for each SST sensor (While & Martin, 2019). The remaining observations are directly assimilated.

When calculating observations-of-bias, observations points from each individual satellite are matched to reference observations within a given radius and time window. The match-up parameters for in-situ and satellite reference observations vary according to the respective spatial and temporal densities: 50 km and 1 hr for observations from drifting buoys, and 2 km and 6 hr for nocturnal observations from SNPP-VIIRS. A limit is set to only allow an

individual reference observation to be used a maximum of five times for each sensor being corrected. However, as matching is performed separately for each sensor, the overall limit for a single reference observation is five times the number of sensors being corrected.

A separate SST bias increment is estimated for each sensor being corrected, each having its own background (sixth) and observation-of-bias (eighth) terms in Equation (5). The relaxation factor $\alpha_i^{\mathrm{SST}} = 0.9$ is used throughout, which is equivalent to an e-folding time of 9.5 days, and the correlation length scale used to define the covariance matrix $\mathbf{O}_i^{\mathrm{SST}}$ is 4° (approximately 400 km at 25° N).

### 3.1.2 | SLA

SLA data are provided by the Copernicus Marine Environment Monitoring Service (Pujol, 2022). These data are assimilated using the Centre Nationale d'Études Spatiales-Collecte Localisation Satellites (CNES-CLS13) MDT, which blends altimetry, gravimetry, and in-situ drifter data over the period 1993–2012 (Rio *et al.*, 2014), to relate SLA to model SSH when calculating innovations. MDT is defined as the difference between mean sea surface, which is a time-averaged SSH, and the geoid, which is the surface the ocean would have if it were at rest (LeGrand *et al.*, 1998), whereas SLA is the difference between SSH and mean sea surface. We note that CMEMS follows the convention that SSH is measured relative to a reference ellipsoid, whereas NEMO defines SSH relative to the geoid, meaning that the NEMO SSH is equivalent to the CMEMS absolute dynamic topography, which is the sum of the MDT and SLA—$\eta$ in Vidard *et al.* (2009). Uncertainties in the geoid lead to inaccuracies in the stationary MDT, which is subtracted from the NEMO SSH obtained from the previous forecast to calculate the model equivalent SLA observations $H^{\mathrm{SLA}}(\mathbf{x}^{\mathrm{f}})$ in Equation (3). This leads to a stationary, but spatially varying, bias in the innovations. The bias-correction framework described in Section 3 is then able to correct for MDT inaccuracies using $\mathbf{b}_{\mathrm{MDT}}^{\mathrm{f,SLA}}$ in Equation (3). Because the error is stationary, no relaxation is applied to this bias ($\alpha_{\mathrm{MDT}}^{\mathrm{SLA}} = 1$). An offline estimate of the bias could be made, but the online scheme is more convenient since it avoids multiple long runs for its estimation and use, and new observations in previously poorly observed areas can produce updates to the bias estimate. The error covariance associated with the bias, $\mathbf{O}_{\mathrm{MDT}}^{\mathrm{SLA}}$, is specified to have a correlation length scale of 0.4° (approximately 40 km at 25° N), allowing the bias vector $\mathbf{b}_{\mathrm{MDT}}^{\mathrm{f,SLA}}$ to address small-scale errors, which are known to be large for MDT (Lea *et al.*, 2008). The diagonal elements of $\mathbf{O}_{\mathrm{MDT}}^{\mathrm{SLA}}$ are derived from the error estimates provided with the latest version of the MDT.

The second bias term associated with SLA data, $\mathbf{b}_{\mathrm{DAC}}^{\mathrm{f,SLA}}$, is a recent development in FOAM and is described here for the first time. The SLA data processed by CMEMS have had ocean tides removed and a further DAC applied, the latter consisting of a low-frequency inverse barometer correction and a high-frequency correction derived from atmospheric wind and pressure forcing. The high-frequency correction, spanning periods from 0.5 to 20 days, is necessary because altimeter measurements alias oceanic signals with frequencies more than the Nyquist frequency, which is twice the orbital repeat frequency, onto low frequencies (Chen & Lin, 2000). To do this, the high-frequency response to ocean winds is calculated by CMEMS using Modèle aux Ondes de Gravité 2-Dimensions pour l'océan Global simulations forced by European Centre for Medium-Range Weather Forecasts (ECMWF) pressure and wind fields (Carrère & Lyard, 2003). The overall DAC mainly benefits high latitudes. A final correction has also been made by CMEMS to the remaining orbital long-wavelength errors, which also include high-frequency errors associated with large spatial scales, following a similar approach to that of Le Traon *et al.* (1998).

The NEMO GO6 configuration used in these experiments does not include tides or the inverse barometer effect, so the removal of these signals from the observations by CMEMS is appropriate. However, the model does represent the remaining DAC high-frequency elements, which we should ideally restore to the SLA data before calculating innovations. The experiments presented here pre-date the availability of correction data. Even if those data were available with the remaining DAC elements included, discrepancies in the modelled and observed estimates of these signals would pose challenges for DA due to the sparse sampling by the observations compared with the high frequency of the signals. Without addressing this inconsistency, the assimilation of SLA data (including DAC) generates SSH increments that are large in both their mean and variability (Figure 1).

The second SLA bias term has been introduced to indirectly address this discrepancy in the barotropic response to pressure and wind forcing by removing a large-scale signal from the innovations using a correlation length scale of 4° (approximately 400 km at 25° N), matching that of the unbalanced (barotropic) SSH background-error covariance matrix. The standard deviation used in the $\mathbf{O}_{\mathrm{DAC}}^{\mathrm{SLA}}$ covariance matrix is small at low latitudes, increasing to 2 cm between 40° and 50° N, and to 3 cm between 40° and 50° S. A relaxation factor $\alpha_{\mathrm{DAC}}^{\mathrm{SLA}} = 0.95$ has also been chosen, equivalent to an e-folding time of 19.5 days, which is

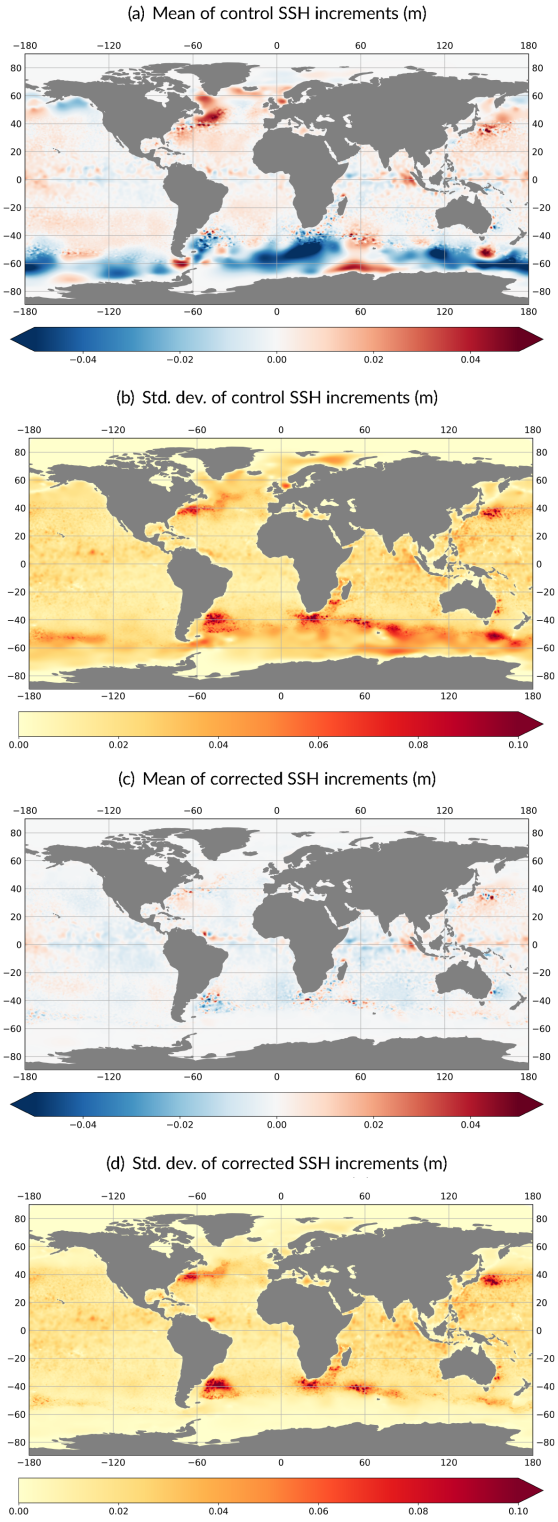


**FIGURE 1** (a, c) Mean (m) and (b, d) standard deviation (m) of daily sea-surface height (SSH) increments over January 2013 for an ORCA025 experiment assimilating sea-level anomaly data with (a, b) only the small-scale mean dynamic topography bias correction term applied and (c, d) another experiment that includes a second large-scale bias term to reduce the discrepancy between the preprocessed sea-level anomaly observations and the Forecast Ocean Assimilation Model arising from the dynamic atmosphere correction. [Colour figure can be viewed at wileyonlinelibrary.com]

similar to the 20-day cut-off period used in the DAC processing (Carrère & Lyard, 2003) and should have a similar effect to the preprocessing of the model counterpart of the observations done by Lellouche *et al.* (2018b).

The effect of the new bias-correction term $\mathbf{b}^{\mathrm{f,SLA}}_{\mathrm{DAC}}$ on the SSH increments is illustrated in Figure 1, where we see significant reductions in the mean and variability of the increments at high latitudes in both hemispheres. The innovation statistics for SSH are largely unchanged, with some decrease in their RMS at high latitudes (not shown).

## 3.2 | Background error

Background-error spatial correlations are modelled using an implicitly formulated diffusion operator (Weaver *et al.*, 2016), whereas multivariate relationships are specified through a linearised balance operator (Weaver *et al.*, 2005). An inverse balance operator $K^{-1}$ is used to transform the state vector $\mathbf{x}$ into a control vector $\mathbf{w}$ of mutually uncorrelated variables comprising temperature, unbalanced salinity, and unbalanced SSH as $\mathbf{w} = K^{-1}(\mathbf{x})$. The inverse of this equation, namely $\mathbf{x} = K(\mathbf{w})$, can be linearised as $\delta\mathbf{x} = \mathbf{K}\delta\mathbf{w}$, where $\mathbf{K}$ is the Jacobian matrix of $K$ evaluated at the prior forecast $\mathbf{x}^{\mathrm{f}}$ (taken at the start of the assimilation time window here), and $\delta\mathbf{w}$ is the increment to control vector $\mathbf{w}$. This enables Equation (5) to be minimised in terms of $\delta\mathbf{w}$ and the bias increments $\delta\mathbf{b}^{\mathrm{xxx}}_{\mathrm{yyy}}$ using the block-diagonal covariance matrix $\mathbf{B}_{\mathbf{w}}$, which is related to the background-error covariance matrix $\mathbf{B}$ from Equation (5) via $\mathbf{B} = \mathbf{K}\mathbf{B}_{\mathbf{w}}\mathbf{K}^{\mathrm{T}}$, where superscript T represents matrix transposition. Apart from temperature, only unbalanced background-error covariances are required. No balance relationships are defined between the ocean and SIC variables, whose increments are obtained from a separate NEMOVAR minimisation.

A detailed description of the background-error covariances for the original FOAM system is provided in Waters *et al.* (2015). Since then, there have been several updates to the background-error correlations. The horizontal correlations are now modelled using an implicit two-dimensional diffusion operator (Weaver *et al.*, 2016) that allows for improved modelling of the correlations around coastlines compared with the previous approach of applying one-dimensional diffusion successively to the two horizontal directions. NEMOVAR has also been updated to allow for multiple length scales in the background-error correlations (Mirouze *et al.*, 2016). In our FOAM implementation, we specify two correlation length scales for temperature and unbalanced salinity. A shorter scale depending on the Rossby radius represents the mesoscale errors, whereas a longer 4° scale represents the errors associated with atmospheric and other larger scale processes.

The ratio of the variances associated with the two scales (which determines the effective overall length scale) varies seasonally and spatially and was estimated using a combination of the National Meteorological Center method of Parrish and Derber (1992) and the innovations method of Hollingsworth and Lönnberg (1986). The scale ratio is modified to only use the short scale when the background (prior forecast) SST falls below a specified threshold, thereby preventing increments from being spread under sea ice. For the other control variables, a single horizontal background-error correlation scale is specified. For unbalanced SSH, this is a 4° length scale (the balanced component of SSH has the same spatial scales as temperature and salinity), whereas SIC uses 25 km everywhere. The vertical length scales for temperature and salinity are unchanged from those of Waters *et al.* (2015) and are parametrised based on the local mixed-layer depth in the background field. Likewise, the background-error variance specification follows Waters *et al.* (2015), with the temperature error standard deviations being related to the vertical temperature gradients in the background field.

The diffusion operator's computational cost is proportional to the ratio of the length scale to the local grid spacing. At high latitudes, this ratio is large for the long length-scale component, so we cap the temperature and salinity length scales at 40 times the DA grid spacing (which affects regions south of 60° S and north of approximately 50° N).

## 3.3 | Pressure correction

A pressure-correction algorithm (Bell *et al.*, 2004) is applied to the model at each time step to improve assimilation at the Equator, where the dominant balance is between the horizontal pressure gradient and the wind stress. When assimilation increments are applied to the model, the pressure gradient no longer balances the wind stress. The pressure correction aims to reduce the impact of systematic biases in the equatorial balance by gradually accumulating potential temperature and salinity assimilation increments that are then converted to a pressure-correction bias, combining the equation of state for sea water with hydrostatic balance. This correction is then applied in the model's equations of motion.

The second recent development to FOAM, which has not been reported elsewhere, is that the pressure correction's gradients have been set to zero at the bottom of the ocean (rather than the top, as done previously) to avoid introducing spurious barotropic motions when applying the pressure correction. Through this small change in the implementation, the RMS of the SSH innovations in the tropical Pacific is reduced from 4.4 cm to 4.1 cm, averaged over 1 year. Figure 2 shows that the RMS temperature profile innovations are reduced over all depths, particularly between 50 m and 200 m. The largest impact is in the western tropical Pacific, where the vertical velocity standard deviation is significantly reduced, as shown in the same figure.

## 3.4 | Freshwater balance

The third change to the FOAM system not described elsewhere is that we include a change in the NEMO configuration to address the issue of freshwater inputs (rivers and precipitation) and outputs (evaporation) not balancing, leading to significant drifts in the model's SSH. We use climatological estimates of freshwater inputs by rivers $R$ and three-hourly estimates of evaporation $E$ and precipitation $P$ calculated using operational Met Office atmospheric NWP fields.[3] Here, we use a NEMO option to prevent the model's SSH from drifting due to freshwater imbalances. The imbalance is calculated on each time step and used to adjust the evaporation or precipitation. When the global $E - P - R$ is negative, the evaporation is increased in regions where the evaporation dominates to restore the total freshwater balance. When the global $E - P - R$ is positive, the precipitation is increased in regions where $P + R$ dominates to restore the balance. The model's global mean SSH can consequently only change in time through the application of SSH assimilation increments.

## 3.5 | Global ocean heat content

The altimetry bias-correction enhancement and freshwater balance update have been partly motivated by the use of FOAM as the ocean component of the seasonal global atmosphere–land–ocean–sea- ice coupled system GloSea (MacLachlan *et al.*, 2015), which is also used to generate a historical reanalysis going back to October 1993, when many fewer temperature and salinity profile observations were available to constrain the model. Altimeter observations impact global ocean heat content (OHC) because the DA projects SSH increments onto the subsurface temperature. ORCA025 experiments have been carried out to examine the impact of assimilating altimeter data on global OHC in an Argo period (for about 15 months starting in July 2012) and to look at the expected impact during the pre-Argo period by running experiments for the same period, but withholding Argo data.

Comparing the "control" and "control_noArgo" results in Figure 3, we see that, when there are no Argo data to constrain the global OHC directly, the altimetry assimilation generates very large drifts and variability in the global OHC when applied in the form described by Waters

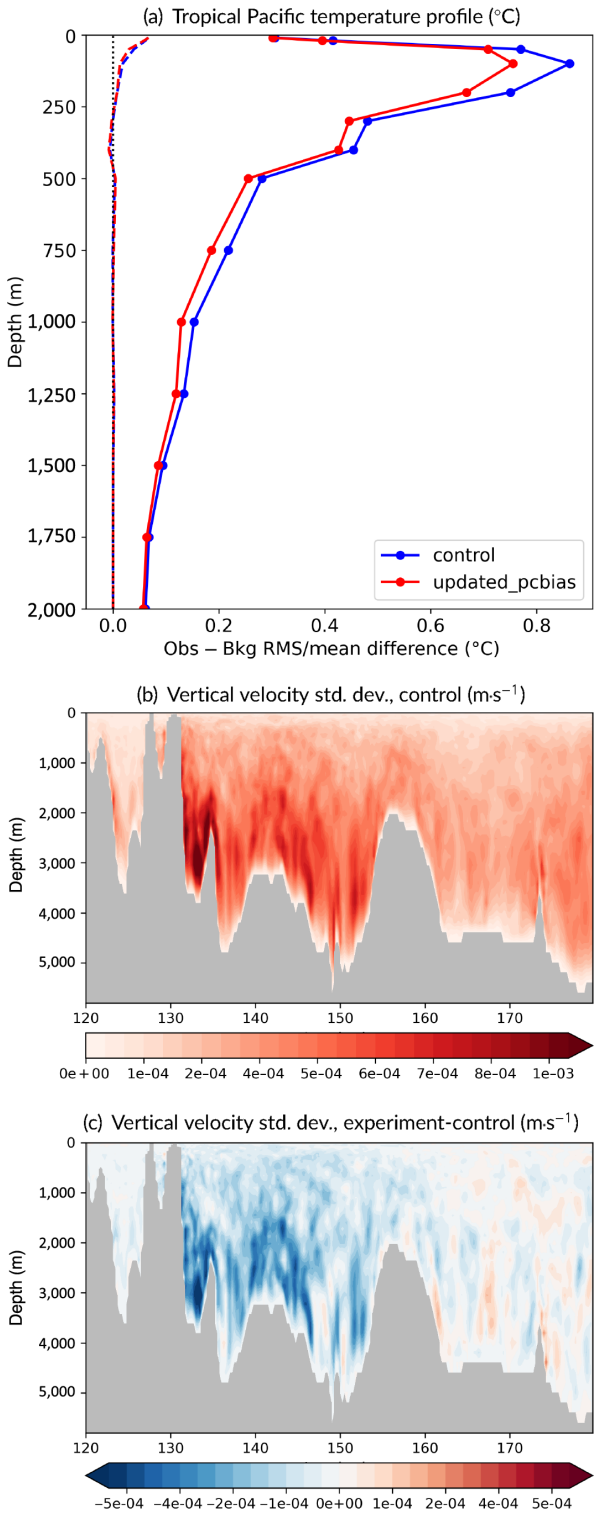


**FIGURE 2** (a) Updated pressure-correction impact on tropical Pacific temperature profile (° C) innovations calculated from February 1, 2015, to January 31, 2016, with dashed lines indicating the mean and solid lines the root-mean-square for the control (blue) and updated pressure-correction experiment (red). (b) Vertical velocity standard deviation ($m \cdot s^{-1}$) depth/longitude cross-section at the western tropical Pacific Equator for the control, and (c) the difference ($m \cdot s^{-1}$) between the experiment and control. [Colour figure can be viewed at wileyonlinelibrary.com]

*et al.* (2015). These drifts are caused partly by the impact of the SLA assimilation at high latitudes due to discrepancies in the representation of the DAC components of the SSH between the model and CMEMS data, and partly by drifts in the SSH due to freshwater imbalances. Local changes in the SSH due to both these issues are seen by the SLA assimilation as an error and projected onto the subsurface temperature through the multivariate balance relationships used in the assimilation.

The impact of applying the altimetry bias correction for the DAC component (described in Section 3.1.2) alone on the global OHC when assimilating Argo data is apparent by comparing the "control" and "newAlt" results, where we see that there is a small reduction in global OHC. In comparison, without Argo data, global OHC is significantly reduced (compare "control_noArgo" with "newAlt_noArgo") and is now much lower than in the experiments that assimilate Argo data. The experiments that assimilate Argo data are more consistent with each other than those that do not, since they are well constrained by the data and are likely to more accurately estimate global OHC.

Without Argo data, the addition of the freshwater balance adjustment from Section 3.4 ("newAltFwb_noArgo") brings the spuriously low global OHC for "newAlt_noArgo" more into line with the experiments that assimilate Argo data ("newAltFwb"). Global OHC variability is still larger than for experiments assimilating Argo data, but the overall values are much improved. We nevertheless note that, when assimilating Argo data, the impact of including the freshwater balance update is smaller (compare "newAlt" with "newAltFwb").

Though the enhancements from Sections 3.1.2 and 3.4 are not expected to significantly affect global OHC when assimilating data from the present-day global observation network, they are expected to produce significant improvements to the quality of global OHC in the pre-Argo period of GloSea reanalyses.

## 3.6 | Implementation in ORCA12

It is common for high-resolution systems to perform the DA at a resolution lower than that of the forecast model in operational NWP and oceanography, such as the Bluelink (Oke *et al.*, 2013), Mercator Ocean International (Lellouche *et al.*, 2018a), and Global Ice Ocean Prediction System Smith *et al.* (2019) ocean DA systems. Though our ORCA025 control runs the DA at the model resolution, the ORCA12 experiment's DA still uses the 1/4° grid, using bilinear interpolation to transform necessary fields between the 1/4° and 1/12° grids. The innovations calculated through Equations 1–3 use the ORCA12 model

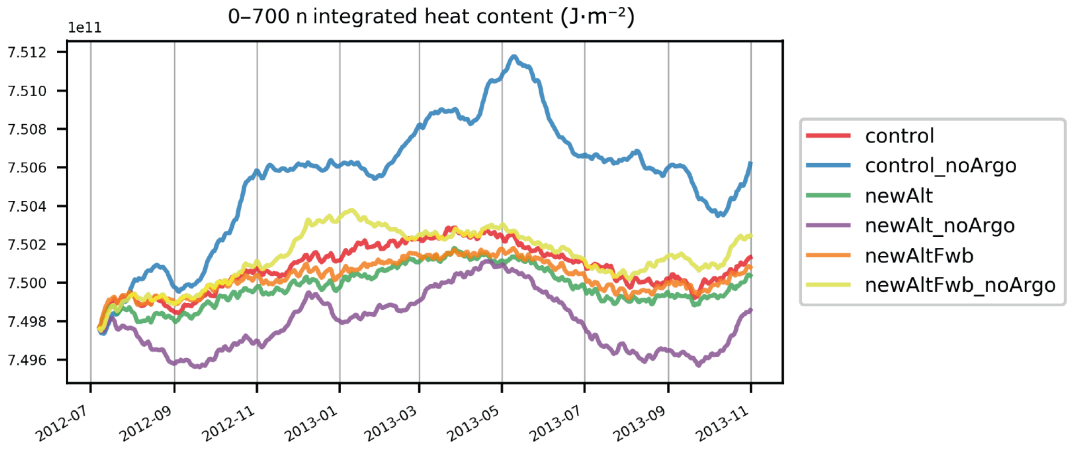


**FIGURE 3** Global average ocean heat content (J·m$^{-2}$) integrated between 0 m and 700 m depth. Control (red): standard Forecast Ocean Assimilation Model, assimilating all standard data, with no change to altimeter assimilation or freshwater budget; newAlt (green): as for control but including changes to the altimeter assimilation described in Section 3.1.2; newAltFwb (orange): as for newAlt but also including the change to the freshwater balance described in Section 3.4; the experiments with "_noArgo" are the same but do not assimilate the temperature/salinity profiles from Argo floats. [Colour figure can be viewed at wileyonlinelibrary.com]

forecast, with the model counterpart of the observations taken from the nearest model time steps to the observation times. The background fields used to estimate multivariate balance fields are output on the ORCA12 grid and then interpolated to the ORCA025 grid used by NEMOVAR. The analysis increments generated by NEMOVAR on the ORCA025 grid are interpolated to the ORCA12 grid before being applied using an IAU in a 1-day model run of the higher resolution model. This approach significantly reduces the overall computational cost compared with running the DA at full resolution and allows us to make use of the well-established 1/4° DA configuration.

In the experiments described later, the DA settings for both ORCA12 and ORCA025 are the same and include all the components described thus far. We have not retuned the error covariances for the ORCA12 experiment, despite the increased model resolution being expected to change the background-error standard deviations and the representation-error component of the observation-error standard deviations whose innovations involve the model forecast, namely those used in $\mathbf{R}$, $\mathbf{R}^{\mathrm{SLA}}$, and $\mathbf{R}_i^{\mathrm{SST}}$ (Lorenc, 1986). Instead, we have retained the same covariances for both the ORCA025 control and ORCA12 experiment. Work is under way to calculate new error covariances for the 1/12° system (Carneiro *et al.*, 2021) using the outputs of the experiments described in Section 5.

## 4 | OBSERVATIONS

The experiments described in Section 5 both assimilate the observations listed in Table 1. SLA and SIC observations have already been reprocessed as global along-track level 3 data and a daily gridded product on a 10 km stereographic projection respectively.

Following evidence that the operational system was overfitting the subsurface data (Ryan *et al.*, 2015), the coupled atmosphere–land–ocean–sea-ice forecast model has restricted profile observations to one per model level since its operational implementation in May 2022 and we have adopted the same approach in our experiments.

In-situ and satellite SST innovations are compared with the observation and background errors prior to assimilation to assess a probability of gross error (Ingleby & Huddleston, 2007), removing observations where this exceeds 50% (about 2% of the total number).

SNPP VIIRS nocturnal observations with quality level 5 are used as reference observations for SST bias correction, following their adoption as reference in the Operational Sea Surface Temperature and Sea Ice Analysis Good *et al.* (2020), augmenting the much smaller and sparser data from in-situ drifting buoys (Section 3.1).

## 5 | EXPERIMENTS

An ORCA025 control and ORCA12 experiment have been run over a 2-year period 2017–2018, running 2-day forecasts for most of the period. For July–December 2018, 5-day forecasts were produced each day. As well as assessing ORCA12 performance against ORCA025, these forecasts enable future recalibration of background-error covariances for the ORCA12 configuration (Section 7).

**TABLE 1** Observations available for assimilation in our experiments.

| Observation | Instrument | Bias corrected | Reference | Source |
|---|---|---|---|---|
| In-situ *T* profile | Argo floats, buoys (moored and drifting), CTD, gliders, marine mammals, thermistor chains, XBT | N | N | MOHC |
| In-situ S profile | Argo floats, buoys (moored and drifting), CTD, gliders | N | N | MOHC |
| In-situ SST | Moored buoys, ships | N | N | WMO GTS |
| In-situ SST | Drifting buoys (GDP) | N | Y | WMO GTS |
| Satellite SST | GCOM-W1 AMSR2 | Y | N | PODAAC |
| | NOAA-18 and NOAA-19 AVHRR | Y | N | PODAAC |
| | MetOp-B AVHRR | Y | N | EUMETSAT |
| | Meteosat-11 and MSG-3 SEVIRI | Y | N | OSI SAF |
| | Sentinel-3A SLSTR (dual view) | Y | N | EUMETSAT |
| | SNPP VIIRS diurnal | Y | N | PODAAC, EUMETSAT |
| | SNPP VIIRS nocturnal | N | Y | PODAAC, EUMETSAT |
| Satellite SLA | SARAL AltiKa | Y | N | CMEMS |
| | Cryosat-2 SIRAL | Y | N | CMEMS |
| | HaiYang-2A | Y | N | CMEMS |
| | JASON-2 and JASON-3 | Y | N | CMEMS |
| | Sentinel-3A and Sentinel-3B SRAL | Y | N | CMEMS |
| Satellite SIC | DMSP SSMIS | N | N | OSI SAF, EUMETSAT |

*Note*: "Y" (yes) or "N" (no) indicates which are bias corrected and/or used as reference observations (Section 3.1). For profiles, *T* denotes temperature and *S* salinity.

Abbreviations: AltiKa, altimeter Ka-band; AMSR, advanced microwave scanning radiometer; AVHRR, advanced very high-resolution radiometer; CMEMS, Copernicus Marine Environment Monitoring Service; CTD, conductivity temperature depth (sonde); DMSP, Defense Meteorological Satellite Program; EUMETSAT, European Organisation for the exploitation of Meteorological Satellites; GCOM-W1, Global Change Observing Mission water "Shizuku"; GDP, Global Drifter Program; GTS, Global Telecommunications System; JASON, Joint Altimetry satellite oceanographic network; MetOP, meteorological operational; MOHC, Met Office Hadley Centre; MSG, Meteosat second generation; NOAA, National Oceanographic and Atmospheric Administration; OSI SAF: Ocean and Sea Ice Satellite Application Facility; PODAAC, Physical Oceanography Distributed Active Archive Centre; SARAL, satellite with Argos and AltiKa; SEVIRI, spinning enhanced visible and infrared imager; SIC, sea-ice concentration; SIRAL, synthetic aperture radar interferometric radar altimeter; SLA, sea-level anomaly; SLSTR, sea and land surface temperature radiometer; SNPP, Suomi National Polar-orbiting Partnership; SRAL, synthetic aperture radar altimeter; SSMIS, special sensor microwave imager/sounder; SST, sea-surface temperature; VIIRS, visible infrared imaging radiometer suite; WMO, World Meteorological Organization; XBT, expendable bathythermograph (probe).

## 5.1 | Surface forcing

Surface forcing is provided by the operational global atmospheric model's archived output in the form of hourly fields of 10-m winds together with three-hourly fields of 10-m temperature, 10-m specific humidity, precipitation fluxes, and radiation fluxes. These fields are interpolated in time and space onto the appropriate ORCA grid. The Unified Model's operational grid spacing reduced during the period covered by these experiments from being approximately 17 km (at 50° N) before July 11, 2017, to approximately 10 km after that date.

When GO6/GSI8.1 was introduced in our operational FOAM system in September 2018, snowfall was halved over sea ice to account for snow blown off the ice, motivated by the findings of Schröder *et al.* (2019). A thicker snow cover leads to increased insulation of the ocean beneath with consequent reduced cooling and sea-ice growth. The halving of snowfall over sea ice was found to bring the analysis sea-ice volumes closer to the Pan-Arctic Ice Ocean Modeling and Assimilation System (Zhang & Rothrock, 2003) reanalysis product (not shown) and has also been included in our experiments.

## 5.2 | Initial conditions

A new ocean experiment is typically started from a fully spun-up state from an earlier experiment or operational forecast. The ORCA025 control took its initial data from an

assimilative experiment whose configuration had minimal differences compared with the control and ran from February 1, 2015, until January 1, 2017. The ORCA12 experiment, on the other hand, was initialised from an ORCA025 assimilative experiment whose output was interpolated to ORCA12 then run for 18 months with a slightly older DA configuration, followed by a further 6 months from June 29, 2016, until January 1, 2017, with the same DA version as described here.

## 6 | RESULTS

To reduce the computational cost of the control and experiment, the first 18 months of the 2017–2018 period have been run using 2-day forecasts to support the DA cycling and future covariance recalibration. Only the last 6 months have been run with 5-day forecasts, with the caveat that this gives a limited insight into the performance of the system over different seasons. As a result, the full 2-year period is studied using innovations statistics comparing the 1-day lead time forecasts with the next day's observations. We have also compared 1-day forecasts of ocean currents, eddy kinetic energy (EKE), and sea-ice volume with reanalysis datasets. Finally, 5-day forecasts for the latter half of 2018 are analysed using neighbourhood verification metrics at different forecast lead times based on the continuous ranked probability score (Hersbach, 2000).

### 6.1 | Innovations statistics

Time series of mean and RMS innovations of in-situ SST, SLA, and SIC for ORCA025 and ORCA12 are shown in Figure 4, averaged over the global ocean. Generally, higher RMS values for ORCA12 appear to suggest a small degradation in performance, but this is due to the intrinsic double-penalty effect that occurs when comparing model forecasts at different resolutions and is not supported by the neighbourhood verification results of Section 6.4. Mean innovations for in-situ SST are consistent for ORCA025 and ORCA12. Poorer SLA performance, measured by higher RMS innovations, is mainly accounted for by discrepancies in the Southern Ocean and the Zapiola anticyclone region in particular, as seen in the time-averaged innovation maps of Figure 6. SIC shows a sharp reduction (improvement) in RMS innovations after March 2017 for ORCA12, and this configuration seems to perform better (with smaller mean innovations than ORCA025) during July and August, which correspond to the Arctic melt season. The March 2017 reduction of RMS

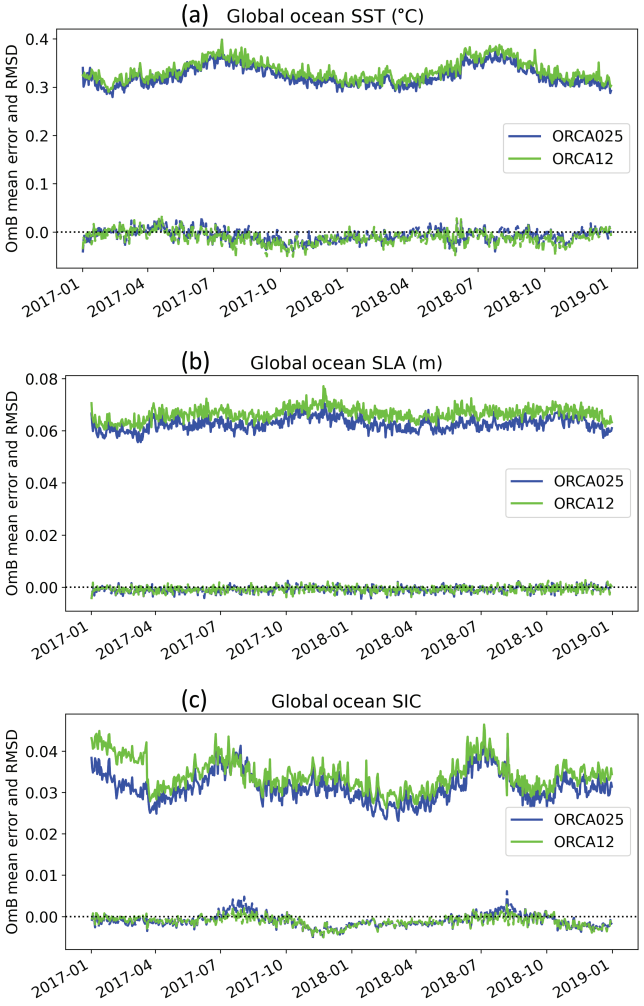


**FIGURE 4** Spatially averaged innovations (observation minus background model equivalent value) time series for the global ocean, with ORCA025 in dark blue and ORCA12 in light green for (a) in-situ sea-surface temperature (SST, °C), (b) sea-level anomalies (SLAs, m), and (c) sea-ice concentration (SIC). The mean values are very close to zero and plotted with dashed lines, whereas root-mean-square deviation (RMSD) values are higher and use solid lines. [Colour figure can be viewed at wileyonlinelibrary.com]

innovations coincides with an update to the observations product from the Ocean and Sea Ice Satellite Application Facility, with additional filters removing some observations along the coasts of Denmark, Iceland, and some Japanese islands.

Figure 5 shows time-averaged innovation profile statistics for temperature and salinity averaged for the global, North Atlantic, and Indian oceans. For the global ocean, the top 250 m of the temperature innovations show that ORCA12 forecasts are less biased compared with observations, but there is an overall increase in their RMS innovations. Temperature statistics are overall less favourable for ORCA12 in the North Atlantic and Indian oceans. For

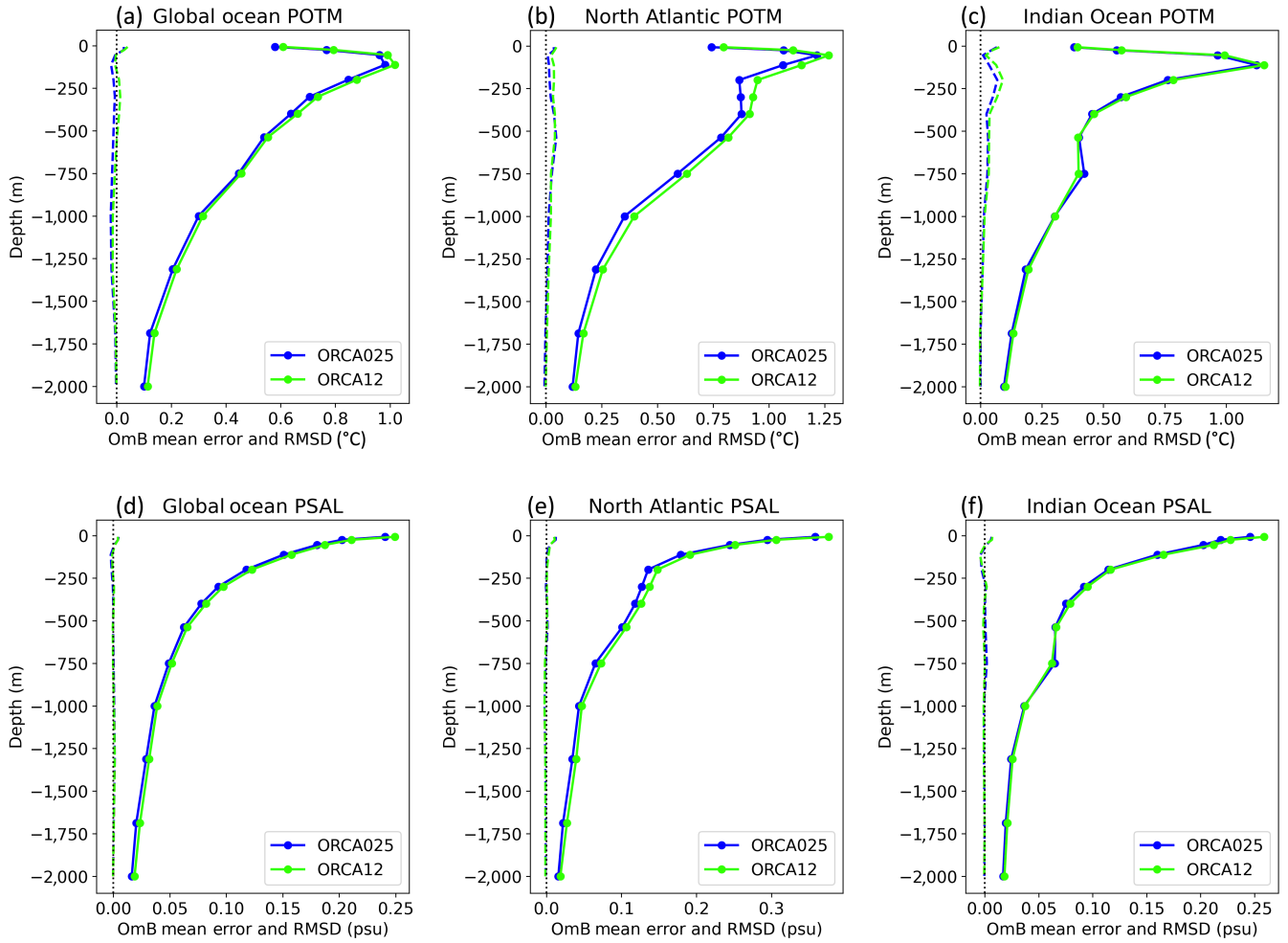


**FIGURE 5** Time-averaged innovation profiles for the (a, d) global, (b, e) North Atlantic, and (c, f) Indian oceans for years 2017–2018: (a–c) temperature $T$ and (d–f) salinity $S$ in ORCA025 (dark blue) and ORCA12 (light green). Mean values are plotted using dashed lines, whereas root-mean-square deviation (RMSD) values use solid lines. [Colour figure can be viewed at wileyonlinelibrary.com]

salinity, the differences between configurations are generally small, but more noticeable in the North Atlantic Ocean.

RMS innovation statistics for geographically binned data are shown in Figure 6 for in-situ SST, SIC, SLA, and temperature and salinity profiles as a percentage change in the ORCA12 minus ORCA025 values relative to the ORCA025 control, averaged over the whole time period. For in-situ SST, ORCA12 shows somewhat higher values in areas of high eddy activity, such as the Antarctic Circumpolar Current (ACC), the Gulf Stream, and the eastern equatorial Pacific. For SIC, it is worth noting that Arctic SIC values are lower for ORCA12 everywhere except along the sea-ice edge, whereas ORCA025 appears to be better overall in the Antarctic. ORCA12 shows larger SLA values for North Atlantic and Kuroshio currents. The Zapiola anticyclone in the southwest Atlantic has an even more pronounced difference of about 75%. This is a confluence region where the northward turn of the ACC after Drake's Passage and the South Brazil Current meet, and there is strong eddy activity as well as strong topography-driven circulation due to the complex bathymetry there (de Miranda *et al.*, 1999). The apparently poor performance of ORCA12 in that region is thought to be due to the DA still using a 1/4° grid, which was defined with a different (and smoother) bathymetry to the one used in ORCA12. This affects the model initialisation in regions where the bathymetry has sharp horizontal gradients. Griffies *et al.* (2015) have also found a striking sea-surface variability over the Zapiola anticyclone, in an independent global coupled model with resolution similar to that of ORCA12, and were unsure if this represented "a limitation of the model or the coarseness of the satellite measurements". The remaining plots of Figure 6 show that RMS temperature and salinity profile innovations are consistently larger for ORCA12 over most of the global ocean.

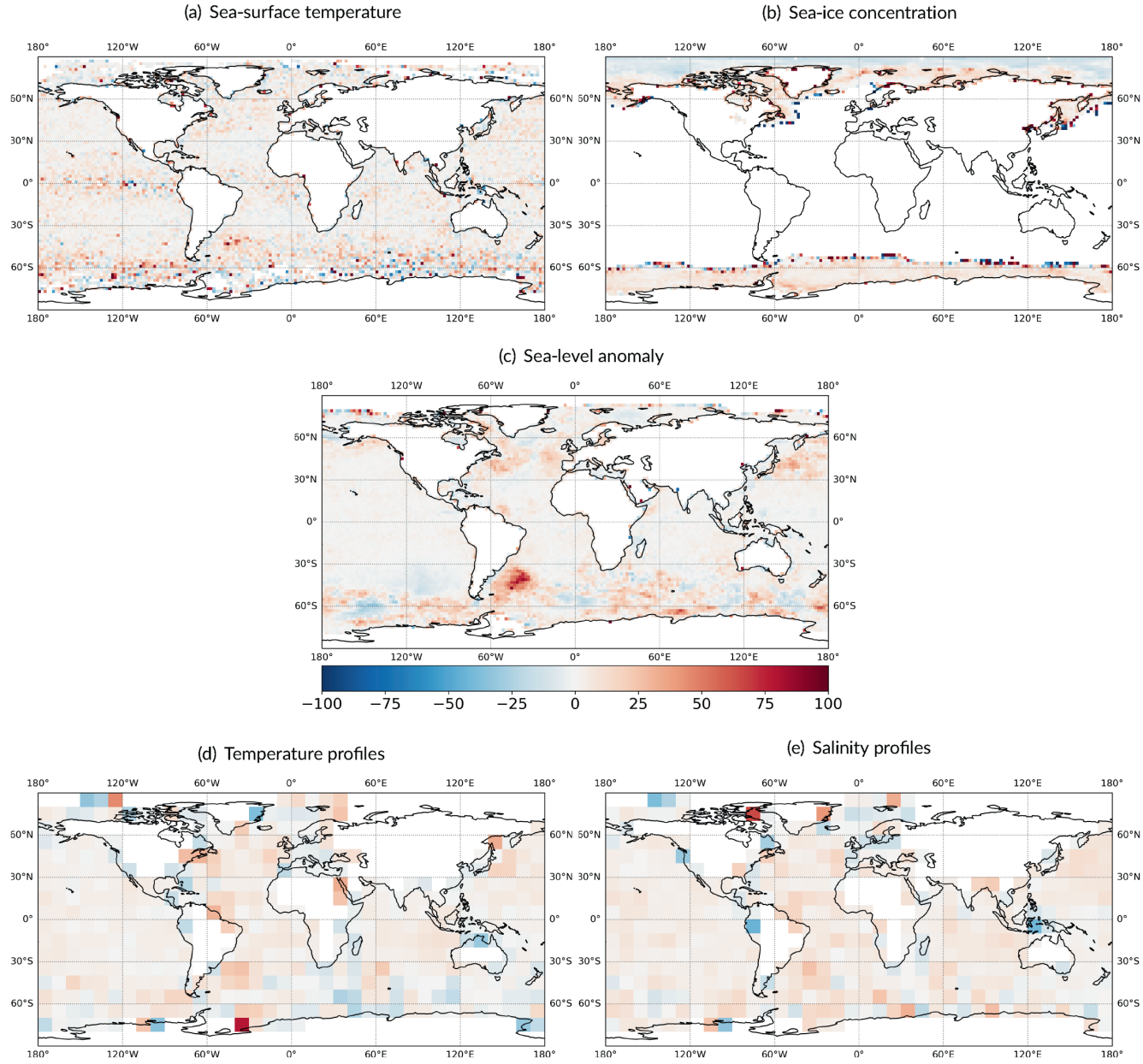


**FIGURE 6** Percentage changes in root-mean-square innovations (100 × (ORCA12 − ORCA025)/ORCA025) averaged over time (2017–2018) and depth (0 m to 2,000 m) for profiles, and binned in latitude–longitude cells. The colour scale is the same for all these plots. The bin sizes are 2° × 2° for (a) sea-surface temperature, (b) sea-ice concentration, and (c) sea-level anomaly, and 10° × 10° for (d) temperature and (e) salinity profiles. [Colour figure can be viewed at wileyonlinelibrary.com]

## 6.2 | Analysis validation

### 6.2.1 | Surface currents

We now compare analyses of surface currents in the regions of the Gulf Stream and Agulhas currents with the Ocean Surface Current Analysis Real-time (Bonjean & Lagerloef, 2002) and GlobCurrent monthly mean gridded products, averaged over July to September 2017 in Figures 7 and 8. Each of these locations is characterised by a highly dynamic boundary current and associated strong eddy activity. OSCAR and GlobCurrent offer observations-based estimates of the surface currents: OSCAR "combines geostrophic, Ekman and Stommel shear dynamics, with a complementary term from the surface buoyancy gradient", whereas GlobCurrent combines surface geostrophic currents derived from altimetry observations with an Ekman component estimated by applying an empirical model to wind stress reanalysis (Rio *et al.*, 2014). The underlying assumption of geostrophy means that both products underestimate the total surface current, being unable to account for the full ageostrophic

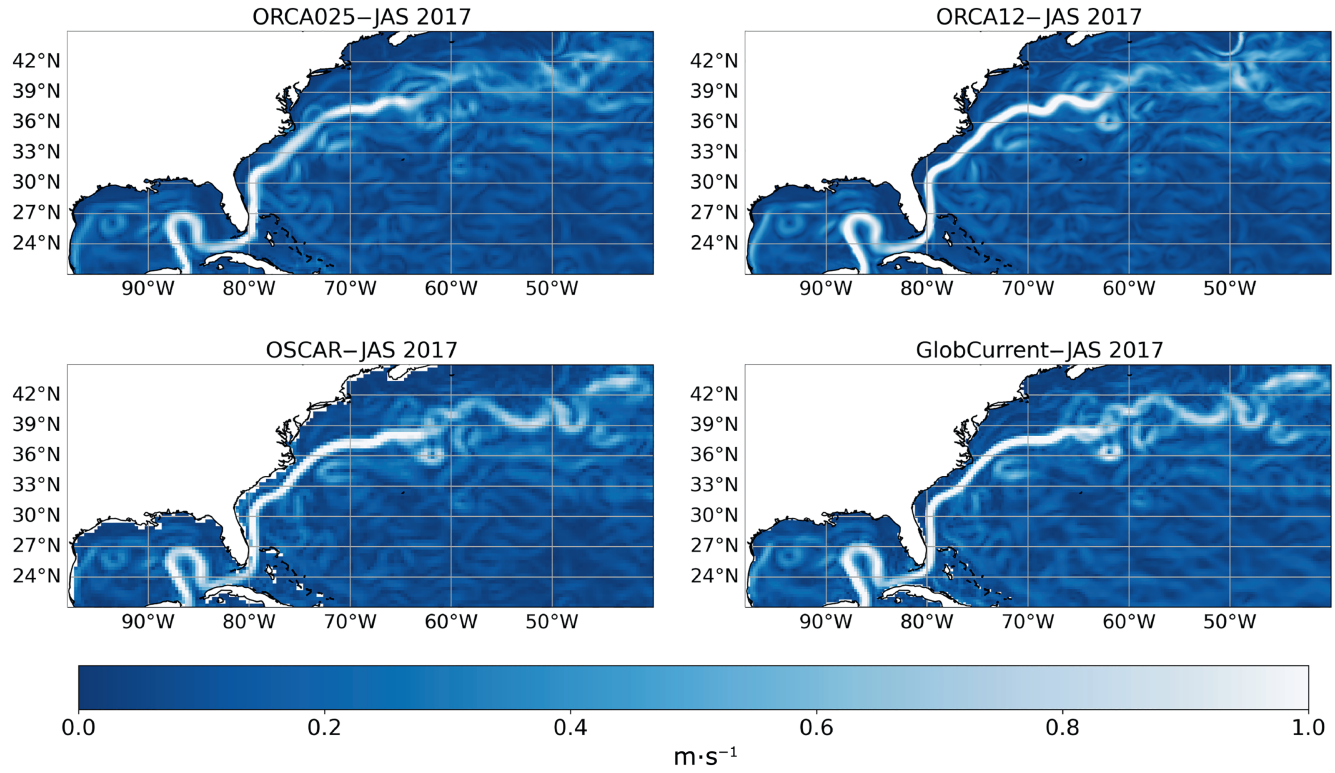


**FIGURE 7** Gulf Stream surface speed averaged for July to September (JAS) 2017: (top left) ORCA025, (top right) ORCA12, and observations from (bottom left) Ocean Surface Current Analysis Real-time (OSCAR) and (bottom right) GlobCurrent. [Colour figure can be viewed at wileyonlinelibrary.com]

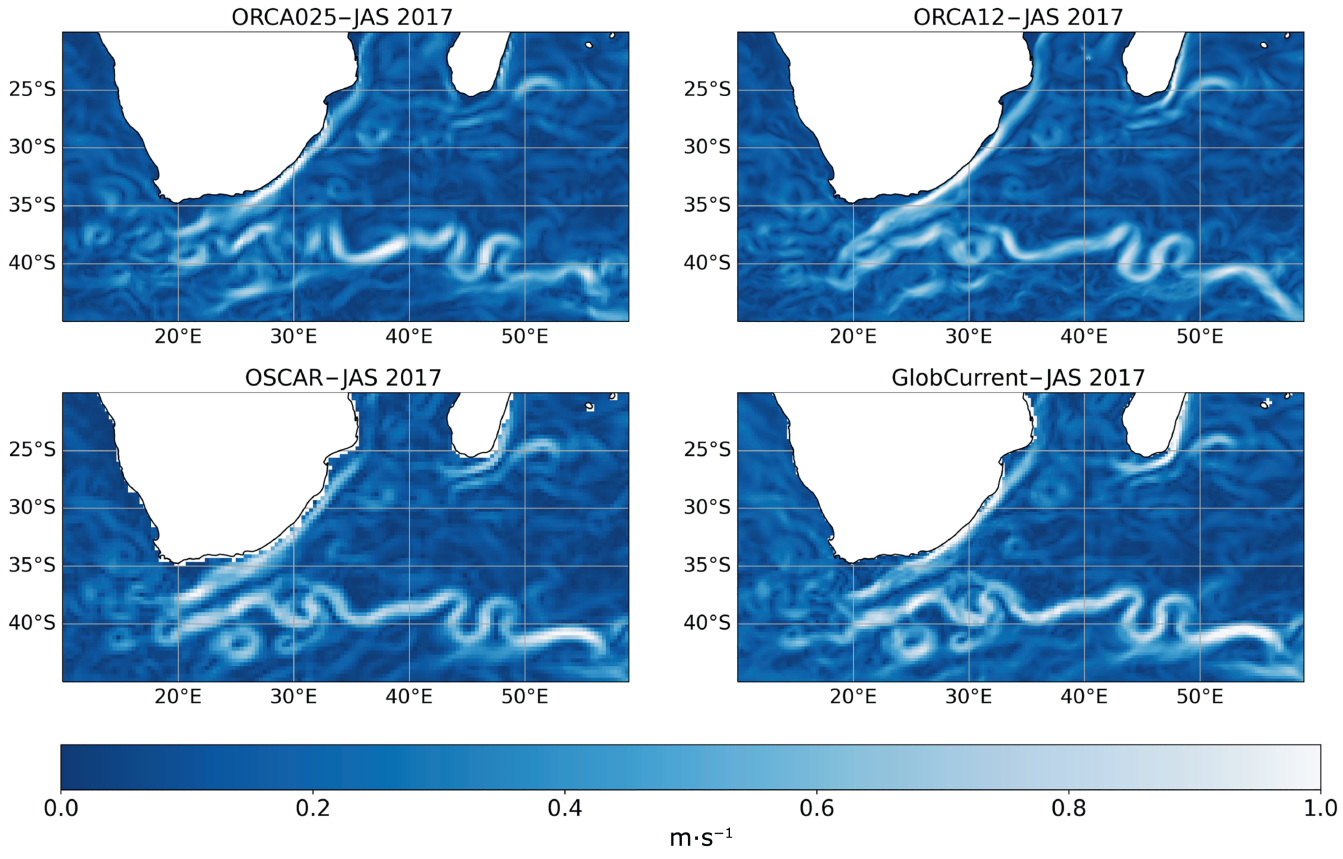


**FIGURE 8** As for Figure 7, but showing the Agulhas Current. [Colour figure can be viewed at wileyonlinelibrary.com]

component of the flow. OSCAR data are averaged over 5 days and presented on a 1/3° latitude–longitude grid, whereas GlobCurrent provides daily and monthly average fields on a 1/4° regular grid.

In general, both ORCA025 and ORCA12 represent the main characteristics of each region well. However, ORCA12 exhibits more details in agreement with observations and displays more eddies than ORCA025. A good example of that can be seen in Figure 7, which shows a prominent eddy south of the Gulf Stream (36° N, west of 60° W) that is very clear in ORCA12 but much weaker in ORCA025. Another example is the Agulhas Current retroflection (near 40° S, 20° E), which appears well defined in both ORCA12 and OSCAR and is also visible in GlobCurrent but is very fragmented in ORCA025 (Figure 8). Still in the Agulhas Current, both OSCAR and GlobCurrent fields show two eddies south of 40° S and 20°–30° E plus a meander at 37°–40° S and 30°–35° E that are roughly captured in ORCA12 but not in ORCA025.

We have also compared the zonal and meridional components of near-surface velocity against drifter observations by interpolating the daily mean model currents with the locations of the drifter-based velocity observations. For this purpose, and to avoid double-penalty effects, the output of ORCA12 at 15 m depth was regridded onto the ORCA025 grid, before interpolating to the locations of the drifter-derived velocities (Blockley *et al.*, 2012). Results for the two configurations are very similar for both velocity components (not shown), with larger mean differences in the zonal component and over the Southern Ocean in particular.

### 6.2.2 | EKE

Specific EKE provides a good proxy to measure the degree of eddy activity by quantifying the anomalies of the velocity field with respect to a mean value. We have computed the specific EKE as $[(u-\bar{u})^2+(v-\bar{v})^2]/2$, where $u$ and $v$ are daily means of the zonal and meridional components of surface velocity respectively, and $\bar{u}$ and $\bar{v}$ are their respective monthly averaged values. Figure 9 shows the mean EKE in 2018 as derived from GlobCurrent daily surface observational data, how it compares with the mean EKE from each model configuration, and also the difference between the two configurations. Not surprisingly, the model EKE values tend to be higher than GlobCurrent along boundary currents, the equatorial region, and ACC, which are regions largely characterised by turbulent flow (Figure 9b,c). This is probably due to the limited temporal and spatial sampling of GlobCurrent. ORCA12 is the most energetic overall, but ORCA025 exhibits higher EKE in some locations, notably in parts of the ACC and

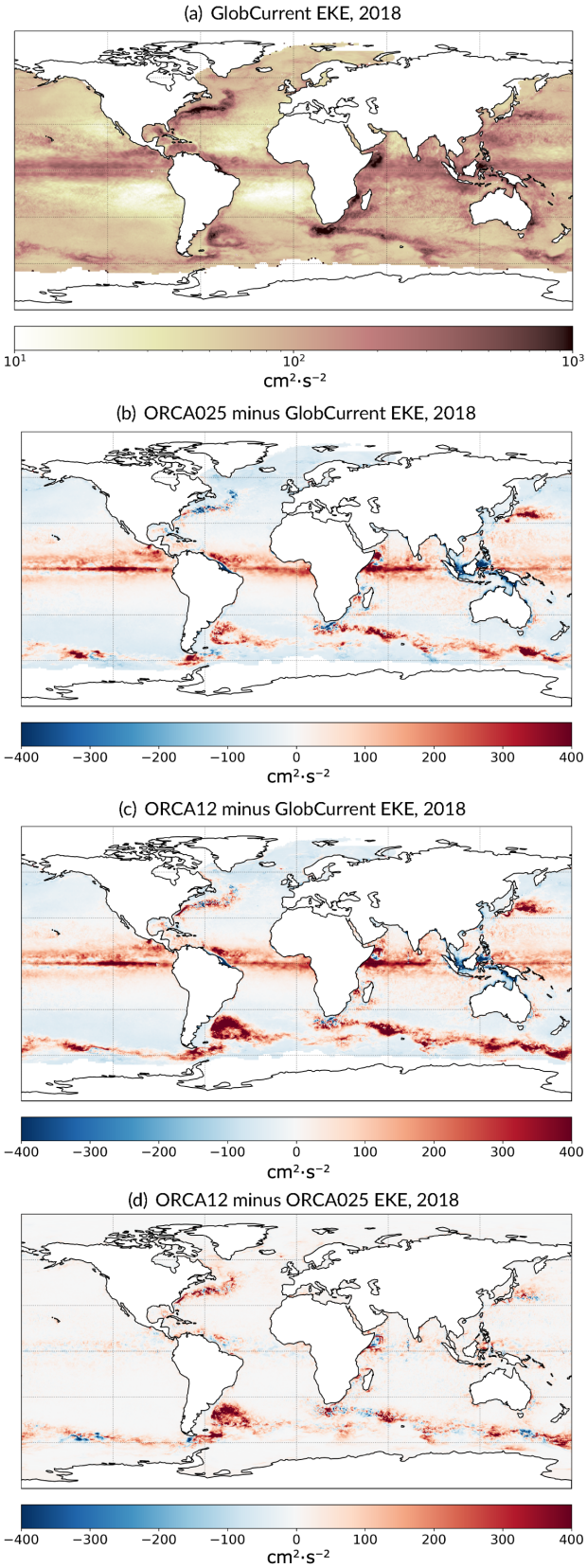


**FIGURE 9** Annual mean specific surface eddy kinetic energy (EKE) in 2018 for (a) GlobCurrent, (b) difference between ORCA025 and GlobCurrent, (c) difference between ORCA12 and GlobCurrent, and (d) difference between ORCA12 and ORCA025. [Colour figure can be viewed at wileyonlinelibrary.com]

near the coast of Somalia (Figure 9d). Elsewhere, in the interior of all the ocean basins (more-quiescent regions), there is widespread underestimation of surface EKE by the models.

For a quantitative comparison of all three datasets, zonal mean EKE estimates (i.e., longitude-averaged monthly mean EKE over 2017–2018) are plotted as a function of latitude in Figure 10. The comparison of zonal mean EKE in the three datasets highlights how ORCA12 outperforms ORCA025 in modelling the Gulf Stream and the Kuroshio Current, getting closer to the local maximum of 200 $cm^2 \cdot s^{-2}$ displayed by the GlobCurrent data

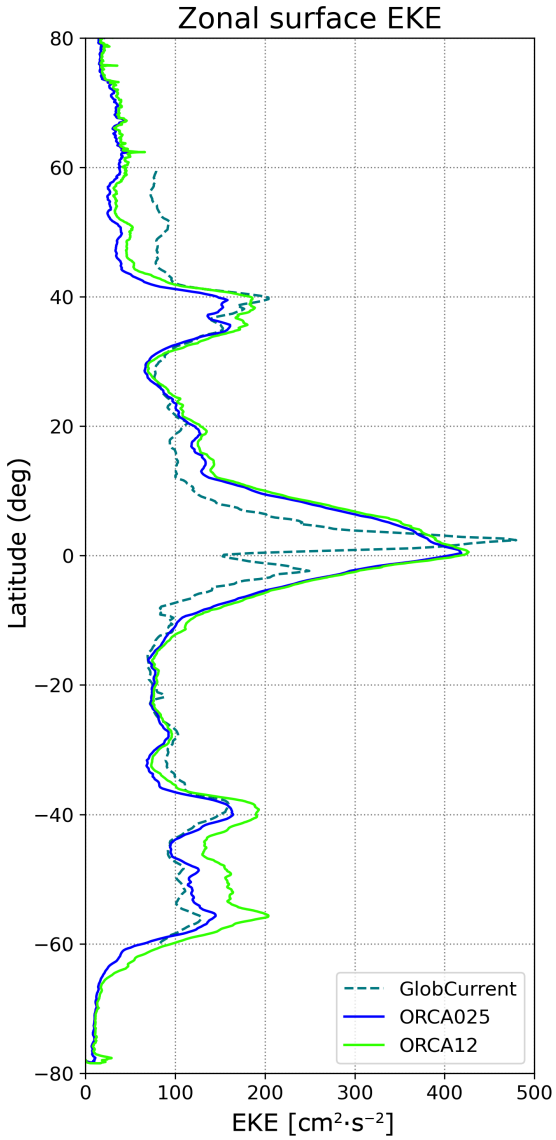


**FIGURE 10** Zonal mean specific eddy kinetic energy (EKE) averaged over 2017–2018 for ORCA12 (solid light green), ORCA025 (solid dark blue), and GlobCurrent (dashed dark green). GlobCurrent data poleward of 60° latitude are discarded here because data coverage is sparse at high latitudes (see Figure 9a). [Colour figure can be viewed at wileyonlinelibrary.com]

at 30°–40° N. North of 40° N, the EKE in both models is about half of the GlobCurrent estimate, probably due to inadequate numerical representation of dynamical processes in the northeast Atlantic Ocean and Northwest European Shelf. Just north of the Equator, zonal GlobCurrent EKE peaks at almost 500 $cm^2 \cdot s^{-2}$, in two locations: southwest of Somalia and west of Indonesia (Figure 9a). We nevertheless note that GlobCurrent has higher errors in equatorial regions, where its geostrophic approximation becomes increasingly invalid.[4] Both models capture the strong equatorial currents as a broad asymmetric ±5° jet shifted slightly north of the Equator, with EKE values between 200 and 400 $cm^2 \cdot s^{-2}$ (Figure 10). ORCA025 and GlobCurrent share similar EKE values in the Southern Hemisphere but not in the Northern Hemisphere. Between 40° S and 50° S, ORCA12's zonal EKE is surprisingly large in comparison with GlobCurrent but in remarkable agreement with the findings of Griffies *et al.* (2015).

### 6.2.3 | Sea-ice volume

The seasonal cycle of the integrated sea-ice volume for each hemisphere, comparing analysed model values (following DA) with PIOMAS reanalyses, is shown in Figure 11. The total volume is computed by multiplying the mean ice thickness of each individual grid cell by its area and summing over the target region. The volume is an indicator of the integrated sea-ice thickness since the sea-ice area is largely constrained by the assimilation of SIC in our experiments. Both models agree well, with ORCA12 showing a tendency for somewhat thicker sea ice during the Arctic melt season from July to September. In the Arctic, both configurations are comparable to the reference reanalysis of PIOMAS, albeit with some differences, such as FOAM having more sea ice than PIOMAS during the winter and less sea ice during the summer. There is no equivalent reference reanalysis product for Antarctic sea ice, but it is known that the ACC is not correctly reproduced in either ORCA025 or ORCA12 (Roberts *et al.*, 2019), and this is expected to have an adverse impact on sea-ice predictions.

Arctic sea-ice volume forecasts are further assessed for July 2018, during the melt season. Figure 12 shows analyses and 5-day forecasts for sea-ice volume and area (aggregate ice fraction multiplied by the grid cell area and summed over the region) starting from each daily analysis. It is apparent that ORCA025 loses sea-ice volume more quickly than ORCA12 does. Although the sea-ice area is similar following SIC assimilation, ORCA025 forecasts a very rapid decrease in sea-ice area in comparison with ORCA12 during this Arctic melting month. This is

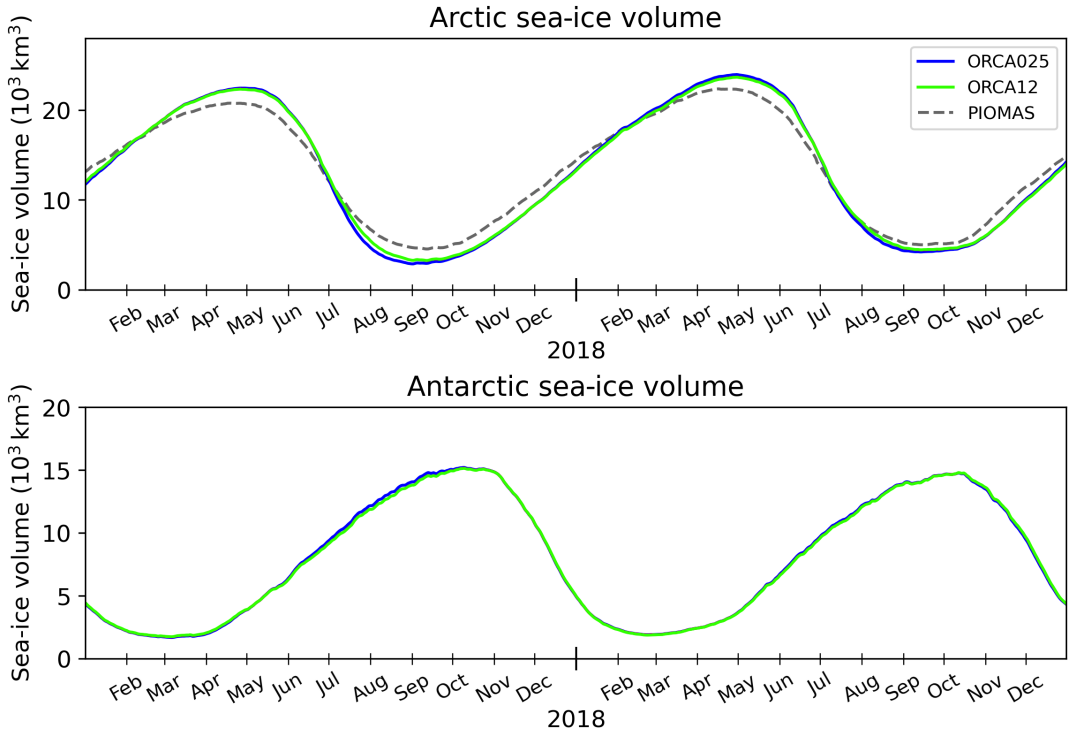


**FIGURE 11** Time series of total sea-ice volume analyses for ORCA025 (solid dark blue) and ORCA12 (solid light green) following data assimilation (DA) in the Arctic (top) and Antarctic (bottom). Pan-Arctic Ice Ocean Modeling and Assimilation System reanalysis data are represented in dashed grey. [Colour figure can be viewed at wileyonlinelibrary.com]

explained by a discrepancy in the total melt pond area, which peaks at around $2 \times 10^5$ km$^2$ for ORCA12, an order of magnitude below the 2017–2018 maximum of nearly $3 \times 10^6$ km$^2$ for ORCA025, as shown in Figure 13. The latter is much closer to an estimate of $2 \times 10^6$ km$^2$ based on moderate-resolution imaging spectro-radiometer observations (Rösel *et al.*, 2012). The total melt pond area was estimated as the sum over all grid cells (in the Northern Hemisphere) of melt pond effective area fraction multiplied by grid cell area. Having lower melt pond area means that ORCA12 will have a higher albedo with reduced absorption of solar radiation, and hence lower sea-ice melt. Despite the improved representation of sea-ice melting, there appears to be an issue in the representation of melt ponds in this ORCA12 configuration. The rate of sea-ice melting in ORCA025 over the 5-day forecast (Figure 12) is nevertheless unrealistic, as shown by the large difference between forecast and analysed values from subsequent cycles.

## 6.3 | Forecast verification

Assessments of forecast accuracy typically compare model forecasts with observations. On global forecasting scales, the traditional approach is to use the bias, mean absolute error (MAE), RMS error (RMSE), and anomaly correlation of large-scale forecast fields to assess forecast accuracy. A common procedure in the ocean community is to use the OceanPredict class 4 methodology (Hernandez *et al.*, 2015), which compares observations and forecasts in observation space using a common set of observations to enable comparison of models from different systems (Aijaz *et al.*, 2023; Divakaran *et al.*, 2015; Ryan *et al.*, 2015).

Figure 14 shows class 4 results obtained from the trial period July to December 2018 for both SST and SLA daily mean forecasts with lead times of 12, 36, 60, 84, and 108 hr. The 3DVar-FGAT assimilation window is from $T-24$ to $T+0$, so $T-12$ is chosen as being a representative time. Of note is the clear trend of ORCA025 having lower RMSE than ORCA12 for all lead times (Figure 14). The SST mean error for ORCA12 is marginally smaller than for ORCA025 and improves with increasing forecast lead time, whereas for SLA there is no noticeable difference between the mean errors, which remain largely unchanged with increasing forecast lead time, possibly due to error cancellation. SST MAE displays similar behaviour to that of RMSE. The forecast error grows at the same rate for both ORCA025 and ORCA12. OceanPredict intercomparisons over a more recent period (not shown) reveal significant differences between different systems in the error growth rate over the first 36 hr. This is perhaps due to the DA fitting the observations in different ways across the independent systems, which is not the case

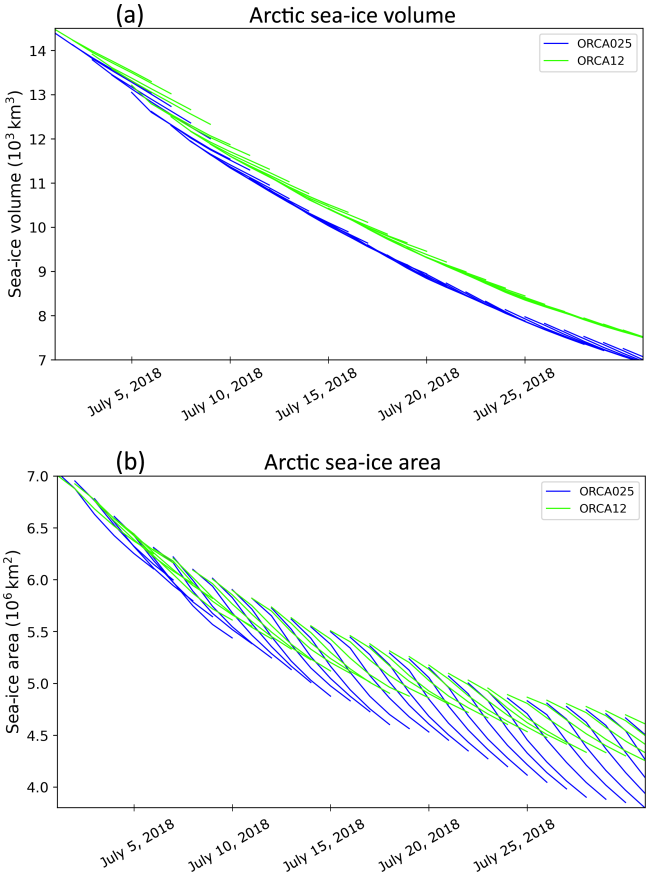


**FIGURE 12** Time series of total sea-ice (a) volume and (b) area in the Arctic for ORCA025 (dark blue) and ORCA12 (light green) over July 2018. Individual lines correspond to daily 5-day forecasts. [Colour figure can be viewed at wileyonlinelibrary.com]

here. However, beyond $T + 36$, growth rates in Ocean-Predict intercomparisons appear very similar for most systems with a mix of resolutions: 1/4°, 1/12°, and 1/16°. We think this is because the statistics are aggregated over a large area. If there was sufficient observational coverage to focus, for example, on Western Boundary Current systems, we might see the impact of different resolutions reflected in different rates of error growth.

ORCA025 SST has lower RMSE than ORCA12 for the majority of the 6-month period for which 5-day forecasts were produced, with quite clear differences in the early and latter stages (Figure 15). Both configurations have a similar warm bias throughout. For SLA, ORCA025 consistently displays lower RMSE than ORCA12 throughout the full 6 months. Both models have similar biases, with a tendency towards forecast values being too low in the final 3 months. Note that the results presented in Figure 15 are for a 3-day forecast (having a lead time of 60 hr relative to the $T - 12$ nominal analysis time), but equivalent results have been produced for other lead times (not shown).

Although the higher resolution ORCA12 forecasts may be better at resolving coastlines, eddies, and bathymetric features, and qualitatively appear more realistic, Figures 14 and 15 illustrate that RMS measures can still show them to be apparently inferior to, or show no obvious improvement over, ORCA025. This is because a lower resolution model produces smoother forecast fields with less variability, leading to smaller errors when measures such as RMSE are applied (Mass *et al.*, 2002). Increasing the resolution makes the model more susceptible to the double-penalty effect (Rossa *et al.*, 2008), where a given detail is forecast to be at the wrong time and/or place, yielding two mismatching contributions to the RMSE. The development of more-appropriate techniques for comparing forecasts from different resolution models and

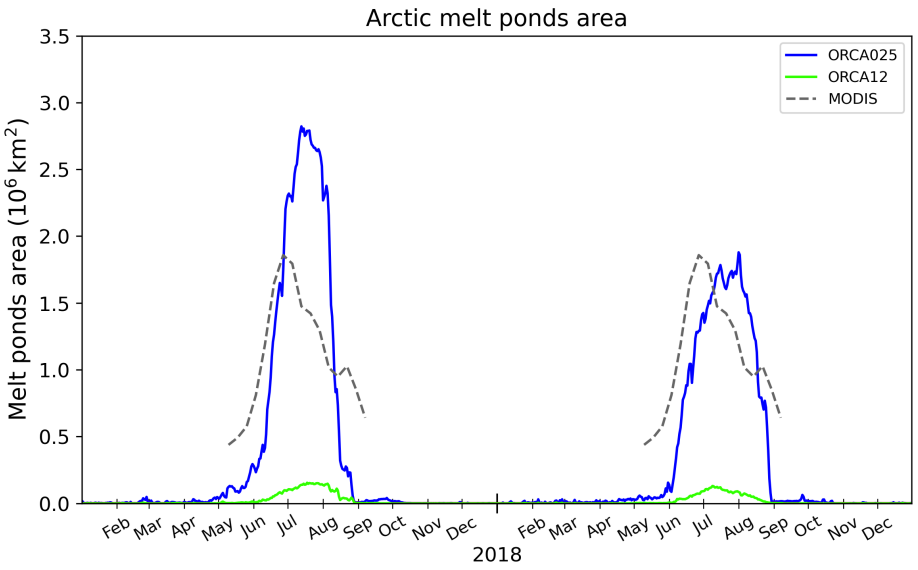


**FIGURE 13** Time series (2017–2018) of the total area of radiatively active melt ponds in the Arctic Ocean: ORCA12 (solid light green), ORCA025 (solid dark blue), and moderate-resolution imaging spectro-radiometer (MODIS) observations (dashed grey). The latter is a multiyear climatology of melt pond area based upon MODIS observations from 2000 to 2011 (Rösel *et al.*, 2012). [Colour figure can be viewed at wileyonlinelibrary.com]

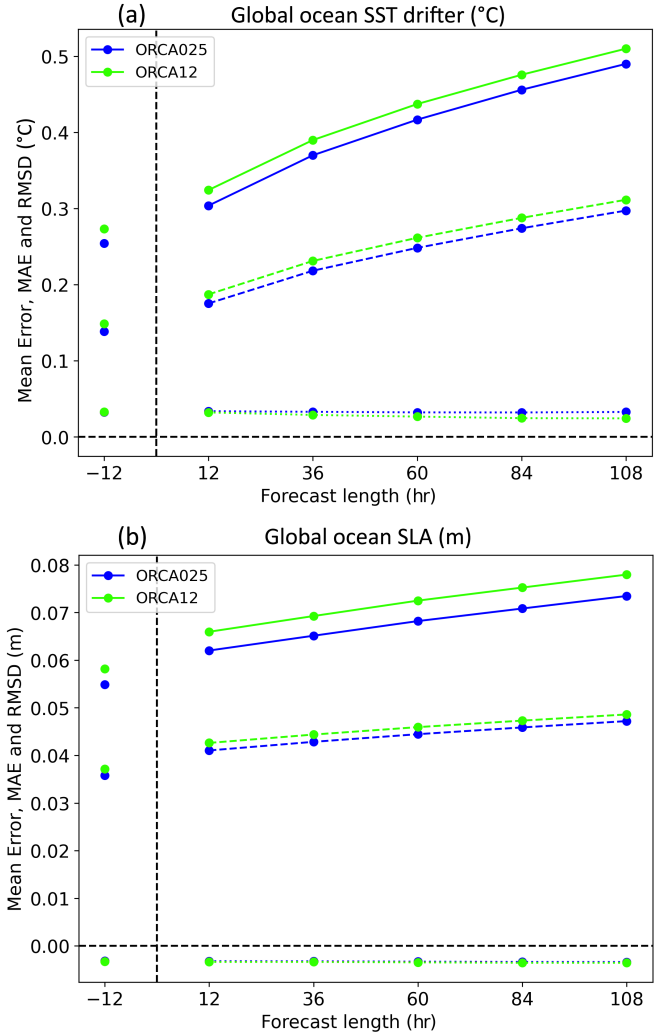


**FIGURE 14** Global ocean (a) sea-surface temperature (SST, °C) drifter and (b) sea-level anomaly (SLA, m) errors ("model minus observation", averaged over July–December 2018) against forecast lead time for ORCA12 (light green) and ORCA025 (dark blue), where −12 represents the nominal analysis time for the assimilation window spanning $T - 24$ to $T + 0$. Dotted, dashed, and solid lines indicate mean error, mean absolute error (MAE), and root-mean-square deviation (RMSD) respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

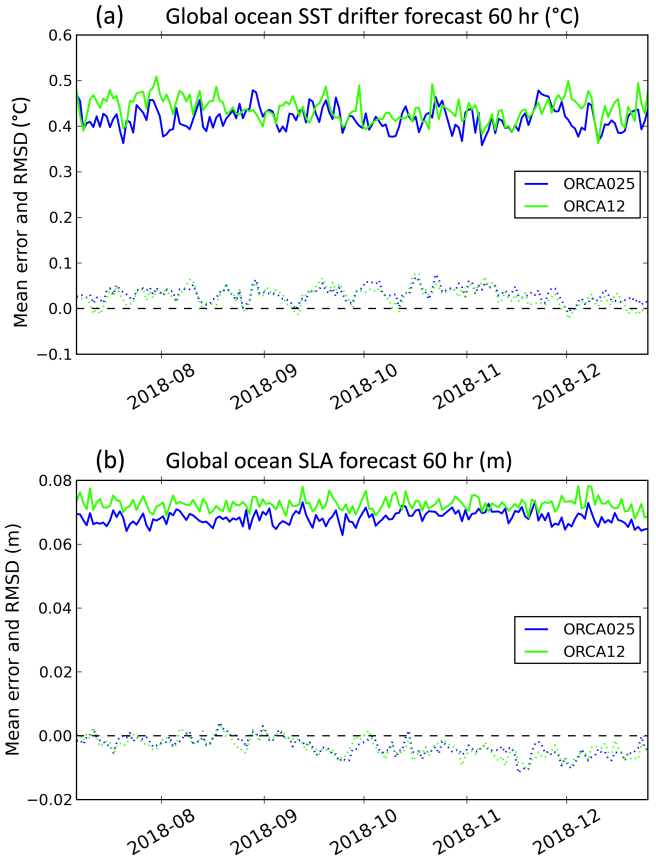


**FIGURE 15** Global ocean (a) sea-surface temperature (SST, °C) drifter and (b) sea-level anomaly (SLA, m) forecast 60 hr time series of "model minus observation" differences for 3-day ($T + 60$ hr) forecasts from ORCA12 (light green) and ORCA025 (dark blue). Both figures show the trend of RMSE (solid lines) and mean error (dotted lines) for July–December 2018, with daily 5-day forecasts. [Colour figure can be viewed at wileyonlinelibrary.com]

demonstrating forecast skill and reliability at the local level has been an active field of research in the atmospheric verification community (Ebert, 2008; Gilleland *et al.*, 2009, 2010), and several new spatial verification methods have been developed recently (Dorninger *et al.*, 2018).

## 6.4 | Neighbourhood verification

Neighbourhood verification can account for the double-penalty effect by regridding forecasts from different resolutions onto a common grid. This approach has gained popularity in the atmospheric community over recent years for both deterministic and ensemble forecasts and has recently been trialled with high-resolution SST forecasts on the European Northwest Shelf (Crocker *et al.*, 2020).

The forecast assessments undertaken here are focussed on location-specific verification using the HiRA spatial method (Mittermaier, 2014), where forecast grid points within neighbourhoods centred on an observing location are treated as being pseudo-ensemble members, so that typical ensemble and probabilistic forecast verification metrics can be used to verify the forecast. An assumption is made that the observation is not only valid at the location at which it is measured but is representative of the surrounding area as well. The CRPS is then applied to quantify the difference between two cumulative distributions, consisting of a forecast distribution formed by ranking the pseudo-ensemble members in a given neighbourhood and a step function describing the observed state. CRPS is a negatively oriented score, in that smaller scores are better,

with zero being perfect. Further information on the CRPS formulation can be found in Hersbach (2000) and Crocker *et al.* (2020).

Figure 16a summarises the HiRA results for SST. For ORCA12, we show results for 0.083° (1 × 1 grid point), 0.25° (3 × 3), 0.75° (9 × 9), and 1.25° (15 × 15) neighbourhoods. For ORCA025, the neighbourhoods start at 0.25° (1 × 1), proceeding to 0.75° and 1.25° for equivalent numbers of grid points (3 × 3 and 5 × 5). To make this clearer, matching line styles have been used for comparable neighbourhood sizes, as shown on the plot legends. At the grid scale (0.25° for ORCA025 and 0.083° for ORCA12), CRPS reduces to the MAE, and ORCA025 has lower scores than ORCA12 across all lead times, consistent with the traditional class 4 metrics shown in Figure 14. At the equivalent grid scale (e.g., neighbourhood size 0.25°), there is a large improvement in the CRPS for ORCA12, which now has lower values than ORCA025. The first comparable neighbourhoods capable of resolving common features are the 3 × 3 ORCA025 and 9 × 9 ORCA12 neighbourhoods, both 0.75°, where it is again evident that ORCA12 has a lower CRPS than ORCA025 across all lead times. As the equivalent neighbourhood sizes increase, ORCA12 generally has lower CRPS than ORCA025, although the differences between the two reduce gradually as the detail becomes increasingly smoothed out. However, for the forecast lead time of 12 hr (a 1-day forecast after the $T - 12$ nominal analysis time) at the next equivalent neighbourhood size (1.25°), the SST CRPS for both ORCA025 and ORCA12 has started to decline in relation to the previous 0.75° neighbourhood, suggesting that at this scale the SST observation no longer represents this larger neighbourhood. All CRPS values degrade (become larger) with increasing lead time for both ORCA025 and ORCA12, as found with traditional RMS metrics.

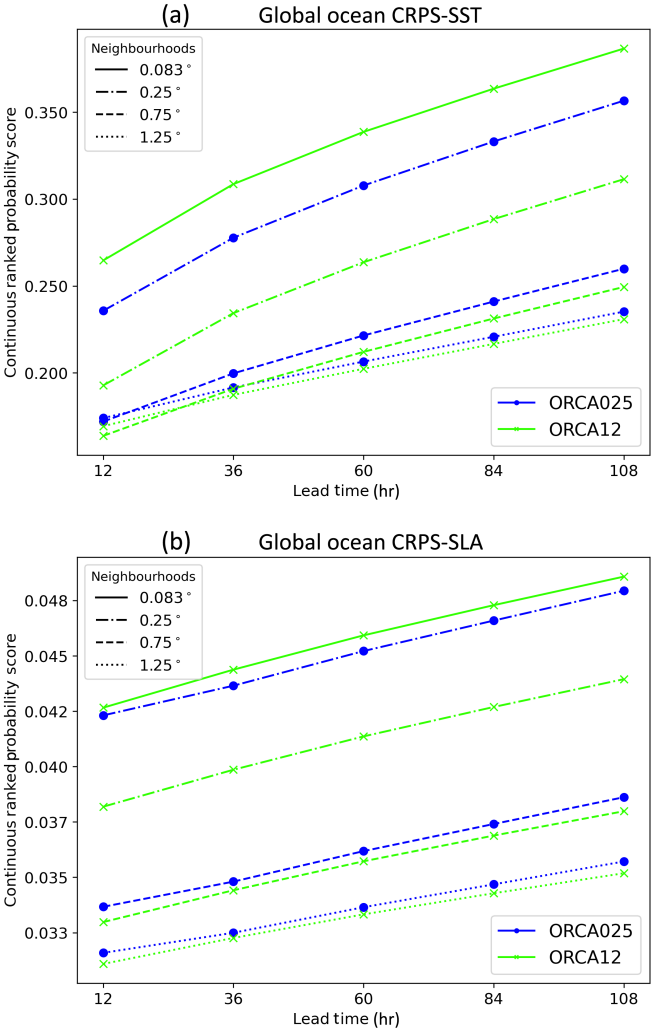


**FIGURE 16** Continuous ranked probability scores (CRPSs) for global ocean (a) sea-surface temperature (SST) and (b) sea-level anomaly (SLA) over July–December 2018, with daily 5-day forecasts, for ORCA12 (light green) and ORCA025 (dark blue). Matching line styles correspond to neighbourhoods with equivalent areas. The legend refers to the grid scale of the neighbourhood defined as a latitude–longitude square. Thus, for example, a grid scale of 0.75° corresponds to the ORCA12 neighbourhood of 9 × 9 = 81 cells and is comparable to the ORCA025 neighbourhood of 3 × 3 = 9 cells. [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 16b shows HiRA results for SLA using the same neighbourhood sizes used to assess SST. At equivalent grid scales, the ORCA12 CRPS is lower than that of ORCA025, and at the first equivalent neighbourhood size for both configurations (0.75°) the ORCA12 neighbourhood has lower CRPS than the equivalent ORCA025 across all lead times. As the neighbourhood sizes increase, ORCA12 generally has lower CRPS than ORCA025 (comparing matching line styles). Again, all CRPSs degrade linearly (become larger) with increasing lead time for both ORCA025 and ORCA12.

## 7 | CONCLUSIONS AND FUTURE WORK

In anticipation of a future upgrade to the ocean and sea ice model resolutions in the Met Office's operational coupled atmosphere/land/ocean/sea ice forecast model, which currently uses a 1/4° ORCA025 grid, we have tested a GO6/GSI8.1 configuration in a forced ocean–sea-ice forecast model driven by uncoupled atmospheric surface boundary conditions using a 1/12° ORCA12 grid. In so doing, we have kept assimilating observations using an ORCA025 analysis grid and have taken this opportunity to provide an updated description of our ocean DA set-up. We have compared the performance of the ORCA025 control and ORCA12 experiment using traditional RMS measures based on the differences between 1-day forecasts and

observations (innovations) and class 4 metrics applied to 5-day forecasts. In addition, we have used neighbourhood verification metrics that allow a fairer comparison of forecasts produced using models run at different resolutions. Application of traditional metrics such as the mean error, RMSE, and MEA have been unable to clearly demonstrate the superiority of ORCA12 over the coarser ORCA025, despite ORCA12 qualitatively having the ability to better resolve features of interest, such as eddies.

Statistically, the performance of the two systems is comparable for SST, but ORCA12 shows some increase in RMS innovations for temperature and salinity profiles, and even larger increases for SLAs and SIC. This could be partly due to the use of background-error covariances that were tuned for the ORCA025 system or the double-penalty effect associated with traditional performance measures. Upgrading the set-up of the ORCA12 DA to use background-error covariances based on 1/12° forecasts might be expected to give a small improvement in comparison with ORCA025, as shown by Lellouche *et al.* (2021). Unsurprisingly, the less favourable results for ORCA12 analyses lead to larger forecast errors overall. We think that the reason for the large degradation in RMS SLA innovations is mainly due to the smoother bathymetry used in the ORCA025 resolution DA, which has significant differences compared with the bathymetry of the ORCA12 model. Also, the ORCA12 configuration has not yet had the benefit of years of development and tuning that have gone into ORCA025. Further tuning of the ORCA12 model settings could also improve the performance of this configuration.

Analysis of surface currents and EKE favours ORCA12 as it tends to represent better strong currents such as the Gulf Stream, Agulhas, and Kuroshio currents. ORCA12 exhibits very large zonal EKE values over the Zapiola anticyclone, which also appears as a conspicuous region with high levels of RMS SLA innovations. Given the strong influence of topography on the dynamics over the Zapiola anticyclone, this region is expected to be better represented in ORCA12 and is a potential candidate for a case study of the benefit of increasing the resolution of the DA in ORCA12.

Integrated sea-ice volumes (and sea-ice thickness) at analysis time are comparable for ORCA025 and ORCA12, with ORCA12 being slightly better in the Arctic summer. Nevertheless, both configurations lose sea ice too quickly during 5-day forecasts (ORCA025 more so than ORCA12), and there are almost no melt ponds in the reported ORCA12 configuration. The latter issue is particularly significant because melt ponds play a critical role in radiative balance by changing the sea-ice albedo, prompting ongoing investigations.

Results obtained from applying the HiRA neighbourhood verification method have shown that the higher resolution ORCA12 has lower errors than ORCA025 (as measured by the CRPS) when equivalent neighbourhood sizes are compared, emphasising the importance of this technique when comparing models run at different resolutions. This approach also has the advantage of being extensible to evaluating the performance of a future ocean and sea-ice ensemble forecasting system (Lea *et al.*, 2022).

Following these initial results, we intend to recalibrate the error covariances used in NEMOVAR DA to better represent ORCA12 forecasts and to also investigate the performance of running DA with an ORCA12 grid, recognising that performance enhancements may be required to make this fast enough for future operational implementation. At the time of running the experiments reported here, our research workflow was still configured to match the 24-hr cycling and assimilation window of the Met Office's 1/4° operational ocean–sea-ice model. Additional flexibility has since been introduced to allow other cycling periods, allowing future ocean–sea-ice development experiments to match the 6-hr cycling and assimilation window of our operational coupled atmosphere–land–ocean–sea-ice model.

The next planned major upgrade to the system will introduce the global ocean and sea- ice configuration GOSI9, based on the updated ocean model configuration NEMO v4, including the thermodynamic equation of seawater TEOS-10 equation of state (Pawlowicz *et al.*, 2012), and changing to use the Sea Ice Modelling Integrated Initiative sea-ice model. We also expect to update the mean dynamic topography configuration used for SSH radar altimeter assimilation.

In conclusion, the adoption of a 1/12° ORCA12 grid has a mixed impact on the performance of our global ocean and sea-ice forecasting system, with some improvements demonstrated using neighbourhood verification methods, and has the potential to benefit our operational coupled atmosphere–land–ocean–sea-ice forecast model through improved representation of high-resolution features.

## ACKNOWLEDGEMENTS

This work was conducted through the Weather and Climate Science for Service Partnership (WCSSP) India, a collaborative initiative between the Met Office, supported by the UK Government's Newton Fund, and the Indian Ministry of Earth Sciences (MoES). We also acknowledge funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 821926 (IMMERSE). Additionally, we have benefited from configurations developed by the Joint Marine Modelling Programme, a partnership between the Met Office, National

Oceanography Centre, British Antarctic Survey, and the Centre for Polar Observation and Modelling.

We acknowledge Anthony Weaver from the Centre Européene de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS) and other NEMOVAR partners at ECMWF and the Institut National de Recherche en Sciences et Technologies du Numérique (INRIA) for developing improvements to the NEMOVAR software. We are also grateful to Doroteaciro Iovino from the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) for assistance configuring the XIOS I/O server for early work to run experiments at ECMWF and to David Storkey, Catherine Guiavarc'h, and Jonah Roberts-Jones from the Met Office for their insight and support.

## DATA AVAILABILITY STATEMENT

Met Office Hadley Centre temperature and salinity quality controlled profile observations EN.4.2.1.f.profiles.g10 were downloaded from https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html in June 2019 and assimilated in the simulations; EN.4.1.1.f.analysis.g10 for years 2011–2015 were downloaded prior to March 2016 and used to produce a reference temperature and salinity climatology. EUMETSAT, Global Drifter Program, and Ocean and Sea Ice Satellite Application Facility data were obtained from https://user.eumetsat.int/data-access/eumetcast-europe, https://www.aoml.noaa.gov/phod/gdp/index.php, and https://osi-saf.eumetsat.int/products/sea-ice-products respectively. The OSCAR 1/3° version 1 PO.DAAC dataset was accessed on April 29, 2020, at https://doi.org/10.5067/OSCAR-03D01. University of Hamburg weekly averages of Arctic sea-ice area are available from https://www.cen.uni-hamburg.de/en/icdc/data/cryosphere/arctic-meltponds.html. This study has been conducted using E.U. Copernicus Marine Service Information: https://doi.org/10.48670/moi-00146 (altimetry data) and https://doi.org/10.48670/moi-00050 (GlobCurrent data).

## ENDNOTES

[1]The multidimensional variables are rearranged and concatenated into a one-dimensional vector.

[2]**H** is closely related to the nonlinear $H$ in Equation (1), but in 3DVar-FGAT **H** ignores any time component and so is not the Jacobian matrix of $H$. Although the observations in **y** occur at specific times within the assimilation window, the innovations are treated as being valid at the middle of the window.

[3]For comparison, Zuo *et al.* (2017) balance the global freshwater budget by applying a spatially uniform freshwater flux calculated using the difference between the change in SSH from the altimeter observations and the steric change in SSH derived from the model's density fields as a result of warming and salinity changes.

[4]See fig. 3 of https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-MOB-QUID-015-004.pdf.

## ORCID

*Ana Barbosa Aguiar* https://orcid.org/0000-0002-1601-7419
*Michael J. Bell* https://orcid.org/0000-0003-1530-6371
*Edward Blockley* https://orcid.org/0000-0002-0489-4238
*Gordon Inverarity* https://orcid.org/0000-0001-8433-0004
*Robert King* https://orcid.org/0000-0002-9573-2567
*Daniel J. Lea* https://orcid.org/0000-0003-1736-628X
*Matthew J. Martin* https://orcid.org/0000-0003-0293-3106
*Martin R. Price* https://orcid.org/0009-0007-2343-5129
*John Siddorn* https://orcid.org/0000-0003-3848-8868
*James While* https://orcid.org/0000-0001-9870-8725

## REFERENCES

Aijaz, S., Brassington, G.B., Divakaran, P., Régnier, C., Drévillon, M., Maksymczuk, J. et al. (2023) Verification and intercomparison of global ocean Eulerian near-surface currents. *Ocean Modelling*, 186, 102241. Available from: https://doi.org/10.1016/j.ocemod.2023.102241

Amante, C. & Eakins, B.W. (2009) ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis. Technical Memorandum NESDIS NGDC-24, NOAA https://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/docs/ETOPO1.pdf

Barnier, B., Madec, G., Penduff, T., Molines, J.-M., Treguier, A.-M., Le Sommer, J. et al. (2006) Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dynamics*, 56, 543–567. Available from: https://doi.org/10.1007/s10236-006-0082-1

Barton, N., Metzger, E.J., Reynolds, C.A., Ruston, B., Rowley, C., Smedstad, O.M. et al. (2021) The navy's earth system prediction capability: a new global coupled atmosphere-ocean-sea ice prediction system designed for daily to subseasonal forecasting. *Earth and space. Earth and Space Science*, 8, e2020EA001199. Available from: https://doi.org/10.1029/2020EA001199

Bell, M.J., Forbes, R.M. & Hines, A. (2000) Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *Journal of Marine Systems*, 25, 1–22. Available from: https://doi.org/10.1016/S0924-7963(00)00005-1

Bell, M.J., Martin, M.J. & Nichols, N.K. (2004) Assimilation of data into an ocean model with systematic errors near the equator. *Quarterly Journal of the Royal Meteorological Society*, 130, 873–893. Available from: https://doi.org/10.1256/qj.02.109

Blockley, E.W., Martin, M.J. & Hyder, P. (2012) Validation of FOAM near-surface ocean current forecasts using Lagrangian drifting buoys. *Ocean Science*, 8, 551–565. Available from: https://doi.org/10.5194/os-8-551-2012

Blockley, E.W., Martin, M.J., McLaren, A.J., Ryan, A.G., Waters, J., Lea, D.J. et al. (2014) Recent development of the met Office operational ocean forecasting system: an overview and assessment

of the new global FOAM forecasts. *Geoscientific Model Development*, 7, 2613–2638. Available from: https://doi.org/10.5194/gmd-7-2613-2014

Bloom, S.C., Takacs, L.L., da Silva, A.M. & Ledvina, D. (1996) Data assimilation using incremental analysis updates. *Monthly Weather Review*, 124, 1256–1271. Available from: https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2

Bonjean, F. & Lagerloef, G.S.E. (2002) Diagnostic model and analysis of the surface currents in the tropical Pacific ocean. *Journal of Physical Oceanography*, 32, 2938–2954. Available from: https://doi.org/10.1175/1520-0485(2002)032<2938:DMAAOT>2.0.CO;2

Carneiro, D.M., King, R., Martin, M. & Aguiar, A. (2021) Short-range ocean forecast error characteristics in high resolution assimilative systems. Forecasting Research Technical Report 645, Met Office https://digital.nmla.metoffice.gov.uk/IO_e084c2c3-dc73-4cf3-acc1-44091ce6ef32

Carrère, L. & Lyard, F. (2003) Modeling the barotropic response of the global ocean to atmospheric wind and pressure forcing - comparisons with observations. *Geophysical Research Letters*, 30, 1275. Available from: https://doi.org/10.1029/2002GL016473

Chen, G. & Lin, H. (2000) The effect of temporal aliasing in satellite altimetry. *Photogrammetric Engineering and Remote Sensing*, 66, 639–644 https://www.asprs.org/wp-content/uploads/pers/2000journal/may/2000_639-644.pdf

Crocker, R., Maksymczuk, J., Mittermaier, M., Tonani, M. & Pequignet, C. (2020) An approach to the verification of high-resolution ocean models using spatial methods. *Ocean Science*, 16, 831–845. Available from: https://doi.org/10.5194/os-16-831-2020

de Miranda, A.P., Barnier, B. & Dewar, W.K. (1999) On the dynamics of the Zapiola anticyclone. *Journal of Geophysical Research*, 104, 21137–21149. Available from: https://doi.org/10.1029/1999JC900042

Divakaran, P., Brassington, G.B., Ryan, A.G., Régnier, C., Spindler, T., Mehra, A. et al. (2015) GODAE OceanView inter-comparison for the Australian region. *Journal of Operational Oceanography*, 8, s112–s126. Available from: https://doi.org/10.1080/1755876X.2015.1022333

Dorninger, M., Gilleland, E., Casati, B., Mittermaier, M.P., Ebert, E.E., Brown, B.G. et al. (2018) The set-up of the MesoVICT project. *Bulletin of the American Meteorological Society*, 99, 1887–1906. Available from: https://doi.org/10.1175/BAMS-D-17-0164.1

Ebert, E.E. (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, 15, 51–64. Available from: https://doi.org/10.1002/met.25

Flocco, D., Feltham, D.L. & Turner, A.K. (2010) Incorporation of a physically based melt pond scheme into the sea ice component of a climate model. *Journal of Geophysical Research*, 115, C08012. Available from: https://doi.org/10.1029/2009JC005568

Flocco, D., Schroeder, D., Feltham, D.L. & Humke, E.C. (2012) Impact of melt ponds on Arctic sea ice simulations from 1990 to 2007. *Journal of Geophysical Research*, 117, C09032. Available from: https://doi.org/10.1029/2012JC008195

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B. & Ebert, E.E. (2009) Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, 24, 1416–1430. Available from: https://doi.org/10.1175/2009WAF2222269.1

Gilleland, E., Ahijevych, D.A., Brown, B.G. & Ebert, E.E. (2010) Verifying forecasts spatially. *Bulletin of the American Meteorological Society*, 91, 1365–1376. Available from: https://doi.org/10.1175/2010BAMS2819.1

Good, S., Fiedler, E., Mao, C., Martin, M.J., Maycock, A., Reid, R. et al. (2020) The current configuration of the OSTIA system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sensing*, 12, 720. Available from: https://doi.org/10.3390/rs12040720

Good, S.A., Martin, M.J. & Rayner, N.A. (2013) EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118, 6704–6716. Available from: https://doi.org/10.1002/2013JC009067

Griffies, S.M., Winton, M., Anderson, W.G., Benson, R., Delworth, T.L., Dufour, C.O. et al. (2015) Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. *Journal of Climate*, 28, 952–977. Available from: https://doi.org/10.1175/JCLI-D-14-00353.1

Guiavarc'h, C., Roberts-Jones, J., Harris, C., Lea, D.J., Ryan, A. & Ascione, I. (2019) Assessment of ocean analysis and forecast from an atmosphere–ocean coupled data assimilation operational system. *Ocean Science*, 15, 1307–1326. Available from: https://doi.org/10.5194/os-15-1307-2019

Hallberg, R. (2013) Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Modelling*, 72, 92–103. Available from: https://doi.org/10.1016/j.ocemod.2013.08.007

Haney, R.L. (1971) Surface thermal boundary condition for ocean circulation models. *Journal of Physical Oceanography*, 1, 241–248. Available from: https://doi.org/10.1175/1520-0485(1971)001<0241:STBCFO>2.0.CO;2

Hernandez, F., Blockley, E., Brassington, G.B., Davidson, F., Divakaran, P., Drévillon, M. et al. (2015) Recent progress in performance evaluations and near real-time assessment of operational ocean products. *Journal of operational. Oceanography*, 8, s221–s238. Available from: https://doi.org/10.1080/1755876X.2015.1050282

Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15, 559–570. Available from: https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

Hewitt, H.T., Bell, M.J., Chassignet, E.P., Czaja, A., Ferreira, D., Griffies, S.M. et al. (2017) Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales? *Ocean Modelling*, 120, 120–136. Available from: https://doi.org/10.1016/j.ocemod.2017.11.002

Hollingsworth, A. & Lönnberg, P. (1986) The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field. *Tellus*, 38A, 111–136. Available from: https://doi.org/10.1111/j.1600-0870.1986.tb00460.x

Holt, J., Hyder, P., Ashworth, M., Harle, J., Hewitt, H.T., Liu, H. et al. (2017) Prospects for improving the representation of coastal and shelf seas in global ocean models. *Geoscientific Model Development*, 10, 499–523. Available from: https://doi.org/10.5194/gmd-10-499-2017

Huang, X.-Y., Mogensen, K.S. & Yang, X. (2002) First-guess at the appropriate time: the HIRLAM implementation and experiments. In HIRLAM Workshop on Variational Data Assimilation and Remote Sensing, Finnish Meteorological Institute, Helsinki, 21–23 January 2002, 28–43.

Hunke, E.C., Lipscomb, W.H., Turner, A.K., Jeffery, N. & Elliott, S. (2015) *CICE: the Los Alamos sea ice model documentation and software user's manual version 5.1*. Technical Report LA-CC-06-012. Los Alamos, NM: Los Alamos National Laboratory.

Ingleby, B. & Huddleston, M. (2007) Quality control of ocean temperature and salinity profiles - historical and real-time data. *Journal of Marine Systems*, 65, 158–175. Available from: https://doi.org/10.1016/j.jmarsys.2005.11.019

Large, W.G. & Yeager, S.G. (2009) The global climatology of an interannually varying air-sea flux data set. *Climate Dynamics*, 33, 341–364. Available from: https://doi.org/10.1007/s00382-008-0441-3

Lawless, A.S. (2010) A note on the analysis error associated with 3D-FGAT. *Quarterly Journal of the Royal Meteorological Society*, 136, 1094–1098. Available from: https://doi.org/10.1002/qj.619

Le Traon, P.Y., Nadal, F. & Ducet, N. (1998) An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technology*, 15, 522–534. Available from: https://doi.org/10.1175/1520-0426(1998)015<0522:AIMMOM>2.0.CO;2

Lea, D.J., Drecourt, J.-P., Haines, K. & Martin, M.J. (2008) Ocean altimeter assimilation with observational- and model-bias correction. *Quarterly Journal of the Royal Meteorological Society*, 134, 1761–1774. Available from: https://doi.org/10.1002/qj.320

Lea, D.J., While, J., Martin, M.J., Weaver, A., Storto, A. & Chrust, M. (2022) A new global ocean ensemble system at the met Office: assessing the impact of hybrid data assimilation and inflation settings. *Quarterly Journal of the Royal Meteorological Society*, 148, 1996–2030. Available from: https://doi.org/10.1002/qj.4292

LeGrand, P., Mercier, H. & Reynaud, T. (1998) Combining T/P altimetric data with hydrographic data to estimate the mean dynamic topography of the North Atlantic and improve the geoid. *Annales Geophysicae*, 16, 638–650. Available from: https://doi.org/10.1007/s00585-998-0638-0

Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R., Garric, G., Melet, A., Drévillon, M. et al. (2021) The Copernicus global 1/12° oceanic and sea ice GLORYS12 reanalysis. *Frontiers in Earth Science*, 9, 698876. Available from: https://doi.org/10.3389/feart.2021.698876

Lellouche, J.-M., Greiner, E., Le Galloudec, O., Garric, G., Régnier, C., Drévillon, M. et al. (2018a) Recent updates to the Copernicus marine service global ocean monitoring and forecasting real-time 1/12° high-resolution system. *Ocean Science*, 14, 1093–1126. Available from: https://doi.org/10.5194/os-14-1093-2018

Lellouche, J.-M., Greiner, E., Le Galloudec, O., Régnier, C., Benkiran, M., Testut, C.-E. et al. (2018b) The Mercator Ocean global high-resolution monitoring and forecasting system. In: Chassignet, E., Pascual, A., Tintoré, J. & Verron, J. (Eds.) *New Frontiers in operational oceanography*. Florida State University Libraries, USA. pp. 563–592. Available from: https://doi.org/10.17125/gov2018.ch20

Lorenc, A.C. (1986) Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112, 1177–1194. Available from: https://doi.org/10.1002/qj.49711247414

MacLachlan, C., Arribas, A., Peterson, K.A., Maidens, A., Fereday, D., Scaife, A.A. et al. (2015) Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141, 1072–1084. Available from: https://doi.org/10.1002/qj.2396

Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D. et al. (2016) NEMO ocean engine, version 3.6 stable. Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace 27, IPSL. https://doi.org/10.5281/zenodo.1472492

Madec, G. & Imbard, M. (1996) A global ocean mesh to overcome the north pole singularity. *Climate Dynamics*, 12, 381–388. Available from: https://doi.org/10.1007/BF00211684

Marsh, R., Ivchenko, V.O., Skliris, N., Alderson, S., Bigg, G.R., Madec, G. et al. (2015) NEMO-ICB (v1.0): interactive icebergs in the NEMO ocean model globally configured at eddy-permitting resolution. *Geoscientific Model Development*, 8, 1547–1562. Available from: https://doi.org/10.5194/gmd-8-1547-2015

Mass, C.F., Ovens, D., Westrick, K. & Colle, B.A. (2002) Does increasing horizontal resolution produce more skillful forecasts? *Bulletin of the American Meteorological Society*, 83, 407–430. Available from: https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2

Mathiot, P., Jenkins, A., Harris, C. & Madec, G. (2017) Explicit representation and parametrised impacts of under ice shelf seas in the $z^*$ coordinate ocean model NEMO 3.6. *Geoscientific Model Development*, 10, 2849–2874. Available from: https://doi.org/10.5194/gmd-10-2849-2017

Mirouze, I., Blockley, E.W., Lea, D.J., Martin, M.J. & Bell, M.J. (2016) A multiple length scale correlation operator for ocean data assimilation. *Tellus A*, 68, 29744. Available from: https://doi.org/10.3402/tellusa.v68.29744

Mittermaier, M.P. (2014) A strategy for verifying near-convection resolving model forecasts at observing sites. *Weather and Forecasting*, 29, 185–204. Available from: https://doi.org/10.1175/WAF-D-12-00075.1

Oke, P.R., Sakov, P., Cahill, M.L., Dunn, J.R., Fiedler, R., Griffin, D.A. et al. (2013) Towards a dynamically balanced eddy-resolving ocean reanalysis: BRAN3. *Ocean Modelling*, 67, 52–70. Available from: https://doi.org/10.1016/j.ocemod.2013.03.008

Parrish, D.F. & Derber, J.C. (1992) The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120, 1747–1763. Available from: https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2

Pawlowicz, R., McDougall, T., Feistel, R. & Tailleux, R. (2012) An historical perspective on the development of the thermodynamic equation of seawater — 2010. *Ocean Science*, 8, 161–174. Available from: https://doi.org/10.5194/os-8-161-2012

Pujol, M.-I. (2022) Product user manual for sea level altimeter products. Copernicus marine product user manual. CMEMS. https://catalogue.marine.copernicus.eu/documents/PUM/CMEMS-SL-PUM-008-032-068.pdf. Issue 7.0

Ridley, J.K., Blockley, E.W., Keen, A.B., Rae, J.G.L., West, A.E. & Schroeder, D. (2018) The sea ice model component of HadGEM3-GC3.1. *Geoscientific Model Development*, 11, 713–723. Available from: https://doi.org/10.5194/gmd-11-713-2018

Rio, M.-H., Mulet, S. & Picot, N. (2014) Beyond GOCE for the ocean circulation estimate: synergetic use of altimetry, gravimetry, and

in situ data provides new insight into geostrophic and Ekman currents. *Geophysical Research Letters*, 41, 8918–8925. Available from: https://doi.org/10.1002/2014GL061773

Roberts, M.J., Baker, A., Blockley, E.W., Calvert, D., Coward, A., Hewitt, H.T. et al. (2019) Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geoscientific Model Development*, 12, 4999–5028. Available from: https://doi.org/10.5194/gmd-12-4999-2019

Rösel, A., Kaleschke, L. & Birnbaum, G. (2012) Melt ponds on Arctic sea ice determined from MODIS satellite data using an artificial neural network. *The Cryosphere*, 6, 431–446. Available from: https://doi.org/10.5194/tc-6-431-2012

Rossa, A., Nurmi, P. & Ebert, E. (2008) Overview of methods for the verification of quantitative precipitation forecasts. In: Michaelides, S. (Ed.) *Precipitation: advances in measurement, estimation and prediction*. Berlin, Heidelberg: Springer, pp. 419–452. Available from: https://doi.org/10.1007/978-3-540-77655-0_16

Ryan, A.G., Régnier, C., Divakaran, P., Spindler, T., Mehra, A., Smith, G.C. et al. (2015) GODAE OceanView class 4 forecast verification framework: global ocean inter-comparison. *Journal of Operational Oceanography*, 8, s98–s111. Available from: https://doi.org/10.1080/1755876X.2015.1022330

Schröder, D., Feltham, D.L., Tsamados, M., Ridout, A. & Tilling, R. (2019) New insight from CryoSat–2 sea ice thickness for sea ice modelling. *The Cryosphere*, 13, 125–139. Available from: https://doi.org/10.5194/tc-13-125-2019

Smith, G.C., Bélanger, J.-M., Roy, F., Pellerin, P., Liu, Y., Chikhar, K. et al. (2019) Global ice ocean prediction system (GIOPS): update from version 2.3.1 to version 3.0.0. Canadian Centre for Meteorological and Environmental Prediction technical note, environment and climate Change Canada. https://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/tech_notes/technote_giops-300_e.pdf

Stewart, K.D., Hogg, A.M.C., Griffies, S.M., Heerdegen, A.P., Ward, M.L., Spence, P. et al. (2017) Vertical resolution of baroclinic modes in global ocean models. *Ocean Modelling*, 113, 50–65. Available from: https://doi.org/10.1016/j.ocemod.2017.03.012

Storkey, D., Blaker, A.T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E.W. et al. (2018) UK global ocean GO6 and GO7: a traceable hierarchy of model resolutions. *Geoscientific Model Development*, 11, 3187–3213. Available from: https://doi.org/10.5194/gmd-11-3187-2018

Storkey, D., Blockley, E.W., Furner, R., Guiavarc'h, C., Lea, D., Martin, M.J. et al. (2010) Forecasting the ocean state using NEMO: the new FOAM system. *Journal of operational. Oceanography*, 3, 3–15. Available from: https://doi.org/10.1080/1755876X.2010.11020109

UNESCO. (1981) Background papers and supporting data on the international equation of state of seawater 1980. Technical Report 38, 194 pp., UNESCO Technical Papers in Marine Science https://unesdoc.unesco.org/ark:/48223/pf0000047363.locale=en

Vidard, A., Balmaseda, M. & Anderson, D. (2009) Assimilation of altimeter data in the ECMWF ocean analysis system 3. *Monthly Weather Review*, 137, 1393–1408. Available from: https://doi.org/10.1175/2008MWR2668.1

Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S. et al. (2017) The met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geoscientific Model Development*, 10, 1487–1520. Available from: https://doi.org/10.5194/gmd-10-1487-2017

Waters, J., Lea, D.J., Martin, M.J., Mirouze, I., Weaver, A. & While, J. (2015) Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. *Quarterly Journal of the Royal Meteorological Society*, 141, 333–349. Available from: https://doi.org/10.1002/qj.2388

Weatherall, P., Marks, K.M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J.E. et al. (2015) A new digital bathymetric model of the world's oceans. *Earth and Space Science*, 2, 331–345. Available from: https://doi.org/10.1002/2015EA000107

Weaver, A.T., Deltel, C., Machu, E., Ricci, S. & Daget, N. (2005) A multivariate balance operator for variational ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131, 3605–3625. Available from: https://doi.org/10.1256/qj.05.119

Weaver, A.T., Tshimanga, J. & Piacentini, A. (2016) Correlation operators based on an implicitly formulated diffusion equation solved with the Chebyshev iteration. *Quarterly Journal of the Royal Meteorological Society*, 142, 455–471. Available from: https://doi.org/10.1002/qj.2664

While, J. & Martin, M.J. (2019) Variational bias correction of satellite sea-surface temperature data incorporating observations of the bias. *Quarterly Journal of the Royal Meteorological Society*, 145, 2733–2754. Available from: https://doi.org/10.1002/qj.3590

Zhang, J. & Rothrock, D.A. (2003) Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Monthly Weather Review*, 131, 845–861. Available from: https://doi.org/10.1175/1520-0493(2003)131<0845:MGSIWA>2.0.CO;2

Zuo, H., Balmaseda, M.A. & Mogensen, K. (2017) The new eddy-permitting ORAP5 ocean reanalysis: description, evaluation and uncertainties in climate signals. *Climate Dynamics*, 49, 791–811. Available from: https://doi.org/10.1007/s00382-015-2675-1

**How to cite this article:** Barbosa Aguiar, A., Bell, M.J., Blockley, E., Calvert, D., Crocker, R., Inverarity, G. *et al*. (2024) The Met Office Forecast Ocean Assimilation Model (FOAM) using a 1/12-degree grid for global forecasts. *Quarterly Journal of the Royal Meteorological Society*, 1–26. Available from: https://doi.org/10.1002/qj.4798