Data assimilation schemes for ocean forecasting: state of the art

Matthew J. Martin[1], Ibrahim Hoteit[2], Laurent Bertino[3], Andrew M. Moore[4]

[1]Met Office, Exeter, UK
[2]Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[3]Nansen Environmental and Remote Sensing Center, Bergen, Norway
[4]Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, California, USA

Correspondence to: Matthew Martin (matthew.martin@metoffice.gov.uk)

Abstract. Data assimilation is a process for integrating models and observations into comprehensive and reliable estimates of the ocean state. It is used to produce near-real time initial conditions (analyses) from which ocean forecasts are produced and to generate reconstructions of the past state of the ocean (reanalyses). Here we provide an overview of the methods currently used in ocean systems for assimilating satellite and in-situ observations, together with a brief review of methods being developed which will be implemented in future operational systems, including the use of machine-learning techniques that provide a way to improve their efficiency. A list of data assimilation software used by most of the global and regional operational ocean forecasting systems is provided, together with its availability. A discussion of practical considerations for employing data assimilation software and techniques operationally is also given, including the types of observations which are commonly used, and the implementation choices made by existing operational systems at global and regional scales is summarized.

1 Introduction

Accurate estimates of the state of the ocean are required for many purposes. Observations provide direct information about the ocean but are sparse in time and space. Numerical models can give information everywhere and describe the time evolution of the ocean but are prone to error. Data assimilation (DA) is the process by which these two sources of imperfect information are combined, taking into account their errors, in order to produce complete and accurate estimates of the ocean (Moore et al., 2019; Hoteit et al., 2018; Alvarez-Fanjul et al. 2022; Stammer et al., 2016; Carrassi et al., 2018). These estimates are used to produce historical reanalyses of the ocean (Storto et al., 2019; Heimbach et al., 2019) and in near real-time to initialize forecasts (Moore et al., 2019).

Data assimilation is used in global, regional and coastal ocean forecasting systems. The characteristics of the models used in each setting can be different including the resolution, processes represented and the model components. Global models are usually coupled physical ocean/sea-ice models, with a strong move at many operational centers to coupled

atmosphere/ocean/sea-ice models. Regional and coastal models usually resolve more of the higher-frequency processes which become more important in shallow seas, and often include coupled physical/biogeochemical components. The observations available for assimilation also often have different characteristics with different technologies needed to measure the ocean closer to the coast. The methods used to initialize forecasts in these different settings have to take into account the characteristics of the model and observations available so that the variability associated with the important processes can be constrained.

Many of the data assimilation methods used in ocean forecasting were originally developed for numerical weather prediction (with the notable exception of the ensemble Kalman Filter). The dominant spatial and temporal scales in the ocean are quite different to the atmosphere though, with the first baroclinic Rossby radius of deformation being a few 10s of kilometers at mid-latitudes (see e.g. Chelton et al., 1998) with temporal scales ranging from days to weeks. To resolve the open-ocean mesoscale at mid-latitudes, model resolutions of the order of at least 1/12° are required (Hewitt et al., 2016) and the aim of many global ocean data assimilation systems is to initialise the ocean state at these scales. Observations of the surface ocean are available at fairly high resolution from satellites, but observations of the sub-surface ocean are much sparser. Sophisticated methods are therefore required to make the most of the observations to constrain models of the 3D ocean on the desired scales. The integration of high-resolution models along with the high computational processing required for implementing an advanced data assimilation method demands computational resources that are available at only a small number of ocean forecasting centers and research institutions worldwide.

Errors in ocean models arise due to approximations in their numerical formulation, errors in the parameterisation of unresolved physics, and errors in the inputs to the model including the surface atmospheric forcing, river inputs, and the lateral boundary conditions for regional systems. The ocean is a chaotic system, so small differences in the initial state grow over time, especially in strongly eddying regions. All these sources of uncertainty contribute to the model forecast error, estimates of which are needed for data assimilation. Observations also contain errors and measure the ocean on different spatial scales (to each other and to the model). Estimates of the errors in the different observations are therefore also needed, including the component due to the measurement itself, as well as the component due to the difference in the representation of the ocean by the observation and model (Janjić et al., 2018).

Here we provide a summary of the status of ocean data assimilation as part of a special issue introduced by Alvarez Fanjul and Bahurel (2024). The next section gives a brief overview of data assimilation theory to put into context the various schemes used in operational ocean forecasting centers. The data assimilation software used at many of the operational centers is also described including community open--source software as well as other code developed and used at some of the main institutes. An overview of the practical considerations needed to apply data assimilation effectively in an operational setting is given. We then describe the current status of data assimilation as applied in many operational ocean forecasting centers, followed by a summary.

2 Data assimilation methodology

A variety of DA methods are being used or currently tested to develop Operational Ocean Forecasting Systems (OOFSs) (Moore et al., 2019). These first followed the 3D formulation of the DA problem (3DDA) in which the ocean state at a given time is estimated based only on the available observations at that time. 3DDA is often cast as a least-squares fitting problem whose solution minimizes a composite objective function involving a data-misfit term and a regularization term representing prior knowledge on the ocean state, called the background/prior and usually taken as the most recent ocean forecast. Both terms of the objective function are generally weighted by their respective observations and background error covariances, which can be also imposed following a (stochastic) Bayesian inverse formulation of the 3DDA problem under the assumption of Gaussian observations and background errors (Moore et al., 2019; Hoteit et al., 2018). When the observational operator relating the ocean state to the observations is linear, the 3DDA problem has an analytical solution, known as the Best Linear Unbiased Estimator (BLUE); when not, this operator is either linearized to compute the Optimal Interpolation (OI) solution, or the objective function is directly minimized using a gradient-based iterative optimization algorithm to compute the 3D variational DA (3DVAR) solution.

The solution of the 4D DA problem is more advanced as it is estimated based on a set of observations that is available over a given period of time (Weaver et al. 2003). It can be computed following a straightforward extension of the 3DVAR problem by formulating an objective function in which the data-misfit term constrains the ocean model prediction to the observations in time. When the ocean model and its forcing are~~is~~ considered perfect, only the ocean state at the start of the observation period needs to be estimated. The resulting strongly constrained (by the ocean model equations) 4DVAR solution is then integrated forward with the model beyond the observations period to compute the ocean forecasts. In contrast, the weak constraint 4DVAR problem considers model errors in the ocean model, which can be then estimated as part of the objective function minimization process. Jointly estimating the ocean initial state and model errors at every time step can quickly become computationally intractable. This was elegantly addressed by moving the optimization in the observation space, which should be of much smaller dimension in this case, using the dual formulation or Representer method (Bennett, 2005). In between the strong and weak constraint 4DVAR, a large variety of different implementations exist, for instance estimating the ocean model parameters (e.g., mixing schemes) and inputs (e.g., atmospheric forcing, open boundary conditions, bathymetry, etc.) as part of the minimization process. This has been successfully demonstrated with the MIT general circulation model (MITgcm) (Forget et al., 2015) and the Regional Ocean Modeling System (ROMS) (Moore et al., 2019). In all 4DVAR methods, the computation of the objective function gradients required for the minimization process can be efficiently implemented through the adjoint model, governed by the adjoint equations to the ocean tangent linear model (Moore et al., 2004; Vidard et al., 2015). Coding and running the adjoint model can be demanding in both human effort and computational resources.

The observational and background error covariances are key in determining the 3D and 4D DA solutions. The first sets the weights of the data-misfit terms and their correlations to avoid overfitting the observations while accounting for redundant information (Moore et al., 2019). The second constrains the DA solution by enforcing some dynamical relationships in the

initial state and/or smoothness on the estimated inputs and parameters to enable a proper propagation of the observations' information into all ocean model variables (Moore et al., 2019).

The DA methods discussed so far are designed to compute a deterministic estimate of the ocean state (the maximum a posteriori of the Bayesian inversion problem), and therefore do not provide a framework to quantify the uncertainties in the ocean forecasts, the covariance of which could be used as the background for the next DA cycle. This sets the stage to the filtering DA methods which sequentially compute the solution of the Bayesian inversion problem by considering the observations as they become available. The filtering formulation of the DA problem allows for model and observational errors and involves computing the probability distribution of the ocean state conditioned on all previous observations. This provides a recursive framework suitable for OOFSs where the model is used for forecasting the ocean state and its error statistics (forecast step), which are then updated with the new incoming observations based on Bayes' rule (analysis step) (Hoteit et al., 2018).

The Bayesian filtering problem can be conceptually solved by the Kalman filter (KF) when the underlying dynamical and observational models are linear and their errors are Gaussian, in which case the forecast and analysis distributions are Gaussian and the analytical form of their mean (state estimate) and covariance are available. Ocean general circulation models are however nonlinear, and the discrete dimension of the underlying ocean state can be very large. This motivated the development of a variety of simplified and extended variants of the KF for ocean DA, either by (i) linearizing the ocean dynamics and enforcing low-rank error covariance matrices (e.g., ~~Reduced Order Extended Kalman – ROEK – and~~ Singular Evolutive Extended – SEEK - filters), or (ii) using the widely celebrated ensemble KF (EnKF) methods (Vetra-Carvalho ~~Hoteit~~ et al., 2018). EnKF methods use samples to compute statistical approximations of the first two moments of the ocean state forecast and analysis distributions. Given an analysis ensemble, an EnKF integrates its members, eventually with perturbed noise to account for model errors, forward with the ocean model for forecasting, and the resulting forecast ensemble statistics are then updated with the incoming observations using the KF analysis step. The latter is referred to as stochastic when the KF analysis step is applied on each forecast ensemble member using perturbed observations, so that the analysis ensemble covariance matches that of the KF, and deterministic (e.g., ETKF, EAKF, SEIK, DEnKF) when the KF analysis step is directly applied on the mean and covariance of the forecast ensemble, after which a deterministic resampling step is needed to resample a new analysis ensemble (Hoteit et al., 2018).

EnKFs are generally integrated with relatively small ensembles (~100 samples) to limit their computational cost, making their sample covariances low-rank and thus necessitating localization/covariance-tapering techniques to confine the spatial range of their correlations (Hoteit et al., 2018). Limited ensemble size can also result in underestimation of the ensemble variance leading to the need for ensemble inflation (Evensen et al., 2022). To further reduce the computational requirements, EnKFs are also often implemented with static ensembles, only using the ocean model to compute the forecast starting from the analysis state (ensemble OI – EnOI - methods), or their ensembles augmented with pre-selected static members (Hybrid EnOI-EnKF methods) (Counillon et al., 2009). On the other side of the spectrum, more sophisticated filtering methods have been also proposed to move beyond the Gaussian error assumption by employing Monte-Carlo approximations of the forecast and analysis distributions, so-called Particle Filters, or through Gaussian mixture approximations which when implemented within

an ensemble framework reduce to some sort of ensemble of EnKFs (van Leeuwen at al., 2019~~5~~). These methods are however still in testing phases and are yet to be applied in operational settings.

4DVAR and EnKFs were proven to provide viable and robust solutions for many ocean DA applications, and most ocean centers are currently developing their operational systems around these approaches. There are benefits and drawbacks in using an EnKF or a 4DVAR (Lorenc, 2003; Kalnay et al., 2007). EnKFs involve flow-dependent ensemble representation of the background, though rank-deficient. On the downside, the EnKF is generally only efficient for moderate model nonlinearity because of its second-order moments approximation of the error statistics. 4DVAR, on the other hand, should better handle nonlinearities, though the optimization of its objective function can be a complex task in the presence of strongly nonlinear dynamics (Moore et al., 2019; Hoteit et al., 2018), and can be implemented with a full-rank, albeit static, background error covariance matrix. 4DVAR further requires coding and maintaining the adjoint of the observation and forecasting models, which is quite demanding. The use of automatic differentiation in distributed HPC environments, which is receiving a renaissance in the context of machine learning, may overcome this limitation (Heimbach et al., 2005). Finally, 4DVAR does not lend itself easily to parallelization, while the important computational cost for computing the forecast ensemble can be drastically mitigated by trivial parallelization.

There have been various attempts to merge the 4DVAR and EnKF approaches in order to combine their strengths, which introduced a new family of Hybrid Ensemble-Variational (EnVAR) methods. This includes (i) considering an ensemble of DA (EnDA) methods to obtain flow-dependent error representations, or (ii) the iterative Ensemble Kalman Filters (iEnKFs) and Smoothers (iEnKSs) which use a forecast ensemble to describe the background statistics and apply a nonlinear optimization to the 4DVAR objective function in the ensemble space (Sakov et al., 2012a), and (iii) the class of 4D Ensemble Variational (4DEnVAR) methods which also performs a set of 4DVAR optimizations in the subspace spanned by the ensemble using a set of perturbed observations (Liu et al., 2012). Different 4DEnVAR versions have been proposed (Bannister, 2017), employing hybrid background covariances, adjoint model or finite differences to compute the gradients, and different types of perturbations.

Recently, machine learning (ML) techniques have been also considered to enhance the efficiency of the DA methods, in terms of both capacity and computations (Cheng et al, 2023). ML techniques harness neural networks' (NNs) potential at approximating highly nonlinear functions, which may enable developing computationally less demanding forecasting models (Barthélémy et al., 2022), and backward models for efficient data fitting. NNs were also proposed as end-to-end replacement of the analysis steps (Beauchamp et al, 2023), and to parameterize and account for model errors (Farchi et al., 2021).

3 Data assimilation software

Data assimilation software packages come in all sizes and flavors. A first distinction needs to be made between educational packages that can be used for methodological developments and operational codes designed for high-performance computers. We will only consider the latter category in this section. A second distinction can be made between software aimed at 4DVAR

methods and those that take the EnKF as their target algorithm. These two types of software differ in their complexity and size, and therefore adopt different development strategies. There are thus several small-sized EnKF packages and a few more ambitious 4DVAR packages on the market. The latter may also include the EnKF as a small addition to their ensemble-variational toolbox. Some of the packages (DART, PDAF, JEDI) have users in other research fields beyond ocean forecasting. See Table 1 for a list of commonly used DA software in ocean prediction systems.

Table 1: Data assimilation software packages.

Software name	Target algorithm(s)	Programming language	Development community	Code availability
JEDI	Variational DA.	C++	JCSDA, NOAA, NASA, US Navy and Air Force, Met Office.	Open source. https://github.com/JCSDA
MITgcm	Variational DA.	Fortran ~~90~~. A version in Julia is under development.	ECCO consortium, GECCO, MIT, Uni. Texas	Open source. https://mitgcm.readthedocs.io/
NEMOVAR	Variational DA.	Fortran ~~90~~	CERFACS, ECMWF, Met Office, INRIA	Not open source.
OceanVar	Variational DA.	Fortran ~~90~~	CMCC, CNR	Not open source.
ROMS	Variational DA.	Fortran ~~90~~	ROMS community	Open source. https://www.myroms.org/
DART	Ensemble DA.	Fortran ~~90~~	NCAR	Open source. https://dart.ucar.edu
EnKF	Ensemble DA.	Fortran ~~90~~	NERSC	Open source. https://github.com/nansencenter/enkf-topaz

EnKF-C	Ensemble DA.	C	Bureau of Meteorology	Open source. https://github.com/sakov/enkf-c
NEDAS	Ensemble DA	Python, parallel	NERSC	Open source. https://github.com/nansencenter/NEDAS
OAK	Sequential DA.	Fortran ~~90~~	U. Liège	Open source. https://github.com/gher-uliege/OAK
OpenDA	Ensemble DA.	Java	TU Delft	Open source. https://www.openda.org
PDAF	Ensemble DA.	Fortran ~~90~~	AWI	Open source. https://pdaf.awi.de~~/trac/wiki~~
SAM2	SEEK filter.	Fortran ~~90~~	Mercator Ocean International, ECCC.	Not open source.
Sequoia	Sequential DA.	Fortran ~~90~~	OMP/LEGOS	Available on demand. https://sirocco.obs-mip.fr/

The above software packages have mainly been used on high-performance computers (HPCs) and some of them on personal computers. The NEMOVAR and MITgcm 4DVAR codes, and the NEDAS ensemble code, are actively being developed for use on GPU-based systems. However, all the DA software listed above have lived long enough to be ported several times to different HPC architectures with different compilers and can be qualified as portable.

4 Practical implementations in operational systems

Several factors dictate the practical implementation of ocean DA systems within an operational environment. The primary controlling factors in any operational environment typically relate to (i) scheduling of the DA analysis and forecast phases with respect to the competing demands of other essential activities (e.g. numerical weather prediction, hydrological forecasts, etc); and (ii) the release of analysis-forecast products in a timely manner so that they are of maximum benefit to the users. These overarching criteria therefore, in turn, dictate the configuration of the forecast model and the data assimilation approach that may be used.

In the case of ensemble approaches, such as the EnKF or EnVar, there may be a trade-off between model resolution and the ensemble size in that computation time increases with resolution. Thus, with limited resources fewer ensemble members can be run within the constraints imposed by (i) and (ii). An advantage of ensemble approaches is that each ensemble member can be computed independently, meaning that in very large HPC environments, many ensemble members can be run

simultaneously. Here again though there can be a trade-off between resolution and ensemble size. While most ocean models scale reasonably well on parallel computing architectures, wall-clock time typically does not scale linearly with the number of cores. Hence, there is a point of diminishing returns whereby it may be better to allocate fewer cores to the business of computing ensemble members at the expense of a longer wall-clock time for each member, rather than dedicating a very large number of cores to a single task.

Unlike ensemble methods, the traditional approaches to variational data assimilation, namely 3DVAR and 4DVAR, are strictly sequential and cannot be parallelized in time. In other words, the inner- and outer-loop iterations of the cost function minimization algorithm must be performed sequentially. The sequential iterative nature of variational approaches therefore imposes a heavy computational burden on the data assimilation phase of the analysis-forecast cycle, especially in the case of 4DVAR. This burden is alleviated in some 4DVAR systems by performing the inner-loop minimization steps at lower model resolution – for example, a reduction of the horizontal resolution by a factor of 2 typically yields a factor of 8 reduction in wall-clock time assuming that the inner-loop time step can be halved also. Performing the inner-loops at lower arithmetic precision (i.e. 32 bit arithmetic versus 64 arithmetic) can lead to further cost savings. In 4DVAR, the inner-loop iterations involve integrations of the tangent linear (TL) and adjoint (AD) versions of the forecast model. Further reductions in computational cost can therefore also be achieved by reducing the complexity of the TL and AD models. Time-parallel formulations of 4DVAR based on a saddle-point algorithm also yield substantial computational savings (Fisher and Gurol, 2017; Moore et al, 2023).

The assimilation strategy employed also depends on the type of observations that are to be assimilated, and their distribution in time. In the case of a Kalman filter, while each observation can be assimilated sequentially at the associated observation time, this may not be an efficient strategy since this might require overly frequent stopping and restarting of the filter computations. Thus, it is often preferable to group together observations that are closely spaced in time and treat them as though they were available at the same time. This approach underpins the strategy of First-Guess at Appropriate Time (FGAT) which is commonly employed in conjunction with both ensemble approaches and 3DVAR. Such approaches necessitate the choice of a time-window over which the observations will be aggregated for assimilation. In between times, the forecast model is run to yield the first-guess or background for the next data assimilation cycle, so the time-window of aggregation also dictates how frequently the analysis-forecast cycle can be performed. For an EnKF, it is sufficient to store observation equivalents from each model ensemble member to calculate asynchronous cross-covariances (Sakov et al. 2010). In the case of 4DVAR, observations are typically assimilated at the actual time of observation. This involves integrations of the TL and AD models forward and backward in time. Since these are based on a linearized version of the forecast model, the validity of the linear assumption through time is an important consideration. In particular, linear instabilities can develop if appropriate care is not exercised. Therefore, while a long-time window in 4DVAR may be preferable so that the analysis is informed by more observations, this must be balanced by the validity of the linear assumptions employed in the TL and AD models, and the added computational burden of the longer assimilation window.

5 Ocean Observations

While there is a common subset of observations from the global ocean observing system (GOOS) that are assimilated into ocean models, additional sources of data may be available for assimilation into regional ocean models that are not appropriate for global models. The GOOS and different types of observations available are discussed in the ETOOFS guide (Alvarez-Fanjul et al., 2022). The mainstay of the GOOS is remote sensing observations of sea surface temperature (SST), sea surface height (SSH), sea surface salinity (SSS) and sea ice concentration. This is supported by the Argo network of profiling floats that provide vertical sections of temperature and salinity (and in some cases biogeochemical variables) mostly over the upper 2000 m of the water column, although deep Argo floats below 2000 m are now also being deployed. In the tropical oceans, the observing system is augmented by networks of buoys that provide profiles of temperature (and in some cases salinity and currents) to depths of ~500 m. Observations from tagged marine mammals also provide useful information in some regions of the world ocean. In coastal regions, other data sources are often available that cannot be readily assimilated into global models because of the disparity in horizontal resolution. These include data from gliders and other autonomous underwater vehicles (AUVs), estimates of surface currents from high-frequency (HF) radars, other tagged marine mammals, moorings, drifters, and in some locations dedicated coastal arrays.

All observations, regardless of their origin, must be subject to strict quality control (QC) standards before they can be assimilated into a model (Good et al., 2023). All operational centers employ sophisticated QC systems for flagging and rejecting erroneous observations and those of poor quality. In addition, the large volume of remote sensing observations from earth orbiting satellites must generally be thinned in space and time. There are three main reasons for this: first, remote sensing observations contain a great deal of redundancy which can be reduced by judicious thinning; second, the sheer volume of remote sensing observations can quickly overwhelm a data assimilation system if not appropriately thinned (particularly in light of the high redundancy); and lastly, accounting for correlated observation errors in data assimilation systems is technically challenging, so thinning the observations is one approach for reducing the degree of correlation. Another important aspect of operational data assimilation systems is the formation of so-called “super observations”. This refers to the procedure for combining multiple observations of the same type that fall within a model grid cell at the same observation time into a single datum (a super observation). This usually entails some simple averaging or aggregation procedure and is necessary in order to improve the numerical conditioning of the data assimilation inverse problem.

The use of observations in data assimilation requires information about their uncertainties. The observation uncertainty consists of a component due to the instrument error and a component related to the different representation of the ocean by the observations and the model (for example representing different spatial and/or time scales; Janjić et al., 2018). Some observation types (e.g. satellite SST) are provided together with information about the expected uncertainty in each measurement and this information can be used directly in the data assimilation. For other observation types, estimates of the uncertainty have to be obtained from the literature. An example list of instrumental uncertainties for different observation types assimilated in a global ocean forecasting system is provided in Table 1 of Lea et al. (2022).

Since the observations are the only, albeit far from complete, measure of the true state of the ocean, they often form the basis for metrics that are used to monitor the performance of data assimilation systems. The statistics of the observation minus background (OmB) and observation minus analysis (OmA) provide information about the fit of the model to observations before and after the observations have been assimilated. The statistics of OmB and OmA provide an important diagnostic check on prior assumptions made about the background error and observation error covariances (Desrosiers et al. 2005). Inconsistencies between the actual and expected error statistics can be used to retune the data assimilation system, regardless of the data assimilation methodology employed. In variational data assimilation systems, continuous monitoring of the cost function and cost function gradient also provide useful diagnostics of system performance. The impact of different components of the observing system can also be quantified and monitored in various ways. This is commonly done in terms of the impact on the skill of forecasts that are initialized from the data assimilation analyses. By continuously monitoring the impact of each component of the observing system on forecast skill, data streams that consistently degrade the forecast skill can be flagged (and removed), and the degradation of any data stream over time can be identified.

6 Current status of data assimilation in operational forecasting systems

An overview of operational ocean data assimilation systems and their characteristics is provided in Figure 1 for global systems and Figure 2 for regional and coastal systems. Not all operational systems are covered here, but the figures provide information about the main choices which have been made by some of the existing operational centers producing near-real time forecasts in the configuration of their data assimilation schemes. The information represents the current operational status, but all centers are continually developing and improving their systems, and many have research configurations which are more sophisticated than those presented.

In general, the global systems use somewhat simpler DA algorithms (though they are still complex in their implementation of those algorithms) than the regional and coastal systems, the exception being the BoM system which uses a hybrid-EnKF with 48 dynamic members and 144 stationary low-mode members (Brassington et al., 2023). Many global forecasting groups use a 3DVAR-FGAT algorithm (Barbosa Aguiar et al., 2024; Zuo et al., 2019; Cummings and Smedstad, 2013; Storto et al., 2016; Ravichandran et al., 2013) with some groups using a SEEK filter or LESTKF with a static ensemble (Lellouche et al., 2018; Smith et al., 2016; Li et al., 2021) ~~or EnOI schemes (Chamberlain et al., 2021)~~. The reason these algorithms are generally simpler is largely due to the large number of grid points, especially in the higher resolution global systems, which restricts the options for more expensive algorithms when timely delivery of forecasts is the main goal. Some groups are testing more sophisticated schemes in research mode though, including those which make use of ensembles, e.g. MOI are testing LETKF, Met Office and ECMWF are testing hybrid-3DEnVAR schemes (Lea et al., 2022; Chrust et al., 2024), ~~BoM are testing use of EnKF,~~ and JMA are implementing 4DVAR (Fujii et al., 2023). The observations assimilated in these systems are fairly consistent across the different systems, with the main difference being whether the systems include sea-ice or atmosphere components. Some of the DA systems are focussed purely on the ocean, many include a sea-ice component, and some now

run with a coupled atmospheric component, though these systems all still use so-called "weakly" coupled DA where the DA in the atmospheric and ocean/sea-ice components are run separately, despite using coupled models (see for example Guiavarch et al., 2019 and de Rosnay et al., 2022). There is a large range of time windows used by the different systems, with the most common time window being 1 day, a short 6-hour window used in the Met Office coupled DA system (to match the time window in the atmospheric DA; Guiavarc'h et al., 2019), and with longer time windows of 5-7 days used by some systems. There is a wider range of DA algorithms employed in regional and coastal forecasting systems from EnOI/static SEEK filters (Carvalho et al., 2019; Ji et al., 2017; Smith et al., 2021; Escudier et al., 2022) and 3DVAR-FGAT schemes (Rahaman et al., 2018; King and Martin, 2021; Coppini et al., 2023) through to the more sophisticated EnKF (Sakov et al. 2012b, Röhrs et al., 2023), LESTKF (Brüning et al., 2021) and 4DVAR algorithms (Moore et al., 2023; Iversen et al., 2023; Hirose et al., 2019; Lee et al., 2018). Many of these regional systems also include biogeochemical DA (see Fennel et al., 2022 for a recent review), and some include coupled sea-ice DA (e.g. ~~Bertino and Xie, 2020~~Sakov et al., 2012b). The range of observations assimilated is also quite varied with some systems only assimilating SST data while others include the full range of available observations including HF radar, gliders and biogeochemical data from satellites and in situ platforms.

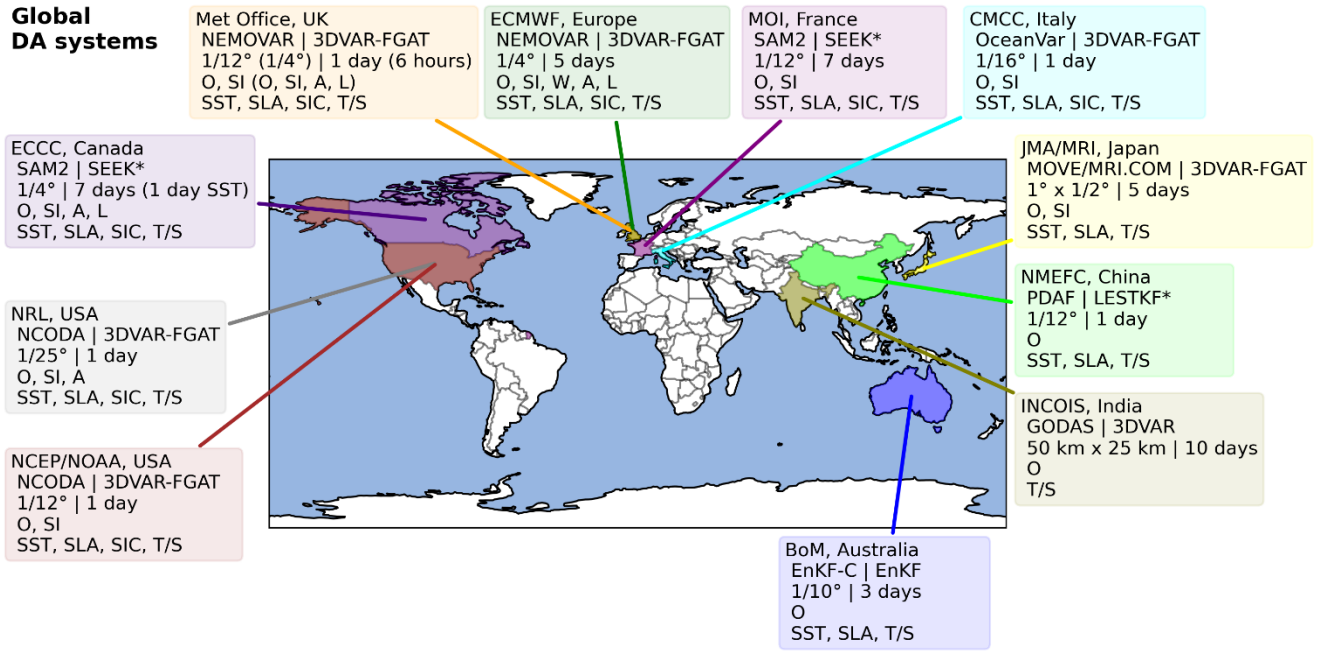

Figure 1: Operational global ocean data assimilation systems. For each institute, the following are listed: the DA algorithm (* indicates fixed basis version of the algorithm) and software; DA resolution and time window; Earth system components (O = physical ocean, SI = sea-ice, A = atmosphere, W = surface waves, BGC = ocean biogeochemistry, L=land); observations assimilated (SST = sea surface temperature, SLA = sea level anomaly, SIC = sea-ice concentration, SID = sea-ice drift, T/S = profiles of temperature and salinity, OC = satellite ocean color, BGC = biogeochemical profile data, HFR = HF radar).

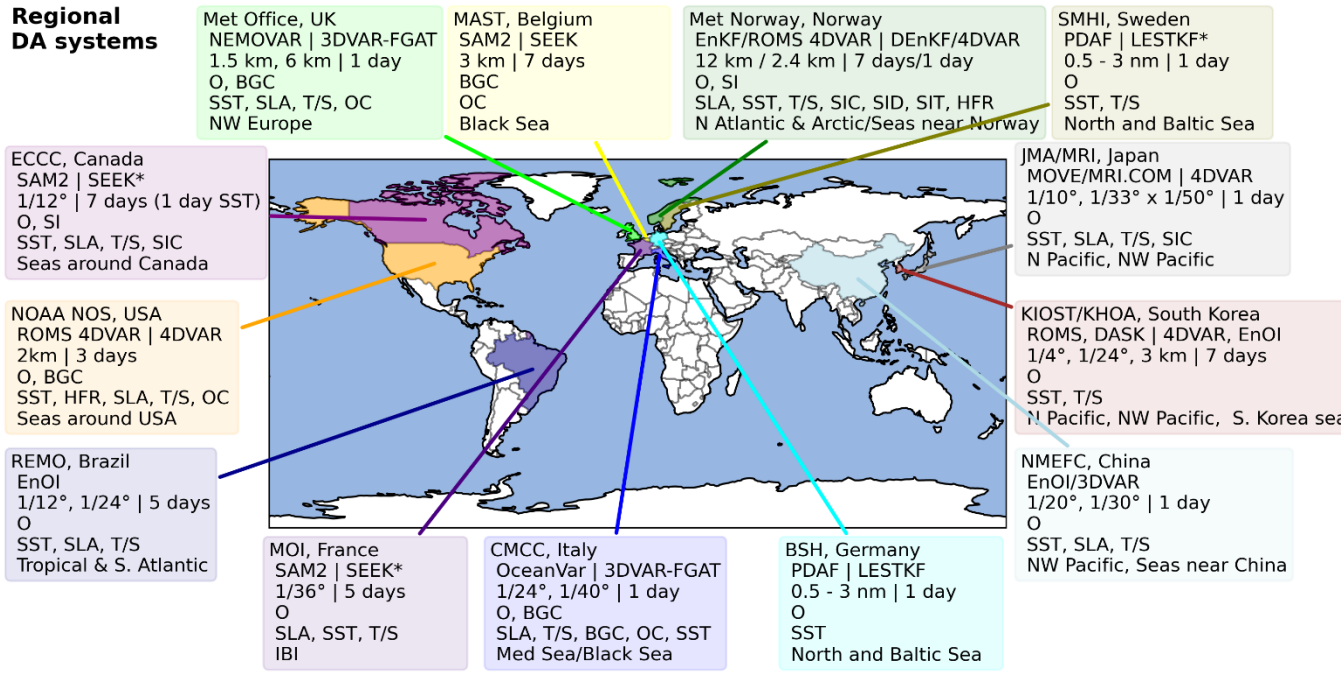

Figure 2: Operational regional and coastal ocean data assimilation systems. See description for Figure 1.

7 Future directions

Operational ocean forecasting systems are under constant development, including the data assimilation component. There is a continued push towards higher resolution at many centers and an increase in the use of ensembles both for improved data assimilation and for providing forecast uncertainty information to users. These directions both require significant additional computational resources so improving the computational efficiency of data assimilation software, particularly on new computer architectures like GPUs, is important to allow more flexibility in the choice of algorithms and resolutions used. While there is evidence that increasing ensemble size provides greater improvements in forecast skill once the important processes are resolved, rather than further increasing model resolution (Thoppil et al., 2021), there is also continued research in improving assimilation methodology to allow sub-mesoscale processes to be constrained where there are sufficient observations (Ying, 2019, Jacobs et al., 2023). New observing systems are being developed and launched, in particular wide-swath altimeter missions such as SWOT (Morrow et al., 2019) which allow improved constraints on mesoscale ocean forecasts (King et al., 2024; Liu et al., 2024; Benkiran et al., 2024). Treatment of spatially correlated observation errors is important to allow the most information to be extracted from such data, and various groups are developing methods to represent these in data assimilation systems (e.g. Guillet et al., 2019; Yaremchuk et al., 2024). Coupled ocean/atmosphere data assimilation is

also an evolving area (de Rosnay et al., 2022) with the development of more strongly coupled data assimilation algorithms requiring the use of consistent software across the different earth system components. The use of machine learning in the ocean forecasting process is also developing quickly with various applications in the context of data assimilation being tested and implemented (Heimbach et al., 2024).

References

Alvarez Fanjul, E. and Bahurel, P.: OceanPrediction Decade Collaborative Center: Connecting the world around ocean forecasting, State Planet Discuss. [preprint], https://doi.org/10.5194/sp-2024-34, in review, 2024.

Alvarez Fanjul, E., Ciliberti, S., Bahurel, P.: Implementing Operational Ocean Monitoring and Forecasting Systems. IOC-UNESCO, GOOS-275. https://doi.org/10.48670/ETOOFS, 2022.

Bannister, R.N.: A review of operational methods of variational and ensemble-variational data assimilation, Q.J.R. Meteor. Soc. 143, 607-633. https://doi.org/10.1002/qj.2982, 2017.

Barbosa Aguiar, A., Bell, M.J., Blockley, E., Calvert, D., Crocker, R., Inverarity, G., et al.: 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. https://doi.org/10.1002/qj.4798, 2024.

Barthélémy, S., Brajard, J., Bertino, L., & Counillon, F.: Super - resolution data assimilation. *Ocean Dynamics*, *72*(July), 661–678. https://doi.org/10.1007/s10236-022-01523-x, 2022.

Beauchamp, M., Febvre, Q., Georgenthum, H., and Fablet, R.: 4DVarNet-SSH: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry- Geosci. Model Dev., 16, 2119-2147. https://doi.org/10.5194/gmd-16-2119-2023, 2023.

Benkiran M, Le Traon P-Y, Rémy E and Drillet Y (2024) Impact of two high resolution altimetry mission concepts on ocean forecasting. Front. Mar. Sci. 11:1465065. doi: 10.3389/fmars.2024.1465065

Bennett, A. F.: Inverse modeling of the ocean and atmosphere. Cambridge University Press. https://doi.org/10.1017/CBO9780511535895, 2005.

~~Bertino, L., Xie, J.: Operational Forecasting of Sea Ice in the Arctic Using TOPAZ System. In: Johannessen, O., Bobylev, L., Shalina, E., Sandven, S. (eds) Sea Ice in the Arctic. Springer Polar Sciences. Springer, Cham. https://doi.org/10.1007/978-3-030-21301-5_9, 2020.~~

Brassington, G. B., Sakov, P., Divakaran, P., Aijaz, S., Sweeney-Van Kinderen, J., Huang, X., and Allen, S.: OceanMAPS v4.0i: a global eddy resolving EnKF ocean forecasting system, in: OCEANS 2023-Limerick, IEEE, 1–8, https://doi.org/10.1109/OCEANSLimerick52467.2023.10244383, 2023.

Brüning, T., Li, X., Schwichtenberg, F., Lorkowski, I. (2021): An operational, assimilative model system for hydrodynamic and biogeochemical applications for German coastal waters. Hydrographische Nachrichten 118. Rostock: Deutsche Hydrographische Gesellschaft e.V.. S. 6-15. https://doi.org/10.23784/HN118-01.

Carrassi, A., Bocquet, M., Bertino, L., Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Clim Change*. 9:e535. https://doi.org/10.1002/wcc.535, 2018.

Carvalho, J. P. S., Costa, F. B., Mignac, D., & Tanajura, C. A. S.: Assessing the extended-range predictability of the ocean model HYCOM with the REMO ocean data assimilation system (RODAS) in the South Atlantic. *Journal of Operational Oceanography*, *14*(1), 13–23. https://doi.org/10.1080/1755876X.2019.1606880, 2019.

~~Chamberlain, M. A., Oke, P. R., Fiedler, R. A. S., Beggs, H. M., Brassington, G. B., and Divakaran, P.: Next generation of Bluelink ocean reanalysis with multiscale data assimilation: BRAN2020, Earth Syst. Sci. Data, 13, 5663–5688, https://doi.org/10.5194/essd-13-5663-2021, 2021.~~

Chelton, D. B., R. A. deSzoeke, M. G. Schlax, K. El Naggar and N. Siwertz, 1998: Geographical variability of the first-baroclinic Rossby radius of deformation. J. Phys. Oceanogr., 28, 433-460.

Cheng, S. B. , Quilodrán-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss, B., Brajard, J., Xiao, D. H., Janjic, T., Ding, W. P., Guo, Y. K., Carrassi, A., Bocquet, M., and Arcucci, R.: Machine learning with data assimilation and uncertainty quantification for dynamical systems: A review. IEEE/CAA J. Autom. Sinica, 10(6), 1361-1387. DOI: 10.1109/JAS.2023.123537, 2023.

Chrust, M., Weaver, A.T., Browne, P., Zuo, H. & Balmaseda, M.A. (2024) Impact of ensemble-based hybrid background-error covariances in ECMWF's next-generation ocean reanalysis system. Quarterly Journal of the Royal Meteorological Society, 1–21. https://doi.org/10.1002/qj.4914

Coppini, G., Clementi, E., Cossarini, G., Salon, S., Korres, G., et al..: The Mediterranean Forecasting System – Part 1: Evolution and performance, Ocean Sci., 19, 1483–1516, https://doi.org/10.5194/os-19-1483-2023, 2023.

Counillon, F., Sakov, P., and Bertino, L.: Application of a hybrid EnKF-OI to ocean forecasting. Ocean Sci., 5, 389-401. https://doi.org/10.5194/os-5-389-2009, 2009.

Cummings J.A., Smedstad O.M.: Variational data assimilation for the global ocean. data assimilation for atmospheric, oceanic and hydrologic applications (Vol. II) chapter 13:303–343, 2013.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131 ,** 3385–3396.

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., et al.: Coupled data assimilation at ECMWF: current status, challenges and future developments. *Quarterly Journal of the Royal Meteorological Society*, 148(747), 2672–2702, https://doi.org/10.1002/qj.4330, 2022.

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., et al. (2022) Coupled data assimilation at ECMWF: current status, challenges and future developments. Quarterly Journal of the Royal Meteorological Society, 148(747), 2672–2702. Available from: https://doi.org/10.1002/qj.4330

Escudier, R., Reffray, G., Hamon, M., Levier, B., and Gutknecht, E.: Improving the assimilation part of the near real time system of IBI, EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-11410, https://doi.org/10.5194/egusphere-egu22-11410, 2022.

Evensen, G., Vossepoel, F.C., van Leeuwen, P.J. (2022). Localization and Inflation. In: Data Assimilation Fundamentals. Springer Textbooks in Earth Sciences, Geography and Environment. Springer, Cham. https://doi.org/10.1007/978-3-030-96709-3_10

Farchi, A., Laloyaux, P., Bonavita, M. & Bocquet, M.: Using machine learning to correct model error in data assimilation and forecast applications. Quarterly Journal of the Royal Meteorological Society, 147(739), 3067-3084. https://doi.org/10.1002/qj.4116, 2021.

Fennel, K., Mattern, J.P., Doney, S.C. et al. Ocean biogeochemical modelling. Nat Rev Methods Primers 2, 76. https://doi.org/10.1038/s43586-022-00154-2, 2022.

Fisher M., and Gurol, S.: Parallelization in the time dimension of four-dimensional variational data assimilation. Quarterly Journal of the Royal Meteorological Society, 143, 1136-1147. https://doi.org/10.1002/qj.2997, 2017.

Fujii Y., Yoshida T., Sugimoto H., Ishikawa I. and Urakawa S.: Evaluation of a global ocean reanalysis generated by a global ocean data assimilation system based on a four-dimensional variational (4DVAR) method. *Front. Clim.* 4:1019673. doi: 10.3389/fclim.2022.1019673, 2023.

Forget, G., Campin, J.-M., Heimbach, P., Hill, C. N., Ponte, R. M. and Wunsch, C.: ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation. Geoscientific Model Development, 8(10), 3071–3104. https://doi.org/10.5194/gmd-8-3071-2015, 2015.

Guillet O, Weaver AT, Vasseur X, Michel Y, Gratton S, Gurol S. Modelling spatially correlated observation errors in variational data assimilation using a diffusion operator on an unstructured mesh. Q J R Meteorol Soc 2019; 145: 1947–1967. https://doi.org/10.1002/qj.3537

Good S., Mills, B., Boyer, T., Bringas, F., Castelão, G., Cowley, R., Goni, G., Gouretski, V., and Domingues, C. M.: Benchmarking of automatic quality control checks for ocean temperature profiles and recommendations for optimal sets. Frontiers in Marine Science, 9, 1075510. https://doi.org/10.3389/fmars.2022.1075510, 2023.

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

Heimbach, P., O'Donncha, F., Garcia-Valdecasas, J. M., Arnaud, A., and Wan, L.: Crafting the Future: Machine Learning for Ocean Forecasting, State Planet Discuss. [preprint], https://doi.org/10.5194/sp-2024-18, in review, 2024.

Heimbach, P., Hill, C. and Giering, R.: An efficient exact adjoint of the parallel MIT General Circulation Model, generated via automatic differentiation. Future Generation Computer Systems, 21(8), 1356–1371. https://doi.org/10.1016/j.future.2004.11.010, 2005.

Heimbach, P., Fukumori, I., Hill, C. N., Ponte, R. M., Stammer, D., Wunsch, C., Campin, J.-M., Cornuelle, B., Fenty, I., Forget, G., Köhl, A., Mazloff, M., Menemenlis, D., Nguyen, A. T., Piecuch, C., Trossman, D., Verdy, A., Wang, O. and Zhang, H.: Putting It All Together: Adding Value to the Global Ocean and Climate Observing Systems With Complete Self-Consistent Ocean State and Parameter Estimates. Frontiers in Marine Science, 6, 769–10. https://doi.org/10.3389/fmars.2019.00055, 2019.

Hewitt, H. T., Roberts, M. J., Hyder, P., Graham, T., Rae, J., Belcher, S. E., Bourdallé-Badie, R., Copsey, D., Coward, A., Guiavarch, C., Harris, C., Hill, R., Hirschi, J. J.-M., Madec, G., Mizielinski, M. S., Neininger, E., New, A. L., Rioual, J.-C., Sinha, B., Storkey, D., Shelly, A., Thorpe, L., and Wood, R. A.: The impact of resolving the Rossby radius at mid-latitudes in the ocean: results from a high-resolution version of the Met Office GC2 coupled model, Geosci. Model Dev., 9, 3655–3670, https://doi.org/10.5194/gmd-9-3655-2016, 2016.

Hirose, N., Usui, N., Sakamoto, K., Tsujino, H., Yamanaka, G., Nakano, H., et al.: Development of a new operational system for monitoring and forecasting coastal and open ocean states around Japan. Ocean Dyn. 69, 1333–1357. doi: 10.1007/s10236-019-01306-x, 2019.

Hoteit, I., Luo, X., Bocquet, M., Köh, A., Ait-El-Fquih, B.: Data assimilation in oceanography: Current status and new directions. In "New Frontiers in Operational Oceanography", E. Chassignet, A. Pascual, J. Tintoré, and J. Verron, Eds., GODAE OceanView, 465-512. https://doi.org/10.17125/gov2018.ch16, 2018.

Iversen, S. C., Sperrevik, A. K., and Goux, O.: Improving sea surface temperature in a regional ocean model through refined sea surface temperature assimilation, Ocean Sci., 19, 729–744, https://doi.org/10.5194/os-19-729-2023, 2023.

Jacobs, G., D'Addezio, J., Bartels, B., DeHaan, C., Barron, C., Carrier, M., Shcherbina, A., Dever, M. Adapting constrained scales to observation resolution in ocean forecasts. Ocean Modelling, 186, 102252, https://doi.org/10.1016/j.ocemod.2023.10225, 2023.

Janjić T, Bormann N, Bocquet M, Carton JA, Cohn SE, Dance SL, Losa SN, Nichols NK, Potthast R, Waller JA, Weston P. On the representation error in data assimilation, Q J R Meteorol Soc. 2018; 144: 1257–1278. https://doi.org/10.1002/qj.3130

Ji, X., Kwon, K.M., Choi, B.J. et al.: Assimilating OSTIA SST into regional modeling systems for the Yellow Sea using ensemble methods. Acta Oceanol. Sin. 36, 37–51. https://doi.org/10.1007/s13131-017-0978-2, 2017.

Kalnay, E., Li, H., Miyoshi, T., Yang, S. C., & Ballabrera-Poy, J. (2007). 4-D-Var or ensemble Kalman filter? Tellus A: Dynamic Meteorology and Oceanography, 59(5), 758–773. https://doi.org/10.1111/j.1600-0870.2007.00261.x

King, R. R., Martin, M. J., Gaultier, L., Waters, J., Ubelmann, C., and Donlon, C.: Assessing the impact of future altimeter constellations in the Met Office global ocean forecasting system, Ocean Sci., 20, 1657–1676, https://doi.org/10.5194/os-20-1657-2024, 2024.

King, R. R. and Martin, M. J.: Assimilating realistically simulated wide-swath altimeter observations in a high-resolution shelf-seas forecasting system, Ocean Sci., 17, 1791–1813, https://doi.org/10.5194/os-17-1791-2021, 2021.

Lea, D.J., While, J., Martin, M.J., Weaver, A., Storto, A. & Chrust, M.: 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(745), 1996-2030. https://doi.org/10.1002/qj.4292, 2022.

Lee, J.H., Kim, T., Pang, I.C. et al.: 4DVAR Data Assimilation with the Regional Ocean Modeling System (ROMS): Impact on the Water Mass Distributions in the Yellow Sea. Ocean Sci. J. 53, 165–178. https://doi.org/10.1007/s12601-018-0013-3, 2018.

Lellouche J.-M., Greiner E., Le Galloudec O., Garric G., Regnier C., Drevillon M., et al.: Recent updates to the copernicus marine service global ocean monitoring and forecasting real-time 1/12° high-resolution system. *Ocean Sci.* 14, 1093–1126. doi: 10.5194/os-14-1093-2018, 2018.

Li Z. J., Wang Z. Y., Li Y., Zhang Y., Zheng J. J. and Gao S.: Evaluation of global high resolution reanalysis products based on the Chinese Global Oceanography Forecasting System Atmos. Oceanic Sci. Lett. 14(5), 2021.

Liu, C., Xiao, Q., and Wang, B.: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formation and preliminary test. Monthly Weather Review, 136: 3363-3373. DOI: 10.1175/MWR-D-12-00130.1, 2008.

Liu G, Smith GC, Gauthier A-A, Hébert-Pinard C, Perrie W and Shehhi MRA (2024) Assimilation of synthetic and real SWOT observations for the North Atlantic Ocean and Canadian east coast using the regional ice ocean prediction system. Front. Mar. Sci. 11:1456205. doi: 10.3389/fmars.2024.1456205

Lorenc, A.C. (2003), The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. Q.J.R. Meteorol. Soc., 129: 3183-3203. https://doi.org/10.1256/qj.02.132

Morrow, R., Fu, L.-L., Ardhuin, F., Benkiran, M., Chapron, B., Cosme, E., d'Ovidio, F., Farrar, J. T., Gille, S. T., Lapeyre, G., Le Traon, P.-Y., Pascual, A., Ponte, A., Qiu, B., Rascle, N., Ubelmann, C., Wang, J., and Zaron, E. D.: Global Observations of Fine-Scale Ocean Surface Topography With the Surface Water and Ocean Topography (SWOT) Mission, Front. Mar. Sci., 6, 232, https://doi.org/10.3389/fmars.2019.00232, 2019.

Moore, A. M., Martin, M. J., Akella, S., Arango, H. G., Balmaseda, M., Bertino, L., Ciavatta, S., Cornuelle, B., Cummings, J., Frolov, S., Lermusiaux, P., Oddo, P., Oke, P. R., Storto, A., Teruzzi, A., Vidard, A., and Weaver, A. T.: Synthesis of ocean observations using data assimilation for operational real-time and reanalysis systems: a more complete picture of the state of the ocean. Frontiers in Marine Science, 6, 90. https://doi.org/10.3389/fmars.2019.00090, 2019.

Moore A.M., Arango, H.G., Wilkin, J., and C.A. Edwards: Weak constraint 4D-Var data assimilation in the Regional Ocean Modeling System (ROMS) using a saddle-point algorithm: Application to the California Current Circulation. Ocean Modelling, 186, https://doi.org/10.1016/j.ocemod.2023.102262, 2023.

Rahaman, H., Venugopal, T., Penny, S. G., Behringer, D. W., Ravichandran, M., Raju, J. V. S., … Sengupta, D.: Improved ocean analysis for the Indian Ocean. Journal of Operational Oceanography, 12(1), 16–33. https://doi.org/10.1080/1755876X.2018.1547261, 2018.

Moore, A.M., H.G. Arango, E. Di Lorenzo, B.D. Cornuelle, A.J. Miller, D.J. Neilson. A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. Ocean Modelling, 7 (2004), 227-258

Ravichandran, M., D. Behringer, S. Sivareddy, M.S. Girishkumar, N. Chacko, R. Harikumar: Evaluation of the global ocean data assimilation system at INCOIS: the tropical Indian Ocean. Ocean Modelling, 69, pp. 123-135, https://doi.org/10.1016/j.ocemod.2013.05.003, 2013.

Röhrs, J., Gusdal, Y., Rikardsen, E. S. U., Durán Moro, M., Brændshøi, J., Kristensen, N. M., Fritzner, S., Wang, K., Sperrevik, A. K., Idžanović, M., Lavergne, T., Debernard, J. B., and Christensen, K. H.: Barents-2.5km v2.0: an operational data-assimilative coupled ocean and sea ice ensemble prediction model for the Barents Sea and Svalbard, Geosci. Model Dev., 16, 5401–5426, https://doi.org/10.5194/gmd-16-5401-2023, 2023.

Sakov, P., Oliver, D. S., & Bertino, L. (2012a). An Iterative EnKF for Strongly Nonlinear Systems. In *Monthly Weather Review* (Vol. 140, Issue 6, pp. 1988–2004). https://doi.org/10.1175/MWR-D-11-00176.1

Sakov, P., Counillon, F., Bertino, L., Lisæter, K. A., Oke, P. R., & Korablev, A. (2012b). TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci.*, *8*(4), 633–656. https://doi.org/10.5194/os-8-633-2012

Sakov, P., Evensen, G., & Bertino, L.: Asynchronous data assimilation with the Ensemble Kalman Filter. *Tellus A*, *62A*, 24–29. https://doi.org/10.3402/tellusa.v62i1.15655, 2010.

Smith, G.C., Roy, F., Reszka, M., Surcel Colan, D., He, Z., Deacu, D., Belanger, J.-M., Skachko, S., Liu, Y., Dupont, F., Lemieux, J.-F., Beaudoin, C., Tranchant, B., Drévillon, M., Garric, G., Testut, C.-E., Lellouche, J.-M., Pellerin, P., Ritchie, H., Lu, Y., Davidson, F., Buehner, M., Caya, A. and Lajoie, M.: Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System. Q.J.R. Meteorol. Soc., 142: 659-671. https://doi.org/10.1002/qj.2555, 2016.

Smith, G. C., Liu, Y., Benkiran, M., Chikhar, K., Surcel Colan, D., Gauthier, A.-A., Testut, C.-E., Dupont, F., Lei, J., Roy, F., Lemieux, J.-F., and Davidson, F.: The Regional Ice Ocean Prediction System v2: a pan-Canadian ocean analysis system using an online tidal harmonic analysis, Geosci. Model Dev., 14, 1445–1467, https://doi.org/10.5194/gmd-14-1445-2021, 2021.

Stammer, D., Balmaseda, M., Heimbach, P., Köhl, A. and Weaver, A.: Ocean Data Assimilation in Support of Climate Applications: Status and Perspectives. Annual Review of Marine Science, 8(1), 1–28. https://doi.org/10.1146/annurev-marine-122414-034113, 2016.

Storto, A., Masina, S. and Navarra, A.: Evaluation of the CMCC eddy-permitting global ocean physical reanalysis system (C-GLORS, 1982–2012) and its assimilation components. Q.J.R. Meteorol. Soc., 142: 738-758. https://doi.org/10.1002/qj.2673, 2016.

Storto, A., Alvera-Azcárate A., Balmaseda, M.A., Barth, A., Chevallier, M., Counillon, F., Domingues, C.M., Drevillon, M., Drillet, Y., Forget, G., Garric, G., Haines, K., Hernandez, F., Iovino, D., Jackson, L.C., Lellouche, J.-M., Masina, S., Mayer, M., Oke, P.R. Penny, S.G., Peterson, K.A., Yang, C. and Zuo, H.: Ocean Reanalyses: Recent Advances and Unsolved Challenges. Frontiers in Marine Science, 6, 418. https://doi.org/10.3389/fmars.2019.00418, 2019.

Thoppil, P.G., Frolov, S., Rowley, C.D. et al. Ensemble forecasting greatly expands the prediction horizon for ocean mesoscale variability. Commun Earth Environ 2, 89 (2021). https://doi.org/10.1038/s43247-021-00151-5

Van Leeuwen PJ, Künsch HR, Nerger L, Potthast R, Reich S. Particle filters for high-dimensional geoscience applications: A review. *Q J R Meteorol Soc*. 2019; 145: 2335–2365. https://doi.org/10.1002/qj.3551~~van Leeuwen, P.J.: Nonlinear Data~~

Assimilation for high-dimensional systems. In: Nonlinear Data Assimilation. Frontiers in Applied Dynamical Systems: Reviews and Tutorials, 2, Springer, Cham. https://doi.org/10.1007/978-3-319-18347-3_1, 2015.

Sakov, P., Oliver, D., Bertino, L.: An iterative EnKF for strongly nonlinear systems. Mon. Weather Rev. 140: 1988–2004, 2012.

Vetra-Carvalho, S., van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., … Beckers, J. M. (2018). State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. Tellus A: Dynamic Meteorology and Oceanography, 70(1), 1–43. https://doi.org/10.1080/16000870.2018.1445364

Vidard, A., Bouttier, P.-A., and Vigilant, F.: NEMOTAM: tangent and adjoint models for the ocean modelling platform NEMO, Geosci. Model Dev., 8, 1245–1257, https://doi.org/10.5194/gmd-8-1245-2015, 2015.

Weaver, A. T., Vialard, J., and Anderson, D.L.T.: Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part I: Formulation, Internal Diagnostics, and Consistency Checks. Monthly Weather Review, 131, 1360-1378. https://doi.org/10.1175/1520-0493(2003)131<1360:TAFVAW>2.0.CO;2, 2003.

Yaremchuk, M.; Beattie, C.; Panteleev, G.; D'Addezio, J. Block-Circulant Approximation of the Precision Matrix for Assimilating SWOT Altimetry Data. Remote Sens. 2024, 16, 1954. https://doi.org/10.3390/rs16111954

Ying, Y. (2019). A multiscale alignment method for ensemble filtering with displacement errors. *Monthly Weather Review*, *147*(12), 4553–4565. https://doi.org/10.1175/MWR-D-19-0170.1

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment, Ocean Sci., 15, 779–808, https://doi.org/10.5194/os-15-779-2019, 2019.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Data and/or code availability

No data or code were used to produce the manuscript, but a list of data assimilation software is provided in Table 1, together with its availability.

Authors contribution

All authors contributed to the writing and review of the manuscript. MM organised the manuscript and led the writing of sections 1, ~~and~~ 6 and 7. IH led the writing of section 2. LB led the writing of section 3. AM led the writing of sections 4 and 5.

Acknowledgements

The authors would like to thank the compilation team of the OceanPrediction Decade Collaboration Centre who organised this special issue. We would also like to thank the reviewers for their suggestions to improve the manuscript.