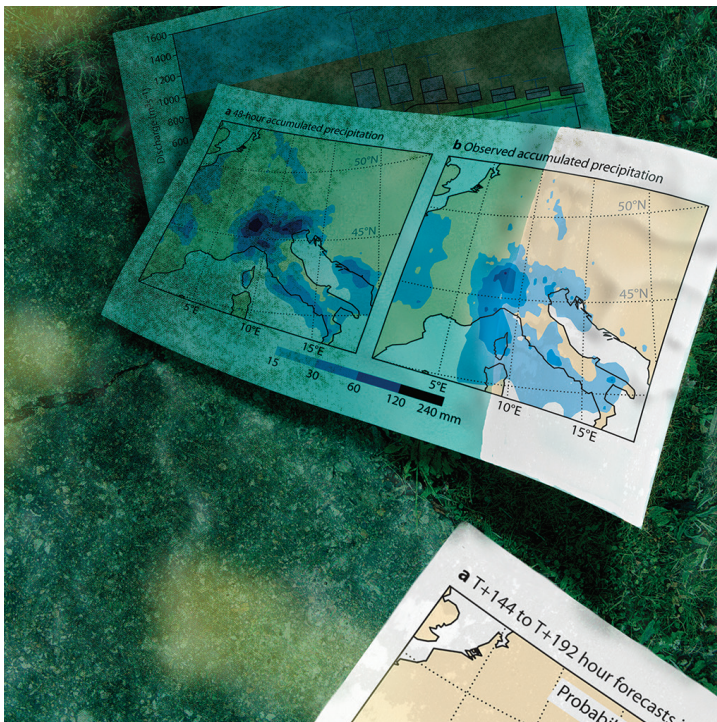


METEOROLOGY

NEMOVAR: A variational data assimilation system for the NEMO ocean model



This article appeared in the Meteorology section of ECMWF Newsletter No. 120 – Summer 2009, pp. 17–21.

NEMOVAR: A variational data assimilation system for the NEMO ocean model

Kristian S. Mogensen, Magdalena Alonso Balmaseda,
Anthony Weaver, Matthew Martin, Arthur Vidard

The ocean-atmosphere coupled model used for monthly and seasonal forecasts requires the modelling and initialization of the three-dimensional global ocean. To this end, ECMWF has been using the HOPE (Hamburg Ocean Primitive Equation) ocean model and an Optimal Interpolation (OI) data assimilation system since the beginning of the coupled-model forecasting activities in 1995. However, with the consolidation of a large part of the European ocean modelling activities around NEMO (Nucleus for European Modelling of the Ocean; Madec, 2008), it was decided that ECMWF should adopt NEMO as the ocean component of its coupled forecasting system.

NEMO is a comprehensive ocean modelling framework consisting of a dynamical ocean model (OPA), sea ice model (LIM) and biochemistry models. ECMWF is primarily concerned with the ocean component of the NEMO system, but is planning to investigate the use of the LIM sea ice model in the future. The next operational seasonal forecast system (System 4) will use the atmospheric model from the Integrated Forecast System (IFS) linked to NEMO via the OASIS coupler that is used to exchange relevant fields. The same coupled models will also be used for monthly forecasting within the ECMWF Ensemble Prediction System (EPS).

Part of the transition from HOPE to NEMO involves setting up a new system for providing ocean initial conditions. Rather than adopting the current OI scheme used by HOPE, it was decided to pursue the development of a variational-based assimilation system (NEMOVAR) in collaboration with CERFACS (European Centre for Research and Advanced Training in Scientific Computation) and the Met Office. Other research institutes, such as INRIA/LJK (Institut National de Recherche en Informatique et Automatique/ Laboratoire Jean Kuntzmann) based in Grenoble, are also making significant contributions to NEMOVAR. Though we usually refer to this system as NEMOVAR, at present it only applies to the dynamical ocean part (OPA) of the NEMO system. This article describes the implementation plan and current status of NEMOVAR.

Development of NEMOVAR

The basis for the NEMOVAR developments is OPAVAR (Weaver et al., 2005), a variational assimilation system developed for a previous version of OPA. The OPAVAR system has been used for several studies as well as European research projects such as DEMETER, ENACT and ENSEMBLES. It supports 4D-Var as well as 3D-Var since it includes the tangent-linear and adjoint codes of the previous version of OPA. It has mostly been used for low-resolution ($\sim 2^\circ$) studies of the global ocean. It is based on an incremental algorithm with outer and inner ‘loops’, similar to that of the IFS 4D-Var. These concepts will be discussed later.

OPAVAR has three major drawbacks.

- It is based on an obsolete version of OPA that is no longer maintained as part of NEMO.
- It is not straightforward to adapt to different ocean model resolutions.
- It cannot be run using distributed memory parallelization.

These points are severe limitations when porting the code to future computer architectures and for higher resolution applications.

For these reasons it was decided to develop NEMOVAR. In the following we summarize the key features of OPAVAR and describe the implementation of NEMOVAR.

AFFILIATIONS

Kristian S. Mogensen, Magdalena Alonso Balmaseda, ECMWF, Reading, UK

Anthony Weaver, CERFACS, Toulouse, France

Matthew Martin, Met Office, Exeter, UK

Arthur Vidard, INRIA/LJK, Grenoble, France

The OPAVAR legacy

OPAVAR is able to assimilate observations from sub-surface profiles of temperature and salinity, along-track sea level anomaly (SLA) observations from satellite altimeters, and sea surface temperature (SST) maps. It has a multivariate background-error formulation for temperature, salinity, horizontal components of velocity, and sea surface height. The multivariate formulation captures the local relationship between temperature and salinity obtained by imposing approximate preservation of water masses, an equation of state relating temperature and salinity to density, geostrophic adjustment of the horizontal components of velocity, and local relationships between sea surface height and vertical density profiles. The details can be found in Weaver et al. (2005) and references therein.

Like most assimilation systems, OPAVAR uses a mechanism where observations within a given time window are collected to compute an analysis. The analysis is obtained by combining observations with a first-guess (background) field, i.e. the previous analysis propagated to the current analysis time using the model. The details are given in Box A.

Development of the current version of NEMOVAR

For the development of NEMOVAR the following tasks were identified:

- Development of observation operators.
- Development of an assimilation system that is based on 3D-Var.
- Implementation of an online, automatic system for the quality control of real-time observations.
- Development of a system to produce an ensemble of analyses to provide estimates of the uncertainty in the ocean initial conditions.
- Development of the tangent-linear and adjoint models of the ocean component of NEMO.
- Development of an assimilation system that is based on 4D-Var.

The first four tasks are needed for the initial operational implementation of NEMOVAR at ECMWF, which in addition will require other developments (see later). Some of the tasks are not completely independent. The 3D-Var system needs the observation operators in order to compute the cost function to be minimized. However, other components such as the quality control of observations and the NEMO tangent-linear and adjoint models are independent and allow for parallel developments by different groups.

The observation operators in NEMO are central to NEMOVAR but are also valuable for model validation since they allow online comparison with a variety of datasets. Presently, observation operators have been developed for temperature and salinity profiles, SLA from satellite altimeters, SST (both from maps and from individual measurements), and currents (e.g. from the TAO (Tropical Atmosphere Ocean) dataset). The observation operator for currents is a new development that is not available in OPAVAR.

The NEMOVAR 3D-Var system has reached a fairly mature state. The first assimilation experiments for extended periods have been performed by assimilating temperature and salinity data from the quality controlled EN3 dataset provided by the Met Office as part of ENSEMBLES (an EU-funded project that aims to construct scenarios of future climate change with ensemble simulations of Earth-system models). This new dataset includes corrections to expendable bathythermographs (XBTs) to compensate for errors in their fall-rates. Some of the results from these experiments will be discussed below.

Work is ongoing to improve the specification of the observation-error and background-error covariances and the models used to represent them in NEMOVAR. In addition, work is being carried out to implement a multivariate model bias correction scheme similar to the one that exists in the current HOPE/OI (System 3). The basic modules for assimilation of SLA and SST data are in place, and testing is underway.

The work on the tangent-linear and adjoint of the NEMO model is in an advanced state and development of a prototype 4D-Var system will start later this year.

The quality control system used by the Met Office (*Ingleby & Huddleston, 2007*) for their operational ocean forecasting system and for producing the ENSEMBLES datasets has been implemented in the NEMOVAR system. The system is currently used for temperature and salinity profiles and will be extended to include SLA and SST data.

Results from some preliminary experiments will now be described.

Variational algorithm implemented in OPAVAR and NEMOVAR

A

The variational algorithm implemented in OPAVAR, which has been migrated to NEMOVAR, consists of the following steps:

- An initialization phase which computes the initial (background) model trajectory and the misfits between the observations and this trajectory. This model integration uses the model state at the end of the previous cycle as the starting point.
- One or several outer loops consisting of:
 - An inner loop that minimizes a quadratic cost function to produce a set of increments (changes to the model field). This cost function depends on the trajectory and the misfits.
 - An update of the trajectory and misfits, obtained by applying the increments to the model. The updated misfits and trajectory can then be used as input for another minimization if multiple outer loops are required.

The analysis is formed by adding the resulting increments to the first guess. This can be performed either directly or by using Incremental Analysis Updates (IAU) where the increments are applied gradually over several time steps. The IAU approach is typically used in 3D-Var to reduce spurious high-frequency adjustment processes resulting from the initialization procedure.

Setup of the experiments

The next ECMWF seasonal forecasting system, based on NEMO, will use a 1° global configuration. This configuration, known as ORCA1, has been chosen for the initial experimentation with NEMOVAR.

The impact of assimilation is assessed by comparing a 3D-Var experiment with an experiment without data assimilation (the control). Both experiments are forced with daily atmospheric data from the ERA-Interim reanalysis. They start with the same initial conditions, obtained after 20 years of spin-up using daily climatological data from ERA-Interim. The model SST is strongly relaxed to external SST analyses (NOAA SST OIv2), and there is a three-year time-scale damping to the climatology of temperature and salinity from the World Ocean Atlas. Rather than using the NEMO ice model, these experiments prescribe the ice cover. The experiments start from an initial state on January 1989 and continue until the end of 2006.

For the control experiment, the model equivalent of the temperature and salinity observations were computed for diagnostic purposes only, whereas for the 3D-Var experiment both the temperature and salinity data were assimilated. An assimilation window of 10 days was used. The 3D-Var experiment was performed with a single outer loop and a fixed set of 40 iterations used for minimization in the inner loop. With the settings used for this experiment the computational cost of the inner loop is roughly equivalent to a 10-day model integration. This means that two thirds of the computer time is spent on integrating the model for the 3D-Var system since two model integrations are needed: one for the background trajectory initialization and one for the application of the analysis increment. For different configurations this cost ratio might change.

Statistics of the model fit to the observations

Figure 1 shows the root-mean-square (rms) error and mean error (or bias) of the model for (a) potential temperature and (b) salinity observations as a function of depth for the global ocean and for a region in the central Equatorial Pacific (NINO3.4: 5°S–5°N, 170°–120°W). Shown are the values for the control experiment and for both the outer and inner loops of the 3D-Var experiment. The biases and rms errors for the 3D-Var experiment are clearly reduced compared to the control experiment.

The bias between the analysis and the observations in the inner loop is very close to zero at all depths. The outer loop results indicate the fit between the model background and observations in the assimilation window before the analysis is performed. If either the model background or observations are biased, the bias in the outer loop will not be zero – this is clearly the case from Figure 1. As expected the rms error of the inner loop departures is lower than that of the outer loop since we are fitting the observations in the analysis. A large discrepancy between the outer and inner loop fits to the observations is an indication that the model is having difficulty in retaining the information brought by assimilation. Figure 1 shows that this is the case for the central regional of the Equatorial Pacific.

The rms error and bias statistics only give part of the picture. Figure 2 shows histograms of the temperature and salinity departures (model minus observations) for the global ocean and for the central Equatorial Pacific for the two experiments. The histograms reveal that for the NEMO model (both with and without assimilation) there is a large warm tail compared to the observations at a depth of 98 metres. In the deeper ocean (below approximately 450 metres), there is a large cold tail in the histograms (not shown), indicating that the thermocline in the model is too diffuse. The histograms for the inner loop are very well centred, but in the outer loop a tail towards the warm side develops. Similarly, for salinity the histogram for the outer loop has a distribution that is too salty.

The surface temperature is well constrained by the strong relaxation to the external Olv2 SST analyses. It is found that the biases in salinity are largest at the surface since there is no such constraint on surface salinity. The results also suggest that the assimilation is less effective in the central Equatorial Pacific compared to the global ocean. In the central Equatorial Pacific the rms error of temperature and salinity grows rapidly from the inner loop to the outer loop (i.e. the analysis and forecasts are inconsistent). In the case of salinity, the information assimilated into the system in a given cycle is lost in the next cycle (see Figure 1b).

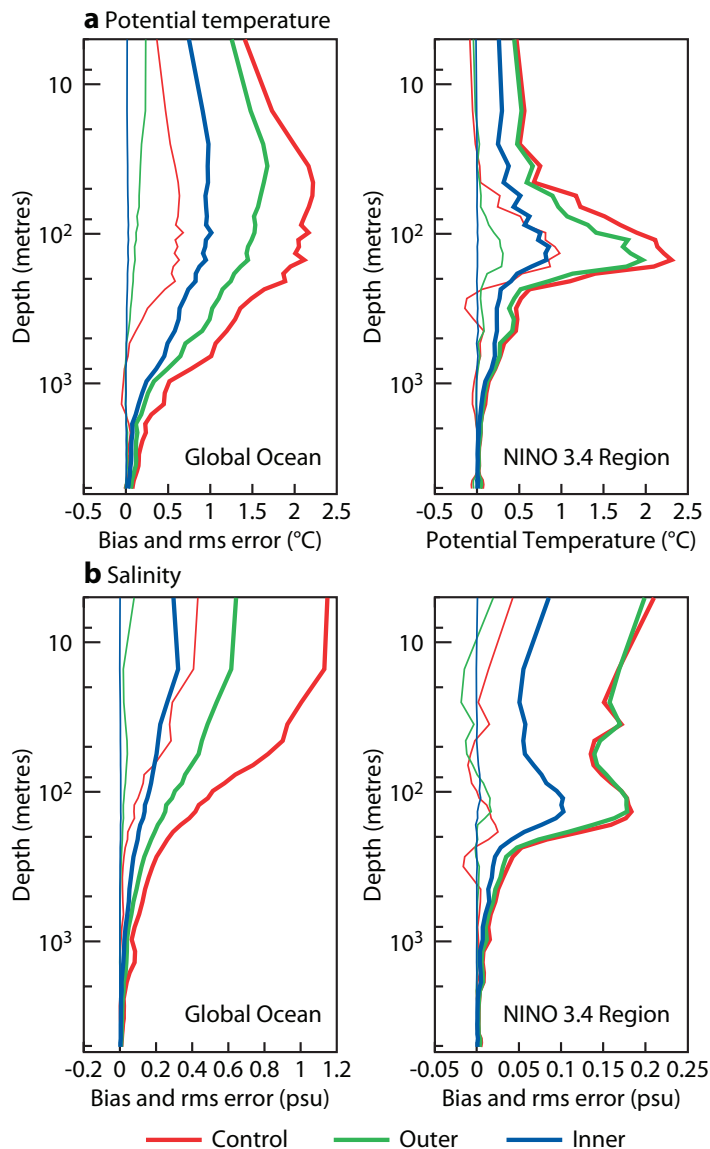


Figure 1 Vertical profiles of the rms error (thick lines) and bias (thin lines) between model and observations for (a) potential temperature (°C) and (b) salinity (practical salinity units, psu) for the global ocean using all observations (left panels) and for observations in the central Equatorial Pacific (NINO3.4) region only (right panels). In the panels, ‘control’ corresponds to the control experiment, ‘outer’ corresponds to the outer loop of the 3D-Var experiment, and ‘inner’ to the inner loop of the 3D-Var experiment. The observations are taken from the EN3 dataset.

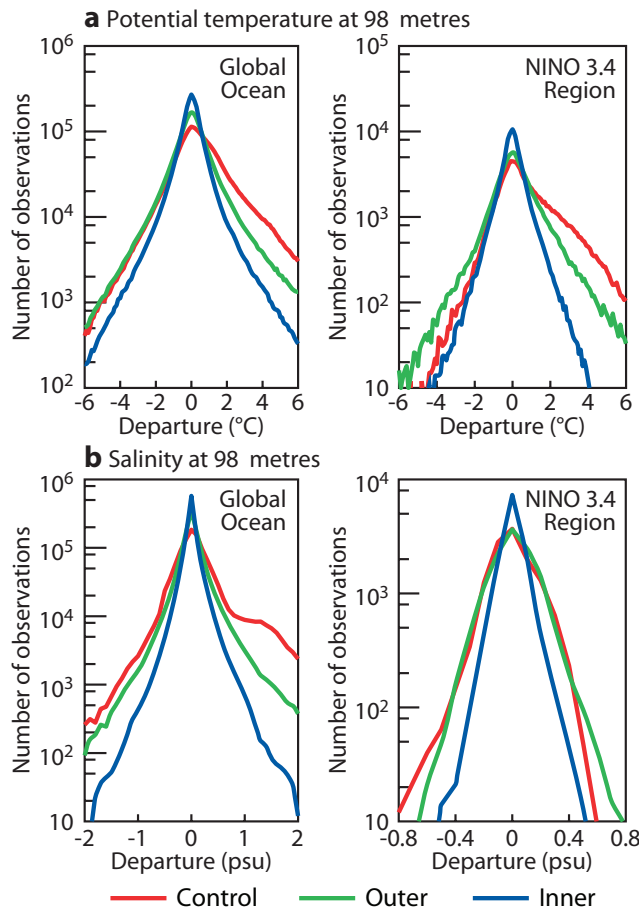


Figure 2 Distribution of model-observation departures at 98 metres depth for (a) potential temperature ($^{\circ}\text{C}$) and (b) salinity (psu) for the global ocean using all observations (left panels) and for observations in the central Equatorial Pacific (NINO3.4) region only (right panels). In the panels ‘control’ corresponds to the control run, ‘outer’ corresponds to the outer loop of the 3D-Var experiment and ‘inner’ to the inner loop of the 3D-Var experiment. The observations are taken from the EN3 dataset.

Geographical distribution of errors

To investigate the geographical distribution of the errors, the temporal mean of the increments for temperature (Figure 3a) and salinity (Figure 3b) at a depth of 98 metres are considered. These results show several interesting features. The largest increments occur in the area of the western boundary currents and in the tropical ocean (within 10° of the equator). The assimilation tries to compensate for an inaccurate position of the boundary currents which the ocean model, with a resolution of 1° , is unable to represent correctly. Especially noticeable is the dipolar structure of the increments in the Gulf Stream separation area. The assimilation is trying to increase the gradient across the Gulf Stream by adding cool/fresh water to the north of the current and warm/salty water to the south.

In the tropics the story is different. The ocean model has enough resolution to resolve the equatorial dynamics, and therefore the source of model error is likely to reside somewhere else. To investigate this in greater detail, a cross-section of the temperature increments at the equator appears in Figure 4a, together with the temperature differences between the control and assimilation experiments in Figure 4b. In the Pacific the assimilation is trying to steepen the east-west gradient of the thermocline (deepening it in the west and shallowing it in the east). In this experiment, equatorial currents are not improved by assimilating temperature and salinity data (not shown). In the western Pacific they can even be degraded. This aspect should be improved with the introduction of the model bias correction scheme.

The east-west dipolar structure of the assimilation increment contrasts with the mean differences between control and assimilation in Figure 4b, which does not show any east-west dipole: in all the basins, the control experiment is too warm above the thermocline and too cold below it. These results demonstrate that the effect of the assimilation is non-local. How the increments in Figure 4a produce the differences shown in Figure 4b is not obvious. It could be achieved by a variety of processes. For instance, in the Equatorial Atlantic, direct cooling by the assimilation in the western part of the basin, and subsequent advective transport of the negative increments via the equatorial undercurrent, could lead to a basin-wide cooling. Another possible reason is the effect on the vertical mixing: the warm increment in the eastern part of the Indian Ocean around 100 metres (just above the thermocline) could increase the stratification, resulting in reduced vertical mixing of heat down to the thermocline. A more careful diagnostic study of the different processes involved in the reduction of the error can help to diagnose model errors.

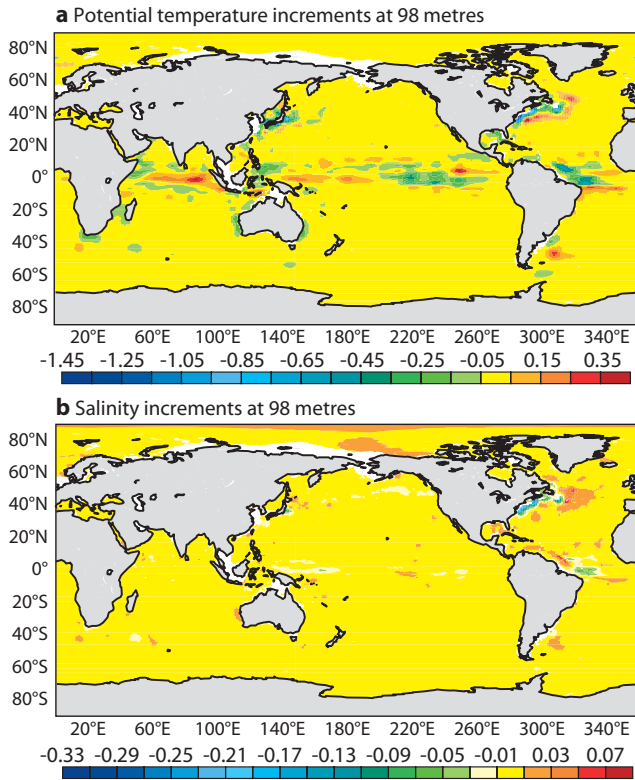


Figure 3 Temporal mean of (a) potential temperature increments (°C) and (b) salinity increments (psu) at 98 metres for all 10-day windows.

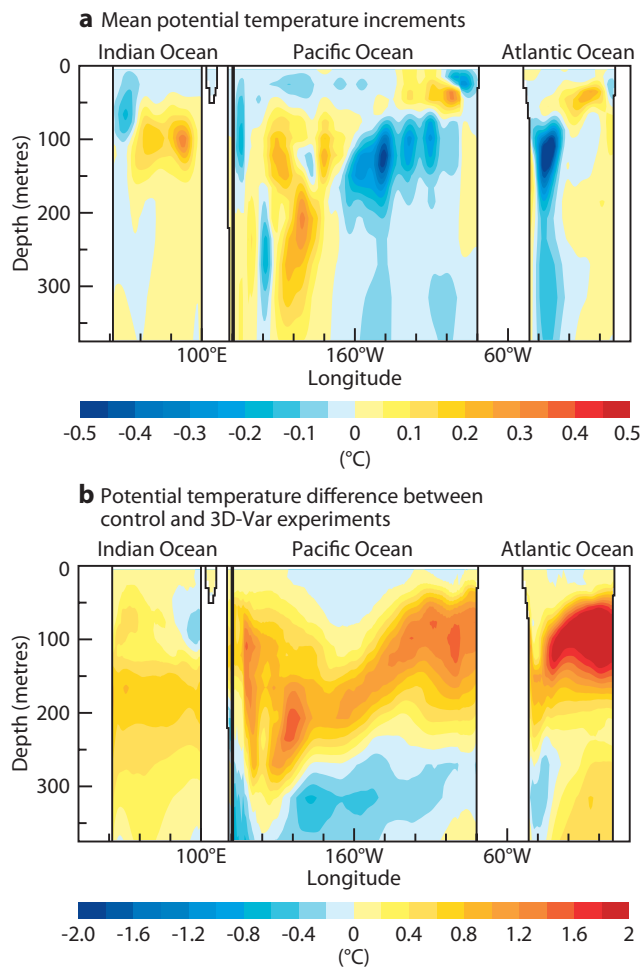


Figure 4 Equatorial cross-section of (a) the temporal mean of all 10-day window potential temperature increments (°C) and (b) the differences in mean potential temperature (°C) between the control and the 3D-Var experiments. The three ocean basins (Indian, Pacific, Atlantic) are illustrated. The averages have been computed for the period of 1989 to 2006. The different spatial structures of the assimilation increment and mean difference demonstrate the non-local effect of the assimilation.

Conclusions and outlook

The preliminary results with the NEMOVAR 3D-Var system are encouraging, but also highlight limitations with the system. Though the assimilation system works correctly according to the assumptions of no model or observation bias, we have demonstrated this is not a valid assumption. The present HOPE OI assimilation system accounts for model bias and the initial results with NEMOVAR seem to indicate that this feature is needed in NEMOVAR as well. Use of information from the assimilation system can be of value in diagnosing model errors and eventually helping to improve the ocean models.

ECMWF plans to use the 3D-Var version of NEMOVAR to produce initial conditions for its coupled ocean-atmosphere modelling system. In the first operational implementation, NEMOVAR will assimilate temperature, salinity and sea level anomaly data in a global ocean configuration at a 1° resolution (the ORCA1 grid mentioned earlier). A first historical reanalysis based on NEMOVAR for the period 1959 to present will be available to selected users in the autumn of 2009.

Several technical and scientific developments are needed for the operational implementation of NEMOVAR, such as the model bias correction scheme, the real-time quality control of the observations, and the assimilation of altimeter-derived sea level anomalies. The operational implementation of the real-time ocean analyses using NEMOVAR is expected by the spring of 2010.

The Met Office plans to use NEMOVAR for their ocean assimilation activities covering short-range ocean forecasting applications, based on the FOAM (Forecasting Ocean Assimilation Model) system, and seasonal forecasting. The FOAM system runs at two different resolutions including a ¼° global configuration with nested 1/12° regional configurations. NEMOVAR will be implemented in these systems with a focus on the initialisation of mesoscale eddies, and it is planned to use the upgraded systems in the MyOcean project. This project aims to set up the infrastructures and services in preparation for the GMES (Global Monitoring for Environment and Security) Marine Services.

Several algorithmic improvements and extensions to NEMOVAR are being made at CERFACS and INRIA/LJK. Multi-incremental versions of NEMOVAR are being developed to allow for reduced resolution in the inner loop. This feature will be particularly important for high-resolution applications of 4D-Var. Work is also being done to improve the specified background-error statistics. The long-term development path of NEMOVAR is to a weak-constraint 4D-Var system in which model error is explicitly accounted for in the 4D-Var algorithm.

NEMOVAR will eventually be released under an open source license and made available to a wider community. The precise details for distribution of the NEMOVAR system are currently being discussed.

The development of NEMOVAR has been an effective collaborative international project with many people from various organisations making significant contributions. These developments have also benefited from the support of the NEMO consortium. It is expected that continuing international collaboration will lead to further developments in ocean modelling and data assimilation thereby enhancing the quality of monthly and seasonal forecasts.

Further reading

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

Madec, G., 2008: NEMO ocean engine. *Note du Pôle de Modélisation, Institut Pierre-Simon Laplace (IPSL)*, No 27.

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

© Copyright 2016

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, England

The content of this Newsletter article is available for use under a Creative Commons Attribution-Non-Commercial-No-Derivatives-4.0-Unported Licence. See the terms at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error or omission or for loss or damage arising from its use.