Options for development of a NEMO Assimilation Component

Mike Bell, Claire Levy, Matt Martin, Kristian Mogensen, Arthur Vidard, Charles-Emmanuel Testut, Anthony Weaver

13 November 2009

Summary

1. Introduction

A meeting between experts engaged in developing data assimilation tools for the NEMO system and members of the NEMO System Team was held at the Institut d'Astrophysique, Paris on 22-23 June 2009. Its overall objective was to design a strategy for implementation of an assimilation component within the NEMO code system and thereby make assimilation tools for NEMO more readily available to the user community.

The meeting participants (listed in appendix A) identified five separate components of the assimilation tools and analysed each of them in some detail in order to propose which assimilation components should be introduced first into the NEMO system and to suggest a sustainable road-map for the longer term development of assimilation components for NEMO.

This report summarises the main points of this analysis and the recommendations which emerged from it. It is intended to be discussed with experts from other groups who were unable to make the meeting (in particular with the INGV, LEGI and Canadian groups), to be discussed by the NEMO Developers Committee and to be presented to the NEMO and NEMOVAR Steering Committees for their consideration.

It should be noted that if these assimilation tools were made readily available as part of the NEMO code that they should be valuable to most users of the NEMO system, not just a few assimilation experts. For example, they can be used to diagnose the differences between model simulations and ocean measurements and to provide valuable information on the robustness of numerical schemes. They can also be used to improve the balance in initial conditions, particularly for shelf-seas models, and to assess the structures of the fastest growing instabilities in any flows.

It is also worth noting that the proposed short-term steps would establish formats for the main inputs to and outputs from the assimilation systems (e.g. of the files containing the observations) which could evolve into standard interfaces or formats for the NEMO system and, in the longer-term, common software tools for handling them. This could greatly facilitate the long-term development of collaboration between groups which are currently working on entirely separate data assimilation code systems.

The work-plan for the incorporation of assimilation components into NEMO should take into account the standards set and lessons learnt by the NEMO System Team during their development of a shared, sustainable, state-of-the-art scientific code system. All code within the NEMO system must be **freely available** under a CeCILL licence agreement. It must also follow the NEMO coding conventions and include in-line and off-line documentation. New code components must be provided with adjustments to the standard reference configurations used by the NEMO System Team. These reference configurations are used to ensure that existing capabilities are safeguarded as the code evolves. They also enable users to test their installation of the NEMO software and provide simple examples illustrating the system's capabilities. Finally, before new code components are included in the NEMO system, additional human resources within the NEMO System Team, with the time and expertise needed to update and develop it, must be identified and agreed.

Each of the following sections of this report is devoted to one of the identified assimilation components. The first two sections address codes which do not read, and in that sense are not dependent on, files containing observations. Section 2 considers the code for the addition of assimilation increments to the model and section 3 addresses the tangent-linear and adjoint codes for the NEMO model. Section 4 discusses the « observation operator » which reads model restart files and files containing observations, calculates the model values corresponding to the observations and outputs these values within the observation files. Sections 5 and 6 consider other codes which make use of files containing observations. Section 5 considers code for quality control of observations and section 6 considers the « analysis kernel » which calculates the assimilation increments to the model from the differences between the observations and their model equivalents.

These sections provide a simple description of the component, discuss the options available, the advantages to be gained from introducing the code into the NEMO system, and hence the relative priority for doing so, the dependencies and interfaces of the component, its requirements for resources (man hours), and the adjustments it requires to the standard NEMO test configurations.

The final section summarises the main conclusions and recommendations to emerge from this analysis and outlines the work required to undertake initial steps towards the recommendations.

2. Application of increment

In most of the assimilation methods relevant to NEMO, the trajectory is controlled by introducing a correction to the model state. This correction or increment is produced by the analysis step on variables linked to a control vector. The increment updates the trajectory of the model either directly (at the model time step equivalent to the analysis date), or in a gradual manner over a time period around the analysis date. This latter approach, usually referred to as the Incremental Analysis Updates (IAU) methodology (Bloom et al. 1996), tends to reduce shock in the model restart stage and to minimize spurious adjustment processes.

From a technical point of view, this latter concept of a correction introduced progressively during the model run is not closely linked to the assimilation methods and could also be seen or interpreted as a use of forcing terms. In practice, these forcing terms correspond to 2D and/or 3D fields (e.g. temperature, salinity, or currents) and are applied directly as additional terms in the equations of the NEMO code. Several issues must be considered with this kind of implementation: I/O requirements; memory overhead; and physical consistency of the forcing terms.

At present, several different implementations (NEMOVAR, SAM,...) are available. The present version of NEMOVAR includes both direct initialisation for 4D-Var applications and the IAU approach for 3D-Var applications. The SAM code developed at Mercator is relatively close to the NEMOVAR implementation for the internal management of the forcing terms, because of their common origin. However, new specificities have been introduced as new fields (ice concentration) or new IAU weight functions. The memory overhead has also been limited using dynamic memory allocation. Note that SAM is also able to manage several corrections. Historically, a direct initialisation also exists in the SAM code but it is no longer maintained. The SAM and NEMOVAR codes differ significantly in their I/O interfaces due to the use of different formats for the exchange with the analysis step and the storage. The SAM implementation uses packed file formats and private libraries developed by the MERCATOR group (i.e. not under a CECILL licence agreement) while the others implementations use a more standard NetCDF file format.

For the internal management of the correction in the NEMO code, the benefits of a convergence of the present codes followed by a shared development are relatively evident in particular for validation or when new data types will be introduced. However, for the I/O component of the increment in the NEMO code, unification of the loading system is not really conceivable because of the private features of the SAM loading libraries.

Within the framework of the existing developments, a pragmatic work-plan could be to start with one of these two codes by integrating it directly and to progressively update it with the main specificities of the second. At present, the NEMOVAR code is sufficiently compatible with the NEMO system to be a good starting point. It uses the NEMO coding convention and it seems that it is not necessary to rewrite the existing routines except, perhaps, for the introduction of a specific key to separate the new code from the rest of the NEMO system. With the NEMOVAR code, I/O can use standard NetCDF file formats, even if some routines should be probably updated during the integration phase in the NEMO reference version. After this first release, a second step could be to modify and update the initial version with some of the missing features which are available and necessary in the SAM2 code like the possibility of using several increments. The I/O FORTRAN routines should probably be modified in order to be easily removable by users wanting their own I/O interface (as in the SAM implementation). In practice, all these developments should be tested and validated within a NEMO reference configuration. One possibility is to use the ORCA2 global configuration which is a system shared between each developing group.

Even if this new NEMO library is not necessary at present due to the presence of existing codes, it is important to define this work-plan and the integration of these developments in the NEMO reference version. However, it is difficult to evaluate precisely the manpower and the time which is required for these tasks (at least a few months). The first step consisting in an integration of the NEMOVAR code could be a contribution of the Met Office while the second step could be a contribution of the MERCATOR group in collaboration with the other groups to the NEMO Assimilation Component.

References

Bloom, S. C., L. L. Takacs, A. M. DaSilva and D. Levina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256-1271.

3. Tangent and Adjoint Model

(Note: for more information about the NEMOTAM development strategy refer to the VODA deliverable D1.1 available at http://voda.gforge.inria.fr/docs/D1.1.pdf)

Tangent-linear and Adjoint Models (TAM) are widely used for variational assimilation applications, but they are also powerful tools for the analysis of physical processes, since they can be used for sensitivity analysis, parameter identification and for the computation of characteristic vectors (singular vectors, Lyapunov vectors, etc.).

NEMOTAM stands for NEMO Tangent-linear and Adjoint Models and is currently an implementation of the TAM for the ocean component of NEMO. As far as we know, this is the only one currently available. Strictly speaking, the NEMO components being not entirely differentiable, it represents *an approximate* tangent-linear model and *its* adjoint for NEMO. For the short term, only the TAM of the dynamical core of the NEMO ocean engine is being developed, but it can be extended to other components of NEMO in the future.

Note that on its own the TAM is not enough: for most of its applications a minimisation environment is required. In the short term such a minimisation environment is not planned to be included with NEMOTAM and the user will have to provide it himself. However, NEMOVAR includes a minimisation framework. Therefore, if at some point NEMOVAR is included in NEMO, it may be used for other applications of NEMOTAM as well.

The only interface required between the direct model and the TAM is the handling of the non-linear trajectory. The addition of a dedicated module in NEMO will be needed. This module is already present in NEMOVAR and is consistent with the NEMO standards. Its implementation in NEMO is then straightforward.

Having NEMOTAM as part of NEMO is important: it will encourage good development practices for the direct model in order to ease the TAM coding and since a modification in the direct model is likely to induce one in the TAM, it will help the TAM to stay consistent with the latest version of the direct model. Due to a lack of resources it is currently not possible to release a new version of TAM at the same time as the standard NEMO version. This puts NEMOTAM at risk of being completely out-of-phase with NEMO in a near future, reducing the interest of such a tool. A possible solution would be to freeze the NEMO developments (at least where a NEMOTAM counterpart is available) while the corresponding release of NEMOTAM is finalized (typically a few weeks). Meanwhile the NEMO-Team could focus on other parts of the code (ice, bio, …), the reference runs, code optimisations, etc. This could be a short/medium term solution; ideally more human resources are needed. Since it is meant to be part of NEMO, NEMOTAM should be validated with the standard NEMO configurations (restricted for now to GYRE and ORCA2 since LIM and LOBSTER are not available in NEMOTAM yet).

NEMOTAM developments are currently centralized in LJK in the framework of the VODA ANR project in close collaboration with the NEMO-Team and other partners of this project (LOCEAN, CERFACS, ECMWF). Once again, this is a short term solution, once the VODA project is over, one will have to find additional resources.

After investigations about the feasibility of using automatic differentiation (AutoDiff) tools to produce the TAM code, it has been decided to follow the hand-coding path, AutoDiff tools not being able to fulfil all our requirements. AutoDiff might be an option in the long term but it would require substantial modifications and strong support from their developers.

4. Observation operator

The main purpose of observation operators is to transform model variables on grid points into observable quantities which can be directly compared with observations. For example, for in situ observations this can simply be interpolation from the model grid to observation points, but more complicated operators involving non-linear transformations of model variables to produce the observable quantity are also possible. Observation operators are needed for assimilation and are for most cases independent of the assimilation kernel used. However, they are also valuable as a diagnostic tool for evaluation of model performance, since they provide the possibility of comparing model variables with observations. Both these aspects of use of observation operators should be taken into account for inclusion of this capability into NEMO to facilitate the broadest possible applications.

The model counterparts of the observations can either be computed on-line as part of the model integration, which is the case for FGAT (First Guess at Appropriate Time) like assimilation schemes, or off-line using a standalone tool. The advantage of the former approach is that the model state closest to the observation time can be used directly without large quantities of input/output whereas the advantage of the latter approach is that observational information is not needed at model execution time. An important issue to consider is the I/O requirement. For on-line observation operators the I/O requirement is fairly moderate since the number of observations is smaller than the number of grid points, whereas for off-line observation operators the I/O requirement for model fields can be quite substantial if large differences in time between the observations and the model state are to be avoided. Ideally, the NEMO observation operators should be made available for both on-line and off-line applications.

At the moment several different implementations (NEMOVAR/SAM/INGV/…) of observation operators for NEMO exist. Each of these implementations has their own standards for interfaces for at least the following issues:

1. input files formats and data types;
2. feedback files (input data and model equivalent) format and content;
3. how and where the observation operators are called.

For each of the different implementations the decisions leading to their current form reflects years of investment and the outcome of the discussions of the meeting was that convergence on a joint set of observation operators is not a short-term possibility. However there was also a clear agreement at the meeting that this should not prevent us from making observation operators available in the NEMO reference version. The proposal from the meeting was to put the observation operators from NEMOVAR into the NEMO reference version as a separate module which can be run both on-line and ideally in off-line mode. The code should be put into a separate directory, possibly called NEMO-OBS. The NEMOVAR observation operators presently consists of operators for temperature, salinity, sea level anomaly, sea surface temperature, sea ice concentration and current data.

Since observation operators from the NEMOVAR code are almost ready for implementation into the NEMO reference system, it is expected that the inclusion of these is relatively straightforward for the on-line application of the observation operators. The off-line application will require additional work, but it would not be difficult to do if time is allocated to the task. It was agreed at the meeting that inclusion of the NEMOVAR observation operators should be done as quickly as possible.

The complete integration into the NEMO reference system will require the definition of an "observation operator" reference configuration to enable testing and verification of the implementation. This configuration will require a comprehensive set of ocean observations testing all available observation operators. Data from the ENSEMBLES data set of temperature and salinity observations is a good candidate for reference T/S observations, but also sea level anomalies data from *e.g.* AVISO, SST data from *e.g.* GHRSST and sea-ice data should be included. The "observation operator" reference configuration should use an already existing reference configuration such as ORCA2_LIM as a basis.

On a slightly longer time scale it would be beneficial if the various assimilation schemes could converge on a common format for feedback files. This would enable us to perform intercomparisons of different assimilation systems from different groups and to share assimilation diagnostic tools where appropriate. It is envisaged that the NEMOVAR feedback format, which is based on NetCDF, can be used as a starting point and refined to fit other users' requirement. Once the NEMOVAR observation operators are integrated into the NEMO reference version, suggested improvements to the feedback format can be implemented.

5. Quality control

Quality controlled observations are a crucial input to model validation studies and data assimilation schemes. If erroneous values are used in validation studies, misleading results will be obtained, whilst assimilation of dubious data can cause serious problems in the analysis and subsequent model forecast. A number of groups produce quality controlled data-sets such as Coriolis, the Met Office, and a number of other international groups (e.g. MEDS, NODC) which are available for validation and assimilation. However, it is useful, particularly for operational data assimilative systems, to run a stand-alone automatic quality control system.

During the meeting between NEMO assimilation developers and the NEMO system team, only one option for an automatic quality control system was presented. It was unclear whether other groups working in this area would be willing to contribute code to the NEMO system. This section therefore concentrates on the system presented at the meeting, but collaboration with other groups working on quality control should continue to be developed.

The Met Office quality control system was developed initially as part of the ENACT/ENSEMBLES European projects (Ingleby and Huddleston, 2007) and is used to generate the EN3 data-set (http://hadobs.metoffice.com/en3/). The system is automated so as to deal with the large data volumes involved. All decisions taken by the system are traceable and the generic checks have a clear theoretical basis in probability theory.

The system has mainly been developed to deal with in situ profile data and includes various checks including:

- applying *a priori* reject lists (e.g. Argo grey list);
- applying fall rate corrections to expendable bathythermograph (XBT) profiles;
- track check;
- profile checks including checks for constant values, spikes/steps, zero values;
- superobbing, duplicate check and thinning;

- stability check which checks for density inversions;
- Bayesian background check to compare the observations with a gridded field (from a model forecast or climatology);
- buddy check which compares observations within a specified radius.

Quality control checks for other data types including sea level anomaly and sea surface temperature are also being incorporated into the code.

The scheme has been further developed recently to enable more flexible input/output options, and has been included in the NEMOVAR system. It is therefore compatible with the file formats used in NEMOVAR, as well as other formats. The QC code is usually applied as a pre-processing step before the data are used in the observation operator code. However, it can be useful to apply some checks (e.g. model background-dependent checks) between the observation operator step and the analysis kernel. The system uses namelist options to control which checks are run, and is flexible enough to be run at these different points during the assimilation process. The code is freely available under the CeCILL licence and comprehensive technical documentation for the QC code is available.

Maintenance and development of the system is currently undertaken at the Met Office and ECMWF. The code could be included within the NEMO system fairly easily. The main dependencies for the code are that it reads in observation files and model background files in specific formats, although there are a number of options for these. The output options are also flexible and include the formats required by the observation operator code in NEMOVAR. The code itself has dependencies with the input/output code for the feedback file format defined in NEMOVAR, but apart from that is stand-alone.

At the meeting between NEMO assimilation developers and the NEMO system team, it was agreed that the inclusion of the QC code in the NEMO base code was not a high priority task. For the model validation applications of the observation operator code, existing quality controlled data-sets can be used. The main applications of the automatic QC code are for data assimilation studies, particularly for operational applications. It was therefore agreed that the higher priority aspects of the data assimilation schemes should be included in NEMO before the QC, but that in one or two years this recommendation should be re-evaluated.

References

Ingleby, B. and M. Huddleston, 2007: Quality control of ocean temperature and salinity profiles - historical and real-time data. *Journal of Marine Systems*, **65**, 158-175 10.1016/j.jmarsys.2005.11.019.

6. Analysis kernel

The data assimilation systems that have been implemented with NEMO are based on methods that can be broadly classified as either variational or sequential Kalman filter (KF). The methods have very different algorithmic characteristics despite their apparent similarities from a theoretical standpoint (both are essentially linear methods derived from statistical estimation theory). This leads to important differences in the technical implementation of each method. Four different assimilation systems (2 variational and 2 sequential KF) are currently under development for NEMO. These include: (1) the NEMOVAR system (Mogensen *et al.* 2009; Daget *et al.* 2009; Weaver *et al.* 2005), a variational assimilation system being developed jointly by CERFACS, ECMWF, Met Office and INRIA/LJK, under the CeCILL licence; (2) the INGV and CMCC variational assimilation system (Dobricic and Pinardi 2008; Dobricic 2009); (3) the Singular Evolutive Extended Kalman (SEEK) filter at LEGI (Pham *et al.* 1998; Brasseur and Verron 1998); and (4) a fixed-basis variant of the SEEK filter (SAM-2) being developed for operational applications at Mercator-Océan (Drillet *et al.* 2008, Tranchant *et al.* 2008). The DFO in Canada is planning to develop a fifth assimilation system for NEMO based on the variational method of Liu and Thompson (2009). Developers from two of the systems (NEMOVAR, SAM-2) were present at the meeting.

The procedure for estimating and evolving the assimilation system-error statistics and for solving the data assimilation equations is referred to here as the *assimilation or analysis kernel*. At the meeting there was general agreement that the inclusion of an assimilation kernel in the NEMO system was not a short-term priority, but that steps could (should) be taken in the short term to facilitate collaboration, inter-comparisons and code-sharing between the different assimilation groups by standardizing certain interfaces with NEMO.

For the variational methods listed above, the assimilation kernel consists of an iterative algorithm to minimize a non-quadratic cost function. The cost function is non-quadratic since some of the constraints (such as the ocean model) are nonlinear. The NEMOVAR and INGV/CMMC systems employ an *incremental algorithm* to minimize the cost function. In the incremental algorithm, nonlinear operators are first linearized about a reference state and the resulting quadratic cost function is minimized iteratively using a conjugate gradient or Lanczos algorithm. A nonlinear feedback loop can be included to update the reference state with improved state estimates produced during minimization. Initially, the reference state is defined as the model background state. The linearized feature of the incremental algorithm means that, in practice, the integration of the NEMO model and the quadratic minimization can be implemented as separate FORTRAN executables within a script, with information required for coupling these components exchanged via simple I/O. This results in little disruption to the NEMO model structure as the number of coupling interfaces is rather limited. Another practical advantage of this approach is that the model and the assimilation kernel can be developed, to some extent, independently. The variational assimilation kernel can also include an ensemble analysis/forecast step (Daget *et al.* 2009) or implicit analysis/forecast error covariance propagation (Dobricic 2009) to allow the background-error covariance matrix to evolve from one assimilation cycle to the next.

For the sequential methods listed above, the assimilation kernel consists of two successive operations: a forecast step to propagate in time the initial probability distribution for the state of the system and an analysis step to update this distribution using available observations. The forecast step can be performed either by propagating the initial covariance using a linearized model operator (as in the original SEEK algorithm of Pham *et al.* (1998)) or by producing an ensemble model forecast starting from a sample of the initial probability distribution. In fixed-basis variants, this forecast step is simplified by prescribing a (time-dependent) forecast covariance (as in SAM-2) or by a simple amplification of the initial covariance. In any case, the analysis step is always performed using the SEEK observational update algorithm, which is especially efficient with large observation vectors and low-rank covariance matrices. It is still a linear algorithm, but extensions exist (with an ensemble forecast step) to account for non-Gaussian distributions using anamorphosis or a truncated Gaussian assumption. From a practical point of view, there is a clear separation between the forecast step (only involving NEMO model integrations) and the analysis step, which can be implemented in separate programs (such as the SESAM software developed at LEGI). The overall system can thus be written as a master program (as a shell script for instance) that cycles forecast and analysis steps with quite simple interfaces between them.

The fundamental input to the assimilation kernel is the vector of differences between the observations and the reference state (the *innovation vector*), as well as the reference state itself if it is needed for certain linearized operators. The fundamental output of the assimilation kernel is a vector of corrections to the background state on the model grid (the *analysis increment*). Typically, these consist of corrections to the model initial state but may also include corrections to other fields such as the surface forcing fields, model tendencies, or system parameters. The analysis increment is used to initialize the model (see section on the Application of the Increment). For budget studies, the increment should also be accounted for in the diagnostic trends. Diagnostic output from the assimilation kernel consists of the difference between the increment at observation points and the innovation vector (the *residual vector*).

An important outcome of the meeting was a recommendation that the different assimilation groups seek agreement on a common (feedback) file format for the observation-space diagnostic output (see section on Observation Operators). The innovation and residual vectors are not only important for assessing the performance of an assimilation system, but are also fundamental for computing background- and analysis-error covariance information which can be used for tuning covariance parameters in the assimilation method (Desroziers *et al.* 2005). Diagnostic software used for manipulating observation-space fields obtained from common format feedback files could be shared between NEMO assimilation groups and made available through a diagnostic library in NEMO. A common file format for the analysis increment would also be desirable.

In some applications (especially at high resolution), the grid used for assimilation may be a low resolution or simplified version of the grid used by the model for computing the innovation vector. In this case, an *interpolation operator* is needed to transform the assimilation increments from the assimilation grid to the model grid. Likewise, for the linearized operators in the assimilation kernel which require a basic state from the nonlinear model, a *simplification operator* is needed to transform the model state from the model grid to the assimilation grid (this operator is needed for the tangent-linear and adjoint models in 4D-Var, for example). The interpolation and simplification operators are related since each can be derived as a generalized inverse of the other. These operators are useful for other applications as well (e.g., initializing low resolution models with high resolution initial conditions, or vice versa) and therefore could be of interest to the wider NEMO community.

Practically, it would be difficult to create a *single* assimilation kernel for NEMO that could accommodate all desired features of variational and KF methods. It may be possible, however, to consolidate *similar* kernels into a single kernel by merging the “best” and complementary parts of both. For example, incremental variational data assimilation systems could share the basic minimization structure of the algorithm but still offer different options for the specification of system components such as the background- or observation-error covariance matrices. A detailed study would be necessary to assess the scientific and technical feasibility of combining existing assimilation kernels.

To include an assimilation kernel in NEMO would require it to be freely available under the CeCILL licence. A reference configuration (or configurations) would be required for testing new developments and for performing “default” assimilation experiments for guiding new users. Special attention would be needed for defining the reference data-sets and for providing sensible estimates of the background- and observation-error covariances. In addition, a portable build environment and set of scripts would need to be developed. Data assimilation systems require complex and generally non-portable scripting procedures to *cycle* the system over long periods. Practically, it would be difficult to develop and maintain a portable scripting environment in NEMO for cycling assimilation systems on different platforms. It would be feasible, however, to include a simpler set of scripts to perform a single cycle assimilation experiment, with multiple cycling being left up to individual groups. Finally, the entire system would need scientific and technical documentation, and a basic users' guide. A minimum of user support would also be required. These tasks would require additional manpower to what is currently available in the NEMO System Team.

References

Brasseur, P. and J. Verron, 2006: The SEEK filter method for data assimilation in oceanography: a synthesis. *Ocean Dynamics*, **56(5-6)**, 650-661.

Daget, N., A. T. Weaver and M. A. Balmaseda, 2009: Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Quart. J. Roy. Meteoror. Soc.,* **135**, 1071-1094.

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

Dobricic, S. and N. Pinardi, 2008: An oceanographic three-dimensional variational data assimilation scheme. *Ocean Modelling*, **22**, 89-105.

Dobricic, S., 2009: A Sequential variational algorithm for data assimilation in oceanography and meteorology. *Mon. Wea. Rev*., **137**, 269–287.

Drillet, Y., C. Bricaud, R. Bourdallé-Badie, C. Derval, O. Le Galloudec, G. Garric, C.-E. Testut and B. Tranchant, 2008: The Mercator Ocean global 1/12° operational system: Demonstration phase in the MERSEA context, The Mercator Ocean Newsletter #29, April 2008.

Liu, Y. And K. R. Thompson, 2009: Predicting mesoscale variability of the North Atlantic using a physically motivated scheme for assimilating altimeter and Argo observations. *Mon. Wea. Rev*., **137**, 2223–2237.

Mogensen, K., M. A. Balmaseda, A. T. Weaver, M. Martin and A. Vidard, 2009: NEMOVAR: A variational data assimilation system for the NEMO model. ECMWF Newsletter, Summer Edition.

Pham, D. T., J. Verron and M. C. Roubaud, 1998: Singular evolutive extended Kalman filter with EOF initialization for data assimilation in oceanography. *J. Marine Sys.* **16 (3-4)**, 323-340.

Tranchant B., Testut C.E., Renault L. and Ferry N., Obligis E., Boone C. and Larnicol G.: Data assimilation of simulated SSS SMOS products in an ocean forecasting system, Journal of operational Oceanography, Vol. 2008, No 2, August 2008., pp 19-27(9).

Weaver, A. T., C. Deltel, E. Machu, S. Ricci and N. Daget, 2005: A multivariate balance operator for variational ocean data assimilation. *Quart. J. Roy. Meteoror. Soc*. **131**, 3605-3625.

7. Conclusions/ Recommendations

It is recommended that the initial steps on the implementation of an assimilation capability within the NEMO code system focus on the three components discussed in sections 2-4. In particular

- The observation operator from the NEMOVAR code system should be included in the NEMO reference version as a separate module as quickly as possible.

- NEMOTAM should be included in the NEMO reference version.

- The NEMOVAR code to apply increments to NEMO should be included in the NEMO reference version.

The work and resources required to introduce the observation operator and IAU code into NEMO are outlined below. Considerably more work would be required to include and then maintain NEMOTAM in the NEMO reference version. The level of resources will of course depend on the level of service provided. A very rough estimate based on comparison with the biological components is that 1 full-time position will be required to develop and maintain the code. We recommend that a careful estimate of the ongoing resources is developed by Arthur Vidard and the NEMO System Team once NEMOTAM has been transitioned to being based on nemoV3.2. We also recommend that future plans for NEMO assimilation are formulated on an annual basis by the team proposing to do the work and linked to the annual work-plan for NEMO.

Once real progress has been made on these issues it would be worth re-considering:

- Incorporation of a quality control component.

- Options for software to handle non-linear trajectories and perform minimisations. These codes are needed to exploit NEMOTAM effectively.

- Consolidation of similar assimilation kernels into a single kernel containing the “best” and complementary parts.

Initial work required to introduce the observation operator and IAU code

The main steps that need to be done in the short term are:

1. Make branches of NEMO so that the NEMOVAR version of OBS and ASM code are split up and made separate to any other changes.
2. Update the OBS and ASM branches from vn3.0 to the latest Paris trunk version.
3. Run the standard NVTK tests.
4. Supply the OBS and ASM code to NEMO systems team.

The Met Office has agreed to take the lead on this work aiming to complete it by the end of the first quarter of 2010. Following on from that, there are some additional tasks:

1. Develop a test environment for OBS.
2. Develop a test environment for ASM.
3. Collate documentation for OBS and ASM and pass to systems team.
4. Help coordinate developments to the IAU code so that the changes from Charles-Emmanuel go into the NEMO base code.

It is estimated that this second set of tasks will require approximately 2 man months of work.

Annex: Authors/contributors/attendees

Contributing Authors

Mike Bell (Met Office)
Claire Levy (NEMO System Team)
Matt Martin (Met Office)
Kristian Mogensen (ECMWF)
Arthur Vidard (INRIA-LJK)
Charles-Emmanuel Testut (Mercator-Ocean)
Anthony Weaver (Cerfacs)

Additional Meeting Attendees

Rachid Benshila (NEMO System Team)
Claire Chauvin (INRIA-LJK)
Charles Deltel (LOCEAN)
Eric Dombrowsky (Mercator-Ocean)
Christian Ethe (NEMO System Team)
Gurvan Madec (NEMO System Team)
Elisabeth Remy (Mercator-Ocean)
Claude Talandier (NEMO System Team)
Franck Vigilant (INIRIA-LJK)