REQUEST FOR A SPECIAL PROJECT 2010–2012

MEMBER STATE: France

Principal Investigator[1]**:** Anthony Weaver

Affiliation: CERFACS

Address:
42, avenue Gaspard Coriolis
31057 Toulouse Cedex 1
France

E-mail: weaver@cerfacs.fr

Other researchers: Isabelle Mirouze, Andrea Piacentini, Thomas Pangaud (from autumn 2009), Amal El Akkraoui (from summer 2009)

Project Title: Variational ocean data assimilation with NEMOVAR

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SP** FRVODA	
Starting year: (Each project will have a well defined duration, up to a maximum of 3 years, agreed at the beginning of the project. For projects started before 2009, please state 2009 as the start year.)	2009	
Would you accept support for 1 year only, if necessary?	YES ☐	NO X

Computer resources required for 2010-2012: (The maximum project duration is 3 years, therefore a continuation project cannot request resources for 2012.)	**2010**	**2011**	**2012**
High Performance Computing Facility (units)	250,000	250,000	
Data storage capacity (total archive volume) (gigabytes)	2000	2000	

An electronic copy of this form ***must be sent*** *via e-mail to:* *special_projects@ecmwf.int*

Electronic copy of the form sent on (please specify date): April 30, 2009

Continue overleaf

---

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide an annual progress report of the project's activities, etc.

Principal Investigator: Anthony Weaver

Project Title: Variational ocean data assimilation with NEMOVAR

Extended abstract

It is expected that Special Projects requesting large amounts of computing resources (500,000 SBU or more) should provide a more detailed abstract/project description (3-5 pages) including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used. The Scientific Advisory Committee and the Technical Advisory Committee review the scientific and technical aspects of each Special Project application. The review process takes into account the resources available, the quality of the scientific and technical proposals, the use of ECMWF software and data infrastructure, and their relevance to the Centre's objectives. - Descriptions of all accepted projects will be published on the ECMWF website.

The aim of this project is to further the development of the variational data assimilation system NEMOVAR for the community ocean model NEMO. The development of NEMOVAR is a collaborative project between CERFACS, ECMWF, the Met Office, and other research institutes such as INRIA/LJK. The NEMOVAR system will form the ocean analysis component of the next operational monthly and seasonal forecasting system at ECMWF (System 4). The Met Office also plans to use NEMOVAR for their ocean assimilation activities, covering short-range ocean forecasting applications using the FOAM system, and seasonal forecasting. Ocean initial conditions from NEMOVAR will be used by several groups (including CERFACS) for decadal predictions in the FP7 COMBINE project. This Special Project will help advance research and development with NEMOVAR. Four research topics will be explored.

- *Ensemble 3D-Var*

  A procedure for computing ensembles of ocean analyses and forecasts has been developed for OPAVAR (Daget *et al.* 2009), the predecessor of NEMOVAR. The ensemble system has a dual purpose. It is used to provide ensembles of ocean analyses for climate studies and forecast initialization, and to obtain estimates of background error for specifying parameters (variances and length scales) of the background-error covariance model in NEMOVAR. The ensemble-generation procedure has many similarities to the Monte Carlo step of the Ensemble Kalman Filter. In our approach, ensembles of ocean analyses and forecasts are generated using random perturbation data-sets for surface windstress, precipitation, sea surface temperature (SST), and the assimilated observations. Work in the next period will focus on adapting the ensemble 3D-Var system to the NEMOVAR framework and to a higher resolution global configuration (the ORCA1 configuration to be used in System 4). Improvements and extensions to the current ensemble generation strategy (e.g., to account for model error) will also be explored. A newly developed 3D correlation model (see below) will be used to evaluate the impact of background-error correlation length scales estimated using the ensemble method. This work will be pursued by Thomas Pangaud (postdoc), starting late 2009.

  References:

  Daget, N., Weaver, A. T. and M. A. Balmaseda, 2009: Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Q. J. R. Meteor. Soc.*, In press.

- *Correlation modelling using implicit diffusion operators*

  The theoretical basis for employing a diffusion equation to represent the action of a correlation operator is described in detail in Weaver and Courtier (2001), Weaver and Ricci (2004) and more recently Mirouze and Weaver (2009). A new 3D correlation operator formulated as a symmetric product of 1D implicit diffusion operators has been developed for NEMOVAR. The new formulation has several advantages over a previous formulation based on an explicit diffusion operator, most importantly by reducing the computational cost of the algorithm at higher resolution and with anisotropic versions of the operator. Implicit diffusion operators require the inversion of a matrix. For the 1D correlation operators, this matrix is small enough that it can be inverted using standard direct methods such as Cholesky factorization combined with a forward elimination and backward substitution algorithm. This direct method is not convenient, however, for massively parallel processor applications with NEMO. Iterative techniques based on Gauss-Seidel (successive-over-relaxation) are better suited for parallel implementations of the implicit diffusion operator, and will be explored here. An important point is that, given our uncertainty in the true correlations of background error, it may not be necessary to solve the implicit equation to a very high precision, and this could be exploited to reduce computational costs. Approximations of this nature will be explored. The correlation algorithms developed here will also be exploited for representing model error in weak-constraint formulations of 4D-Var (see below). This work will be conducted by Isabelle Mirouze (PhD student).

References:

Mirouze, I. and A. T. Weaver, 2009: Correlation modelling using a one-dimensional implicit diffusion equation. *In preparation.*

Weaver, A. T. and S. Ricci, 2004: Constructing a background-error correlation model using generalized diffusion operators. In Proceedings of the ECMWF Seminar Series on "*Recent developments in atmospheric and ocean data assimilation*", ECMWF, Reading, U. K., 8 –12 September 2003, pp. 327-340.

Weaver, A. T. and P. Courtier, 2001, Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.*, **127**, 1815-1842.

- *Observation-space minimization and weak-constraint 4D-Var*

An incremental 4D-Var system for NEMOVAR is being developed in collaboration with INRIA/LJK, and a prototype system will be ready later this year. The initial version will be based on a strong-constraint formulation in which only the initial conditions are taken as control variables (the model is assumed to be perfect). Weak-constraint formulations of 4D-Var will be developed in a second stage. In the weak-constraint formulation, model error will be accounted for in the assimilation algorithm by including extra (error) terms in the control vector. This will enable longer assimilation windows but will result in a significant increase in the size of the control vector and thus will require huge memory storage for the preconditioning and reorthogonalization procedures used by the CONGRAD minimization algorithm (Fisher 1998; Tshimanga *et al.* 2008) in NEMOVAR. An observation-space minimization of weak-constraint 4D-Var will be developed to reduce memory requirements. Compared to model-space formulations, conventional observation-space formulations based on PSAS (Courtier 1997) can result in large discrepancies between the quadratic and non-quadratic cost functions on outer iterations, especially when few iterations are used in the inner loop (El Akkraoui *et al.* 2008). As an alternative to PSAS, here we will exploit an observation-space preconditioned conjugate gradient (PCG) algorithm recently proposed by Gratton and Tshimanga (2009). This algorithm has identical convergence properties as standard model-space PCG algorithms. Initially, the new observation-space algorithm will be implemented and tested in a 3D-Var version of NEMOVAR. Work in this area will be carried out by Amal El Akkraoui (postdoc), starting in the summer 2009.

References:

Courtier, 1997: Dual formulation of four-dimensional variational assimilation. *Q. J. Roy. Meteorol. Soc.*, **123**, 2449-2462.

El Akkraoui, A., Gauthier, P., Pellerin, S. and S. Buis, 2008: Intercomparison of the primal and dual formulations of variational data assimilation. *Q. J. Roy. Meteorol. Soc.*, **124**, 1015-1025.

Gratton, S. and J. Tshimanga, 2009: An observation-space formulation of variational assimilation using a restricted preconditioned conjugate gradient algorithm. *Submitted to Q. J. Roy. Meteorol. Soc.*.

Tshimanga, J., Gratton, S., Weaver, A. T. and A. Sartenaer, 2008: Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation. *Q. J. Roy. Meteorol. Soc*., **134**, 753-771.

Fisher, M., 1998: Minimization algorithms for variational data assimilation. In Proceedings of the ECMWF Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling, pp. 364-385, Reading, England, 7-11 September 1998.

- *Representation of correlated observation error in the absolute dynamic topography*

Correlated observation error is an important but under-researched aspect of observation error. In practical ocean data assimilation systems, observation-error correlations are usually neglected even when this assumption is known to be incorrect. One example is satellite altimeter data, which represents one of the most important sources of ocean data for assimilation. In practice, altimeter data must be considered as anomalies with respect to a long-term mean. This is necessary to remove the poorly known geoid from the altimeter measurement. For assimilation, a Mean Dynamic Topography (MDT) must be added to the anomalies to produce measurements of Absolute Dynamic Topography (ADT), which can then be related to the ocean model sea-surface height variable. MDTs used to reference the anomalies are typically based on gridded products, derived either from a long-term integration of the ocean model or from an analysis of *in situ* and gravity measurements (Rio *et al.* 2004). Uncertainty in the MDT is currently the main source of error in the ADT (Knudsen *et al.* 2006) and this uncertainty must be carefully taken into account in the assimilation method. By construction, errors in gridded MDT products are highly correlated in space. Furthermore, the stationary nature of MDT errors will induce a time-correlated component in the ADT, which cannot be neglected. The purpose of this task will be to build an error correlation model for ADT data that can account for both of the aforementioned components of correlated error.

The spatial error correlation model for the gridded MDT will exploit the implicit diffusion developments used in the background-error specification. The full ADT error covariance matrix, $\mathbf{R}_{\mathrm{AD}}$, will be a function of the spatial MDT error covariance matrix, the observation operators that relate the gridded MDT field to the altimeter data points, and the altimeter anomaly error covariance matrix. In NEMOVAR, it is the inverse of this matrix $(\mathbf{R}_{\mathrm{AD}})^{-1}$ that must be defined. This makes the problem particularly challenging in view of the large size of the $\mathbf{R}_{\mathrm{AD}}$ matrix (10^5 x 10^5) and the fact that $\mathbf{R}_{\mathrm{AD}}$ is only available in operator-form. An explicit representation of the inverse matrix is not required, however, as the iterative variational algorithm only requires the application of the inverse to a given vector (i.e., an operator $(\mathbf{R}_{\mathrm{AD}})^{-1}$). Different methods (iterative or direct) will be explored for determining the inverse operator. Both numerical efficiency and the need to ensure positive definiteness of the inverse operator will be of crucial importance in selecting the appropriate method. This work will be carried out by Andrea Piacentini (research engineer).

References:

Knudsen, P. and co-authors, 2006: Final report of the GOCINA (Geoid and Ocean Circulation in the North Atlantic) project. Danish National Space Center. Tech. Rep. No. 5.

Rio, M.-H. and F. Hernandez, 2004: A Mean Dynamic Topography computed over the world ocean from altimetry, in situ measurements and a geoid model. {\em J. Geophys. Res.}, {\bf 109}, C12032.