

Quantifying the Amplitude, Structure and Influence of Model Error during Ocean Analysis and Forecast Cycles

Andrew M. Moore
Department of Ocean Sciences, 1156 High St.
University of California
Santa Cruz CA 95064
Phone: (831) 459-4632 fax: (831) 459-4882 email: ammoore@ucsc.edu

Chris Edwards
Department of Ocean Sciences, 1156 High St.
University of California
Santa Cruz CA 95064
Phone: (831) 459-46323734 fax: (831) 459-4882 email: cedwards@ucsc.edu

Ralph F. Milliff
Colorado Research Associates
3380 Mitchell Lane
Boulder CO 80301
Phone: (303)-415-9701 fax: 720-836-5609 email: milliff@cora.nwra.com

Award Number: N00014-10-1-0476

LONG-TERM GOALS

The long-term scientific goals of this research project are to:

1. Understand and quantify the sources of error in ocean models that fundamentally limit the practical predictability of the coastal ocean circulation.
2. Use information about model error to improve ocean circulation estimates obtained using weak constraint data assimilation methods.

OBJECTIVES

The primary objective of the proposed research is to develop a weak constraint, 4-dimensional variational (4D-Var) data assimilation capability for the Regional Ocean Modeling System (ROMS) with application to the California Current System (CCS). The CCS is of considerable socio-economic and strategic significance to the United States, and ROMS CCS has transitioned to a near real-time analysis system in support of the U.S. west coast components of the Integrated Ocean Observing System (IOOS). This project is therefore very timely given the limiting nature of model errors on coastal ocean prediction.

APPROACH

The proposed research activities seek to quantify the amplitude, structure and influence of model error in ocean analysis and forecast systems using two approaches. The first uses the ROMS 4D-Var data assimilation systems (Moore et al, 2010a,b,c) in combination with Bayesian hierarchical modeling to identify sources of model error and their characteristics. The second approach uses Generalized Stability Analysis (GSA) (Moore et al, 2004) to explore the fastest growing model error structures, how these relate to different ocean circulation regimes, and the implied bounds on model error growth and predictability.

The project involves three major research tasks:

Proposed Research Task #1:

(i) A sequence of strong constraint 4D-Var experiments for ROMS CCS spanning several years will be used to identify geographical areas where model error influences surface forcing adjustments imposed by 4D-Var. (ii) A surface flux BHM will be used to identify periods when model error identified in (i) is a significant factor. The basis for the CCS BHM will be the BHM of Milliff et al (2011) for the Mediterranean surface winds which will be reconfigured for the ROMS CCS domain and expanded to include surface fluxes of heat and momentum. The BHM data stage will use QuikSCAT surface vector winds, and COAMPS surface winds and fluxes, while the process model stage will utilize the bulk surface flux model subcomponent of ROMS and COAMPS standard 10 m atmospheric boundary layer variables. (iii) A second sequence of weak constraint 4D-Var experiments using the surface flux BHM and strong constraint flux increments to inform the model error *prior* will be used to build and test various different hypotheses about model error. (iv) A detailed analysis of the spatio-temporal corrections for model error from each weak constraint 4D-Var assimilation cycle will be performed to identify the nature of the model errors.

Proposed Research Task #2:

A complete characterization of model error during each 4D-Var cycle is provided by the *posterior* covariance. ROMS 4D-Var permits computation of the leading eigenvectors (EOFs) of the *posterior* error covariance matrix. Spatio-temporal variations in the structures of the leading EOFs of *posterior* error will yield *posterior* information about the efficacy of the resulting 4D-Var circulation estimates. Therefore a detailed analysis of the leading EOFs of the *posterior* model error covariance matrix will also be performed for the assimilation sequences performed during Task #1(i) for strong constraint and Task #1(iii) for weak constraint.

Proposed Research Task #3:

(i) The stochastic optimals (SOs) and forcing singular vectors (FSVs) of any time evolving circulation can be computed using the ROMS GSA tool-kit described by Moore et al (2004). A systematic study of the FSVs and SOs associated with model error will be performed for ROMS CCS using hindcasts for the same period as Task 1(i). Hindcast initial conditions will be generated using weak constraint 4D-Var during Task 1. Since the leading FSVs and SOs are the most damaging model errors for the hindcast interval, they yield upper bounds on model error growth, and the loss of predictability of the circulation during each hindcast interval due to the growth of model error. (ii) The projection of the *posterior* model errors diagnosed from the weak constraint 4D-Var experiments of Task #1 onto the FSVs and SOs will also be examined.

WORK COMPLETED

In addition to the PIs, other personnel involved in this project include: Dr. Polly Smith, a post-doctoral scholar hired to work on this project; and Mr. Kevin Smith, a graduate student researcher hired to work on this project. Polly Smith left the project in June 2012 to take up a position in the U.K., and she has been replaced by another post-doctoral scholar Dr. Emilie Neveu.

Task #1:

Work is well underway for Task 1 activities. Much of this activity has been led by Dr. Polly Smith. A sequence of strong constraint 4D-Var analyses have been generated for the period 2002-2004, a period that has already been extensively studied (Moore et al, 2011b, 2011c). In these experiments, the ROMS 4D-Var control vector comprises the initial conditions, surface forcing and open boundary conditions. The surface forcing fields are derived from output from COAMPS which is known to perform well in this region (Doyle et al, 2009). Therefore any significant departures of the 4D-Var adjusted surface fluxes from the COAMPS *priors* is a strong indication of the influence of model error. The significance of the 4D-Var surface flux adjustments is assessed using a surface flux BHM developed by PI Milliff. So far the BHM has been developed only for the surface wind stress, but will be further developed later to include surface fluxes of heat and fresh water. The BHM provides an estimate of the probability distribution for the surface winds which can be used to quantify the efficacy of the ROMS 4D-Var wind adjustments. An example calculation is shown in the Results section.

By treating the COAMPS surface flux fields as “perfect,” a series of model integrations spanning each data assimilation cycle and forced with the 4D-Var adjusted fields has been performed and used to assess the impact of model errors on the resulting circulation estimates. These calculations were used to estimate the parameters necessary for the model error covariance model that is required for weak constraint data assimilation. The parameters so derived were then used for Task 1(iii). In addition, some additional refinements were made to the ROMS 4D-Var code including the addition of a quality control check on the observations based on departures from the background, and the addition of a temporal decorrelation function to the model error covariance matrix \mathbf{Q} . The latter was found to be particularly critical. Some issues still remain, however, in particular poor performance during some assimilation cycles which appears to be related to the spatio-temporal distribution of the observations. Identification of the nature of the model errors described in Task 1(iv) is ongoing.

Task #2:

Work has yet to begin on Task 2.

Task #3:

Work is also well underway for Task #3 which is being led by PI Moore in conjunction with a graduate student Kevin Smith. Since the bulk of the work for this component of the project will form the basis of Kevin Smith's thesis research, we have adopted a pedagogic approach to the problem. First, we have explored the structure of the initial condition singular vectors (SVs), the FSVs and SOs for an analytical zonal jet in a periodic channel using ROMS. The purpose of this was two-fold: first it is an excellent illustrative test-bed for introducing students to the mathematical and physical ideas that underpin singular value decomposition (SVD) for geophysical flows; second it served as a test of the ROMS codes to identify any bugs that may have crept into the GSA toolkit over time. Some bugs were identified and have been fixed. The zonal jet examples have revealed that the dynamical structures that are optimal for stimulating error growth measured in terms of perturbation energy are remarkably insensitive to the nature of the error, be it growth due to errors in the initial conditions as described by the SVs, growth due to errors in the surface forcing described by the FSVs, or growth due to model error described by the SOs. This is mostly likely because the fundamental dynamical mechanisms that are available for perturbation energy growth are limited primarily to barotropic and baroclinic instability so perturbation structures that capitalize on energy release from the basic state jet, regardless of how they are excited, will win out. To test these ideas and observations, we repeated the SVD calculations for a baroclinically unstable jet in a larger domain with high horizontal resolution. In this case also, the dynamical structures that are optimal for stimulating perturbation energy growth are insensitive to the nature of the errors they represent. Some examples are shown in the Results section.

The focus of this proposal is the influence of model error on ocean predictability. Therefore, a more appropriate choice of norm for the SVD analyses is one based on the analysis error covariance matrix, \mathbf{E}^a , and the forecast error covariance, \mathbf{E}^f . For the initial condition SVs, $(\mathbf{E}^a)^{-1}$ is used as the norm at initial time, while \mathbf{E}^f is used as the norm at final time. Because of the intimate connection between \mathbf{E}^a and the 4D-Var cost function Hessian, the SVs so-defined are usually referred to as Hessian singular vectors (HSVs) (Barkmeijer et al, 1998). During the last year we have developed the code necessary to compute the HSVs in ROMS. We have also extended the same ideas to stochastic forcing representing model error to yield what we refer to as Hessian stochastic optimals (HSOs). To our knowledge the idea of HSOs is new. Calculations have been performed for the unstable baroclinic jet and for the CCS. Like the energy norm calculations, the HSV and HSO structures appear to be very similar for the few cases that we have looked at so far. More work is required to confirm the robustness of this result. Some examples are shown in the Results section.

We have also been developing some new ideas in relation to SVD. The first of these involves using the entire 4D-Var control vector (i.e. initial condition error, surface forcing error, open boundary condition error, and model error) as the state-vector for the SV calculations. The advantage of this is that many of the assumptions that are required to compute the SOs, such as separable space-time correlations and errors that are Markovian in time, are no longer necessary so the resulting analyses will be more general. We refer to these new singular vectors as control singular vectors. The second idea that we are developing relates to SVD using \mathbf{E}^a and \mathbf{E}^f to define the norms. The computation of both covariance matrices is non-trivial, and in numerical weather prediction it is usually assumed that a good alternative for \mathbf{E}^f is the energy norm. It remains to be seen if this is true in the ocean. However, as part of this project we have developed a method for computing both \mathbf{E}^a and \mathbf{E}^f using the adjoint of the entire 4D-Var system (Moore et al., 2012). By combining the adjoint of 4D-Var with the GSA tool-kit, we will be

able to use the true expected forecast error covariance matrix as the appropriate norm for SVD. Finally, a considerable reduction in the dimension of the SVD problem is afforded by using the Lanczos vectors derived during 4D-Var to define the state-vector for the HSV and HSO calculations referred to above. To our knowledge these are all new developments in the practical application of SVD to large geophysical circulation problems.

Restricted Preconditioned Conjugate Gradient – Lanczos Formulation

One additional and very important research activity that has been undertaken as part of this project is the implementation of a Lanczos formulation of the Restricted Preconditioned Conjugate Gradient (RPCG) method described by Gratton and Tshimanga (2009). This is an extremely important development for ROMS 4D-Var, and overcomes the poor convergence behavior of the dual 4D-Var algorithms reported by Moore et al (2011b) for ROMS but also widely acknowledged in the field. RPCG is formulated so as to guarantee the same rate of convergence as the primal formulation. What makes this so significant is that weak constraint 4D-Var is only practically possible in the dual formulation, so the fast convergence enjoyed by the strong constraint primal approach can now be achieved during weak constraint 4D-Var also. During the last year, RPCG has been implemented in ROMS using the equivalent Lanczos formulation, and sample 4D-Var ROMS calculations, along with a comparison to 3D-Var using the European NEMO model, are presented in Gürol et al. (2012).

RESULTS

Model error diagnosed from a BHM wind model

An example of the type of analysis involved in Tasks 1(i) and 1(ii) above is shown in Fig. 1 for the surface wind stress for the 7 day period 20-27 Feb, 2003. The fan of red vectors represents several realizations of the 7 day average surface wind stress vectors derived from the wind BHM. The mean wind from the BHM is also indicated. Also shown is the 7 day average *prior* and *posterior* wind stress from the ROMS strong constraint 4D-Var data assimilation cycle spanning the same period. In this particular proof-of-concept example, ROMS is configured with 30 km horizontal resolution and 30 σ -levels in the vertical. All available observations in form of satellite SST, sea surface height and *in situ* hydrographic data are also assimilated as described in Moore et al. (2011b). When the *posterior* wind stress estimates from 4D-Var fall within two standard deviations of the BHM distribution, this is taken as an indication that the computed wind corrections are realizable. However, when the *posterior* stress estimates fall outside this limit then we take this to be an indication of the influence of model error which is manifested in the surface forcing adjustments because no account is taken of model error by the assimilation procedure in this case. Figure 1 shows that in some areas of the model domain, departures of the *posterior* wind stress from the BHM distribution are large.

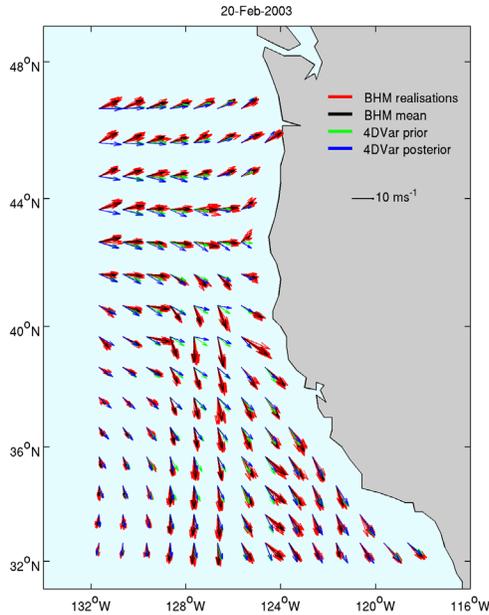


Figure 1: Vectors of 7 day average surface wind stress over the ROMS CCS model domain for the period 20-27 Feb, 2003. The wind roses in red represent several realizations of surface wind stress from the wind BHM and can be thought of as a vector representation of the wind stress probability distribution function (pdf) at each point. The black vector represents the mean of the pdf. The prior and posterior wind stress for the ROMS 4D-Var cycle spanning the same period are shown in green and blue respectively. For clarity, only every 4th grid point is shown.

Singular Value Decomposition

An example of SVD for the baroclinically unstable jet discussed earlier is shown in Fig. 2, and shows one member of the SO spectrum for a 2 day time interval. The panel on the left illustrates the sea surface height structure of the SO superimposed on the basic state circulation which takes the form of two well developed meanders on an unstable jet flowing from left to right. Physically, the SO structure shown represents the sea surface height component of a field of stochastic forcing that is efficient for exciting energy variance in the circulation. In this case, the perturbation takes the form of a wave train that favors one arm of the left meander of the basic state. As noted above, these structures are found to be insensitive to the temporal nature of the perturbation (i.e. an initial condition perturbation versus a random perturbation in time). In addition, in the case of the SOs the perturbation structure is insensitive to the assumed decorrelation timescale of the stochastic forcing. In the example shown, a decorrelation time of 2 days was used. Figure 2 also shows the three dimensional structure of the temperature field component of the same SO from two different aspects. These figures show very clearly the upstream tilt of the wave train indicating that perturbations excited by this particular SO will grow primarily as a result of baroclinic energy release from the underlying basic state circulation.

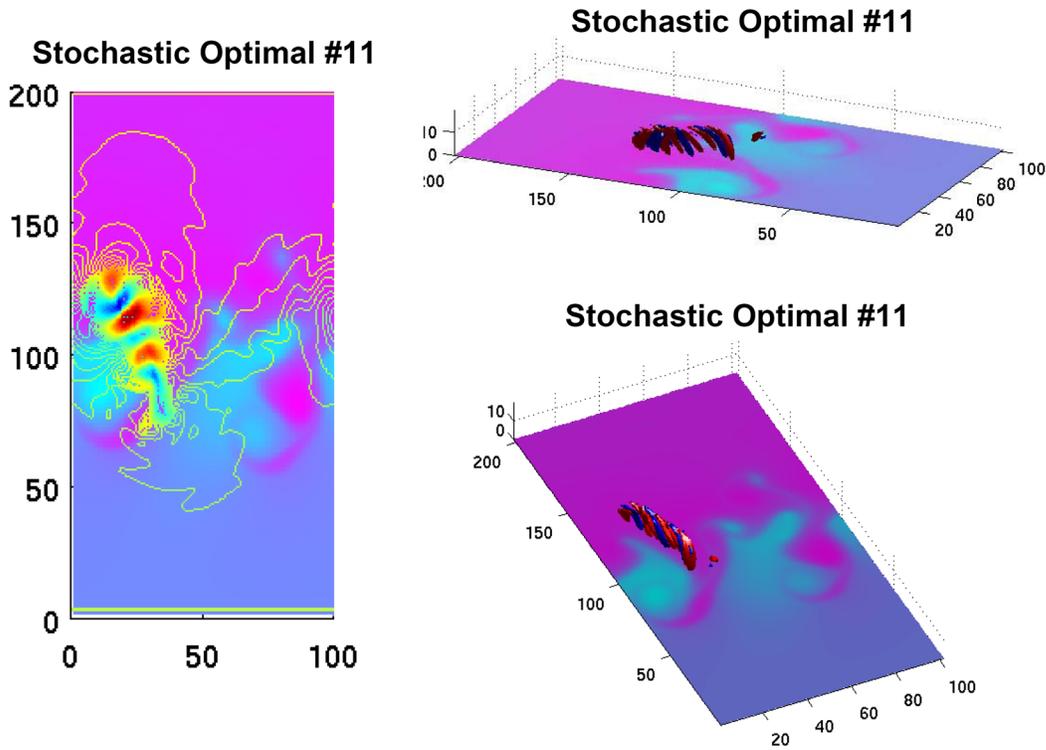


Figure 2: *The panel on the left shows contours (red and blue) of the sea surface height component associated with one member of the SO spectrum for a 2 day time interval. This is superimposed on the basic state sea surface height (blue and purple solid colors) which is an excellent surrogate for the surface stream function. The latter shows a clear signature of a meandering circulation. The two panels on the right show two different 3-dimensional views of the temperature structure of the same SO in relation to the basic state circulation (again represented by the solid purple and blue colors).*

Hessian Singular Vectors and Stochastic Optimals

As noted above, a more relevant norm for the predictability problem is based on the analysis error covariance matrix and forecast error covariance matrix. Using the latter to define the final time norm and the inverse of the former to define the initial norm yields the so-called Hessian singular vectors (HSVs). An extension of these ideas can also be applied to the SOs to yield the Hessian SOs (HSOs). Some example calculations are illustrated in Fig. 3 for the ROMS CCS model with 30 km resolution and 30 σ -levels for a randomly chosen 7 day assimilation cycle. The *posterior* sea surface height field resulting from 4D-Var is presented in the upper right panel and shows a well developed California Current punctuated by large eddies. The sea surface height component of the leading member of the HSV and HSO spectra for the same 7 day period are also shown. In the case of the HSOs, two cases are shown corresponding to stochastic forcing in the form of white noise and red noise with a 2 day correlation time. The structures of the leading HSV and HSOs are clearly very similar, suggesting that the structures of the errors that lead to the largest growth in forecast error are robust to the temporal nature of the errors (i.e. initial condition errors versus errors that are continuous and random in time).

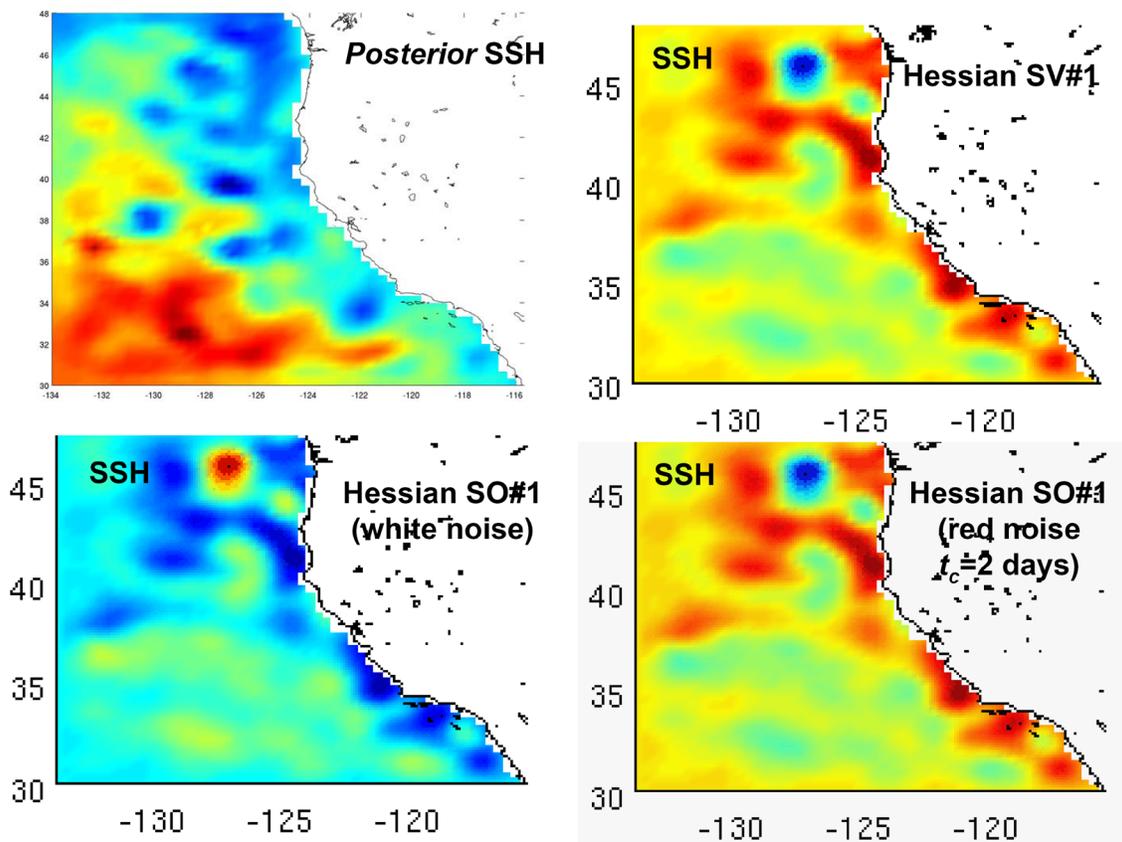


Figure 3: The upper right panel shows the posterior sea surface height field for a representative 7 day 4D-Var data assimilation cycle. The remaining panels show the sea surface height component of the leading members of the HSV spectrum and the HSO spectra for the same period. In the case of the HSOs, two cases are shown corresponding to stochastic forcing that is either white in time or in takes the form of red noise with a 2 day decorrelation time. The sign of each HSV and HSO is arbitrary.

Restricted Preconditioned Conjugate Gradient (RPCG)

As mentioned earlier, an important development in the ROMS weak constraint 4D-Var system in the last year has been the implementation of the Lanczos version of RPCG which guarantees that the primal and dual formulations of ROMS 4D-Var will converge at the same rate. This is illustrated in Fig. 4 which shows time series of the 4D-Var cost function J from ROMS CCS this time configured with 10 km resolution and 42 σ -levels. The panel on the left shows results from a representative 7 day period using several different configurations of the strong constraint 4D-Var system. In one case the primal formulation preconditioned by the background error covariance matrix is used in conjunction with the Lanczos formulation of the conjugate gradient algorithm, and the convergence of J is both rapid and monotonic (P Lanczos). Also shown are the results from three different calculations using the dual formulation. Traditionally dual 4D-Var is preconditioned by the observation error covariance matrix but under these circumstances the convergence rate using the traditional Lanczos formulation of conjugate gradient is very slow and non-monotonic (D Lanczos). This problem is well known and one

remedy proposed by El Akkraoui and Guathier (2010) is to use the minimum residual method in place of the usual conjugate gradient approach. In Fig. 4 this approach is referred to as D MINRES and it is indeed successful at speeding up convergence which is also mostly monotonic. However, the primal formulation (P Lanczos) is still superior. The last dual calculation shown in Fig. 4 (D RLanczos) uses the Lanczos formulation of RPCG as described by Gürol et al. (2012) and confirms that the dual and primal approaches yield the same sequence of cost function values in agreement with theory. The right hand panel in Fig. 4 shows results from three weak constraint 4D-Var calculations and demonstrates the superior performance of RPCG in this case also. The computational cost of weak constraint 4D-Var in the primal formulation is prohibitive.

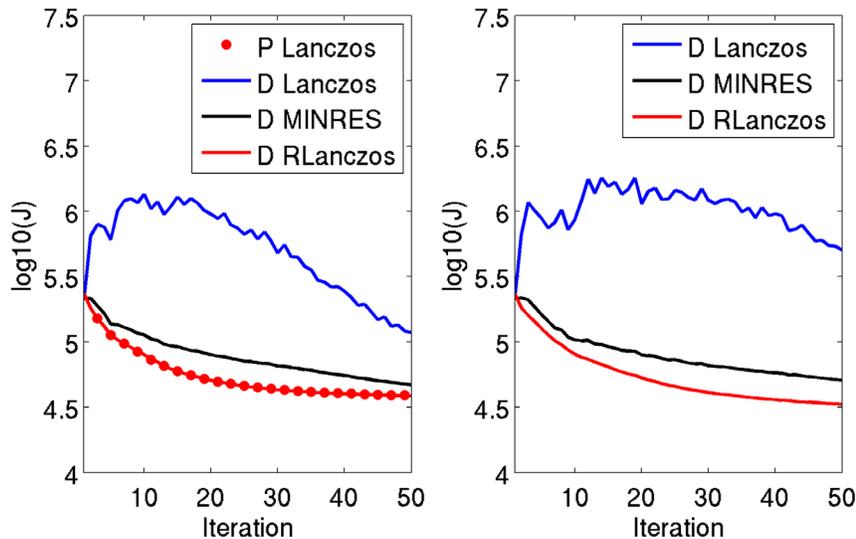


Figure 4: The panel on the left shows time series of cost function J versus inner-loop iteration number for four different configurations of ROMS strong constraint 4D-Var in the CCS (10 km resolution, 42 σ -levels) for a representative 7 day period 29 March – 4 April, 2003. In one case the primal formulation is used in conjunction with a Lanczos formulation of the conjugate gradient algorithm preconditioned by the background error covariance matrix (P Lanczos). In a second case, the dual formulation and a Lanczos algorithm preconditioned by the inverse observation error covariance are used (D Lanczos). In the third experiment the dual formulation is employed in conjunction with the minimum residual method (D MINRES). The last case, the dual formulation is used in conjunction with the Lanczos formulation of RPCG preconditioned by the background error covariance matrix (D RLanczos). The panel on the right shows a similar sequence of weak constraint dual 4D-Var experiments.

IMPACT/APPLICATIONS

This project contributes significantly to the functionality and utility of ROMS, a widely used and important community model and resource. ROMS is unique in that of all the community ocean models that are available, it is the only model that possesses such a wide range of 4D-Var algorithms, analysis tools, and diagnostic capabilities. As part of this project, we have already added to the utility of ROMS

by addition of the Lanczos formulation of the RPCG algorithm, time correlations for model error in weak constraint 4D-Var, and two new drivers for computing Hessian SVs and Hessian SOs.

TRANSITIONS

The new ROMS utilities developed as part of this project will be freely available from the ROMS web site <http://myroms.org> and will be actively used and further developed by other research groups in the U.S. and elsewhere as user competence increases.

RELATED PROJECTS

The work described here is closely related to the following ONR supported projects:

“A community Terrain-Following Ocean Model (ROMS)”, PI Hernan Arango, grant number N00014-08-1-0542.

“Bayesian Hierarchical Models to Augment the Mediterranean Forecast System”, PI Ralph Milliff, grant number N00014-05-C-0198.

“Understanding Predictability of the Ocean”, PI Brian Powell, grant number N00014-09-1-0939.

“Bayesian Hierarchical Model Characterization of Model Error in Ocean Data Assimilation and Forecasts”, PI Ralph Milliff, grant number N00014-10-C-0354.

In addition, the development, implementation and testing of the Lanczos formulation of the RPCG algorithm in ROMS dual 4D-Var is work that has been done in close collaboration with researchers at CERFACS in Toulouse. This project has benefited considerably from direct interactions with not only with CERFACS scientists, but also scientists at ECMWF.

REFERENCES

- Barkmeijer, J., M. van Guzen and F. Bouttier, 1998: Singular vectors and estimates of the analysis-error covariance metric. *Q. J. Roy. Meteorol. Soc.*, **124**, 1695-1713.
- Doyle, J.D., Q. Jiang, Y. Chao, and J. Farrara, 2009: High-resolution atmospheric modeling over the Monterey Bay during AOSN II. *Deep Sea Res. II*, **56**, 87-99.
- El Akkraoui, A. and P. Gauthier, 2010: Convergence properties of the primal and dual forms of variational data assimilation. *Q. J. Roy. Meteorol. Soc.*, **136**, 107-115.
- Gratton, S. and J. Tshimanga, 2009: An observation-space formulation of variational assimilation using a restricted preconditioned conjugate gradient algorithm. *Q. J. Roy. Meteorol. Soc.*, **135**, 1573-1585.
- Gürol, S., S. Gratton, A.T. Weaver, A.M. Moore, A. Piacentini and H.G. Arango, 2012: Dual-space minimization algorithms for variational data assimilation. *Q. J. Roy. Meteorol. Soc.*, In final preparation.

- Milliff, R.F., A. Bonazzi, C.K. Wikle, N. Pinardi and L.M. Berliner, 2011: Ocean ensemble forecasting, Part I: Mediterranean winds from a Bayesian hierarchical model. *Q. J. Roy. Meteorol. Soc.*, **137**, 858-878.
- Moore, A.M., H.G. Arango, E. Di Lorenzo, B.D. Cornuelle, A.J. Miller and D.J. Neilson, 2004: A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. *Ocean Modelling*, **7**, 227-258.
- Moore, A.M., H.G. Arango, G. Broquet, B.S. Powell, J. Zavala-Garay and A.T. Weaver, 2011a: The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems. Part I: System overview and formulation. *Progress in Oceanography*, **91**, 34-49..
- Moore, A.M., H.G. Arango, G. Broquet, C. Edwards, M. Veneziani, B. Powell, D. Foley, J. Doyle, D. Costa and P. Robinson, 2010b: The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems. Part II: Performance and application to the California Current System. *Progress in Oceanography*, **91**, 50-73.
- Moore, A.M., H.G. Arango, G. Broquet, C. Edwards, M. Veneziani, B. Powell, D. Foley, J. Doyle, D. Costa and P. Robinson, 2010c: The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems. Part III: Observation impact and observation sensitivity in the California Current System. *Progress Oceanography*, **91**, 74-94.
- Moore, A.M., H.G. Arango and G. Broquet, 2012: Analysis and forecast error estimates derived from the adjoint of 4D-Var. *Mon. Wea. Rev.*, In press.

PUBLICATIONS

- Gürol, S., S. Gratton, A.T. Weaver, A.M. Moore, A. Piacentini and H.G. Arango, 2012: Dual-space minimization algorithms for variational data assimilation. *Q. J. Roy. Meteorol. Soc.*, In final preparation.
- Moore, A.M., H.G. Arango and G. Broquet, 2012: Analysis and forecast error estimates derived from the adjoint of 4D-Var. *Mon. Wea. Rev.*, **140**, 3183-3203.
- Moore, A.M., C. Edwards, J. Fiechter, P. Drake, H.G. Arango, E. Neveu, S. Gürol and A.T. Weaver, 2012: A Prototype for an Operational Regional Ocean Data Assimilation System. In “*Data Assimilation for Atmospheric, Oceanic and Hydrological Applications, Vol II*”, Liang Xu and Seon Park, Eds. Springer, In press.