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
Cooperative Institute for Research in Environmental Sciences (CIRES)
University of Colorado
Boulder CO 80304
Phone: (303)-492-2013 fax: 303-492-1149 email: ralph.milliff@colorado.edu

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:

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.

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.

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 took-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 was replaced by another post-doctoral scholar Dr. Emilie Neveu. Unfortunately, Dr. Neveu also left the University in early 2013 to take up another position in France. In addition to the loss of critical research personnel, the project also ran short of funds because of the recent sequestration by the federal government. As a result Mr. Smith, the graduate student funded to work on this project, was unsupported during the spring and summer quarters of 2013. Despite these extenuating circumstances, work has continued on the project during the last fiscal year, albeit at a slower rate. We are actively looking to hire a new post-doctoral researcher to work on this project.

Task #1:

Work has continued on Task 1 activities following on from the work led and initiated by Dr. Polly Smith. Dr. Smith generated a sequence of strong constraint 4D-Var analyses for the period 2002-2004, and in these experiments, the ROMS 4D-Var control vector comprised the initial conditions, surface forcing and open boundary conditions. The surface forcing fields were derived from output from COAMPS which is known to perform well in this region (Doyle et al, 2009). Our working hypothesis is therefore that 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 was 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.

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 were performed and used to assess the impact of model errors on the resulting circulation estimates. These calculations were then used to estimate the parameters necessary for the model error covariance matrix that is required for weak constraint data assimilation. The parameters so derived were then used for Task 1(iii).

In order to prove the concept of our approach, Dr. Smith performed all of the aforementioned experiments using a low resolution (~30 km grid-spacing) version of ROMS CCS (known as WC13), which has for several years served as an efficient test case for all of the ROMS 4D-Var-based algorithms. Unfortunately, in this case some significant issues relating to the convergence of the weak constraint 4D-Var algorithm were discovered during several of the 4D-Var cycles, calling into question the validity of our results. These issues appear to be related to assimilating hydrographic observations close to the coast and in relatively shallow water. Despite repeated attempts, we were unable to mitigate the problem in the low resolution WC13, so a decision was

made to use a higher resolution version of ROMS CCS (known as WC12) with ~10 km grid-spacing. This model has been the workhorse of a recent 30 year sequence of historical analyses for the CCS, and a near real-time 4D-Var and prediction system currently run an UCSC in support of CeNCOOS. This model configuration is well behaved during weak constraint 4D-Var experiments, and does not exhibit any of the issues found in WC13. The sequence of experiments outlined above is now being repeated using a subset of years from the 30 year historical analysis. Identification of the nature of the model errors described in Task 1(iv) will be explored.

Task #2:

Due to the technical delays associated with Task #1, and the lack of research personnel that could be supported on this award during the last fiscal year, work has yet to begin on Task 2.

Task #3:

Work has continued on Task #3 and led by PI Moore in conjunction with a graduate student Kevin Smith. Following some pedagogic applications of singular value decomposition (SVD) to a steady analytical zonal jet in a periodic channel using ROMS, Smith's calculations have been extended to a complex time-dependent case in a larger domain with high horizontal resolution. The basic set-up of the model is the relaxation of a zonally uniform meridional temperature gradient and the subsequent development of a baroclinically unstable jet. The initial focus of these experiments was to explore the structure and energetics of singular vectors conditioned on the growth of the perturbation energy norm. However, since the ultimate focus of the proposed research is the influence of model error on ocean predictability, a more appropriate choice of norm for the SVD analyses is one based on the analysis error covariance matrix, $\mathbf{E^a}$, of a forecast initial condition and the forecast error covariance, $\mathbf{E^f}$, at the forecast time. Following Ehrendorfer and Tribbia (1995), $(\mathbf{E^a})^{-1}$ is used as the norm at initial time, while $\mathbf{E^f}$ is used as the norm at final time. As shown by Ehrendorfer and Tribbia (1998) the appropriately rescaled, time-evolved SVs computed in this case are the EOFs of the expected forecast error covariance matrix. Because of the intimate connection between $\mathbf{E^a}$ and the 4D-Var cost function Hessian, the SVs that evolve into the EOFs are often also referred to as Hessian singular vectors (HSVs) (Barkmeijer et al, 1998). During the last year we have continued to develop the mathematical framework and methodology necessary to compute the HSVs in ROMS, and have 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. Another aspect of our work which is new is that we compute the HSVs (and HSOs) *only* in the subspace that is spanned by the 4D-Var increments to the *prior* circulation estimate. This results in significant computational savings, and is in contrast to related calculations in numerical weather prediction where the HSVs are typically identified in the full state-space of the model. An extensive sequence of HSV calculations have been performed for the unstable baroclinic jet and for the CCS by Smith and Moore, and we are in the process of analyzing them with the view of submitting a peer-reviewed publication in late 2013 or early 2014.

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 surrogate 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. A key element to advancing these activities has been the development and testing of the tangent linear and adjoint of the Lanczos formulation of the Restricted Preconditioned Conjugate Gradient (RPCG) algorithm for ROMS. As reported last year, the Lanczos form of RPCG was implemented and tested in ROMS as part of this project and is described in Gürol et al. (2013).

Augmented 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 Augmented Restricted Preconditioned Conjugate Gradient (RPCG) method described by Gratton and Tshimanga (2009). The implementation of Lanczos-RPCG in ROMS was described in last year's annual report and is documented by Gürol et al. (2013) for the case of a single outer-loop. In the case of multiple outer-loops, the augmented form of the algorithm is required to ensure that the resulting *posterior* increments lie in the space spanned by the previous outer-loops. The augmented Lanczos-RPCG algorithm has been successfully implemented and tested in ROMS, and will be released to the ROMS community in the near future. This is an very significant 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 practical in the dual formulation, so the fast convergence enjoyed by the strong constraint primal approach using multiple outer-loops can now be achieved during multiple outer-loops of weak constraint 4D-Var also. An example calculation is shown in the Results section.

RESULTS

Hessian Singular Vectors

As noted above, the relevant norms for the predictability problem for SVD are 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). The rescaled time evolved HSVs are the EOFs of the forecast error covariance matrix. Some example calculations are illustrated below.

Figure 1 shows the SSH component of the EOFs (i.e. the time evolved HSV) for a 6 day forecast of a southern hemisphere unstable baroclinic jet in a zonal channel (1000 km $\times$ 2000 km) for one member of a sequence of identical twin experiments. The forecast was initialized using an ocean

state estimate derived using strong constraint 4D-Var and imperfect simulated observations of the upper-ocean temperature every 24 hours, and a data assimilation window of 48 hours. The forecast SSH at the initial and final forecast times is also shown, along with the percentage of the expected forecast error variance accounted for by each EOF. The first 3 members of the EOF spectrum account for almost ~60% of expected forecast error variance.
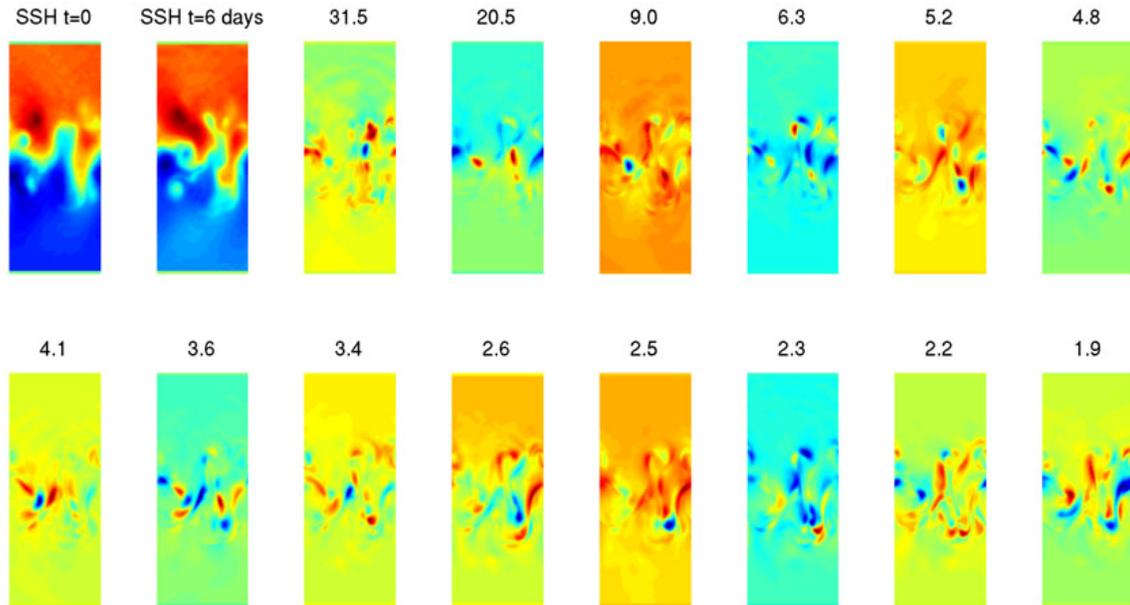

Figure 1: The upper left panels show the sea surface height field at initial and final forecast time for a representative 6 day forecast of a southern hemisphere baroclinically unstable jet in a zonal channel (1000 km × 2000 km). The percentage of the expected forecast error variance explained by each EOF is also indicated. The EOFs are the time evolved HSVs computed using the inverse of the expected analysis error covariance as the initial time norm and the expected forecast error covariance matrix as the final time norm. In order to account for the different units of the forecast state-vector elements, the final time forecast error is defined in terms of the perturbation energy of the deviations of the forecast from the truth.

Figure 2 shows an example calculation from ROMS CCS (WC12). In this example actual satellite and in situ hydrographic observations were assimilated into ROMS CCS using strong constraint 4D-Var and a 8 day assimilation window (see http://oceanmodeling.ucsc.edu). In this case the first three members of the EOF spectrum account for ~70% of the expected forecast error covariance.

At the present time we are actively exploring the link between the properties of the EOF spectrum derived directly from the HSVs and the actual forecast skill of the model, both in the case of the baroclinically unstable jet and for the CCS. This may provide a means for predicting the expected level of forecast error.
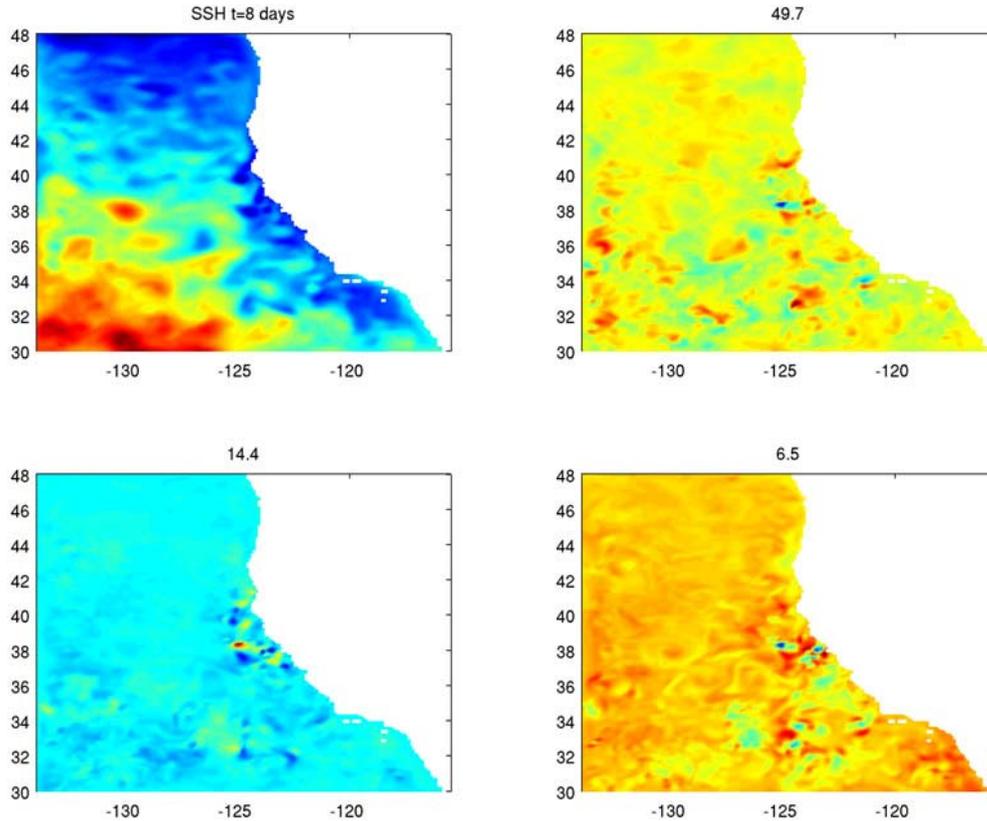
Figure 2: The upper left hand panel shows the 8 day forecast sea surface height field initialized from an 8 day 4D-Var data assimilation cycle. The remaining panels show the day 8 sea surface height component of the leading members of the EOF spectrum computed from the time evolved HSVs, and in each case the percentage of the expected forecast error variance explained by each EOF is indicated.

Augmented 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 augmented RPCG which guarantees that the primal and dual formulations of ROMS 4D-Var will converge at the same rate when using multiple outer-loops. This is illustrated in Fig. 3 which shows time series of the 4D-Var cost function J from ROMS CCS (WC12) from three experiments. In one experiment, the primal form of the strong constraint 4D-Var algorithm is used in conjunction with 5 outer-loops and 15 inner-loops, and the cost function convergence during each outer-loop is relatively fast and monotonic. The jumps that occur in J at the start of each outer-loop are primarily due to two sources of non-linearity in the model associated with (i) 3rd-order upwind finite differencing, and (ii) GLS vertical mixing. In a second experiment, the data assimilation was repeated using the dual form of strong constraint 4D-Var in conjunction with the Lanczos version of augmented RPCG. Figure 3 shows that in agreement with the theory, the primal and dual cost functions track each other perfectly across all inner- and outer-loops. In the third experiment, dual weak

7

constraint 4D-Var was applied in conjunction with Lanczos augmented RPCG to demonstrate the superior performance of the system in this case. Recall that weak constraint is possible in ROMS only using the dual form of 4D-Var.
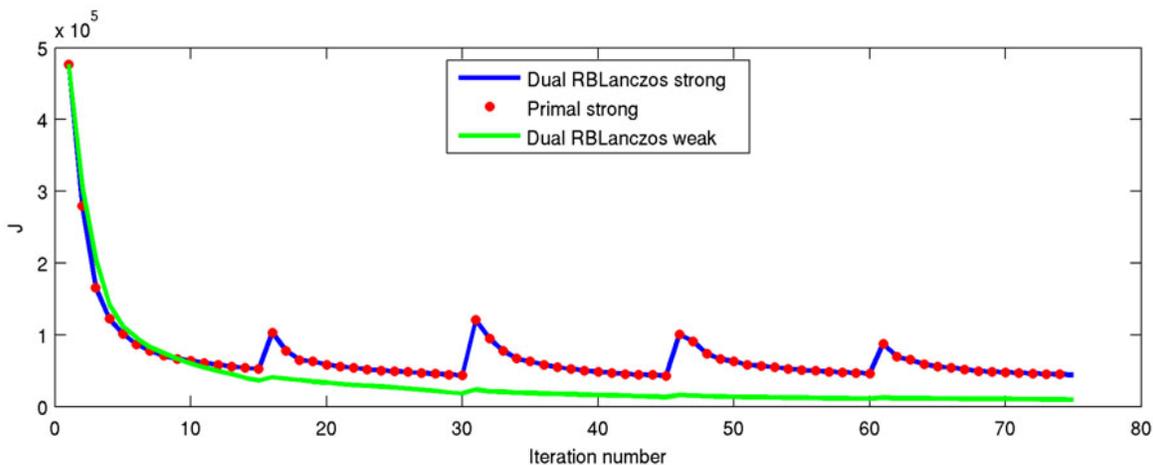

Figure 3: A time series of the cost function *J* versus the total number of iterations (i.e. number of inner-loops × number of outer-loops) for three different ROMS 4D-Var calculations in the CCS (10 km resolution, 42 σ-levels) for a representative 7 day period 29 March – 4 April, 2003, each using 5 outer-loops and 15 inner-loops. In one case the strong constraint primal formulation was used (red dots), while in a second case the strong constraint dual formulation was used in conjunction with the Lanczos formulation of the augmented RPCG algorithm (blue curve). As required the dual formulation using augmented Lanczos-RPCG yields identical iterates to the primal formulation. In the third case, the weak constraint dual 4D-Var algorithm was applied in conjunction with augmented Lanczos RPCG (green curve).

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 the addition of the Lanczos formulation of the augmented RPCG algorithm, time correlations for model error in weak constraint 4D-Var, and three new drivers for computing Hessian SVs, Hessian Forcing 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 Miliff, 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 augmented 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.

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., A.T. Weaver, A.M. Moore, A. Piacentini, H.G. Arango and S. Gratton, 2013: B-preconditioned minimization algorithms for variational data assimilation with the dual formulation. *Q. J. Roy. Meteorol. Soc.*, doi:10.1002/qj.2150.

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, 2011b: 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, 2011c: 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.*, **140**, 3183-3203..

PUBLICATIONS

Gürol, S., A.T. Weaver, A.M. Moore, A. Piacentini, H.G. Arango and S. Gratton, 2013: B-preconditioned minimization algorithms for variational data assimilation with the dual formulation. *Q. J. Roy. Meteorol. Soc.*, doi:10.1002/qj.2150.

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, 2013: A Prototype for an Operational Regional Ocean Data Assimilation System. Chapter 14 In "*Data Assimilation for Atmospheric, Oceanic and Hydrological Applications*, *Vol II*", Liang Xu and Seon Park, Eds. Springer, pp345-366.