Highly-Scalable Algorithms for Ensemble Data Assimilation

Jeffrey Anderson, Nancy Collins, and Helen Kershaw

Institute for Mathematics Applied to Geophysics,
National Center for Atmospheric Research, USA, Email: jla@ucar.edu.

A complete ensemble data assimilation algorithm can be composed into three basic computational tasks: advancing an ensemble of model forecasts, computing ensembles of forward operators for available observations, and assimilating the observations to modify the ensemble of model states. Each of these tasks requires a fundamentally different pattern of communication when implemented on current generation supercomputers.

Large geophysical forecast models are usually parallelized by decomposing the domain into a number of physically contiguous subdomains. Typically, communication is only to adjacent subdomains and is generally facilitated by associating a halo of redundant state variables around the boundaries of each subdomain. The ensembles are independent for the purposes of the forecasts, which makes their calculation embarrassingly parallel.

Computing the forward operators for in situ observations like a radiosonde temperature generally requires a spatial (and possibly temporal) interpolation between spatially contiguous state variables. The communication patterns here are similar to those required for the forecast. However, some observations, for instance a radio occultation phase delay, are functions of many state variables that span large distances in the horizontal and vertical, potentially far beyond any halo region. The communication required to compute these forward operators is not compatible with the haloed subdomains used for the forecast.

Finally, the assimilation requires two additional communication patterns. Sample means, variances and covariances must be computed for model state variables and observation priors, all of which require communication across different ensembles. In addition, information from each observation must be distributed so that all related state variables (most commonly those within a given distance) can be updated.

We begin by describing currently implemented algorithms in which global data transposes and broadcast communication are used to facilitate the different communication patterns [1]. Algorithms that avoid transposes for particular applications are also reviewed [2]. We then describe several novel algorithms that can avoid global transposes. The first will decompose the entire ensemble assimilation problem into a set of physically contiguous subdomains. A second will use data flow methods for the assimilation part of the computation. The computation, communication, and memory requirements of the existing and proposed algorithms as a function of model size, ensemble size, and processor count will be discussed. Implications for implementation on machines with computational accelerators will also be noted.

References:

[1] Anderson, J. and N. Collins. "Scalable implementations of ensemble filter algorithms for data assimilation," *J. Atmos. Oceanic Technol.*, **24**, 1452–1463, 2007.

[2] Wang, Y., Y. Jung, T. Supinie and M. Xue. "A Hybrid MPI/OpenMP Parallel Algorithm and Performance Analysis for an Ensemble Square Root Filter Designed for Multi-scale Observations," To appear in *J. Atmos. Oceanic Technol.*, doi: http://dx.doi.org/10.1175/JTECH-D-12-00165.1, 2013.

An Iterative Ensemble Kalman Smoother

Marc Bocquet[a,b] and Pavel Sakov[c]

[a] *Université Paris-Est, CEREA joint laboratory École des Ponts ParisTech and EdF R&D, France*
[b] *INRIA, Paris Rocquencourt research center, France*
[c] *Bureau of Meteorology, Melbourne, Australia*

Two main classes of data assimilation methods have taken the lead in geophysical data assimilation [1]. As a nonlinear smoother, 4D-Var is a powerful method found to outperform filters in strongly nonlinear conditions. But its background statistics suboptimally rely on a climatology. Besides, it technically requires the long endeavor of building adjoints of the models.

The ensemble Kalman filter (EnKF) easily propagates the uncertainty with the use of an ensemble of model trajectories, which gives it an advantage over variational methods in a sequential data assimilation context. Yet, it is not a smoother and might not handle nonlinearity within a data assimilation window as well as 4D-Var would. Besides, the finite-size of the ensemble technically often requires the use of inflation and localization to counteract sampling errors.

We have recently introduced the iterative ensemble Kalman smoother (IEnKS) that has the potential of getting the best of both methods [2,3,4]. It is not an hybrid method as it does not run two distinct data assimilation systems. Like 4D-Var, as a nonlinear smoother, it solves for an underlying variational problem, but without the use of the tangent linear and adjoint model. Like the EnKF, it is a flow-dependent method and propagates the uncertainty.

This ensemble variational (EnVar) method will be explained. Differences with other EnVar systems (Environment Canada, Météo-France, Met-Office) and hybrid approaches will be discussed. We will show on low-order meteorological models that, at a scalable computational cost, the method systematically outperforms 4D-Var, the EnKF and even a standard ensemble Kalman smoother. Not only does it lead to a much better re-analysis, but it also leads to a better analysis and forecast, even in the typical mild nonlinearity of current synoptic meteorological models.

However, as an ensemble method, it is still plagued by sampling errors. We shall discuss the obstacles to a well-suited localization scheme, as it is a fundamental issue in such EnVar context.

References

[1] Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J. "4D-Var or ensemble Kalman filter?", *Tellus A*, 59, 758-773, 2007.

[2] Sakov, P., Oliver, D. and Bertino, L. "An iterative EnKF for strongly nonlinear systems", *Mon. Wea. Rev.*, 140, 1988-2004, 2012.

[3] Bocquet, M. and Sakov, P. "Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems", *Nonlin. Processes Geophys.*, 19, 383-399, 2012.

[4] Bocquet, M. and Sakov, P. "The iterative ensemble Kalman smoother", *submitted to Q. J. Roy. Meteor. Soc.*, 2013.

Accounting for Model Error in Data Assimilation

Alberto Carrassi[a], and Stephane Vannitsem[b]

[a] Institut Català de Ciències del Clima (IC3), Barcelona, Spain, alberto.carrassi@ic3.cat, [b] Institut Royal Météorologique de Belgique, Bruxelles, Belgique

The treatment of model error in Data Assimilation (DA) procedures is still done, in most instances, following simple assumptions. This is in part justified by the fact that on the time scale of NWP, where most of the geophysical DA advancements have occurred, the influence of model is reasonably considered small as compared to the initial condition error that grows in view of the chaotic nature of the dynamics. Nevertheless, the improvement in DA techniques and observational networks on the one hand, and the recent growth of interest in seasonal-to-decadal (s2d) prediction on the other, has placed model error as a main concern and a key priority.

In the present contribution we describe a new approach, referred to as deterministic model error treatment, in which the deterministic evolution of the model error is described based on a short-time approximation suitable for realistic applications and used to estimate the model error contribution in the state estimate, i.e. covariance and correlations. We have distinguished two situations: first assuming that model error originates only from a misspecification of the parameters, and second from the presence of unresolved scales, and has been applied in the context of sequential and variational approach, for state and parameter estimation. The duration of the short-time approximation depends on the type and size of the model error. In the case of uncertain parameters, it scales with the parametric error but also with the functional dependence of the model dynamics on the uncertain parameters. When the error comes from the presence of unresolved scales, the short-time model error evolution in the resolved scales depends on the difference between the truth (unknown) velocity field and the modelled one: a solution to estimate this term in practice has been proposed.

The deterministic model error treatment has been proven competitive in a number of different applications with prototypical chaotic dynamics, in the framework of sequential [1, 3] and variational schemes [2] as well as for parameter estimation [4, 5]. Research is currently undergone for the application of this approach in soil data assimilation where the soil temperature and moisture content is estimated on the basis of atmospheric observations close to the ground.

References

[1] Carrassi, A., S. Vannitsem, and C. Nicolis, Model error and sequential data assimilation: A deterministic formulation, *Quart. J. Roy. Meteor. Soc.*, **134**, 1297–1313, 2008.

[2] Carrassi A. and S. Vannitsem, Accounting for Model Error in Variational Data Assimilation: A Deterministic Formulation, *Mon. Wea. Rev.*, **138**, 3369–3386, 2010.

[3] Carrassi, A. and S. Vannitsem, Treatment of the error due to the unresolved scales in sequential data assimilation, *IJBC*, **21**, 3619–3626, 2011.

[4] Carrassi, A. and S. Vannitsem, State and parameter estimation with the extended Kalman filter: An alternative formulation of the model error dynamics, *Q. J. Roy. Meteorol. Soc.,* **137** 435–451, 2011.

[5] Carrassi, A., R. Hamdi, P. Termonia and S. Vannitsem, Short time augmented extended Kalman filter for soil analysis: a feasibility study, *Atmos. Sci. Let.* (2012), DOI: 10.1002/asl.394

4D-En-Var : link with weak-constraint 4D-Var and different possible implementations

G. Desroziers[a], J.-T. Camino[a] and L. Berre[a]

[a] *CNRM-GAME, Météo-France and CNRS, France, gerald.desroziers@meteo.fr*

The 4D-Var formulation is currently used by most operational Numerical Weather Prediction centres. However, background error covariances have to be modelled at the initial time of the 4D-Var assimilation period, and their space and time variations are often approximated. Furthermore, the 4D-Var is known to be poorly scalable on computers with a large number of processors, due to its use of low-resolution versions of the linearized model.

The Ensemble Kalman Filter (EnKF) formalism has received a great attention and has initiated a lot of work on the use of ensembles in data assimilation. While the use of localization in observation space raises issues (e.g. for the use of satellite data), it allows indeed background error covariances to evolve in space and time. It was also shown that the EnKF can be combined with the variational approach.

The 4D-En-Var formulation has also received a considerable attention in recent years. The interest in this formulation relies on different nice properties: it allows 4D flow-dependent background error covariances to be represented, it avoids (in contrast to EnKF) the localization of background error covariances in observation space, it is potentially highly parallel on new computer architectures, and it also gets rid of the development, maintenance and cost of tangent-linear and adjoint models.

The aim of this paper ([1], [2]) is to point out the link between the weak-constraint 4D-Var formulation and its 4D-En-Var counterpart. Furthermore, the 4D-En-Var formulation is relatively easy to precondition, and it can potentially include a larger class of representation of model errors. The paper also aims at showing that the so far proposed implementations of the 4D-En-Var implicitly rely on a preconditioning of the variational problem using the square-root of the localized 4D ensemble covariance matrix $\mathbf{B}^e$. Two other possible formulations are proposed, both relying on a preconditioner given by the complete $\mathbf{B}^e$ matrix. One of them performs the minimization in the dual space (with a size given by the number of observations).

It is shown that the avoidance of the need to manipulate the square-root of the $\mathbf{B}^e$ matrix allows a larger flexibility in the specification of the covariance localization needed in 4D-En-Var. The use of a hybrid covariance matrix combining a modelled $\mathbf{B}^c$ matrix and an ensemble $\mathbf{B}^e$ matrix is also discussed. An application of the proposed implementations of 4D-En-Var is shown with the Burgers' model and compared to the use of 4D-Var.

References

[1] G. Desroziers, J.-T. Camino and L. Berre. "A possible implementation of the 4D-Var based on a 4D-ensemble", *Presentation at the International Conference on Ensemble Methods in Geophysical Sciences*, Toulouse, France, 12-16 November. Available online at www.meteo.fr/cic/meetings/2012/ensemble.conference/presentations/session06/2.pdf .

[2] G. Desroziers, J.-T. Camino and L. Berre. "4D-En-Var: link with weak-constraint 4D-Var and different possible implementations", *Q. J. R. Meteor. Soc.*, submitted.

An Iterative Greedy Algorithm for Observation Network Design with an Ensemble Kalman Filter

Denis Dreano[a], Mohamad El Gharamti[b], and Ibrahim Hoteit[b,c]

[a] *Applied Mathematics and Computational Sciences, King Abdullah University of Science and Technology, Saudi Arabia, denis.dreano@kaust.edu.sa,* [b] *Earth Sciences and Engineering, King Abdullah University of Science and Technology, Saudi Arabia,* [c] *Applied Mathematics and Computational Sciences, King Abdullah University of Science and Technology, Saudi Arabia.*

Sequential assimilation systems use incoming observations to guide the trajectory of dynamic models. When only a limited number of sensors can be deployed, which often the case in practice, an important question is then where to locate the observation sensors so that a large amount of information can be retrieved.

In a sequential assimilation setting, it has been proposed to choose an observation network that will ensure a reasonable reduction of the forecast error variance [1]. To achieve this goal, one could perform, at every assimilation cycle, an optimization problem to maximize the gain between the forecast and the analysis statistics. This would generally result in a "sub-optimal" observation network using some greedy algorithms.

In this study, we propose an alternative algorithm that is expected to produce "more optimal" observation networks. We use for assimilation the well-known Singular Evolutive Interpolated Kalman (SEIK) filter. The method works by pushing the results of the greedy algorithm more towards optimality through backward iterations on the sub-optimal network until convergence. Convergence to some network usually happens in a finite number of steps. Numerical results from synthetic experiments using the wave equation and the Lorenz-96 model will be presented. We compare the performance of the filter when using fixed observations, adaptive observations with a greedy algorithm and adaptive observations with an iterative greedy algorithm.

References

[1] T. M. Hamill and C. Snyder. "Using improved background-error covariances from an ensemble Kalman filter for adaptive observations," *Monthly weather review*, vol. 130, pp. 1552-1572, 2002.

Inter-comparison of static and flow-dependent estimates of background error covariances

Amal El Akkraoui[ab], and Ricardo Todling[b]

[a] *Science System and Applications, Inc., USA,amal.elakkraoui@nasa.gov,* [b] *Global Modeling and Assimilation Office, NASA, USA.*

Variational data assimilation techniques rely on a good estimate of the background error covariance matrix (B) to provide reliable atmospheric analyses. The so-called NMC method is traditionally used to estimate a climatological B from the statistics of different-length differences of forecasts verified at the same time. This approach parameterizes the background error covariance as homogeneous and isotropic, and requires explicit tuning of its covariance coefficients. Recently, introduction of hybrid assimilation methodologies combines the static structure of B with flow-dependent estimates derived from an available ensemble of forecasts running in parallel with the data assimilation system. Ensemble estimates of B have the additional advantage of implicitly obtaining correlations among all control variables, but suffer from being low-rank due to limited size of the typical ensemble. Localization is normally applied to remedy this issue so long as the scales are adequately tuned. This work plans to examine the structure of B derived in these two ways when using forecasts from the Global Modeling and Assimilation Office (GMAO) hybrid three-dimensional variational (3D-Var) data assimilation system.

Motivated by a possible reformulation of the Incremental Analysis Update procedure of Bloom et al. (1996) more consistent with extensions of 3D-Var to weak-constraint 4D-Var, this work also plans to derive error covariance estimates from model forecasts tendency differences (e.g. Trémolet 2007). Both the NMC- and ensemble-based error covariance estimation methodologies mentioned above will be used to characterize climatological and flow-dependent components of tendency-based error covariance matrices to aid the GMAO variational assimilation procedures.

Preliminary inter-comparison results will be shown for these four error covariance estimates. The correlation lengths and structures will be evaluated, and analysis increments of corresponding twin experiments will be examined in the context of the Grid-point Statistical Interpolation system used at GMAO. Results from cycling experiments will be the subject of a companion work also presented at this Symposium (see Todling et al.).

References

[1] Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. Mon. Wea. Rev., 124, 1256-1271.

[2] Trémolet, Y., 2007: Model error estimation in 4D-Var, Q. J. R. Meteorol. Soc., 133, 1267-1280.

The role of additive and multiplicative noise in filtering complex dynamical systems

John Harlim

Department of Mathematics, North Carolina State University, USA, jharlim@ncsu.edu

Covariance inflation is an ad-hoc treatment that is widely used in practical real-time data assimilation algorithms to mitigate covariance underestimation due to model errors, nonlinearity, or/and, in the context of ensemble filters, insufficient ensemble size. Here, we systematically derive an effective "statistical" inflation for filtering multi-scale dynamical systems with moderate scale gap, $\varepsilon = O(10^{-1})$, to the case of no scale gap with $\varepsilon = O(1)$, in the presence of model errors through reduced dynamics from rigorous stochastic subgrid-scale parametrizations.

We will demonstrate that for linear problems, an effective covariance inflation is achieved by a systematically derived additive noise in the forecast model, producing superior filtering skill. For nonlinear problems, we will study an analytically solvable stochastic test model, mimicking turbulent signal in regimes ranging from a turbulent energy transfer range to a dissipative range to a laminar regime. In this context, we will show that multiplicative noise naturally arises in addition to additive noise in a reduced stochastic forecast model. Subsequently, we will show that a "statistical" inflation factor that involves mean correction in addition to covariance inflation is necessary to achieve accurate filtering in the presence of intermittent instability in both the turbulent energy transfer range and the dissipative range.

Reference:

[1] G. Gottwald and J. Harlim, "The role of additive and multiplicative noise in filtering complex dynamical systems**"** ***Proc. R. Soc. A*****, 2013. (in press)**

Square Root and Perturbed Observation Ensemble Generation for Kalman and Quadratic Ensemble Filters

Daniel Hodyss[a] and William F. Campbell

[a] *Division of Marine Meteorology, Naval Research Laboratory, Monterey, California, USA, Email: daniel.hodyss@nrlmry.navy.mil*

Ensemble-based Kalman Filter (EBKF) algorithms have proven highly effective in a wide-variety of applications in the geosciences. However, the application of the EBKF to the highly nonlinear physical process found in geoscience applications may lead in certain situations to non-Gaussian prior distributions and therefore suboptimal or even degenerative behavior. A new extension of the EBKF will be described here that allows for the explicit effects of skewness in the prior distribution to be accounted for in the data assimilation algorithm. The algorithm operates as a global-solve (all observations are considered at once) using a minimization technique and Schur/Hadamard (element-wise) localization. The central feature of this technique is the squaring of the innovation and the ensemble perturbations so as to create an extended state-space that accounts for the second, third and fourth moments of the prior distribution. Both square-root and perturbed observation ensemble generation techniques will be shown to be implementable within the new framework and will be compared against the corresponding technique within the traditional EBKF. These new techniques will be illustrated on the Lorenz (1963) attractor as well as in a Boussinesq model of $O(10^4)$ variables configured to simulate nonlinearly evolving Kelvin-Helmholtz waves in shear flow. It is shown that ensemble sizes of at least 100 members are needed to adequately resolve the third and fourth moments required for the algorithm. For ensembles of this size it is shown that this new technique is superior to the EBKF in situations with significant skewness, otherwise the new algorithm reduces to the performance of the EBKF.

Assimilation of wind speed and direction observations: a new formulation and results from idealized experiments

Xiang-Yu Huang[a], Feng Gao[a,b], Neil Jocobs[c] and Hongli Wang[a]

[a] *National Center for Atmospheric Research, Boulder, CO, USA,* [b] *Nanjing University of Information Science & Technology, Nanjing, China,* [c] *AirDat LLC, Morrisville, NC, USA.*

We present a new methodology for assimilating wind observations in their observed form of speed and direction, while taking into account both speed and direction error. It ensures the analyzed speed and direction will be consistent with their background and observed values. The new formulation is implemented in the Weather Research and Forecasting Data Assimilation system, and idealized experiments are used to demonstrate the potential benefit. The results suggest that analyses from the new formulation are more reasonable when compared to the conventional methodology. The forecasts generated in these idealized experiments also demonstrate the value of this new formulation.

References

Huang, X.-Y., Gao, F., Jacobs, N. and Wang, H. 2013. "Assimilation of wind speed and direction observations: a new formulation and results from idealized experiments". *Tellus A*, 65, 19936, http://dx.doi.org/10.3402/tellusa.v65i0.19936.

LETKF-based method for Hybrid Data Assimilation

Kayo Ide[a]

[a]*University of Maryland, College Park, MD, USA, ide@umd.edu*

A hybrid data assimilation method based on the Local Ensemble Transform Kalman Filter (LETKF) is formulated with three schemes to enhance performance. First, by taking advantage of the cost function representation of LETKF, this hybrid method fully integrates the VAR and the EnKF elements without the necessity to re-center ensembles while the analysis ensemble perturbation represents the ensemble spread associated with the hybrid analysis. Second, using the maximum likelihood approach, the weights between dynamic (ensemble) and static (VAR) background covariances as well as the inflation parameters are estimated adaptively. Third, multiscaleness is addressed by assimilating successively from large to smaller scales. Using the 960-variable Lorenz model, the hybrid method is evaluated in the 3D- and 4D-hybrid modes and advantages of the three supplemental schemes are discussed.

Mass Conservation and Positivity Preservation with Ensemble-type Kalman Filter Algorithms

Tijana Janjic [a], Dennis B. McLaughlin [b], Stephen E. Cohn [c] and Martin Verlaan [d]

[a] *Hans Ertel Centre for Weather Research, DWD, Germany, tijana.janjic-pfander@dwd.de,*
[b] *Massachusetts Institute of Technology, USA,* [c] *NASA Goddard Space Flight Center, USA,*
[d] *Deltares/TU Delft, Netherlands.*

Maintaining conservative physical laws numerically has long been recognized as being important in the development of numerical weather prediction (NWP) models. Numerical conservation of the total mass of air, water in its different phases, chemical species, etc., including sign (positivity) preservation in each grid volume, is sometimes taken to be the first of a number of fundamental design principles for NWP models. In the broader context of data assimilation, concerted efforts to maintain conservation laws numerically and to understand the significance of doing so have begun only recently.

In this talk we consider the use of equality and inequality constraints in ensemble-based Kalman filter algorithms to enforce mass conservation and positivity preservation. We show that the analysis steps of ensemble transform Kalman filter (ETKF) algorithm and ensemble Kalman filter algorithm (EnKF) can conserve the mass integral, but do not preserve positivity. Further, if localization is applied or if negative values are simply set to zero, then the total mass is not conserved either. In order to ensure mass conservation, a projection matrix that corrects for localization effects is constructed. In order to maintain both mass conservation and positivity preservation through the analysis step, we construct two data assimilation algorithms based on quadratic programming and ensemble Kalman filtering.

We examine the extent to which imposing mass conservation and positivity preservation in data assimilation changes the assimilation results and we demonstrate the benefits of imposing these constraints in two simple experimental setups: a solid body rotation experiment and idealized problems of environmental pollution. The results show clear improvements in both analyses and forecasts, particularly in the presence of localized features.

References

[1] Janjic, T., D. B. McLaughlin, S. E. Cohn: "Preservation of physical properties with ensemble based Kalman filter algorithms", *Mathematical and Algorithmic Aspects of Atmosphere-Ocean Data Assimilation, Mathematisches Forschungsinstitut Oberwolfach*, Report No. 58/2012, 17–20, doi:10.4171/OWR/2012/58.

[2] T. Janjic, D. B. McLaughlin, S. E. Cohn: "Conservation of mass and preservation of positivity with ensemble type Kalman filter algorithms", submitted to *Monthly Weather Review*.

The Ensemble Variational Assimilation and Bayesian Estimation

M. Jardak & O. Talagrand
Laboratoire de Météorologie Dynamique
École Normale Supérieure
Paris, France

Abstract

The purpose of assimilation can be stated as to determine the probability distribution for the state of the observed system, conditioned by the available data. In the case of linearity and additive Gaussian errors, a sample of independent realizations of the (Gaussian) conditional probability distribution can be obtained by repeating the following process. Perturb the data according to their own error probability distribution, and compute the Best Linear Unbiased Estimator (BLUE) corresponding to those perturbed data.

This approach has been implemented on two small dimension nonlinear chaotic systems, the Lorenz'96 and the Kuramoto-Sivashinsky equations, in the form of **Ensemble Variational Assimilation**, or **EnsVar**, in which the BLUE is determined by standard Variational Assimilation. The Bayesian character of a probability distribution cannot be in general objectively verified, and the weaker property of reliability (statistical consistency between predicted probabilities and observed frequencies of occurrence) is used instead.

The main conclusion is that EnsVar produces almost perfectly reliable, and accurate, ensembles. This conclusion remains valid for long assimilation periods, either through the use of Quasi-Static Variational Assimilation, in which the length of the assimilation window is progressively increased (in the case of a perfect model), or through weak-constraint assimilation (in the case of an imperfect model). In addition, non-Gaussianity of the errors has no significant impact.

Comparisons with Ensemble Kalman Filter and Particle Filters produce ensembles with a significantly lower reliability and lesser accuracy. On the other hand, the cost of EnsVar is higher.

The significance of those results is discussed in the general perspective of Bayesian estimation.

A practical ensemble technique for the estimation of spatially inhomogeneous, model-dependent observational errors

Alicia R. Karspeck,

Climate and Global Dynamics, National Center for Atmospheric Research, USA, aliciak@ucar.edu.

Most least-squares algorithms for data assimilation require estimates of both background error covariances and observational error variances. The relative size of observational and background errors guides the influence of an observational increment, making the specification of these errors an important part of designing an assimilation system. Observational error estimates typically lump together measurement (and instrumental) error sources and representativeness error. When doing data assimilation with a coarse-resolution dynamical model, the errors of representation can be the dominant contribution to observational errors. Unlike resolved model errors, which influence the background error estimates, representativeness errors are associated with unresolved scales and physical processes in the model. The size of these errors will vary by geographic region and will also vary from model to model.

Here we present a practical approach for estimating model-dependent, spatially varying observational error variances. The method uses ensemble model simulations to compute an expected value and uncertainty associated with the observational error variance. We illustrate the method with a low-order model and show its application in the POP2 global ocean general circulation model.

Ensemble Data Assimilation without Ensembles with Application to Ocean Data Assimilation

Christian L. Keppenne[a] Michele M. Rienecker[b], Robin M. Kovach[a] and Guillaume Vernieres[a]

[a] *Science Applications International Corporation, USA, Christian.keppenne@nasa.gov,*
[b] *Global Modeling and Assimilation Office, NASA, USA.*

Two methods to estimate background error covariances for data assimilation are introduced. While both share properties with the ensemble Kalman filter (EnKF), they differ from it in that they do not require the integration of multiple model trajectories. Instead, all the necessary covariance information is obtained from a single model integration. The first method is referred-to as SAFE (Space Adaptive Forecast error Estimation) because it estimates error covariances from the spatial distribution of model variables within a single state vector. It can thus be thought of as sampling an ensemble in space. The second method, named FAST (Flow Adaptive error Statistics from a Time series), constructs an ensemble sampled from a moving window along a model trajectory. The underlying assumption in these methods is that forecast errors in data assimilation are primarily phase errors in space and/or time.

SAFE and FAST are applied to the assimilation of Argo temperature profiles into version 4.1 of the Modular Ocean Model (MOM4.1) coupled to the GEOS-5 atmospheric model and to the CICE model developed at Los Alamos National Laboratory. The results are validated against un-assimilated Argo salinity data. They show that SAFE and FAST are competitive with the ensemble optimal interpolation (EnOI) used to produce the latest ocean analysis by the Global Modeling and Assimilation Office (GMAO). Further, when only temperature data are assimilated, FAST is better able to improve the model salinity field than SAFE or EnOI.

Stochastic Galerkin method for dynamic data assimilation using Wiener's polynomial chaos

S.Lakshmivarahan and $JunjunHu$
School of Computer Science, University of Oklahoma(Varahan@ou.edu)

Let $X(t,\omega)$ be a real valued square integrable stochastic process. It is well known that this process can be represented using the so called polynomial chaos expansion developed by Wiener in 1938 and is given by $X(t,\omega) = \sum_{k=0}^{\infty} x_k(t)H_k(\xi)$ where ξ is the standard normal random variable, $H_k(x)$ is the Hermite polynomial of degree k in x. It is well known that $H_0(\xi) = 1$, for $k \geq 0$, $E[H_k(\xi)] = 0$, and $E[H_n(\xi)H_m(\xi)] = h_n\delta_{nm}$. That is, the Hermite polynomials are orthogonal with respect to the standard Gaussian weighting function.Let $X_N(t,\omega)$ be an approximation to $X(t,\omega)$ obtained by keeping only the first (N+1) terms of the infinite expansion. It can be shown that $X_N(t,\omega)$ converges (in mean square) to $X(t,\omega)$ as N $\rightarrow \infty$.

Let $X^f(t,\omega)$ be the solution of a stochastic dynamical system with (a)random initial condition, (b) random forcing, and (c) random parameters. Indeed, we can represent all the random variables and random processes including the solution using the (N+1) mode polynomial chaos expansion in the Hermite basis. Substituting this representation into the dynamical model and taking inner product with $H_k(\xi)$ for $0 \leq k \leq N$ and exploiting the orthogonality property of Hermite polynomials, we perform the standard Galerkin projection of the dynamics onto the (N+1) dimensional subspace. This results in a system of (N+1) systems of coupled ODE in the time varying coefficients $x_k^f(t)$. By solving this system, we can reconstruct $X_N^f(t,\omega) = \sum_{k=0}^{N} x_k^f(t)H_k(\xi)$. By drawing M independent samples $\xi(j)$ from the standard Gaussian distribution, we can obtain forecast ensemble $X_N^f(t,\omega)(j)$. Using this forecast ensemble, we can readily compute the forecast covaraiance matrix. Notice that, unlike the other contemporary methods, we solve the forecast dynamics only once and be able to generate the forecast ensemble rather easily by drawing samples from the standard normal distribution [1]. Let $Z(t) = x(t) + v(t)$ the observation with the known covaraince R. We can assimilate the observation either using (a) the stochastic framework with perturbed observations or (b) the deterministic ensemble transform methods to create the analysis ensemble. We illustrate the above framework using nonlinear model given by the stochastic differential equation $dx(t) = -4x(1-x^2)dt + \sigma dw(t)$ where dwt is the standard Brownian incremental process with random initial conditions.

[1] D. Xiu (2010) Stochastic Computations, Princeton University Press

Pontryagin's Minimum Principle
Applied to Meteorological Data Assimilation

S. Lakshmivarahan[a] and John M. Lewis[b]

[a] School of Computer Science, University of Oklahoma, USA (varahan@ou.edu)
[b] National Severe Storms Laboratory, USA

Pontryagin's Minimum Principle (PMP) came into prominence during the1950s – 1960s when space travel became a dominant theme [1]. A typical problem centered on minimizing the travel time or energy used by a space vehicle between two specified points by adding feedback control to the dynamical constraints — essentially taking account of the difference between the vehicle's position and a desired position determined *a priori*. Instead of the classical Lagrangian approach which describes the optimal trajectory as the solution of a second-order PDE, PMP using the Hamiltonian approach describes the optimal trajectory as the solution of first-order equations. Computationally, the optimal control and the optimal trajectory are obtained as a solution of a nonlinear two point boundary value problem (TPBVP) which can be solved iteratively by one of many known methods such as the shooting method, gradient methods, to name a few.

When the model and observations are linear, it can be shown using the sweep method, that the TPBVP reduces to two initial value (IVP) problems - one of which takes the form of a nonlinear matrix Riccati equation which is independent of the data and the other is a linear vector equation that depends on the data and uses the solution of the Riccati equation. Since the Riccati equation is independent of data, its solution can be computed off-line and can be used repeatedly.

In this research, we describe the first attempt at adapting the PMP to meteorological data assimilation [2] by optimally determining a forcing term (as a feedback control term) that minimizes the quadratic fit of model to noisy observations. Numerical tests of fit are performed on the spectral dynamics of the Burgers' equation — a linear form of the equation that captures the development of an extreme gradient and frontal-like structure. Results indicate that the feedback control terms produce excellent fit of model to observations. Further, the methodology is reconfigured to give information on model error [2].

References

[1] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E .Mischenko,(1962) *The Mathematical Theory of Optimal Control Processes*; Interscience Pub.,
[2] S. Lakshmivarahan, J. Lewis and D. Phan (2013) "Data Assimilation as Problem in Optimal Tracking: Application of Pontyagin's Minimum Principle to Atmospheric Sciences", **Journal of Atmospheric Sciences**, Vol 70, April Issue, pp1257-1277

Error Dynamics and Instability in Data Assimilation

Alexander J.F. Moodey[a], Amos S. Lawless[a], Roland W.E. Potthast[ab] and Peter Jan van Leeuwen[a]

[a] *School of Mathematical and Physical Sciences, University of Reading, U.K.* a.s.lawless@reading.ac.uk
[b] *Deutscher Wetterdienst, Germany.*

In many applications of data assimilation the observation operator, which maps from state space to observation space, leads to an ill-posed equation, meaning that the assimilation problem is unstable to small perturbations in the data. The use of a prior or background estimate acts to regularize the problem and thus ensure its stability at each assimilation time. This can be shown to be equivalent to a form of Tikhonov regularization as used in traditional inverse problems. Several common data assimilation algorithms, including variational methods and Kalman filters, can be considered in this framework.

In operational weather forecasting the data assimilation problem is usually cycled, so that the background state for one assimilation time is provided by a forecast of the analysis from the previous assimilation time. In this work we examine the stability of the error in a sequence of analyses as the assimilation process is cycled in time. Since the data assimilation problem is often solved in very high dimensional systems (of order 10^8 and higher), we derive theory using an infinite-dimensional framework.

We show that for a certain class of linear model dynamics it is possible to guarantee the stability of the analysis error in time by applying a multiplicative inflation to the background error variances. In the case of time-varying dynamics the inflation factor can be chosen adaptively at each assimilation time to ensure stability. However, as the inflation factor is increased, the assimilation problem at each time is less well-conditioned and the bound on the analysis error increases [1]. For nonlinear dynamics similar stability results are obtained under certain Lipschitz continuity and dissapitivity assumptions on the dynamical operator [1], [2]. The theory is illustrated with numerical results.

References

[1] Moodey, A.J.F. "Instability and regularization for data assimilation," *PhD thesis*, Department of Mathematics and Statistics, University of Reading, 2013. *(Submitted)*
[2] Moodey, A.J.F., Lawless, A.S., Potthast, R.W.E. and van Leeuwen, P.J. "Nonlinear error dynamics for cycled data assimilation methods," *Inverse Problems*, vol 29, pp. 025002, 2013

Posterior covariance and analysis error covariance in variational data assimilation

Francois-Xavier Le Dimet[a], Victor Shutyaev[b], and Igor Gejadze[c]

[a] *MOISE project, LJK, University of Grenoble, France, Email: Francois-Xavier.Le-Dimet@imag.fr,*
[b] *Institute of Numerical Mathematics, Russian Academy of Sciences, Russia,* [c] *Department of Civil and Environmental Engineering, University of Strathclyde, UK.*

The problem of variational data assimilation for a nonlinear evolution model is stated as an optimal control problem to estimate the analysis (initial condition) [1]. The data contain errors (observation and background errors), hence there is an error in the analysis. For mildly nonlinear dynamics, the analysis error covariance can be approximated by the inverse Hessian of the cost functional in the auxiliary data assimilation problem [2], whereas for stronger nonlinearity the 'effective' inverse Hessian approach may be used [3, 4]. However, the analysis error covariance is not the posterior covariance from the Bayesian perspective. While these two are equivalent in the linear case, the difference may become significant for nonlinear problems [5]. For the proper Bayesian posterior covariance a new approximation via the Hessian of the original cost functional is derived and its 'effective' counterpart is introduced. An approach for computing the analysis error covariance and posterior covariance in the matrix-free environment using Lanczos method with preconditioning is suggested. Numerical examples which validate the developed theory are presented for the model governed by the Burgers equation with a nonlinear viscous term.

The authors acknowledge the funding through the Natural Environment Research Council (NERC grant NE/J018201/1), the Russian Foundation for Basic Research (project 12-01-00322), the Ministry of Education and Science of Russia, the MOISE project (CNRS, INRIA, UJF, INPG) and Région Rhône-Alpes.

References

1. Le Dimet F.X., Talagrand O. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, 1986, v.38A, pp.97-110.
2. Gejadze I., Le Dimet F.-X., Shutyaev V. On analysis error covariances in variational data assimilation. *SIAM J. Sci. Computing*, 2008, v.30, no.4, pp.184-1874.
3. Gejadze I.Yu., Copeland G.J.M., Le Dimet F.-X., Shutyaev V. Computation of the analysis error covariance in variational data assimilation problems with nonlinear dynamics. *J. Comp. Phys.*, 2011, v.230, pp.7923-7943.
4. Shutyaev V., Gejadze I., Copeland G.J.M., Le Dimet F.-X. Optimal solution error covariance in highly nonlinear problems of variational data assimilation. *Nonlin. Processes Geophys*., 2012, v.19, pp.177-184.
5. Gejadze, I., Shutyaev, V.P., Le Dimet, F.-X. Analysis error covariance versus posterior covariance in variational data assimilation. *Q. J. R. Meteorol. Soc.*, 2012, 1-16. DOI: 10.1002/qj.2070.

Validation of a New Algorithm for Empirical Localization of Observations for Ensemble Kalman Filter Data Assimilation in Global Atmospheric Models

Lili Lei and Jeff Anderson

National Center for Atmospheric Research, USA, lililei@ucar.edu

Localization in the ensemble Kalman filter (EnKF) is a technique to reduce the sampling error in the statistical relations between observations and model state variables. Localization is required for good results in applications of the EnKF to large atmospheric and oceanic models. The empirical localization algorithm described here uses the output from an observing system simulation experiment (OSSE) and constructs localization functions that minimize the root mean square (RMS) difference between the truth and the posterior ensemble mean for state variables. This algorithm can automatically provide an estimate of the localization function and does not require tuning of the localization scale. Moreover, the algorithm can compute an appropriate localization function for any potential observation type and state variable kind.

The empirical localization algorithm is investigated in the dynamical core of the Geophysical Fluid Dynamics Laboratory (GFDL) B-grid model and the Community Atmospheric Model version 5 (CAM5). In the B-grid model, the empirical localization is computed for every observation type and state variable kind. The empirical localizations are Gaussian-like functions and have detailed structures for observations and state variables that are relatively close to each other. The empirical localization outperforms the best Gaspari and Cohn (GC) localization that is obtained by tuning the GC localization cutoff. In CAM5, the empirical localization function is computed for the horizontal and vertical separately, thus the vertical localization is explored. The horizontal empirical localizations are similar to the GC localization, but the vertical empirical localizations are much broader and have a nearly linear relationship with the natural logarithm of pressure for small separations between observations and state variables. The empirical localization in CAM also produces smaller RMS error than the GC localization that is routinely used for assimilation with CAM.

Application of Physical Filter Initialization Scheme on WRFVAR

Xudong Liang[a], Shuiyong Fan[a], Xiang-Yu Huang[b] , and Xin Zhang[b]

[a] *Institute of Urban Meteorology, CMA, Beijing, China xdliang@ium.cn,* [b] *National Center for Atmospheric Research, Boulder, Colorado, USA*

The 3DVar and 4DVar has come a long way since its beginning in the late 1980s and early 1990s. In variational data assimilation scheme, usually, the analysis fields are not in balance because of poorly defined background error covariance, insufficient observations, imperfect modal, and observation error. Some researches have been focused on using weak constraints to reduce the dynamic imbalance between model variables based on the idea that unbalanced initial conditions often generate high-frequency oscillations with amplitude larger than those observed in nature. One of the approaches is digital-filter initialization (DFI) proposed by Lynch and Huang (1992)[1]. Another approach is to use some physical constraint such as temporal and spatial smoothness penalty functions. Liang et al. (2007)[2] proposed a model constrained 3DVar (MC-3DVar) technique to apply the full physics and dynamics of numerical model as constraints in 3DVar. In MC-3DVar, the full physics and dynamics of numerical model act as low-pass filter (physical filter) which can dramatically reduce the high frequency oscillations in the analyses fields. In this study, the physical filter initialization (PFI) scheme used in 3DVar is extended to four-dimension and implemented in WRF-4DVar system with the cost function as

$$J_4 = [\mathbf{x}-\mathbf{x}_b]^T \mathbf{B}^{-1}[\mathbf{x}-\mathbf{x}_b] + \sum_i^N [\mathrm{H}(\mathbf{x}_i)-\mathbf{y}_i]^T \mathbf{O}^{-1}[\mathrm{H}(\mathbf{x}_i)-\mathbf{y}_i] + \sigma \sum_j^M [\frac{\Delta \mathbf{x}_j}{\Delta t}]^T \mathbf{R}^{-1}[\frac{\Delta \mathbf{x}_j}{\Delta t}] \quad (1).$$

The first term of Eq.1 is as same as that in 3DVar. The second term is for distance between analysis and observations at time $i = 1,2...N$. The third term is the penalty defined at time $j = 1,2...M$.

Idealized experiment and real case study were performed in this study using WRF-4DVar with PFI scheme. The results indicate that the high frequency oscillations are obviously eliminated using PFI scheme and the spin-up time is shortened. The analysis and forecast fields are analyzed carefully and comparision with general 4DVar and 4DVar with DFI scheme in this study.

References

[1]Lynch P, Huang XY. Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.* **120**, 1019–1034, 1992.

[2]Liang Xudong, B. Wang, CL Chan, et al. Tropical cyclone forecasting with model-constrained 3D-Var. I: Description. *Quart. J. Roy. Meteor Soc.*, **133**, 147-153, 2007.

Handling of Position Errors in Variational and Hybrid Ensemble/Variational Data Assimilation Using Image Registration

Tomas Landelius[a], Nils Gustafsson[a], Magnus Lindskog[a], Jelena Bojarova[b]

[a] *The Swedish Meteorological and Hydrological Institute, SMHI, SE-601 76, Norrköping, Sweden, email: Magnus.Lindskog@smhi.se*
[b] *The Norwegian Meteorological Institute, P.O. Box 43, Blindern, NO-0313 Oslo, Norway*

It is hard to forecast the position of localized weather phenomena such as clouds, precipitation, and fronts. Moreover, cloudy areas are important since this is where most of the active weather occurs. Position errors, also known as phase or alignment or displacement errors, can have several causes; timing errors due to deficient model physics, inadequate model resolution, etc. Furthermore, position errors have been shown to be non-additive and non-Gaussian, which violates the error model most data assimilation methods rely on.

Remote sensing data contain coherent information on the weather development in time and space. By comparing structures in radar or satellite images with the forecast model state it is possible to get information about position errors. We use an image registration (optical flow) method to find a transformation, in terms of a displacement field, that aligns the model state with the corresponding remote sensing data. The estimated displacement field is used differently in variational and in hybrid ensemble/variational data assimilation.

In the variational setting the displacement field is used as a mapping function to obtain a new, better aligned, background state from the old one by means of interpolation (warping). To reduce the effect of imbalances, the aligned first guess is not used as is. Instead it is used for generation of pseudo observations that are assimilated in a first step to get an aligned and balanced first guess. This step reduces the non-additive errors due to mis-alignment and is followed by a second step with a standard variational assimilation to compensate for the remaining additive errors.

In the hybrid ensemble/variational data assimilation a displacement field is estimated for each ensemble member and is used as a distance measure. In areas where a member has a smaller displacement (smaller position error) than the control it is given an increased weight in the subsequent assimilation. It is also possible to first obtain an aligned and balanced background state based on variational assimilation of superobservations from a composite consisting of information from the best member (least position error) in each grid point.

Results are presented from experiments done with both variational and ensemble data assimilation systems using satellite data from the SEVIRI instrument.

A Unified Framework for Four-Dimensional Ensemble-Variational Hybrid Data Assimilation

Chengsi Liu[1] and Ming Xue[1,2]
[1]Center for Analysis and Prediction of Storms and [2]School of Meteorology
University of Oklahoma, Norman Oklahoma 73072, U.S.A
cliu@ou.edu

Four ensemble-variational data assimilation (DA) algorithms that can incorporate the time dimension (four-dimensional or 4D) and combine static and ensemble-derived background error covariances (hybrid) are formulated in general form and their relationships are discussed. These algorithms are: (1) the standard 4D variational (4DVar) employing ensemble-derived covariance (En4DVar), (2) the 4D ensemble-variational with non-time-propagating control variable (4DEnVar-NPC), (3) the 4D ensemble-variational with propagating localization (4DEnVar-PL), and (4) the 4DEnVar with non-propagating localization (4DEnVar-NPL). Those algorithms can be derived, with generalization in some cases, from existing ensemble and/or variational algorithms with or without adjoint (Lorenc 2003, Liu et al 2009, Zhang et al 2009, Buehner et al 2010). For comparison purpose, 4D ensemble Kalman filter (EnKF) with flow-following localization (4DEnKF-FL) and non-flow-following localization (4DEnKF-NFL) are also introduced. All of the above algorithms can be made hybrid by combining static and ensemble covariances, either through the extended control variable method or explicit method.

It is shown mathematically that for a linear system, (1) hybrid En4DVar, 4DEnVar-PL and 4DEnKF-FL are equivalent, while 4DEnVar-NPC, 4DEnVar-NPL and 4DEnKF-NFL are equivalent; (2) the adjoint model used in En4DVar and the huge perturbation matrix used in 4DEnVar-PL for localization allow for flow-following time evolution of covariance localization but carry high software development and/or computational costs; (3) Non-propagating localization in 4DEnVar-NPC and 4DEnVar-NPL avoid adjoint and tangent linear models, but at the expense of losing flow-following covariance localization.

Numerical DA experiments are performed with a one-dimensional advection model to confirm the mathematical equivalence among various algorithms, and to demonstrate the effects of non-propagating or non-flow-following localization.

Buehner, M, P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. Mon. Wea. Rev., 138, 1550–1566.

Liu, C., Q. Xiao, and B. Wang, 2009: An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with Advanced Research WRF (ARW). Mon. Wea. Rev., 137, 1687-1704.

Lorenc, A., 2003: The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. Quart. J. Roy. Meteor. Soc., 129, 3183-3204.

Zhang, F., M. Zhang and J. A. Hansen, 2009: Coupling ensemble Kalman filter with four-dimensional variational data assimilation, Advances in Atmospheric Sciences, 26, 1-9.

Ensemble Bayesian filtering with Residual Nudging

Xiaodong Luo[a], and Ibrahim Hoteit[b]

[a] *IRIS, Norway, xiaodong.luo@iris.no,* [b] *AMCS, KAUST, Saudi Arabia.*

An auxiliary data assimilation technique, called residual nudging, is proposed in order to monitor and, if necessary, adjust the distance (residual norm) between the predicted observation and the true observation.

A rule for choosing the threshold of the residual norm is suggested, and two different methods are introduced in order to make the residual norm no larger than the specified threshold. In the first method, the objective of residual nudging is achieved by (potentially) introducing certain modification to the analysis mean after the observation is assimilated. Conceptually, this method is applicable to various data assimilation methods, including, for instance, the ensemble Kalman filter (EnKF) [3] and the particle filter [2].

The second method is specific for the EnKF. The objective of residual nudging is achieved by modifying the mean update formula before the information content of the observation is incorporated. Such a modification is equivalent to tuning the covariance inflation factor in the EnKF. It is shown that, in order for the residual norm to be bounded in a certain interval, the corresponding covariance inflation factor should be bounded in an interval as well. The bounds of the covariance inflation interval are explicitly derived, and some implications of the analytic results are highlighted [1].

References

[1] X. Luo and I. Hoteit. "Covariance inflation in the ensemble Kalman filter: a residual nudging perspective and some implications," *Mon. Wea. Rev.*, submitted.

[2] X. Luo and I. Hoteit. "Efficient particle filtering through residual nudging," *Quart. J. Roy. Meteor. Soc.*, in press (preprint http://arxiv.org/abs/1303.2698).

[3] X. Luo and I. Hoteit. "Ensemble Kalman filtering with residual nudging," *Tellus A*, vol. 64, 17130, 2012 (http://www.tellusa.net/index.php/tellusa/article/view/17130/html).

Non-Gaussian Ensemble Data Assimilation using a Deterministic Approach

Sammy Metref[a], Emmanuel Cosme[a], Chris Snyder[b] and Pierre Brasseur[a]

[a] *MEOM, LGGE, Grenoble, France, sammy.metref@hmg.inpg.fr,* [b] NCAR, Boulder, *Colorado, USA.*

Various methods implementing non-Gaussian data assimilation are described in the literature. But all of them are either extensions of Gaussian-based methods (e.g. EnKF with anamorphosis) or Monte-Carlo based filters still inapplicable to high dimensional systems (e.g. particle filters) as mentioned in Snyder et al. [2008].

In this presentation, we introduce an ensemble filter, the Multivariate Rank Histogram Filter (MRHF), which extends the Rank Histogram Filter (RHF) of Anderson [Anderson, 2010] in order to apply Bayes theorem to solve the full non-Gaussian problem. While the RHF takes into account non-Gaussianity only in the observed variables and the observations, the MRHF also deals with the joint non-Gaussianity of all variables. The data assimilation problem is solved by splitting the problem of estimating a joint density into a series of smaller problems of estimating conditional, one-dimensional densities. The MRHF analysis scheme is completely deterministic and reproducible, meaning no randomness is involved in the sampling process. The MRHF also presents the advantage to perform an intrinsic localization, which considerably decreases the number of particles needed. However the curse of dimensionality still strikes when the observations are overly scattered.

We describe a series of experiments on several small case benchmarks, each of them corresponding to different levels of non linearity and non-Gaussianity, and show that, in each case, the MRHF provides an appropriate correction to the prior ensemble. In the medium term, the MRHF will be implemented to solve the strongly non-Gaussian problem of the assimilation of ocean color data in physical-biogeochemical models, in the framework of the European project SANGOMA.

References

[Anderson, 2010] J. L. Anderson. "A Non-Gaussian Ensemble Filter Update for Data Assimilation", *Mon. Wea. Rev.*, vol. 138, pp. 4186-4198, June 2010.

[Snyder et al., 2008] C. Snyder, T. Bengtsson, P. Bickel and J. Anderson. "Obstacles to High-Dimensional Particle Filtering", *Am. Met. Soc.*, vol. 136, pp. 4629-4640, May 2008.

Scalable Sequential Data Assimilation with the Parallel Data Assimilation Framework PDAF

Lars Nerger, Wolfgang Hiller, and Jens Schröter

Alfred Wegener Institute, Bremerhaven,Germany.
Email: Lars.Nerger@awi.de

Data assimilation applications with high-dimensional numerical models exhibit extreme requirements on computational resources. Good scalability of the assimilation system is necessary to make these applications feasible. Sequential data assimilation methods based on ensemble forecasts, like ensemble-based Kalman filters, provide such good scalability, because the forecast of each ensemble member can be performed independently. However, this parallelism has to be combined with the parallelization of both the numerical model and the data assimilation algorithm. In order to simplify the implementation of scalable data assimilation systems based on existing numerical models, the Parallel Data Assimilation Framework PDAF [1,2,3] has been developed. PDAF is suitable for educational use with toy models but also for high-dimensional applications and operational use [4]. PDAF is distributed as open source software.

PDAF provides a framework for implementing a data assimilation system with parallel ensemble forecasts and parallel numerical models. For maximum efficiency, a single assimilation program can be built that includes both the model and the analysis step. A well-defined interface connects PDAF to the model as well as to the observations. To compute the analysis, PDAF provides several optimized parallel filter algorithms and smoothers. Included are ensemble filters like the Local Ensemble Transform Kalman Filter (LETKF) [5] and the Error Subspace Transform Kalman Filter (ESTKF) [6].

We discuss the philosophy behind PDAF as well as features and scalability of data assimilation systems based on PDAF on the example of data assimilation with the finite element ocean model FEOM.

References

[1] http://pdaf.awi.de

[2] L. Nerger, W. Hiller. "Software for Ensemble-based Data Assimilation Systems - Implementation Strategies and Scalability", *Computers and Geosciences*, 55, 110-118, 2013.

[3] L. Nerger, W. Hiller, J. Schröter. "PDAF - The Parallel Data Assimilation Framework: Experiences with Kalman Filtering", in *Use of high performance computing in meteorology : Proceedings of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology*, Reading, UK, 25 - 29 October 2004 / Eds.: Walter Zwieflhofer; George Mozdzynski, Singapore: World Scientific, 63-83, 2005.

[4] S. N. Losa, S. Danilov, J. Schröter, L. Nerger, S. Massmann, F. Janssen. "Assimilating NOAA SST data into the BSH operational circulation model for the North and Baltic Seas: Inference about the data", *Journal of Marine Systems*, 105, 152-162, 2012.

[5] B. R. Hunt., E.J. Kostelich, and I. Szunyogh. „Efficient data assimilation for spatio-temporal chaos: A local ensemble transform Kalman filter", *Physica D,* 230, 112–126, 2007

[6] L. Nerger, T. Janjić, J. Schröter, W. Hiller. "A unification of ensemble square root Kalman filters", *Monthly Weather Review*, 140, 2335-2345, 2012.

The Hybrid Local Ensemble Transform Kalman Filter

Stephen G. Penny[a]

[a] *Department of Atmospheric and Oceanic Science, University of Marlyand, USA, Steve.Penny@noaa.gov*

Hybrid data assimilation methods combine elements of ensemble Kalman filters (EnKF) and variational methods. While most approaches have focused on augmenting an operational variational system with dynamic error covariance information from an EnKF [1][2][4][5][8], we take the opposite perspective of augmenting an operational EnKF with information from a simple 3D-Variational (3D-Var) method [7]. We wish to determine which aspects of the variational methods are necessary for successful application of a hybrid method. To this end we have developed the Hybrid Local Ensemble Transform Kalman Filter (Hybrid-LETKF), which improves analysis errors when using small ensemble sizes and low observation coverage versus either LETKF [3] or 3D-Var used alone on a Lorenz-96 model [6]. The results imply that for small ensemble sizes, allowing a solution to be found outside of the space spanned by ensemble members provides robustness in the hybrid method compared to LETKF alone. Finally, the simplicity of the Hybrid-LETKF design implies that this method can be applied operationally while requiring almost no modification to an existing operational 3D-Var system. Further tests exploring this potential using a global scale ocean model are underway.

References

[1] M. Buehner. "Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational NWP setting," *Q.J.R. Met. Soc.*, 131, 1013–1043, 2005.

[2] T.M. Hamill and C. Snyder. "A hybrid ensemble Kalman filter-3D variational analysis scheme." *Mon. Wea. Rev.*, 128, 2905-2919, 2000.

[3] B.R. Hunt, E. J. Kostelich, and I. Szunyogh. "Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter," *Physica D*, 230, 112-126, 2007.

[4] D.T. Kleist, "An Evaluation Of Hybrid Variational-Ensemble Data Assimilation For the NCEP GFS," University of Maryland College Park, (Doctoral Dissertation), 2012.

[5] A.C. Lorenc. "The Potential of the Ensemble Kalman Filter for NWP – a comparison with 4D-Var," *Q.J.R. Meteorol. Soc.*, 129, 3183-3203, 2003.

[6] E.N. Lorenz. "Predictability – A problem partly solved," *In Proceedings of the Seminar on Predictability*, volume 1. ECMWF: Reading, UK. 1–18, 1996.

[7] S.G. Penny. "The Hybrid Local Ensemble Transform Kalman Filter," *Mon. Wea. Rev.*, (Submitted), 2013.

[8] X. Wang. "Incorporating Ensemble Covariance in the Gridpoint Statistical Interpolation Variational Minimization: A Mathematical Framework," *Mon. Wea, Rev*, 138, 2990-2995, 2010.

Error Analysis and Adaptive Localization for Ensemble Methods in Data Assimilation

África Periáñez[a,b] , Hendrik Reich[a], Roland Potthast [a,b]

[a]*Research and Development section, German weather service (DWD), Germany, africa.perianez@dwd.de*
[b] *Department of Mathematics,University of Reading, UK*

Ensemble data assimilation systems are increasingly popular for a variety of applications, as they provide a flexible alternative to large-scale variational systems. A number of different versions of ensemble filters have been suggested and tested over the last two decades.

Even though Ensemble-DA has high potential, its theoretical justification is still under development, both on the application side, testing the operational setup for different data types, as well as on the mathematical analysis, looking into approximation properties with ensemble-size small compared to the number of degrees of freedom in the model.

Localization is often used to reduce the effect of spurious long-range correlations and to increase the effective size of the ensemble space. It is the goal of our work [1] to understand the basic properties of localization. We derive deterministic error estimates for the EnKF and study its dependence on localization both for the one-step analysis case and for the assimilation of two kind of observations. We also derive the formula for an optimal localization radius depending on the observation density and the observation error. The validity of the theory is demonstrated by numerical experiments for several simple models using LETKF [2].

References

[1] Periáñez Á., Reich H., Potthast R., "Error Analysis and Adaptive Localization for Ensemble Methods in Data Assimilation", *Submitted to Inverse Problems*, 2013
[2] Hunt B.R., Kalnay E., Kostelich E.J., Szunyogh I., "Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter", *Physica D*, vol. 230, pp 112-126, 2007

Evaluation of Model Errors using Data Assimilation

Chiara Piccolo[a] and Mike Cullen[a]

[a] *UK Met Office, United Kingdom, chiara.piccolo@metoffice.gov.uk*

We wish to demonstrate how data assimilation techniques can be used to quantify model errors. Model error on a case by case basis is unknowable, as the governing equations cannot be solved completely and some processes cannot be represented by equations. However, information about the statistics of model error can be inferred from the observations.

In order to achieve this, we incorporate a stochastic representation of model error which can be calibrated. This can be physically based, for instance the stochastic backscatter scheme, or be represented by a simple covariance model as used in the operational data assimilation. We then carry out a stochastic data assimilation using an ensemble DA method. The model in the ensemble is augmented by the stochastic term. The ensemble is then run on into a forecast. The predictions are verified by testing the hypothesis that the truth is a member of the population represented by the ensemble.

Results are shown using an ensemble of 4dVars and comparing the stochastic backscatter and other physically based model error representations with that given by using a simple covariance model. These results can be used to inform the design of ensembles. They also give a measure of the quality of the deterministic model because the size of the model error is estimated independently of the uncertainty in the initial data.

Inter-comparison and Coupling of Ensemble and Variational Data Assimilation Approaches for Regional-Scale Modeling

Jonathan Poterjoy[a] and Fuqing Zhang[a], and Meng Zhang[b]
[a] *Department of Meteorology, Pennsylvania State University, United States, jpoterjoy@psu.edu,*
[b]*IBM Research, Beijing 100101, China*

Ensemble, variational, and hybrid data assimilation systems are compared using the WRF model for a set of regional and storm-scale data assimilation experiments. Five different data assimilation schemes are considered, including standard three-/four-dimensional variational methods (3D/4DVar), an ensemble Kalman filter (EnKF), and coupled ensemble-variational methods (E3DVar and E4DVar). The hybrid methods require that EnKF and 3D/4DVar run independent of one another with two coupling steps; the variational component uses the ensemble forecast mean and perturbations at the beginning of the observation time window, and the minimizing solution replaces the posterior EnKF mean after assimilating the available data. The coupled approach uses the ensemble-estimated error covariance to incorporate flow-dependent information into the variational analysis, while considering a full-rank climatological covariance estimate. Among the tested data assimilation methods, E4DVar is the most flexible. This approach can assimilate high volumes of asynchronous observations during the analysis, while taking into account flow-dependent information from the ensemble perturbations and adjoint model. These benefits provide E4DVar with theoretical advantages over the other tested data assimilation methods, especially at the mesoscale.

In a month-long coarse-resolution cycling data assimilation experiment over the continental US, E4DVar significantly outperforms all other data assimilation methods in terms of 48-h root-mean square forecast error [1,2]. Most of the forecast improvements are observed in the lower-model levels for both dynamic and thermodynamic variables. E3DVar proved to be the second most effective method, followed by EnKF, 4DVar, and 3DVar over the month-long study. The EnKF, 4DVar, and E4DVar methods are also compared in a weeklong convection-permitting experiment that uses the WRF model to simulate the genesis and rapid intensification of Hurricane Karl (2010). Routinely collected observations are assimilated, as well as dropsondes from the Pre-Depression Investigation of Cloud Systems in the Tropics field campaign. The E4DVar method produces significantly more accurate 12 – 72 h forecasts in the vicinity of the developing tropical wave, when compared to the benchmark systems. Improvements in the wind and moisture fields translate into more accurate forecasts for the timing and location of genesis for this event. Furthermore, the hybrid variational scheme reduces the number of inner-loop iterations by about 30% for multi-incremental 4DVar, owing to the smooth first guess and ensemble estimation of background error covariance. This study demonstrates the effectiveness of coupled ensemble-variational data assimilation methods at the synoptic and mesoscale, while considering possible computational cost savings in the new system.

References

[1] Zhang, M., and F. Zhang, "E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model," Mon. Wea. Rev., vol. 140, pp. 841-859, 2012.

[2] Zhang, F., M. Zhang, and J. Poterjoy, "E3DVar: Coupling an ensemble Kalman filter with three-dimensional variational data assimilation in a limited-area weather prediction model and comparison to E4DVar," *Mon. Wea. Rev.*, vol. 140, pp. 900-917, 2013.

Aposteriori Error Estimates in Variational Data Assimilation

Adrian Sandu[a]

[a]*Computational Science Laboratory, Computer Science Department, Virginia Tech, Blacksburg, VA 24060.*

In this talk we focus on variational data assimilation, specifically, on both weak and strong constraint 4D-Var.

The analyses provided by data assimilation are affected by errors in the data and in the model, as well as by errors in the misspecification of various error covariances. We develop aposteriori error estimates that quantify the effect of each of these errors on the analysis. The resulting estimates are of the form

$$E[\Delta x_0^a] = < \alpha, \Delta y > \; + < \beta, \Delta M > \; + < \gamma, \Delta R > \; + < \theta, \Delta B >$$

where E is a scalar metric of accuracy for the analysis, and <a, Δb>are dot products in appropriate spaces. The rigorous procedure to find the weights of each of the errors is a new methodology that will be presented in this talk.

The application of the new error estimation methodology to relevant testcases will also be presented.

Improving Reanalyses using Physically Based Stochastic Parameterizations

Prashant D. Sardeshmukh[1,2] , Cecile Penland[2], Jian-Wen Bao[2],
Gilbert P. Compo[1,2], and Jeffrey S. Whitaker[2]

[1] CIRES, University of Colorado, Boulder Colorado
[2] NOAA/Earth System Research Laboratory, Boulder, Colorado
Prashant.D.Sardeshmukh@noaa.gov

Abstract

Reanalysis datasets are a crucial part of climate research, offering consistent sets of climate fields for use in the monitoring, simulation and prediction of climate variability. The quality of reanalysis fields depends strongly on the quality of background "first guess" fields and the proper specification of their expected errors relative to observational errors. Many modern Kalman filter based reanalysis systems use an ensemble of short-range forecasts as first guess fields and their spread to specify their expected errors. In principle, such systems have a superior flexibility to make these specifications state-dependent, i.e. "to go with the flow", in a physically consistent manner. In practice, however, much of this potential advantage is lost because the spread of the forecast ensemble is often unrealistically low, necessitating an artificial covariance inflation to produce reasonable results. To what extent such misspecifications of the error covariances compromise the quality of existing reanalysis datasets is completely unclear at present.

In this presentation, we will explore the potential impact on reanalyses of including physically based stochastic parameterizations (SPs) of atmospheric convection and other physical processes in the forecast model used for generating the first-guess fields. Proper inclusion of the SPs should substantially reduce the need for the artificial inflation of first-guess error covariances and lead to an improved relative weighting of the observations and first guesses in the reanalysis algorithm. We will discuss strategies of how the development of such SPs could be guided not only by physical considerations but also by assessing to what extent they lead to better probabilistic weather predictions and better representations of climate variability in long climate runs of the same forecast model.

Adaptive ensemble Kalman filtering of nonlinear systems

Tyrus Berry[a] and Timothy Sauer[a]

[a] *Department of Mathematical Sciences, George Mason University, USA, tsauer@gmu.edu.*

A necessary ingredient of an ensemble Kalman filter is covariance inflation [1], used to control filter divergence and compensate for model error. There is an ongoing search for inflation tunings that can be learned adaptively. Early in the development of Kalman filtering, Mehra [2] enabled adaptivity in the context of linear dynamics with white noise model errors by showing how to estimate the model error and observation covariances. We propose an adaptive scheme, based on lifting Mehra's idea to the nonlinear case, that recovers the model error and observation noise covariances in simple cases, and in more complicated cases, results in a natural additive inflation that improves state estimation. It can be incorporated into nonlinear filters such as the Extended Kalman Filter (EKF), the Ensemble Kalman filter (EnKF) and their localized versions. We test the adaptive EnKF on a 40-dimensional Lorenz96 model and show the significant improvements in state estimation that are possible. We also discuss the extent to which such an adaptive filter can compensate for model error, and demonstrate the use of localization to reduce ensemble sizes for large problems.

References

[1] Anderson, J. and S. Anderson. "A Monte-Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts." Mon. Wea. Rev., 127, 2741– 2758, 1999.

[2] Mehra, R. "On the identification of variances and adaptive Kalman filtering." IEEE Trans. Auto. Cont., 15, 175–184, 1970.

Application of a Spectral Transform on Cubed-Sphere Grids to Representation of Forecast Errors for Variational Data Assimilation

Hyo-Jong Song[a], Jihye Kwun[b], and Sang-Yoon Jun[c]

[a] *Data Assimilation Team, Korea Institute of Atmospheric Prediction Systems (KIAPS), Republic of Korea, hj.song@kiaps.org*, [b] *Data Assimilation Team, KIAPS, Republic of Korea,* [c] *Systems Development Team, KIAPS, Republic of Korea.*

The assumed lack of covariance between parameters has reduced the amount of statistical information needed to draw a background error covariance. The auto-covariance of each parameter between different positions in space needs to be handled. There are usually in excess of 10^6 components of a field and so the spatial error covariance cannot be dealt with explicitly [1]. The spatial transform is the next stage of the covariance model. Spectral transform achieves approximately the non-correlation of errors by assuming that errors between spectral modes are uncorrelated.

We commonly use spherical harmonics as a basis function for spectral transform [2]. To obtain the spectral coefficients, we conventionally take Fourier transformation and Legendre transformation which are line integrals on the sphere. In a cubed-sphere grid system based on an equi-angular coordinate, the unit vectors are not orthogonal [3]. If using that grid system, we should take a surface integral of the multiplication of spherical harmonics and a scalar function on the sphere. In this presentation, we show the detailed algorithm of the spectral transform on a cubed-sphere grid. We apply it to represent forecast errors for data assimilation. We present the resultant structure of forecast errors and the result of application to a three-dimensional variational system.

References

[1] R. N. Bannister. "A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, no. 637, pp. 1971-1996, October 2008.

[2] P. Courtier, E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, and M. Fisher. "The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation," *Quarterly Journal of the Royal Meteorological Society*, vol. 124, no. 550, pp. 1783-1807, July 1998.

[3] R. D. Nair, S. J. Thomas, and R. D. Loft. "A discontinuous Galerkin transport scheme on the cubed sphere," *Monthly Weather Review*, vol. 133, no. 4, pp/ 814-828, April 2005

A New Observation Error Probability Model for Nonlinear Variational Quality Control and Applications within the NCEP Gridpoint Statistical Interpolation.

Xiujuan Su[a], and R. James Purser[a]

[a]IMSG at NOAA/NWS/NCEP/EMC, USA, xiujuan.su@noaa.gov

Nonlinear variational quality control provides a methodology for treating observations within a variational data assimilation in a way that accounts for the fact that, in practice, their effective errors are distributed according to probability densities whose tails are significantly heavier than those of a Gaussian. In this way the rare but potentially damaging gross errors infecting typical atmospheric measurements can be recognized and down-weighted automatically within the iterative variational assimilation algorithm without recourse to a separate procedure. Unfortunately, an undesirable consequence of some of the simplest popular models of non-Gaussian error is an implied cost function characterized by multiple-minima. This means that it is easy for the assimilation to become effectively "locked" into false solutions close to the initial background state of the iterations that either accepts observations that should actually be severely down-weighted, or fails to attribute sufficient weight to valid observations at variance from a badly misleading background to permit these good data from adequately correcting the bad background. A new probability model for representing realistic measurement errors, which generalizes the "logistic" distribution, corrects the defective characteristics of traditional nonlinear quality control by ensuring that the negative-log-posterior distribution preserves the property of convexity possessed by the negative-log-prior, and is therefore free of multiple minima. We report on parallel tests of the new quality control procedure within the operational gridpoint statistical interpolation scheme at NCEP, and address the problem of estimating the parameters of the new probability model from the diagnostics of the assimilation that are available to us.

MODELLING FORECAST ERROR STATISTICS using GAUSSIAN ANAMORPHOSIS

C.E. Testut[a], G. Garric[a], C. Perruche[a], L. Parent[a], E. Greiner[b] and J.-M. Brankart[c]
[a]*Mercator-Ocean, Ramonville Saint Agne, France,* [b]*CLS, Ramonville Saint Agne, France,* [c] *LGGE/CNRS, Grenoble, France, ctestut@mercator-ocean.fr*

Mercator-Ocean has developed a hierarchy of ocean analysis and forecasting systems designed to simulate the global ocean circulation (from 1/4° to 1/12°). These systems are based on the NEMO ocean/sea ice coupled model and the multivariate data assimilation system SAM2 (Système d'Assimilation Mercator V2). An extension of the coupling to the biogeochemical component using the PISCES model is also under development. However, up to now, in the real time operational applications, only the ocean component is controlled by the assimilation using in situ Temperature and Salinity vertical profiles, satellite Sea Surface Temperature and Altimetry. In practice, this assimilation scheme is based on a reduced order Kalman filter derived from the SEEK filter. The background error of the SAM2 method is represented by an ensemble of multivariate state vectors defining a subspace of the control space. In the future, Mercator-Ocean has then planned to extend the assimilation control both to the sea ice component and to the biogeochemical component. For the ice model or the biogeochemical model, the prior ensemble defining the forecast error statistics could be built from a free simulation as it is done for the oceanic component. However, the distributions of variables like the sea ice concentration or the chlorophyll are particularly not Gaussian, breaking the assumption of the linear analysis concerning the background error. This property could be restored by application of an anamorphic transformation based on the prior ensemble. This study presents the transformation algorithm in the context of an application on the sea ice or the biogeochemical quantities. The method will be illustrated by multivariate representers showing the extrapolation from observed to unobserved non-Gaussian variables like the sea ice thickness. Results using simple analysis scheme will be also presented in order to evaluate the impact on the statistical estimation step.

Particle Filters for Numerical Weather Forecasting

Peter Jan van Leeuwen[a]

[a] *Department of Meteorology, University of Reading, United Kingdom,* p.j.vanleeuwen@reading.ac.uk

With ever increasing model resolution, complexity of physical parameterisations, and complexity of observation operators the data-assimilation problem becomes more and more nonlinear. Traditional data-assimilation methods like 4DVar and Ensemble Kalman Filters rely on Gaussian assumptions for the prior that are often not justifiable. Modern extensions to hybrid formulations do not solve this problem. Fully nonlinear data-assimilation methods are available, but typically prohibitively expensive in high-dimensional settings. Perhaps most promising are particle filters, although the so-called 'curse of dimensionality' leads to degenerate filters and has hampered their use in the geosciences [1,2].

Recently a breakthrough in efficient particle filters [3, see also e.g. 4] has let to successful applications in systems of dimensions up to 65,000 [5]. Moreover, the formulation is such that degeneracy will not occur by construction, removing one of the barriers of widespread application of particle filters in the geosciences.

In this paper I will discuss what needs to be done to make particle filters useful for the real job, numerical weather forecasting, based on experience implementations in large-scale ocean models, coastal ocean models, and climate models. For instance, the importance of allowing for errors in the model equations (they are present and substantial, after all), and how to implement them will be discussed. Also, the practical advantages of using particle filters will be addressed. Specifically the fact that the quality of the initialization does not depend on the quality of the error covariances of the model state will be highlighted. This point should not be underestimated. Tens (hundreds?) of man/woman years have gone into developing accurate and efficient coding of the background error covariance in 4DVar. Ensemble Kalman filters rely on estimates of state covariances from a relatively small ensemble, and suffer from ad hoc modifications such as localization and inflation. Particle filters do not have this problem, and activity can be redirected to where it is most badly needed, the model errors. This would have the additional advantage of making data-assimilation a much more useful tool for model improvement.

References

[1] C. Snyder et al. "Obstacles to high-dimensional particle filtering". *Mon. Wea. Rev.*, 136: 4629-4640, December 2008.

[2] P.J. Van Leewen. "Particle filtering in geophysical systems". *Mon. Wea. Rev.*, 137: 4089-4114, 2009.

[3] P.J. van Leeuwen. "Nonlinear data assimilation in geosciences: an extremely efficient particle filter". *Q. J. R. Meteorol. Soc.*, 136: 2010.

[4] Chorin AJ, Tu X. 2009. "Implicit sampling for particle filters", Proceedings of the National Academy of Sciences 106(41): 17 249– 17 254.

[5] Van Leeuwen, P.J., and M. Ades. "Efficient fully nonlinear data assimilation for geophysical fluid dynamics" Computers and Geosciences, 10.1016/j.cageo.2012.04.015, 2012.

Multi Resolution Variational Data Assimilation Schemes With Application to a Realistic Ocean Model.

Arthur Vidard, Laurent Debreu, and Emilie Neveu

INRIA / LJK, France, vidard@imag.fr.

The convergence of variational data assimilation algorithms for high dimensional non-linear systems is an important issue. The approach currently in use in most of the operational applications is known as Multi-Incremental 4D-Var or Perturbed Gauss Newton (PGN) where the inner problem of the Gauss Newton algorithm is solved with a succession of reduced resolution / simplified physics approximations of increasing complexity of the original problem. However this algorithm is known to converge toward a minimum that is different from that of the original problem. As a result of this the number of iterations has to be limited in order to avoid the risk of divergence.

By controlling the error made between successive approximations one can derive a more robust algorithm that will converge toward the original minimum provided the right conditions are fulfilled. These conditions can easily be verified and the additional cost is modest compared to PGN. Applying this algorithm to the outer iteration of the Gauss Newton algorithm instead can bring a further improvement by reducing the risk of getting stuck in a local minimum.

These algorithms are part of recent developments of the NEMOVAR system, which is a state-of-the-art variational data assimilation system dedicated to NEMO the European ocean community model. NEMOVAR is used in two major operational centres in its 3D-Fgat configuration.
The general NEMOVAR framework and the different algorithms will be presented, they will be applied to realistic ocean configurations and their relative merits will be compared. Moreover difficulties related to the specificities of the ocean will be discussed since both algorithms requires the use of operators allowing to project system states from one resolution to another, which could be delicate for ocean applications due to complex boundaries

Application of Ensemble Sensitivity to Data Assimilation

Shu-Chih Yang[1] and Eugenia Kalnay[2]

1: National Central University, Taiwan
2: University of Maryland, USA

Given an ensemble of forecasts, it is possible to determine the ensemble sensitivity (ES), i.e., a linear combination of the forecasts that, given the choice of the perturbation norm and forecast interval, will maximize the growth of the perturbations (Enomoto, 2007, Matsueda et al., 2010). Because the leading ES vector indicates the directions of the fastest growing forecast errors, we explore the potential of applying the ES in Ensemble Kalman Filter for correcting fast growing errors. We construct the ES within a quasi-geostrophic multi-level channel model and use the Local Ensemble Transform Kalman Filter (LETKF) for the data assimilation experiments. We confirm that even during the early spin-up the final ES with a 6-h forecast window is strongly related to the background errors. Thus the ES has the potential to improve data assimilation methods through the correction of fast growing errors. A positive impact is found when using the initial ES as the additive perturbations to inflate the analysis ensemble compared with optimal random perturbations.

When an Ensemble Kalman Filter (EnKF) is "cold-started" from random perturbations, it has a long spin-up period. The "running in place" (RIP) method has been introduced to accelerate the EnKF spin up by improving the accuracy of the mean state and the structure of the flow-dependent error covariance. Under the LETKF-RIP framework of the QG model, the ES is applied to rotate the ensemble perturbations to align along the fast growing errors. Results suggest that using the ES further improves the accuracy of the LETKF-RIP analysis without enlarging the ensemble size, by effectively projecting the ensemble perturbations onto the fast growing directions.

A Hybrid Variational-Ensemble Data Assimilation Method with an Implicit Optimal Hessian Preconditioning

Milija Zupanski

Cooperative Institute for Research in the Atmosphere, Colorado State University, U.S.A.,
Milija.Zupanski@colostate.edu

Hybrid variational-ensemble data assimilation has been the focus of research and applications in recent years. The advantages of hybrid methods are generally due to using an error covariance that has both full rank and flow dependence, as well as being able to account for nonlinearities of the model and observation operators. A common characteristic of hybrid methods is that nonlinearity is addressed through its variational component that typically employs an iterative minimization to numerically obtain the optimal solution. However, since variational data assimilation is searching for the solution in model or observation spaces, both of very high dimensions, the mathematically correct solution equation has to be approximated. This eventually results in numerous minimization iterations with undesirable impact on efficiency and accuracy of the analysis solution.

In this presentation we will present a new hybrid-variational data assimilation method that has a potential to overcome the above deficiencies of current hybrid methods. The reduced efficiency of iterative minimization in variational methods can be traced back to Hessian preconditioning. The approach adopted here is to develop a hybrid method that includes an optimal Hessian preconditioning based on the inverse square root of the Hessian matrix. This preconditioning method is implicit, being embedded in data assimilation algorithm.

The new hybrid method will be evaluated in the context of cloud-resolving data assimilation with the Weather Research and Forecasting (WRF) model. A special attention will be given to the interaction of hybrid error covariance and Hessian preconditioning, as well as to examining the structure of hybrid error covariance. The presentation will also address the issues relevant to future development of hybrid variational-ensemble data assimilation.

Ensembles of 4DVars and Their Use Within a Particle Filter Context

Javier Amezcua[a], Peter Jan van Leeuwen[a]

[a] *Department of Meteorology, University of Reading, United Kingdom, j.AmezcuaEspinosa@reading.ac.uk*

The standard implementation of particle filters (using sequential importance resampling) is unfeasible in large dimensional geophysical systems. The ensemble of particles is known to collapse (degeneracy phenomenon) when the number of independent observations is large [1,3]. An alternative is to use proposal densities other than the likelihood. In particular, [2] and [3] present an efficient particle filter which combines two elements (a) nudging towards future observations and (b) a step for the stabilization of the particle weights.

In this work we explore another type of proposal density in which the nudging is 'optimal'. In this case, a 4DVar is performed in each one of the particles. Starting with the linear case, we explore the statistical characteristics of these families of 4DVars and their use within the particle filter setting. In the nonlinear case, these characteristic are not straightforward; we try to shed light in these issues.

We start in a perfect model setting using ensembles of strong constraint 4DVars. We then introduce model error and study the characteristics of families of weak constraint 4DVars, but also strong constraint 4DVars where the model error is ignored in the cost function (as is often done in practice). The consequences on the weights of the particles are analyzed, and recommendations on the use of 4DVars as proposal densities will be presented.

References

[1] C. Snyder et al. "Obstacles to high-dimensional particle filtering". *Mon. Wea. Rev.*, 136: 4629-4640, December 2008.

[2] P.J. van Leeuwen. "Nonlinear data assimilation in geosciences: an extremely efficient particle filter". *Q. J. R. Meteorol. Soc.*, 136: 1991-1999, October 2010 Part B.

[3] M. Ades and P.J. Van Leeuwen. "An exploration of the equivalent weights particle filter". *Q. J. R. Meteorol. Soc.*, 2012.

Wavelet Covariance Approximations for Wildfire Data Assimilation using SFIRE

Jonathan D. Beezley[a] and Jan Mandel[b]

[a] *CERFACS, Toulouse, France, jon.beezley@gmail.com,* [b]*Department of Mathematical and Statistical Sciences, University of Colorado Denver, United States*

With wildfires burning tens of millions of acres per year and threatening the lives and property of numerous people. There is increasing interest in predicting the spread and intensity of these outbreaks with precision. New data sources from satellites and airborne cameras provide an opportunity to improve the fire forecasts with accuracy unattainable using computational modeling alone. Naturally, researchers have begun exploring the application of data assimilation to these models to improve their accuracy in real time predictions.

While robust data assimilation methods have been developed for traditional weather forecasting domains, these methods tend to fail when applied to wildfire models due to the unique nature of the error statistics. In particular, standard localized sample covariance estimates can be shown to provide particularly poor representations of the true error statistics using reasonable sized ensembles. We show that one can improve sample correlation estimates dramatically by representing the covariance matrix as diagonal in a wavelet basis similar to the method presented in [2]. This representation provides the same benefits as covariance tapering at a significantly reduced computational expense. In addition, the wavelet basis can better reproduce local correlations with fewer samples than tapering alone. It can also be demonstrated that the wavelet method is capable of resolving non-uniform error distributions and local features of the fields making it much more suitable for covariance estimation of wildfire models than other related methods based on the Fourier transform.

The wavelet-based covariance estimate can be used in the ensemble Kalman filter (EnKF) in place of the sample covariance to produce a method with many of the same benefits as other localized filtering methods[1]. However, the wavelet EnKF is much more computationally efficient and easier to implement because the covariance is represented by a simple block diagonal structure. In this form, the EnKF becomes simply a manipulation of diagonal matrices for each pair of variables in the model.

The results obtained here use a large number randomly generated fires from the non-coupled SFIRE wildfire model[3]. Each sample is generated from a randomly chosen ignition point and model parameters. We compare the covariance and cross-covariance estimates derived from various methods to a large sample estimate calculate the representation error for a variety of cases. Finally, we apply this covariance representation to the wavelet EnKF as in in a simple experimental framework showing benefits compared to standard tapered covariance estimates.

References

[1] J. D. Beezley, J. Mandel, L. Cobb, "Wavelet Ensemble Kalman Filters," ***Proceedings of IEEE IDAACS 2011*****, vol. 2, pp. 514–518, 2011.**

[2] A. Deckmyn and L. Berre, "A wavelet approach to representing background error covariances in a limited-area model," ***Monthly Weather Review*****, vol. 133, no. 5, pp. 1279–1294, 2005.**

[3] J. Mandel, J. D. Beezley, and A. Kochanski, "Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011," ***Geoscientific Model Development*****, vol. 4, pp. 591–610, 2011.**

Ensemble Kalman Filtering Without the Intrinsic Need for Inflation

Marc Bocquet[a,b]

[a] *Université Paris-Est, CEREA joint laboratory École des Ponts ParisTech and EdF R&D, France*
[b] *INRIA, Paris Rocquencourt research center, France*

Efficient implementation of the ensemble Kalman filter (EnKF) requires the use of ad hoc techniques such as inflation and localization. They are meant to compensate for the misspecification of the background errors (from a previous ensemble forecast) that often leads to underestimated analyzed errors. Aside from external sources of noise, such as model error, the main intrinsic source of error in the EnKF is sampling error. To account for those sampling errors, it is possible to derive a new prior in place of the usually assumed Gaussian prior of the traditional EnKF. This defines the so-called finite-size EnKF (EnKF-N). It has been tested on a wide range of 1D and 2D-toy models, where it does remove the need to tune and apply inflation to cope with sampling errors [1].

A dual formulation of the EnKF-N has also been derived [2]. It shows that EnKF-N is very similar to a traditional deterministic EnKF, with a multiplicative inflation specified by an algebraic equation. This shows in return that sampling errors are better dealt with by multiplicative inflation (or the EnKF-N). The dual EnKF-N can also be seen as an adaptive inflation scheme and it is compared to other adaptive inflation schemes.

In addition to the new light this approach sheds on the EnKF, it is also practically useful when developing and testing new methods on low-order data assimilation systems because it avoids the burden of tuning inflation on systematic numerical tests.

References

[1] Bocquet, M. "Ensemble Kalman filtering without the intrinsic need for inflation", *Nonlin. Processes Geophys.* 18, 735-750, 2011.

[2] Bocquet, M. and Sakov, P. "Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems", *Nonlin. Processes Geophys.*, 19, 383-399, 2012.

A Smooth Ensemble Kalman Filter for Parameter estimation

Felix Carrasco[a], Juan Ruiz[a], Axel Osses[b], Celeste Saulo[a],

[a]*Oceanic and Atmospheric Research Center (CIMA), Buenos Aires, Argentina, UMI-IFACE*
felix.carrasco@cima.fcen.uba.ar

[b]*Center for Mathematical Modelling (CMM), Santiago, Chile,*

Parameter estimation is a promising extension of data assimilation techniques that provides an optimal value for the uncertain model parameters, based on the information provided by the observations and the sensitivity of the model to the parameters. One of the main issues of this technique is that model errors can produce noisy oscillations in the value of the estimated parameters. In this work the Lorenz's 1963 model is used to study the performance of parameter estimation in the context of an imperfect model. The Ensemble Kalman Filter (EnKF) is used to estimate one of the three parameters that appears in the model equations while the other two parameters are set to an incorrect value in order to simulate the effect of model error (imperfect model). Under these conditions, the estimated parameter shows high temporal variability during the assimilation cycle.

In further experiments an Smooth Ensemble Kalman Filter (SEnKF) [1] has been applied by adding a new restriction over the time derivative of the estimated parameter which results in a reduction of its temporal variability. The analysis and short range forecast error has been quantified in order to evaluate the impact of the SEnKF upon the assimilation cycle. Results shows that SEnKF is a promising approach that can improve the short range forecast by reducing the effect of model error upon the estimated parameters.

References

[1] Craig J. Johns and Jan Mandel. "A two stage ensemble Kalman filter for smooth data," *Environmental and Ecological Statistics*, 15, 101-110, Marzo 2008. (Journal Article)

Observational Quantification of Non-Gaussian Errors Within a Humidity-Temperature 1DVAR Retrieval System over Japan.

Anton J. Kliewer[a], Steven J. Fletcher[a], Andrew S. Jones[a], and John M. Forsythe[a].

[a] *Cooperative Institute for Research in the Atmsophere, Colorado State University, USA.*
steven.fletcher@colostate.edu

The major assumption made in both variational and ensemble based data assimilations systems is that the background and observational errors are Gaussian distributed. 1DVAR humidity retrieval systems also make this assumption. One such system is the CIRA 1-Dimensional Optimal Estimator (C1DOE). This is a satellite microwave brightness temperature retrieval system that is designed to find the best Gaussian estimation for temperature and moisture. However, we know that moisture variables are not always Gaussian distributed throughout the year or across the whole domain of the Earth (e.g., humidity > 0%). However, assuming a Gaussian distribution for positive definite variables can allow unphysical negative values as a possible solution to the 1DVAR scheme. A logarithmic transform can be used to overcome this, but may not be consistent with the modal solution of the maximum likelihood estimation problem. In this work, we test three different approaches over Japan against a GPS network for validation, investigating seasonal variability. The three approaches are: 1) assumed Gaussian, 2) logarithmic transform and 3) the mixed Gaussian-lognormal distribution. Using each approach and the independent GPS validation data source, we quantify the observed non-Gaussian error behaviors.

Adaptive Localization in Ensemble Kalman Filter Methods by Controlling the Observation Space

Paul Kirchgessner[a], Lars Nerger[a], and Angelika Bunse-Gerstner [b]

[a] *Alfred Wegener Institute for Polar and Marine Research, Germany, paul.kirchgessner@awi.de,*
[b]*University of Bremen, Germany.*

In data assimilation using ensemble Kalman filter methods localization is necessary to make the method work within the constraints of geophysical systems. For the LETKF, domain localization (DL) and observation localization (OL) are typically used. Even though localization is widely used, no criteria for the choice of an optimal localization radius are known yet.

In idealized twin experiments with the Lorenz-96 model and a wind-driven shallow water model we investigate the dependence of the localization radius to the ensemble size for different localization functions. It turned out that for DL, if the ensemble size is small compared to the state dimension, there is a linear dependence of the optimal localization radius to the ensemble size. This can be related to other localization functions by considering the effective observation dimension o_{eff}. This quantity is defined as the sum of the observation weights and is taken as a measure of the degrees of freedom that are available in the ensemble for assimilation.

Based on the results, we propose an adaptive localization method. This method aims to keep o_{eff} constant and adapts the localization radius accordingly. Thus, in dense regions the localization radius can be reduced and in sparse regions increased. In order to account for regions with very sparse observations, one has to limit the maximal localization radius, so that observations beyond a certain distance are not used any more.

We present first results of this method for different observation densities and show the effect of the adaptivity.

Correlations of Control Variables for Representing Forecast Errors on Cubed-Sphere Grids

Jihye Kwun[a], Hyo-Jong Song[b], Junghan Kim[c]

[a] *Data Assimilation Team, Korea Institute of Atmospheric Prediction Systems (KIAPS), Republic of Korea, jh.kwun@kiaps.org,* [b] *Data Assimilation Team, KIAPS, Republic of Korea,* [c] *System Development Team, KIAPS, Republic of Korea.*

The background error covariance is essential in data assimilation for spreading out information spatially especially in data-sparse areas, providing statistically consistent and dynamically balanced increments at the neighbouring grid points and levels of the model [1]. The full representation of the matrix is impossible because of the huge size, so the matrix is constructed implicitly by means of a variable transformation.

Background state is derived from a short numerical forecast, and so background errors share properties with forecast errors [2]. Forecast error statistics were based on Community Atmosphere Model-Spectral Element (CAM-SE) model runs every 6 hours for 31 days. CAM-SE is built upon the cubed-sphere grid, where the grid points are located at Legendre-Gauss-Lobatto (LGL) points on each local element of 6 faces on the sphere. We used the cubed-sphere geometry based on the spectral element method which is better for parallel application to apply control variable transform.

The variable transformation from model variables to a set of control variables whose errors were assumed to be uncorrelated was developed on the cubed sphere-using Galerkin method. The motivation of the control variable transform is to capture the properties of B without the need for an explicit matrix [2]. Winds were decomposed into rotational part and divergent part by introducing stream function and velocity potential as control variables. The dynamical constraint for balance between mass and wind were made by applying linear/nonlinear balance operators. The statistical structure of forecast errors will be presented for the basic model variables (T, u, v, etc.) and control variables (stream function, velocity potential and unbalanced geopotential height, etc.).

References

[1] J. Derber and F. Bouttier. "A reformulation of the background error covariance in the ECMWF global data assimilation system," Tellus, vol. 51A, no. 2, pp. 195-221, 1999.

[2] R. N. Bannster. "A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, no. 637, pp. 1971-1996, October 2008.

Correcting for Position Errors in Variational Data Assimilation

Thomas Nehrkorn[a], Bryan Woods[a], Thomas Auligne[b] and Ross Hoffman[a]

[a] *Atmospheric and Environmental Research, USA, tnehrkor@aer.com,*
[b] *National Center for Atmospheric Research, USA*

Forecasts used in variational data assimilation schemes for the atmosphere and ocean often exhibit significant position errors. These errors are particularly difficult to correct in data assimilation because they are usually not normally distributed. In addition, in the case of atmospheric data assimilation of cloud-related quantities, complications arise from the fact that cost function gradients are zero in cases where the first guess does not produce clouds or precipitation, making it difficult to correct cases where observed and predicted areas of clouds or precipitation do not coincide. A number of approaches have been proposed to measure and correct position or phase errors, ranging from techniques originating from image processing such as "image warping" or "morphing" [1] and "optical flow" [2], object-oriented verification measures [3], to alignment approaches aimed at data assimilation improvements [4].

We present an implementation of a displacement scheme to correct phase errors based on the feature calibration and alignment procedure described in Grassotti et al. [5]. In its original formulation, a set of two-dimensional displacement vectors is applied to forecast fields to improve the alignment of features in the forecast and observations. These displacement vectors are obtained by a nonlinear minimization of a cost function that measures the misfit to observations, along with a number of additional constraints (e.g., smoothness and non-divergence of the displacement vectors) to prevent unphysical solutions. Results from this implementation will be compared with a more recent implementation within the WRF-Var algorithm, in which the nonlinear minimization is replaced by the (linear) conjugate gradient inner-loop minimization combined with outer loop nonlinear adjustments, and the ad-hoc penalty function constraints are replaced by an error-covariance representation of the displacement vectors (analogous to the regularization proposed by Nehrkorn et al. [6]). Approaches to deal with questions of model imbalance will be described for the example of the WRF model.

Additional URLs:
http://www.mmm.ucar.edu/wrf/users/workshops/WS2012/ppts/9.4.pdf
http://www.emc.ncep.noaa.gov/GEFS/prod-review/posters/Nehrkorn_phase-errors.pdf

References

[1] Beezley, J. D. and J. Mandel. "Morphing ensemble Kalman filters". *Tellus A*, 60 (1), 131–140, 2008.

[2] Marzban, C. and S. Sandgathe. "Cluster analysis for object-oriented verification of fields: A variation". *Mon. Wea. Rev.*, 136 (3), 1013–1025, 2008.

[3] Ahijevych, D., E. Gilleland, B. G. Brown, and E. E. Ebert. "Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts". *Wea. Forecasting*, 24 (6), 1485– 1497, 2009.

[4] Ravela, S., K. Emanuel, and D. McLaughlin. "Data assimilation by field alignment". *Physica D: Nonlinear Phenomena*, 230 (1-2), 127–145, 2007.

[5] Grassotti, C., H. Iskenderian, and R. N. Hoffman. "Fusion of ground-based radar and satellite-based rainfall data using feature calibration and alignment". *J. Appl. Meteor.*, 38, 665–676, 1999.

[6] Nehrkorn, T., R. N. Hoffman, C. Grassotti, and J.-F. Louis. "Feature calibration and alignment to represent model forecast errors: Empirical regularization". *Quart. J. Roy. Meteor. Soc.*, 129 (587), 195–218, 2003.

The Error-Subspace Transform Kalman Filter

Lars Nerger, Wolfgang Hiller, and Jens Schröter

Alfred Wegener Institute, Bremerhaven,Germany.
Email: Lars.Nerger@awi.de

Ensemble square-root Kalman filters are currently the computationally most efficient ensemble-based Kalman filter methods. In particular, the Ensemble Transform Kalman Filter (ETKF, [1]) is known to provide a minimum ensemble transformation in a very efficient way. A similar filter algorithm is the Singular Evolutive Interpolated Kalman (SEIK, [2]) filter. In contrast to the ETKF, the SEIK filter solves the estimation problem of the Kalman filter directly in the in the error-subspace that is represented by the ensemble. Further, the matrix square roots computed in the analysis step of the filters are distinct. Comparing both filters shows small differences between their ensemble transformations, while their state estimates are identical if the same forecast ensemble is used for the analysis. To obtain the minimum transformation also with a filter using directly the error subspace, a new projection onto the error subspace is introduced. It leads to the Error-Subspace Transform Kalman Filter (ESTKF, [3]). The ESTKF shows a slightly lower computational cost compared to the ETKF. Using numerical experiments with the parallel data assimilation system PDAF [4,5, http://pdaf.awi.de], we show that also the SEIK filter can profit from the use of a symmetric matrix square root. In addition, the effect of the new projection in the ESTKF is shown.

References

[1] C. H. Bishop, B. J. Etherton, S.J. Majumdar. „Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects“, *Mon. Wea. Rev.,* 129, 420–436, 2001

[2] D. T. Pham. „Stochastic methods for sequential data assimilation in strongly nonlinear systems“, *Mon. Wea. Rev.,* 129, 1194–1207, 2001.

[3] L. Nerger, T. Janjić, J. Schröter, W. Hiller. “A unification of ensemble square root Kalman filters”, *Monthly Weather Review*, 140, 2335-2345, 2012.

[4] L. Nerger, W. Hiller. “Software for Ensemble-based Data Assimilation Systems - Implementation Strategies and Scalability”, *Computers and Geosciences*, 55, 110-118, 2013.

[5] L. Nerger, W. Hiller, J. Schröter. “PDAF - The Parallel Data Assimilation Framework: Experiences with Kalman Filtering”, in *Use of high performance computing in meteorology : Proceedings of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology*, Reading, UK, 25 - 29 October 2004 / Eds.: Walter Zwieflhofer; George Mozdzynski, Singapore: World Scientific, 63-83, 2005.

Application of Normal Mode Strong Constraint for Initialization of WAM

Miodrag Rancic[a], Houjun Wang[b], David F. Parrish[c], Daryl T. Kleist[c], and Jun Wang[a]

[a]*EMC,IMSG,USA, miodrag.rancic@noaa.gov,* [b]*CIRES, University of Colorado,USA,* [c]*EMC,NOAA,USA.*

Whole Atmosphere Model (WAM), developed by extensions of the vertical domain and functionality of NCEP's Global Spectral Model (e.g. [1]), in preliminary free runs showed a remarkable ability for research and study of the middle and upper atmosphere. The model was combined with corresponding extensions of NCEP's data assimilation system and successfully applied for simulation of phenomena such as sudden stratospheric warming [2]. One of the problems noted in this, as well as in other similar studies (e.g. [4]), is that techniques of non-incremental filtering applied for balance of the initial field result in an excessive damping of the atmospheric tides, which represent an important feature of the upper atmosphere. At the time, this problem was addressed by application of technique of incremental analysis update [2,3,4].

In the present study an alternative approach is considered, based on modifications of the balance imposed on the analysis increment of NCEP's 3DVAR data assimilation system, through a tangent-linear normal mode strong constraint [5]. A straightforward application of this constraint in WAM, where the upper part of the domain does not have enough observations, results in an erroneous adjustment of the analysis. This situation is attributed to the large amplitudes of the vertical modes in the upper portion of the domain, which are presumable consequence of a physically unjustifiable reflection from the upper lid boundary and a very low density of the upper atmosphere. The problem is solved by damping of the vertical modes in the formulation of strong constraint and by a selective application of digital filtering at the beginning of integration. The effect of these interventions on initialization of the lower atmosphere and on preservation of tides in the upper is analyzed.

Since NCEP, as many other centers, is planning to increase height of the vertical boundary of its operational global model in the near future, a study of various approaches to initialization in situation when the upper atmospheric perturbations are present has a general significance.

References

[1] R. A. Akmaev, T. J. Fuller-Rowell, F. Wu, J. M. Forbes, X. Zhang, A. F. Anghel, M. D. Iredell, S. Moorthi, and H.-M. Juang. "Tidal variability in the lower thermosphere: Comparison of Whole Atmosphere Model (WAM) simulations with observations from TIMED". *Geophysical Research Letters,* vol. 35, L03810, doi: 10.1029/2007GL032584, February 2008.

[2] H. Wang, T. J. Fuller-Rowell, R. A. Akmaev, M. Hu, D. T. Kleist, and M. D. Iredell. "First simulations with a whole atmosphere data assimilation and forecast system: The January 2009 major sudden stratospheric warming". *Journal of Geophysical Research,* vol. 116, A12321, doi: 10.1029/2011JA017081, December 2011.

[3] Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina. "Data assimilation usingincremental analysis updates". *Monthly Weather Review,* vol. 124, no. 6, pp 1256-1271, June 1996.

[4] D. Sankey, S. Ren, S. Polavarapu, Y. J. Rochoh, Y. Nezlin, and S. Beagley. "Impact of data assimilation filtering methods on the mesosphere". *Journal of Geophysical Research,* vol. 112, D24104, doi: 10.1029/2007JD008885, December 2007.

[5] D. T. Kleist, D. F. Parish, J. D. Derber, R. Treadon, R. M. Errico, and R. Yang. "Improving Incremental Balance in the GSI 3DVAR Analysis System". *Monthly Weather Review,* vol. 137, no. 3, pp. 1046-1060, March 2009.

Performance Bounds for Particle Filters Using the Optimal Proposal

Chris Snyder[*] and Thomas Bengtsson[#]

[*]*NCAR, Boulder, Colorado, USA,* [#]*Genentech, San Francisco, California, USA*

Particle filters may be cast in the form of sequential importance sampling and thus allow the choice of a proposal density. This choice is known to be crucial to the performance of the algorithm. We show that the asymptotic arguments of Bengtsson et al. (2008) and Snyder et al. (2008) apply directly to the "optimal" proposal density, in which new particles are drawn at each observation time from a distribution that is conditioned on the new observations and which minimizes the variance of the particle weights over different realizations of particles drawn from the proposal density. The asymptotic results demonstrate that the ensemble size required for the optimal proposal to avoid degeneracy of the particle-filter update (i.e., the situation in which a single particle receives a weight very close to 1) will grow exponentially with an appropriately defined measure of the problem size. At the same time, the asymptotic results indicate that the ensemble size needed may be dramatically smaller for the optimal proposal than for the standard proposal, in which new particles are drawn by evolving particles from the previous time under the system dynamics. For linear, Gaussian systems, an explicit expression for the exponent can be derived and the exponent for the standard proposal is always larger than that for the optimal proposal. The degree to which a smaller ensemble is feasible with the optimal proposal depends crucially on the magnitude of the system noise.

References:

Bengtsson, T., P. Bickel, and B. Li, 2008: Curse of dimensionality revisited: Collapse of the particle filter in very large scale systems. ***Probability and Statistics: Essays in Honor of David A. Freedman,*** **D. Nolan and T. Speed, Eds., Vol. 2, Institute of Mathematical Statistics, 316–334.**

Snyder, C., T. Bengtsson, P. Bickel and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. ***Mon. Wea. Rev.,*** **136, 4296-4640.**

An Adjoint-Based Adaptive Ensemble Kalman Filter

Hajoon Song[a], Ibrahim Hoteit[b], Bruce D. Cornuelle[c] , Xiaodong Luo[b,d] and Aneesh C. Subramanian[c]

[a] *Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A., hajsong@mit.edu,* [b] *King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia,* [c] *Scripps Institution of Oceanography, University of California, San Diego, U.S.A.,* [d] *International Research Institute of Stavanger, 5008 Bergen, Norway.*

We introduce a new hybrid EnKF (ensemble Kalman filter)/4D-VAR (four-dimensional variational) approach to mitigate background covariance limitations in the EnKF. The work is based on the adaptive EnKF (AEnKF) method which bears a strong resemblance to the hybrid EnKF/3D-VAR (three-dimensional variational) method. In the AEnKF, the representativeness of the EnKF ensemble is regularly enhanced with new members generated after back projection of the EnKF analysis residuals to state space using a 3D-VAR (or OI - optimal interpolation) scheme with a preselected background covariance matrix. The idea here is to reformulate the transformation of the residuals as a 4D-VAR problem, constraining the new member with model dynamics and the previous observations. This should provide more information for the estimation of the new member and reduce dependence of the AEnKF on the assumed stationary background covariance matrix. This is done by integrating the analysis residuals backward in time with the adjoint model.

Numerical experiments are performed with the Lorenz-96 model under different scenarios to test the new approach and to evaluate its performance with respect to the EnKF and the hybrid EnKF/3D-VAR. The new method leads to the least root-mean- squared estimation errors as long as the linear assumption guaranteeing the stability of the adjoint model holds. It is also found to be less sensitive to choices of the assimilation system inputs and parameters.

Use of Green's function in spatial covariance analysis

Xudong Sun

Centre for Australian Weather and Climate Research Bureau of Meteorology, 700 Collins St, Melbourne Vic 3000. Australia

For numerical weather prediction, covariance plays a very important role in both ensemble and variational data assimilation. This research discusses possible applications of Green's function to covariance analysis. Since Green's function is an integral kernel that it can be used to solve partial differential equations for analysing the spatial statistical behaviour of forecasting error. By analysing error spectral statistics of Green's function, the linkage between the Green's function and covariance can be established. For example, from the Helmholtz differential equation for forecast error statistics, the covariance is the square of the Green's function after it performs the spectral transform. The possible use of Green's function for variational and ensemble data assimilation will also be discussed.

The Object-Oriented Prediction System — a Flexible Framework for Data Assimilation

Yannick Trémolet[a] and Mike Fisher

[a] *European Centre for Medium Range Weather Forecasts, UK, yannick.tremolet@ecmwf.int*

Data assimilation algorithms can typically be expressed mathematically by a few generic equations involving a small set of abstract vectors and operators. These equations do not depend on the details of a specific numerical forecast model, and can be applied to many different systems. By contrast, the implementations of most data assimilation algorithms are highly model specific, so that it is impossible to take an algorithm implemented for one model and apply it to another.

The ECMWF Integrated Forecasting System (IFS) is a world-leading data assimilation system for NWP. It has developed organically, in collaboration with Météo-France, over more than 20 years. During this time, the data assimilation algorithm has become so intertwined with the numerical model, and with the particular representation of the state and observations, that it would be impossible to separate the algorithm from the model. Moreover, the code makes many implicit assumptions, for example about the order in which certain calculations are performed. This makes it extremely difficult to implement new algorithms that may, for example, require computations to be performed in a different order. The sheer complexity of the code has become a barrier to scientific development and collaboration.

To address these issues, ECMWF is developing the Object-Oriented Prediction System (OOPS). The aim is to use modern computer techniques and languages (specifically, object-oriented programming and C++) to provide a clean, and above all flexible framework for data assimilation that separates the algorithm from its specific implementation in any one model. We believe that such a framework will improve productivity, encourage scientific collaboration, and provide a valuable tool that will allow the rapid transfer of research developments into practical applications.

OOPS currently incorporates a selection of data assimilation algorithms, including strong- and weak-constraint 4D-Var, 4D-Ens-Var and 3D-Var, together with two different simple models: the Lorenz 1996 model and a quasi-geostrophic channel model. OOPS has already proved itself to be a useful tool with which to investigate novel data assimilation algorithms and minimization techniques. It has also clearly demonstrated that algorithms developed using one model can be readily transferred to another model without any recoding. Work to incorporate ECMWF's IFS model into the OOPS framework is well underway.

We will present an overview of the OOPS system. We will demonstrate how, through careful design, data assimilation algorithms can be separated from the details of any particular model. We will also show how OOPS allows models implemented in Fortran to be interfaced in a clean, machine independent way, to an object-oriented data-assimilation layer written in a different computer language.

On Observability of Atmospheric Model Tendency Errors

Michael Tsyrulnikov[a] and Vadim Gorin [b]

[a] *Meteorological Data Assimilation Department, HydroMetCenter of Russia, Russia, tsyrulnikov@mecom.ru,* [b] *Department of Mathematics, Massachusetts Institute of Technology, USA.*

Objective knowledge on probability distributions of atmospheric model tendency errors is needed both in ensemble prediction and ensemble data assimilation. In order to identify and estimate a model for model errors, a sample of real model errors (the training sample) is needed. The question whether such a sample can be obtained (at observations locations) by comparing model tendencies with observed tendencies is addressed in this study.

In predictability experiments with simulated model errors and the limited-area COSMO model, reproducibility of model errors from finite-time model-minus-observed tendencies is studied. It is found that in 1-h to 6-h tendencies, model errors appear to be too heavily contaminated by noises due to, first, initial errors and, second, trajectory drift as a result of model errors themselves. The resulting reproducibility error is far above the acceptable level. The conclusion is drawn that accuracy and coverage of current routine observations are far from being sufficient to reliably estimate model errors.

Ensemble Kalman filtering with regularization through a graphical model

Genta Ueno[a]

[a] *The Institute of Statistical Mathematics, Institution, Japan, gen@ism.ac.jp*

It is known that the ensemble Kalman filter (EnKF) can estimate state variables properly using a relatively small ensemble size. Using an ensemble of much smaller size, however, the EnKF requires regularization of an ensemble covariance matrix that reduces spurious correlations between variables defined on distant grid points. One of the methods is localization of the ensemble covariance matrix using a correlation function that has a compact support.

The present study proposes an alternative method for regularizing the ensemble covariance matrix. The method assumes conditional independence between variables which is realized by imposing zero elements in the inverse matrix of the original covariance matrix. The EnKF is reformulated on the basis of the inverse covariance matrix. Since a graphical model is used in modeling the conditional independence, I call the present filtering method a graph-based EnKF (GEnKF). A numerical experiment using the Lorenz 96 model demonstrates that the GEnKF surely reduces spurious correlations and provides reasonable state estimates.

OpenDA Data Assimilation Toolbox

Martin Verlaan[a,b], Julius Sumihar[b], and Nils van Velzen,[b,c]

[a] *Department , Delft University of Technology, Netherlands, M.Verlaan@tudelft.nl,* [b] HOS, *Deltares, Netherlands,* [c] Vortech, *Netherlands.*

OpenDA is a data-assimilation toolbox that aims to be useful for a wide range of applications. It has been applied to a range of applications ranging from calibration of river flow models, to Kalman filtering for spectral wave-models. The object-oriented approach makes the data-assimilation algorithms fully independent of the model, which allows for a very easy exchange of research developments from one application to another.

Several methods have been developed for coupling models to OpenDA. The easiest method for coupling uses the input files and output files of the model. A small number of additional subroutines to read and write the model specific formats are all that is needed. Parallel computing of ensemble members can be configured without additional programming effort. Asynchronous assimilation reduces the overhead of the restarts needed at analysis times in this approach.
On the other end of the scale, an in-memory coupling with use of MPI for parallel computing makes better use of a larger number of nodes. Parallel computing of ensemble members as well as parallel domains can be employed simultaneously if desired.

Several applications will highlight the range of applications, algorithms and coupling methods available.

The latest release, version 2.1 is freely available on our website http://www.openda.org. We encourage a free exchange of ideas and code in this open-source project.

References

[1] F. Zijl, M. Verlaan, H. Gerritsen. "Improved water level forecasting for the Northwest European Shelf and North Sea through direct modelling of tide, surge and non-linear interaction" submitted to Ocean Dynamics.
[2] M. Verlaan "Wave data-assimilation for SWAN using OpenDA" MyOcean science days 2012, 19-21 November, Geesthacht Germany.
[3] A.W. Heemink, R.G. Hanea, J.Sumihar, M. Roest, N. Velzen, M. Verlaan, "Data Assimilation Algorithms for Numerical Models", in Advanced Computational Methods in Science and Engineering, Springer, Series: Lecture notes in computational science and engineering, p.107-142
[4] N. Van Velzen and M. Verlaan, COSTA a problem solving environment for data assimilation applied for hydrodynamical modelling, Meteorologische Zeitschrift, Vol. 16, No. 6, 777-793 (December 2007), DOI: 10.1127/0941-2948/2007/0241

B-Preconditioned Minimization Algorithms for Variational Data Assimilation with the Primal and Dual Formulations

Selime Gürol[a, b], Anthony T. Weaver[c], Serge Gratton[a, d], Andrew M. Moore[e], Andrea Piacentini[c] and Hernan G. Arango[f]

[a] *CERFACS, France,* [b] *ECMWF, UK,* [c] *CERFACS / SUC URA 1875, France, weaver@cerfacs.fr,*
[d] *ENSEEIHT, France,* [e] *Department of Ocean Sciences, University of California at Santa Cruz, USA,*
[f] *Institute of Marine and Coastal Sciences, The Rutgers University of New Jersey, USA*

Variational data assimilation (Var) problems in meteorology and oceanography require the solution of a regularized nonlinear least-squares problem. Practical solution algorithms are based on the incremental approach which involves the iterative solution of a sequence of linear least-squares (quadratic minimization) sub-problems. Each sub-problem can be solved using a primal approach where the minimization is performed in a space spanned by vectors of the size of the model control vector, or a dual approach where the minimization is performed in a space spanned by vectors of the size of the observation vector. The dual formulation can be advantageous for two reasons. First, the dimension of the minimization problem with the dual formulation does not increase when additional control variables, such as those accounting for model error in a weak-constraint formulation, are considered. Second, whenever the dimension of observation space is significantly smaller than that of the model control space, the dual formulation can reduce both memory usage and computational cost.

This presentation describes recent work in developing the dual approach for two operational ocean Var systems. One is a global ocean 3D-Var system for the NEMO model, while the other is an eddy-resolving regional 4D-Var system for the ROMS model. NEMO employs the Restricted **B**-preconditioned Conjugate Gradient (RBCG) method ([2]), while ROMS employs the Restricted **B**-preconditioned Lanczos (RBLanczos) method ([4]). RBCG and RBLanczos, and the corresponding **B**-preconditioned Conjugate Gradient (BCG) and Lanczos (BLanczos) algorithms used in the primal approach, generate mathematically equivalent iterates and require the same number of applications of the standard operators $\mathbf{H}$, $\mathbf{H}^T$, $\mathbf{B}$ and $\mathbf{R}^{-1}$ needed in Var. Furthermore, all these algorithms can be implemented without the need for a square-root factorization of **B**, which is convenient with general **B** formulations such as those used with ensemble and hybrid ensemble Var. Numerical results comparing the dual and primal algorithms in NEMO and ROMS will be presented. Extensions of this work to account for multiple outer-loop iterations and second-level preconditioners based on Limited-Memory Preconditioners such as Quasi-Newton and Ritz ([1], [3]) will also be described.

References

[1] Gratton, S., Gürol, S, and P. L. Toint, "Preconditioning and globalizing conjugate gradients in dual space for quadratically penalized nonlinear least squares problems", *Comput. Optim. Appl.*, vol. 54, pp. 1-25, 2013.

[2] Gratton, S. and J. Tshimanga, "An observation-space formulation of variational assimilation using a Restricted Preconditioned Conjugate-Gradient algorithm", *Q. J. R. Meteorol. Soc.*, vol. 135, pp. 1573-1585.

[3] Gürol, S. "Solving regularized non-linear least squares problems in dual space with application to variational data assimilation", PhD thesis, Institute National Polytechnique de Toulouse, June 2013.

[4] Gürol, S., Weaver, A. T., Moore, A. M., Piacentini, A., and H. G. Arango. "B-preconditioned minimization algorithms for variational data assimilation with the dual formulation", *Q. J. R. Meteorol. Soc.*, in press, 2013.

Effects of Assimilation Window Length on Diurnal Features in a Mars Atmosphere Reanalysis

Yongjing Zhao[a], Eugenia Kalnay[a], Steven J. Greybush[a], Ross N. Hoffman[b], and R. John Wilson[c]

[a] *Department of Atmospheric and Oceanic Science, The University of Maryland, College Park, MD, USA, yzhao@atmos.umd.edu,* [b] *Atmospheric and Environmental Research, Lexington, MA,USA,* [c] *GFDL, Princeton, NJ, USA.*

Effective simulation of diurnal features is an important aspect of many geophysical data assimilation systems. For the Martian atmosphere, thermal tides are particularly prominent and contribute much to the martian atmospheric circulation and dust transport. To study the Mars diurnal features (or thermal tides), data assimilation based on the GFDL Mars Global Climate Model (MGCM) with the 4D-Local Ensemble Transform Kalman Filter (4D-LETKF) is used to perform a reanalysis of satellite observed temperature retrievals. Since the traditional 6-hr assimilation cycle induces spurious resonance in the Kelvin waves represented in both surface pressure and mid-level temperature, different assimilation window lengths are introduced in 4D-LETKF. In order to compare the performances of different assimilation window lengths, 6-hr and 12-hr forecasts based on the hour 00 and 12 reanalysis are evaluated and compared. The shorter windows show improved forecast root mean square difference with respect to observations, and not only removes the spurious resonance but also improves the analysis of the 50-pa temperature, wave number 2, and westward propagating semi-diurnal wave .

Use of Radiances in the CNMCA (Italian National Meteorological Center) Operational Ensemble Data Assimilation System

Lucio Torrisi[a] and Francesca Marcucci[b]

[a] *CNMCA-Italian National Meteorological Center, Italy, torrisi@meteoam.it,* [b] *CNMCA-Italian National Meteorological Center, Italy*

Localization is essential in an ensemble-based Kalman filter data assimilation scheme, due to the practical necessity of using a limited number of ensemble members.

The vertical localization of satellite radiances is less straightforward to implement than for other observation types in a local algorithm such as LETKF [1] because these data are an integrated measure sampling different layers of the atmosphere.

The use of radiances from AMSU-A and MHS sensors has been investigated in the CNMCA operational data assimilation system based on the LETKF algorithm [2,3]. The assimilation of these satellite radiances has been tested applying different localization strategies. Other aspects of data assimilation (e.g. quality control, radiance bias correction, etc.) has been handled in order to obtain positive impact in the operational NWP system.

References

[1] Hunt, B. R., E. Kostelich, I. Szunyogh. " Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter," *Physica D*, 230, 112-126, 2007.

[2] Bonavita M, Torrisi L, Marcucci F. "The ensemble Kalman filter in an operational regional NWP system: Preliminary results with real observations", *Q. J. R. Meteorol. Soc.*, 134, 1733-1744, 2008.

[3] Bonavita M, Torrisi L, Marcucci F. "Ensemble data assimilation with the CNMCA regional forecasting system", *Q. J. R. Meteorol. Soc.*, 136, 132-145, 2010

Lagrangian Data Assimilation: A Hybrid Particle-Ensemble Kalman Filter and Applications to Geophysical Fluid Flows

Laura Slivinski[a], Bjorn Sandstede[a], Elaine Spiller[b] and Amit Apte[c]

[a] *Division of Applied Mathematics, Brown University, USA, laura_slivinski@brown.edu,* [b] *Department of Mathematics, Statistics and Computer Science, Marquette University, USA,* [c] *TIFR Centre for Applicable Mathematics, Bangalore, India.*

Lagrangian data assimilation involves using observations of the positions of passive drifters in a flow in order to obtain a probability distribution on the underlying Eulerian flow field; for example, positions of rafts released into the ocean can be assimilated in order to estimate the currents in the ocean. This framework has several complications that differ from some traditional data assimilation frameworks. First, as in many geophysical applications, the dimension of the flow variable is very high (10^6). However, unlike many atmospheric applications, the dimension of the observations (e.g. the drifter variable) is relatively low dimensional; if we have, say, 10 drifters, then the dimension of the drifter variable would be 20. The second main complication is the highly nonlinear, chaotic nature of the Lagrangian trajectories of the drifters. Even if the flow itself evolves linearly, the evolution of the drifters will be highly nonlinear.

These complications lead to disadvantages with traditional sequential data assimilation methods. The first traditional method is the ensemble Kalman filter (EnKF). This method works well for high-dimensional problems when localization is included, and when the true posterior distribution is close to Gaussian. However, its disadvantage becomes clear when dealing with strongly non-Gaussian distributions (which arise from the nonlinear evolution of the drifters): the posterior distribution of the EnKF will be closer to a Gaussian than the true (Bayes) posterior. The second traditional method considered here is the particle filter. This method approximates the true Bayes posterior when enough particles are used; however, the necessary number of particles has been shown to increase exponentially in a variable related to the system dimension [1]. Therefore, the particle filter is intractable when applied to a problem with a high-dimensional state vector.

To avoid these disadvantages, we propose a hybrid particle-ensemble Kalman filter (based on the hybrid grid-particle filter of Salman [2]). Generally, the state vector will include both the flow variable and the drifter variable, and the distributions will be approximated by a weighted sum of particles. The distribution on the drifter variables will be updated according to the particle filter analysis step, but the distribution on the flow variables will be updated according to the ensemble Kalman filter. This method is more computationally intensive than a basic EnKF, but will approximate the highly non-Gaussian distributions on the drifters better than an EnKF. We present results of this hybrid method applied to the shallow water equations, a common model used in testing Lagrangian data assimilation schemes.

References

[1] C. Snyder et al. "Obstacles to high-dimensional particle filtering," *Monthly Weather Review*, vol. 136, pp. 4629-4640, December 2008. (Journal Article)

[2] H. Salman. "A hybrid grid/particle filter for Lagrangian data assimilation. I: Formulating the passive scalar approximation," *Q. J. R. Meteorol. Soc.*, vol. 134, pp. 1539-1550, 2008. (Journal Article)

Revisiting hybrid 3D-Var error covariance for IAU-based assimilation systems and exploiting extensions to weak-constraint 4D-EnVar

Ricardo Todling[a], Amal El Akkraoui[ab], Wei Gu[ab] , and Jing Guo[ab]

[a] *Global Modeling and Assimilation Office, NASA, USA, ricardo.todling@nasa.gov,* [b] *Science System and Applications, Inc., USA.*

As a first step toward weak-constraint 4D-Var/4D-EnsVar, the Global Modeling and Assimilation Office (GMAO) is revisiting its 3D-Var formulation using the Incremental Analysis Update (IAU) procedure of Bloom et al. (1996). Typical applications of IAU to 3D-Var are inconsistent in that the variational procedure assumes the model to be unbiased and seeks for a correction to the initial condition, while IAU leaves the initial condition untouched and requires a constant forcing (tendency) term to be applied to the model during a 6-hour filtering period as if the model were biased within this period. The benefits from the filtering properties of IAU are undeniable, but the structures of the tendency forcing terms provided to the model by the 3D-Var solutions appear suspiciously large and are not quite consistent with typical model dynamical tendencies. The spatial structures of the 3D-Var solutions are implied by the intrinsic structures of its underlying background error covariance matrix. With IAU, it seems more natural to base the error covariance structures of 3D-Var on model tendency differences (e.g., Trémolet 2007) instead of forecast differences as currently done.

This work will show results from replacing the current error covariance formulation of the GMAO hybrid 3D-Var with the tendency-based estimates derived in a companion paper presented at this Symposium (see El Akkraoui & Todling). Results reformulating both the climatological (static) and ensemble estimates of the hybrid Grid-point Statistical Interpolation analysis system will be presented, as well as a preliminary evaluation of extensions to 4D-EnsVar.

References

[1] Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. Mon. Wea. Rev., 124, 1256-1271.

[2] Trémolet, Y., 2007: Model error estimation in 4D-Var, Q. J. R. Meteorol. Soc., 133, 1267-1280.