Banff International Research Station
for Mathematical Innovation and Discovery

Mathematical Advancement in Geophysical Data Assimilation

February 3-8, 2008

http://www.birs.ca/birspages.php?task=displayevent&event_id=08w5096

ABSTRACTS
(Alphabetic order by surname of first author)

Marc Bocquet (Université Paris-Est, France)
Non-Gaussian Data Assimilation: Application to Inverse Modelling of Atmospheric Tracers

On the practical side, the goal of this talk is to demonstrate that data assimilation techniques can be used to identify the source of an accidental release of pollutant into the atmosphere, and forecast (possibly in real-time) the subsequent dispersion plume. On the methodological side, the aim is to show that there are circumstances when the use of non-Gaussian techniques in data assimilation is profitable.

A first method is based on the principle of maximum entropy on the mean and will be briefly reviewed. A second approach, which has not been applied in this field yet, is based on an exact Bayesian approach, through a maximum a posteriori estimator. The methods share common grounds, and both perform equally well in practice. When specific prior hypotheses on the sources are taken into account such as positivity, or boundedness, both methods lead to purposefully devised cost-functions, thanks to non-linear convex analysis. These cost-functions are not necessarily quadratic because the underlying assumptions are not Gaussian. As a consequence, several mathematical tools developed in data assimilation on the basis of quadratic cost functions in order to establish a posteriori analysis, need to be extended to this non-Gaussian framework. Concomitantly, the second-order sensitivity analysis needs to be adapted, as well as the computations of the averaging kernels of the source and the errors obtained in the reconstruction. All of these developments are applied to a real case of tracer dispersion: the European Tracer Experiment [ETEX]. Examples will also be given on the Chernobyl accident. Comparisons are made between a least squares cost function (similar to 4DVar) approach and a cost function which is not based on Gaussian hypotheses. Besides, the information content of the observations which is used in the reconstruction is computed and studied on the application case.

Alternatively these methods can be interpreted as weakly constrained 4DVar-like approaches with a non-Gaussian formalism for model error. Here, model error would be the pollutant source field, which is very uncertain but still the main forcing of the plume dynamics. Possible generalizations of these methods to non-linear physics will be sketched.

Mark Buehner (Environment Canada)
Towards an Improved Use of Flow-Dependent Background Error Covariances
in a Variational Data Assimilation System

Since early 2005, Environment Canada has been operationally running both a global four-dimensional variational analysis system (4D-Var) and an ensemble Kalman filter (EnKF).

Both systems employ the same atmospheric model, though with different spatial resolution. The co-existence of these two operational systems at a numerical weather prediction (NWP) centre provides a unique opportunity. In this talk approaches for incorporating background-error covariances estimated from the EnKF ensembles in the variational assimilation system are discussed. The techniques of spatial and spectral localization are briefly described and demonstrated with a simple one-dimensional problem. Then, the impact of localizing ensemble-based covariances in the variational system are shown. The practical challenge of simultaneously applying localisation in both spectral and spatial domains within a realistic NWP data assimilation system is also presented.

Gérald Desroziers (Météo-France)
Use of Observations in Data Assimilation Schemes
Modern operational data assimilation schemes rely on the use of a large range of types of observations. This is made possible via the implementation of observation operators that allow to go from model space to observation space and that especially permit the direct assimilation of satellite radiances. Because most data assimilation schemes rely on the theory of estimation, they is also a need to diagnose and specify observation error covariances. Some observations, such as satellite data, are known to be biased. In that case, a particular procedure such as variational bias correction has to be implemented. Interestingly, ensembles of perturbed assimilations are classically based on a perturbation of observations. Such ensembles allow to document background error covariances that are a key ingredient in a data assimilation scheme. On another hand, the variances given by such ensembles have most often to be inflated. Such an inflation can be tuned via the comparison of ensemble variances to the variances deduced from statistics of the innovation vector. Other aspects will be discussed such as the different ways to measure the impact of observations on analyses and subsequent forecasts.

Mike Dowd (Dalhousie University, Canada)
Sequential Monte Carlo Approaches for Parameter and State Estimation
Statistical methodologies for estimating state and parameters for time dependent stochastic dynamic systems are now well established. For nonlinear dynamics and non-Gaussian observations these data assimilation approaches rely on Monte Carlo solutions implemented via particle or ensemble methods. In this talk, I will overview new developments in such state estimation methods which rely on resampling/bootstrap and MCMC (or combinations thereof). The associated parameter estimation problem for nonlinear stochastic DE based numerical models is also considered using likelihood and state augmentation approaches.Ê Simple toy models of ocean biogeochemistry using non-Gaussian ecological measurements from a coastal ocean observatory will be used for illustration throughout. Challenges for adaptation to data assimilation in large dimension PDE based systems are discussed.

Martin Ehrendorfer (University of Reading, UK)
Ensemble-Based Data Assimilation
Ensemble-based data assimilation methods related to the fundamental theory of Kalman filtering have been explored in a variety of mostly non-operational data assimilation contexts over the past decade with increasing intensity. While promising properties have been reported, a number of issues that arise in the development and application of ensemble-based data assimilation techniques, such as in the basic form of the ensemble Kalman filter (EnKF), still deserve particular attention.
The necessity of employing an ensemble of small size represents a fundamental issue which in turn leads to several related points that must be carefully considered. Attempts to reduce effectively the sampling error due to small ensembles and at the same time maintaining an

ensemble spread that realistically describes error structures has led to the development of variants of the basic form of the EnKF. In this presentation, several of the above-mentioned issues will be discussed and illustrated together with a brief review of the methodology that has been developed by varying the basic form of the EnKF.

Pierre Gauthier (Université du Québec à Montréal, Canada)
Mathematical Problems Associated with
Atmospheric Data Assimilation and Weather Prediction

The duality principle presented by Courtier (1997) states that the 3D-Var and the Physical-Space Analysis System (PSAS) (or dual form) are just different algorithms to solve the same problem. It was also shown that the two algorithms with their own specific preconditioning can be expected to converge at a similar rate as their Hessian matrices have the same condition number. A 4D-Var algorithm being merely an extension of 3D-Var, the so-called 4D-PSAS should also be equivalent. In this presentation, results will be presented to establish the equivalence of the two approaches. To tackle the nonlinear problem, an incremental formulation is introduced and the equivalence will be established to stress the role of the outer loop process. Results from El Akkraoui et al. (2008) indicate some convergence problems that were encountered with the dual form. Moreover, when cycling an assimilation system, the Hessian from the previous minimization can be used to precondition the next assimilation. This is beneficial but cannot be done as easily in a PSAS system. An approach is proposed that takes advantage of the equivalence between the eigenvectors of the dual and primal problems. The interest for the dual form is motivated by the fact that it permits to compute the observation impact through adjoint sensitivities with respect to observations (Langland, 2004). An intercomparison experiment is being carried out at the moment to assess the robustness of the method when used at different NWP centres. Finally, the weak-constraint 4D-Var can also be formulated in either its primal or dual form, the latter having the advantage of using a control variable with a considerably lower dimension. This avenue is now being examined in the context of a simplified barotropic model.

Kayo Ide (University of California, Los Angeles, USA)
Guillaume Vernieres (University of North Carolina at Chapel Hill, USA)
Christopher K.R.T. Jones (UNC-CH, USA and U. Warwick, UK)
Lagrangian Data Assimilation:
Methods, Observing System Design, and Application to Gulf of Mexico

The Lagrangian data assimilation (LaDA) is a method for the direct assimilation of Lagrangian observations. Lagrangian instruments in the oceans, such as drifters and floats, are often designed to sample ocean properties while remaining on a two-dimensional surface in the three-dimensional ocean except when descending to or ascending from the desired depth. By augmenting the model state vector with the coordinates of the instruments, the LaDA can assimilate the Lagrangian positions without the need for any commonly used approximations to transform Lagrangian observations into Eulerian (i.e., velocity) observations.

In the first part of this talk, we describe the LaDA method and the observing system design for optimal deployment of Lagrangian instruments. Using the judicious design of the deployment strategy, the LaDA is strikingly efficient in tracking the local coherent structures, such as ocean eddies, as well as estimating the large-scale ocean circulation.

In the second part of this talk, we demonstrate effectiveness of LaDA in a realistic setting for ocean-eddy tracking in Gulf of Mexico. We evaluate three types of observations for ocean eddy tracking: the measurement of velocities at fixed station, the horizontal position of surface drifters, and the three dimensional position of isopycnal floats. By considering the

"volume of influence", we examine how and to what extent the LaDA propagates the information vertically to estimate the three-dimensional ocean structure. We show that as little as one judiciously placed drifter or isopycnal float is needed to recover an eddy being shed by the loop current.

Christopher K.R.T. Jones (UNC-CH, USA and U. Warwick, UK)
Bayesian Approach to Lagrangian Data Assimilation
Lagrangian data arise from instruments that are carried by the flow in a fluid field. Assimilation of such data into ocean models presents a challenge due to the potential complexity of Lagrangian trajectories in relatively simple flow fields. We adopt a Bayesian perspective on this problem and thereby take account of the fully nonlinear features of the underlying model. In the perfect model scenario, the posterior distribution for the initial state of the system contains all the information that can be extracted from a given realization of observations and the model dynamics. We work in the smoothing context in which the posterior on the initial conditions is determined by future observations. This posterior distribution gives the optimal ensemble to be used in data assimilation. The issue is then sampling this distribution. We develop, implement, and test sampling methods, based on Markov-chain Monte Carlo (MCMC), which are particularly well-suited to the low-dimensional, but highly nonlinear, nature of Lagrangian data. We compare these methods to the well-established ensemble Kalman Filter (EnKF) approach. It is seen that the MCMC based methods correctly sample the desired posterior distribution whereas the EnKF may fail due to infrequent observations or nonlinear structures in the underlying flow. This is joint work with Amit Apte (TIFR-Bnagalore) and Andrew Stuart (Warwick).

Eric Kostelich (Arizona State University, USA)
Recent Results of the Local Ensemble Transform Kalman Filter (LETKF)
I will outline the latest results of the Maryland/ASU implementation of the Local Ensemble Transform Kalman Filter (LETKF) to the Global Forecast System of the National Centers for Environmental Prediction (NCEP). Measures of forecast accuracy and comparison with with operational NCEP analyses will be described. The computational efficiency of our implementation of the LETKF also will be assessed, including issues of scaling, load balancing, and data transport.

Art Krener (Naval Postgraduate School, USA)
Eulerian and Lagrangian Observability of Point Vortex Flows
We study the observability of one and two point vortex flow from one or two Eulerian or Lagrangian observations. By observability we mean the ability to determine the locations and strengths of the vortices from the time history of the observations. An Eulerian observation is a measurement of the velocity of the flow at a fixed point in the domain of the flow. A Lagrangian observation is the measurement of the position of a particle moving with the fluid. To determine observability we introduce the observability and the strong observability rank conditions and compute them for the various flows and observations. We find that vortex flows with Lagrangian observations tend to be more observable then the same flows with Eulerian observations.
We also simulate extended Kalman filters for the various flows and observations and find that they perform poorly when the observability rank condition or the strong observability rank condition fails to hold.

Zhijin Li (NASA Jet Propulsion Laboratory, USA)
Yi Chao (NASA Jet Propulsion Laboratory, USA)
James C. McWilliams (University of California, Los Angeles, USA)
Kayo Ide (University of California, Los Angeles, USA)
Development of Data Assimilation Schemes in Support of
Coastal Ocean Observing Systems

The Integrated Ocean Observing System (IOOS) is an emerging national program, whose goal is to develop and sustain a network of integrated coastal and ocean observations. One major component of IOOS is the Regional Coastal Ocean Observing Systems (RCOOSs). For RCOOSs, a real-time data assimilation and forecasting system has been considered as an integrated component. We have developed a three-dimensional data assimilation (3DVAR) scheme in association with the Regional Ocean Modeling System (ROMS) in support of coastal ocean observing systems. The formulation will be described, and several challenges will be addressed. The major challenges are those related to complexity of topography, uncertainties of atmospheric and fresh water runoff forcing, and observability.

Andrew Lorenc (UK Met Office)
Research Issues in Data Assimilation for Operational NWP

Operational NWP has improved predictive skill by about a day per decade since the 1970s. Much of this is due to four-dimensional data assimilation methods, which fit the best available numerical forecast model to observations, allowing for the error characteristics of both. Statistically based methods are needed because observations, although plentiful in the satellite era, are incomplete and relatively inaccurate, compared to the information carried forward by a good forecast model from past observations.

Variational methods use Gaussian models to characterise the errors in the forecasts and observations; the cost and complexity of good forecast models means that the true errors are unknowable. Error modelling assumptions are essential, and the Gaussian assumption leads to computationally tractable linear systems. Recently, in order to continue to improve, it has become necessary to allow for the flow dependence of forecast errors. Statistically based four-dimensional variational methods (4D-Var) are used in most of the best synoptic-scale NWP systems, while ensemble methods for characterising errors are increasingly important. Another issue topic is the modelling of the effect of forecast model errors, rather than initial condition errors, on forecast errors. The development of these methods is the subject of much current research.

Gaussian assumptions, with a linearisable forecast model, lead to methods related to the Kalman filter (KF). Although the KF itself is not affordable, its linear equations enable approximations such as 4D-Var and the Ensemble KF which are feasible even for complex NWP models; it is likely that operational NWP will be based on such methods for some time. Nevertheless nonlinear relationships and the related non-Gaussian distributions are important; methods must be adapted to perform well in such cases (sometimes in an ad hoc way). One possible consequence of nonlinearity is that the Kalman filter Òblows upÓ if there are chaotic features which are so poorly observed that they remain chaotic in the assimilation system. This can happen in the upper stratosphere, and will happen in a global convective-scale model. In the real world, and in realistic nonlinear models, there is a climate attractor. Nonlinear processes damp perturbations (from the climate mean state) which are large compared to the climate variability. In the Kalman filter, and in the linear models used in 4D-Var, this damping needs to be added explicitly. Another consequence of this attractor is the existence of coherent structures such as fronts, inversions and cyclones. Forecasts often have plausibly structures, in the wrong place. This can lead to a significantly non-Gaussian and biased error distributions. Adaptations include the incremental approach and nonlinear initialisation. Much harder to deal with is a model with errors which mean its attractor does not match the atmosphere's.

Increases in computer power mean we can now develop convective-scale NWP systems; a

major challenge is to extend the above approaches. Except perhaps in limited regions with good Doppler radar coverage, nonlinear effects will be more important; we need to spin-up realistic convective cells which match available observations. Luckily for our ability to make forecasts, convection is often largely determined by larger-scale and topographic forcings, so multi-scale assimilation methods are needed which preserve large-scales from an optimal global system, while still fitting detailed observations when they exist.

Richard Menard (Environment Canada)
Model Error as an Unobserved Variable:
What Do We Know From Estimation Theory

Model error can be accounted as an unknown tendency in the dynamics equation. Since most observation operators relates to state variables as opposed to their tendencies, model error can be accounted as an "unobserved variable". Its impact however built up with time. We review several of the known schemes, making the links between variational and Kalman filter and smoother formulations. In a simple model we present some results and discuss the issue observability of the model error estimation problem. Some ideas about the use of lagged innovations to obtain critical error statistics are also presented. Finally, using an information filter formulation, we note some specific properties arises in the estimation of unobserved variable and speculate that it could be used to distinguish it from observation error.

Robert N. Miller (Oregon State University, USA)
Estimation of Representation Error in Ocean Models

N. Sri Namachchivaya (University of Illinois at Urbana-Champaign, USA)
Target Detection in Multi-Sensor and Multi-Scale Environments

Nancy Nichols (University of Reading, UK)
Amos S. Lawless (University of Reading, UK)
Caroline Boess (University of Bremen, FRG)
Angelika Bunse-Gerstner (University of Bremen, FRG)
Use of Reduced Order Models
in Incremental Four-Dimensional Variational Data Assimilation

Incremental methods used operationally for 4D-variational data assimilation aim to solve a sequence of linear approximations to the full nonlinear problem. These methods consist of inner iteration loops for solving each linear approximation via an adjoint procedure and an outer iteration loop in which the nonlinear analysis is updated and the problem is re-linearized about the current estimate of the analysis. In order to increase the computational efficiency of these methods, low rank approximations to the inner linear systems are used, thereby reducing the work needed in each inner iteration loop. Low resolution models derived from spatial or spectral truncations of the full system commonly provide the reduced rank approximations. Convergence of these procedures depends on how accurately the low order models approximate the full system model.

New techniques for finding reduced rank models based on balanced truncation methods developed for large scale control system design are presented here with application to the incremental 4DVar procedure. Reduced models determined by these methods are *optimal* in the sense of producing the best match to the frequency response of the full system. More of the dynamical information of the full system is therefore retained by these reduced models than by a low resolution system. Results obtained for simple shallow water test cases

illustrate the superior performance of these reduced models and show that the same accuracy in the analysis can be obtained more efficiently with a much lower dimensional reduced order approximation than with a low resolution model.

Olivier Pannekoucke [*,] (Météo-France)**
Loik Berre [*] (Météo-France)
Gérald Desroziers [*] (Météo-France)
Sébastien Massart [**] (CERFACS, France)
[*: Part I; **: Part II]

Background Error Correlation Modeling:
Representation of the Local Length-Scale From ("small") Ensemble

Ensembles of perturbed assimilations and forecasts are now a well known method. They mimic the time evolution of the dispersion of possible states in model space. However this method is costly and only a few assimilations can be computed. The issue of how to extract some robust information from an ensemble is discussed. In a first part, some features about background error correlation functions are reminded. In particular, it is shown how diagnoses of local correlation length-scale can illustrate the geographical variations of correlation functions. A second part of the talk is focused on the representation of geographical variations through the wavelets diagonal assumption. In particular, some interesting properties of this formulation are illustrated: it allows the representation of length-scale variations, and it offers interesting filtering properties. The last part of the talk deals with an alternative way to represent geographical variations of correlation functions via a modelisation based on the diffusion equation.

Chris Snyder (National Center for Atmospheric Research, USA)

Obstacles to Particle Filtering in High Dimensions

Particle filters are ensemble-based assimilation schemes that, unlike the ensemble Kalman filter, employ a fully nonlinear and non-Gaussian analysis step. Evidence is presented that, if the observation log-likelihood has an approximately Gaussian distribution, the ensemble size required for a successful particle filter scales exponentially with the variance of the observation log-likelihood, rather than the state dimension per se. Asymptotic results, following the work of Bengtsson, Bickel and collaborators, are provided for two cases: one in which each prior state component is independent and identically distributed, and one in which both the prior pdf and the observation errors are Gaussian. I will also discuss "effectively low dimensional" situations in which the observation log-likelihood is far from Gaussian despite large state dimension and large numbers of observations.

Istvan Szunyogh (University of Maryland, USA)

Flow Dependence of the Performance of
an Ensemble Based Analysis-Forecast System

Data assimilation is problem at the intersection of dynamical systems theory and mathematical statistics. In this talk, we focus on the dynamical systems aspects of the problem. In particular, we argue that most techniques of dynamical systems theory, which have already been applied to geophysical fluid dynamical systems, have a solid theoretical foundation only for low-dimensional systems. Since geophysical fluid dynamical systems are inherently high-dimensional, a systematic approach to extend the theoretical machinery to increasingly more complex systems would be highly desirable. In this talk, we outline one potential approach to address this issue: the high-dimensional system is viewed as the collection of local systems; the local state vector is defined by the variables of the original high-dimensional system from a local neighborhood of each physical location; and properties that smoothly vary with the location are computed based on the local state vectors. We

illustrate this approach by using it to explain the spatio-temporal variability of the performance of an ensemble-based analysis-forecast system. This system consists of the Local Ensemble Transform Kalman Filter data assimilation scheme and a reduced resolution version of the model component of the Global Forecast System of the National Centers for Environmental Prediction.

Olivier Talagrand (Laboratoire de Meteorologie Dynamique, France)
A Few Future Perspectives for Assimilation

A number of questions relative to the theory and practice of assimilation are discussed.

- Can fully Bayesian method be developed and practically implemented? In the case of ensemble methods, the main difficulty seems to be the large dimension of atmospheric and ocean models, which have so far been prohibitive for algorithms such as particle filters.
- Ensemble estimation, which aims at defining probability distribution, is of a different nature than deterministic estimation, which aims at determining a point estimate. What are the ways for properly evaluating ensemble estimation?
- Model errors, and different types of observational erros too, are certainly correlated in time. Sequential assimilation cannot be optimal in the presence of time-correlated errors. What are the ways to deal with temporal correlation?
- Turbulence is associated with complex interaction between spatial scales. What is the impact of those interactions on assimilation? For instance, is it possible to reconstruct the small scales of the motion from observations of the large scales?

Youmin Tang (University of North British Columbia, Canada)
Advanced Data Assimilation in Strongly Nonlinear Systems

Performance of advanced derivativeless, sigma-point Kalman filter (SPKF) data assimilation schemes in a strongly nonlinear dynamical model is investigated. The SPKF data assimilation scheme is compared against traditional Kalman filters such as extended Kalman filter (EKF) and ensemble Kalman filter (EnKF) schemes. Three particular cases, namely the state, parameter and joint estimation of states and parameters simultaneously, from a set of discontinuous noisy observations were studied.

The celebrated Lorenz model with highly nonlinear condition is used as the test bed for data assimilation experiments. The results of SPKF data assimilation schemes were compared with those of traditional EKF and EnKF where a highly nonlinear chaotic case is studied.

Andrew Tangborn (NASA Goddard Space Flight Center, USA)
Assimilation of Vorcore Polar Balloons

25 long lived constant volume stratospheric balloons were flown over Anarctica from September 2005 - January 2006. They produced position, pressure and temperature data every several minutes. Wind vectors were derived from the position measurements. We present the results of assimilating Vorcore balloon winds into the GEOS-4 data assimilation system. Improvements to the wind field are found by comparing with an independent data set. The analyzed winds are used to transport ozone fields during the same period in the GMAO ozone assimilation system. Modest improvements to the ozone forecasts are found. We also discuss current plans for assimilation in the GEOS-5 assimilation system and the upcoming Concordiasi balloon campaign.

Zoltan Toth (National Centers for Environmental Prediction, USA)
Malaquias Pena (National Centers for Environmental Prediction, USA)
Mozheng Wei (National Centers for Environmental Prediction, USA)
Yucheng Song (National Centers for Environmental Prediction, USA)
Issues Related to the Use of Ensembles in Data Assimilation and Targeting

The assimilation of observations into numerical models of dynamical systems ideally builds on both dynamical and statistical principles. This presentation will focus on the interface between dynamics and statistics, exploring some aspects from both fields that may either be critical, or could possibly be compromised for achieving a balanced solution, leading to successful data assimilation (DA). Questions and issues explored will include: The role of a numerical first guess in traditional and ensemble-based DA; Alternative ways of generating first guess fields for ensemble-based DA; Statistical and dynamical considerations when estimating the background error covariance; How to assimilate data with an imperfect model? Can information from highly non-linear forecasts be used for targeting observations to improve such forecasts? What other issues one must consider for an integrated observing, data assimilation and forecast system for chaotic systems?

Keith R. Thompson (Dalhousie University, Canada)
Yimin Liu (Dalhousie University, Canada)
Predicting Mesoscale Variability of the North Atlantic
Using a Simple, Physically Motivated Scheme For Assimilating Altimeter and Argo Data

A computationally-efficient scheme is described for assimilating sea level measured by altimeters and vertical profiles of temperature and salinity measured by Argo floats. The scheme is based on a transformation of temperature, salinity and sea level into a new set of physically-motivated variables for which it is easier to specify spatial covariance functions. The scheme allows for sequential correction of temperature and salinity biases, and online estimation of time-dependent background error covariance parameters. Two North Atlantic application, both focused on predicting mesoscale variability, are used to assess the effectiveness of the scheme. In the first application the background field is a monthly temperature and salinity climatology and skill is assessed by how well the scheme can recover Argo profiles that have not been assimilated. In the second application the backgrounds are forecasts made by an eddy permitting model of the North Atlantic. Skill is assessed by the quality of forecasts with lead times of 1 to 60 days. For both applications it is shown that the scheme has useful skill. The benefits of using physical constraints to reduce the computational cost of assimilation will be discussed in general, and compared to other cost-reducing approaches such as those used in the SEEK filter.

Richard Todling (NASA Goddard Space Flight Center, USA)
Yannick Tremolet (NASA Goddard Space Flight Center, USA)
Catching up to the World: The GMAO 4d-Var and its Adjoint-based Tools

The fifth generation of the Goddard Earth Observing System (GEOS-5) Data Assimilation System (DAS) is a 3d-var system that uses the Grid-point Statistical Interpolation (GSI) system developed in collaboration with NCEP, and a general circulation model developed at Goddard, that includes the finite-volume hydrodynamics of GEOS-4 wrapped in the Earth System Modeling Framework and physical packages tuned to provide a reliable hydrological cycle for the integration of the Modern Era Retrospective-analysis for Research and Applications (MERRA). This MERRA system is essentially complete and the next generation GEOS is under intense development. A prototype next generation system is now complete and has been producing preliminary results. This prototype system replaces the GSI-based Incremental Analysis Update procedure with a GSI-based 4d-var which uses the adjoint of the finite-volume hydrodynamics of GEOS-4 together with a vertical diffusing scheme for

simplified physics. As part of this development we have kept the GEOS-5 IAU procedure as an option and have added the capability to experiment with a First Guess at the Appropriate Time (FGAT) procedure, thus allowing for at least three modes of running the data assimilation experiments.

The prototype system is a large extension of GEOS-5 as it also includes various adjoint-based tools, namely, a forecast sensitivity tool, a singular vector tool, and an observation impact tool, that combines the model sensitivity tool with a GSI-based adjoint tool. These features bring the global data assimilation effort at Goddard up to date with technologies used in data assimilation systems at major meteorological centers elsewhere.

Various aspects of the next generation GEOS will be discussed during the presentation at the Workshop, and preliminary results will illustrate the discussion.

Peter Jan van Leeuwen, (University of Utrecht, the Netherlands)
Particle Filtering in Large-Scale Systems: Problems and Potential Solutions

Starting from Bayes theorem, it is shown how the Kalman filter and 4D-Var can be derived. Then we concentrate on methods that use a particle representation of the model probability density function (pdf). The Ensemble Kalman Filter (EnKF) is derived, and we shortly touch upon the Ensemble Kalman Smoother. For applications in large-scale problems the number of particles is limited to typically 10-100 because of computational limitations. Given the number of independent unknowns, of the order of 1 million, the ensemble size is way too low. In that case the general formulation does not work, and adjustments have to be made. One of them is to do the analysis locally, i.e., use only observations close to a certain gridpoint to determine the analysis there. This procedure increases the effective ensemble size by a substantial factor, sometimes a factor 1000. While this is still too low, first-order statistics can be determined with reasonable accuracy. A serious drawback is that the EnKF assumes that the pdf is Gaussian, which can be bad for strongly nonlinear systems. In that case a so-called particle filter can be used in principle, in which the analysis consists of weighting the particles according to their 'distance' to the observations (importance weighting/sampling). Again, the affordable number of particles is too low for most geophysical applications. The localization cannot be applied directly in the particle filter, because it tends to break local geophysical balances, so other solutions have to be found. We discuss several possibilities, such as reduced-space solutions, merging particle filters, marginal pdf's, localization guided by the EnKF etc. Although all these ideas have potential, the final answer is still to be found.

Tomislava Vukicevic (University of Colorado, Boulder, USA)
Analysis of the Impact of Model Nonlinearities, Modeling Errors and Gaussian Prior
in Inverse Problem Solving

In this study, the relationship between nonlinear model properties and inverse problem solutions is analyzed using a numerical technique based on the inverse problem theory formulated by Mosegaard and Tarantola. According to this theory, the inverse problem and solution are defined via convolution and conjunction of probability density functions that represent stochastic information obtained from the model, observations and prior knowledge in a joint multidimensional space. This theory provides an explicit analysis of the nonlinear model function, together with information about uncertainties in the model, observations, and prior knowledge through construction of the joint probability density, from which marginal solution functions can then be evaluated. The numerical analysis technique derived from the theory computes the component probability density functions in discretized form via a combination of function mapping on a discrete grid in the model and observation phase space, and Monte Carlo sampling from known parametric distributions. The efficacy of the numerical analysis technique is demonstrated through its application to two well known simplified models of Atmospheric physics: Damped oscillations and Lorenz' 3-component

model of dry cellular convection. The major findings of this study are: a) Use of a nonmonotonic forward model in the inverse problem gives rise to the potential for a multimodal posterior pdf, the realization of which depends on the information content of the observations, and on observation and model uncertainties, b) Cumulative effect of observations, over time, space or both, could render unimodal final posterior pdf even with the nonmonotonic forward model, c) A greater number of independent observations are needed to constrain the solution in the case of a nonmonotonic nonlinear model than for a monotonic nonlinear or linear forward model for a given number of degrees of freedom in control parameter space, d) A nonlinear monotonic forward model gives rise to a skewed unimodal posterior pdf, implying a well posed maximum likelihood inverse problem, e) The presence of model error greatly increases the possibility of capturing multiple modes in the posterior pdf with the nonmonotonic nonlinear model, and f) In the case of a nonlinear forward model, use of a Gaussian approximation for the prior update has a similar effect to an increase in model error, which indicates there is the potential to produce a biased mean central estimate even when observations and model are unbiased.

Fuqing Zhang (Texas A&M University, USA)
Meng Zhang (Texas A&M University, USA)
James A. Hansen (Naval Research Laboratory, Monterey, USA)
Coupling Ensemble Kalman Filter with Four-Dimensional Variational Data Assimilation
This study examines the performance of coupling deterministic four-dimensional variational assimilation (4DVAR) with an ensemble Kalman filter (EnKF) to produce a superior hybrid approach for data assimilation. The coupled assimilation scheme (E4DVAR) benefits from using the state-dependent uncertainty provided by EnKF while taking advantage of 4DVAR in preventing filter divergence. The 4DVAR analysis produces posterior maximum likelihood solutions through minimization of a cost function about which the ensemble perturbations are transformed, and the resulting ensemble analysis can be propagated forward both for the next assimilation cycle and as a basis for ensemble forecasting. The feasibility and effectiveness of this coupled approach are demonstrated in an idealized model with simulated observations. It is found that the E4DVAR is capable of outperforming both 4DVAR and the EnKF under both perfect- and imperfect-model scenarios. The performance of the coupled scheme is also less sensitive to either the ensemble size or the assimilation window length than that for standard EnKF or 4DVAR implementations.

Milija Zupanski (Colorado State University, USA)
Dynamical Approach to Nonlinear Ensemble Data Assimilation
We present an overview of ensemble data assimilation methods by focusing on their approach to nonlinearities. This subject is introduced from the dynamical, rather than a typical statistical point of view. As such, most problems in ensemble data assimilation, and in data assimilation in general, are seen as means of producing an optimal state that is in dynamical balance, rather than producing a state that is optimal in a statistical sense. Although in some instances these two approaches may produce the same results, in general they are different. Details of this difference will be discussed, and also related to variational data assimilation.
Nonlinearities are of special interest in realistic high-dimensional applications, which implies the need for considering the number of degrees of freedom. The means for increasing the degrees of freedom in ensemble data assimilation will be briefly discussed, in particular their impact on dynamical balance.
An algorithm named the Maximum Likelihood Ensemble Filter (MLEF) is presented as a prototype nonlinear ensemble data assimilation method. Some results with the MLEF are shown to illustrate its performance, including the assimilation of real observations with the Weather Research and Forecasting (WRF) model for the hurricane Katrina.