NSF Collaboration in Mathematical Geosciences, Award No. 0930265
June 2, 2011

Annual Report for Year 2: September 1, 2010 – August 31, 2011

Milija Zupanski
CIRA, Colorado State University
Fort Collins, CO 80523-1375

Project Title: Collaborative Research: CMG--Ensemble Data Assimilation for Nonlinear and Nondifferentiable Problems in Geosciences
Status: Year 2
Duration: 3 years
Participants: Milija Zupanski (PI)

Project summary and motivation:

Ensemble Data Assimilation (EnsDA) is an advanced filtering method that simultaneously estimates the optimal state of a system and quantifies its uncertainty. Among most challenging problems remaining for EnsDA applications in geosciences is the problem of highly nonlinear and nondifferentiable processes and observations. Examples of such processes are cloud, aerosol and precipitation processes, as well as remote sensing (e.g., satellite and radar) observations.

Unfortunately, a typical EnsDA analysis equation is linear, being based on the Kalman filter equations. Although nonlinearity of the analysis has been addressed in the past, there were no attempts to include nondifferentiability, much less to combine nonlinear and nondifferentiable components. This fundamentally prevents EnsDA from extracting maximum information from such observations, and ultimately limits its applicability to outstanding geosciences problems such as hurricane prediction and high-resolution climate simulation.

In order to overcome the limitation of existing EnsDA methodologies in application to highly nonlinear and nondifferentiable geoscience problems, we develop a new strategy based on including a nondifferentiable minimization algorithm in EnsDA. We also explore the potential for improving EnsDA by strengthening its link with four-dimensional variational (4D-Var) data assimilation. This research represents a synergy between control theory, estimation theory, and data assimilation. The anticipated outcome is a development of new EnsDA system with nonlinear and nondifferentiable capability. The research has the following objectives:
1) *Evaluate nondifferentiable minimization methods suitable for EnsDA,*
2) *Examine the value of hybrid EnsDA methods for nonlinear and nondifferentiable applications,*
3) *Develop and evaluate a nonlinear/nondifferentiable EnsDA method designed to quantify uncertainty in realistic high-dimensional geosciences applications.*

Management Overview of the Year 2:

We continue supporting a graduate student from CSU (Biljana Orescanin, a student of Prof. Scott Denning, Atmospheric Science Department), and advising a Ph.D. student from FSU (Jeff Steward, a student of Prof. Navon). In addition to regular contacts with Jeff Steward regarding his research, Dr. Milija Zupanski serves as an outside member of his Ph.D. committee. We also developed several independent collaborations that address nonlinear and nondifferentiable problems in geosciences, in particular the snow assimilation and prediction, and assimilation of atmospheric chemical constituents to predict air pollution. These applications will have an indirect positive impact on this NSF-funded research, since they are good examples of nonlinear and nondifferentiable data assimilation.

Major milestones and accomplishments (Year 2):

The NSF project web page can be accessed from our research web page (http://www.cira.colostate.edu/projects/ensemble/research.php), with a link to "NSF Collaboration in Mathematical Geosciences 2009-2012"

Milestones and accomplishments

1) Decide about most viable option(s) for nonlinear/nondifferentiable minimization algorithm and for hybrid filter development (*completed*)

One of the main requirements for a nonlinear/nondifferentiable minimization algorithm is good convergence characteristics coupled with computational/memory feasibility. Our results, also included in the submitted manuscript by Steward et al. (2011), suggest that Limited memory BFGS (LBFGS) quasi-Newton algorithm and the Limited Memory Bundle Method (LMBM) respond well to nondifferentiable problems, as opposed to the nonlinear conjugate gradient (CG) algorithm.

Our conclusion is that the most viable algorithm for nonlinear/nondifferentiable minimization is the LMBM algorithm, but also that the generalized LBFGS quasi-Newton algorithm described in Zupanski et al. (2008) can be a very good less expensive option.

To illustrate some of the results, we consider a set of nondifferentiable observation operators used in the experiments with shallow water model, given by Eqs. (1) and (2):

$$H(x) = \begin{cases} H_1(u) \\ H_2(v) \\ H_2(v) \end{cases} \tag{1}$$

where

$$H_1(u) = \begin{cases} u^3/u_{\min}^2 & u < u_{\min} \\ u^2/u_{\max}^2 & u \ge u_{\max} \\ u & else \end{cases} \quad H_2(v) = \begin{cases} \log(v+\delta) & v \ge 0 \\ \log(-v+\delta) & v < 0 \end{cases} \quad H_3(\phi) = \begin{cases} \phi & \phi < H_{\max} \\ \phi^2/H_{\max} & \phi \ge H_{\max} \end{cases} \tag{2}$$

The functions corresponding to (2) are plotted in Fig.1, and their first derivatives are shown in Fig.2. One can see that there are considerable jumps in the gradient indicating a nondifferentiable cost function and minimization problem.

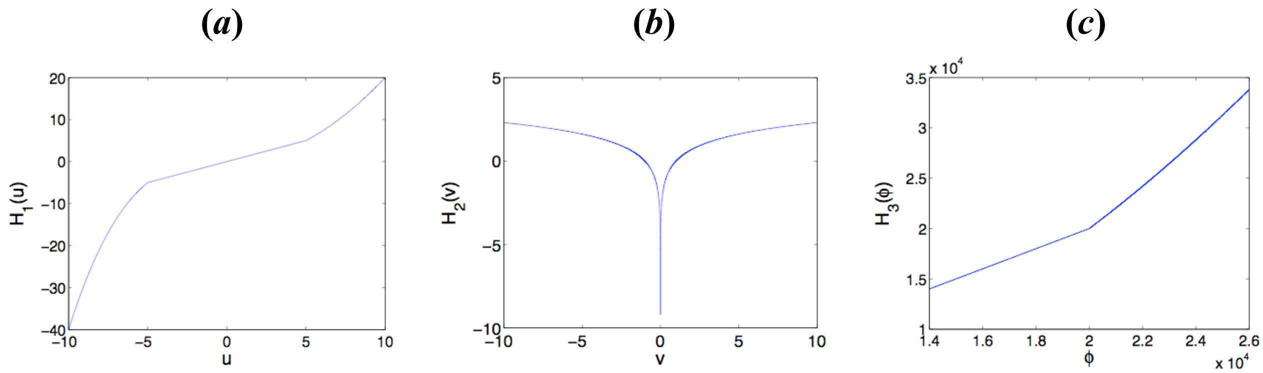

***Figure* 1**. Non-differentiable observation operator branches used in data assimilation: (a) $\mathcal{H}_1$, (b) $\mathcal{H}_2$, and (c) $\mathcal{H}_3$. The function is continuous, but it does include sharp changes.

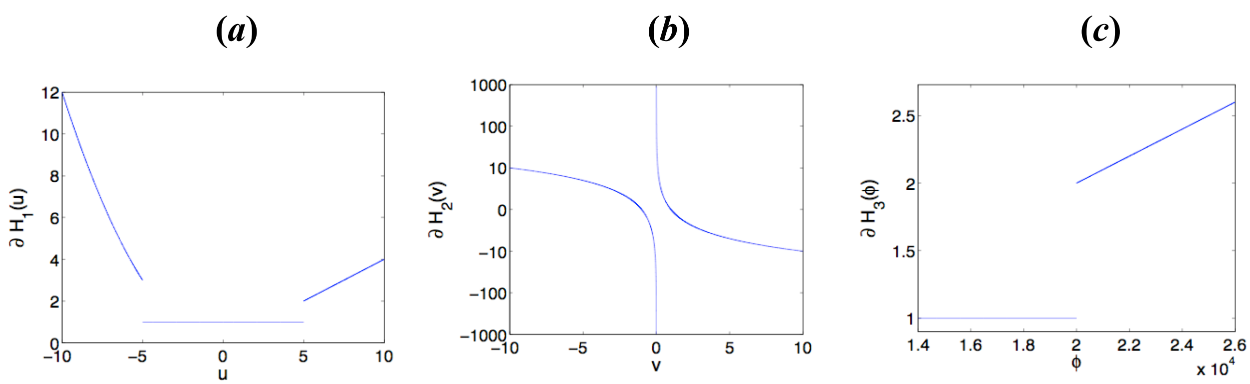

***Figure* 2**. Corresponding gradients of the observation operator branches from Fig.1. Each of the branches (a, b, c) has a jump in the first derivative making it a challenging minimization problem for data assimilation.

We applied two minimization algorithms with nondifferentiable observation operators (1)-(2): (*i*) generalized nonlinear conjugate-gradient (CG), and (*ii*) Limited Memory Bundle Method (LMBM). The CG algorithm is implemented in the derivative-free form of Zupanski et al. (2008), while the LMBM algorithm is adopted from Karmitsa (2007). The results of minimization with MLEF are shown in Fig. 3, for several degrees of difficulty ranging from less difficult (Exp 1) to most difficult (Exp 4). Both experiments show a generally satisfactory analysis RMS error. However, the LMBM algorithm has a clear advantage, especially in the left side of the domain for more difficult experiments 3 and 4.

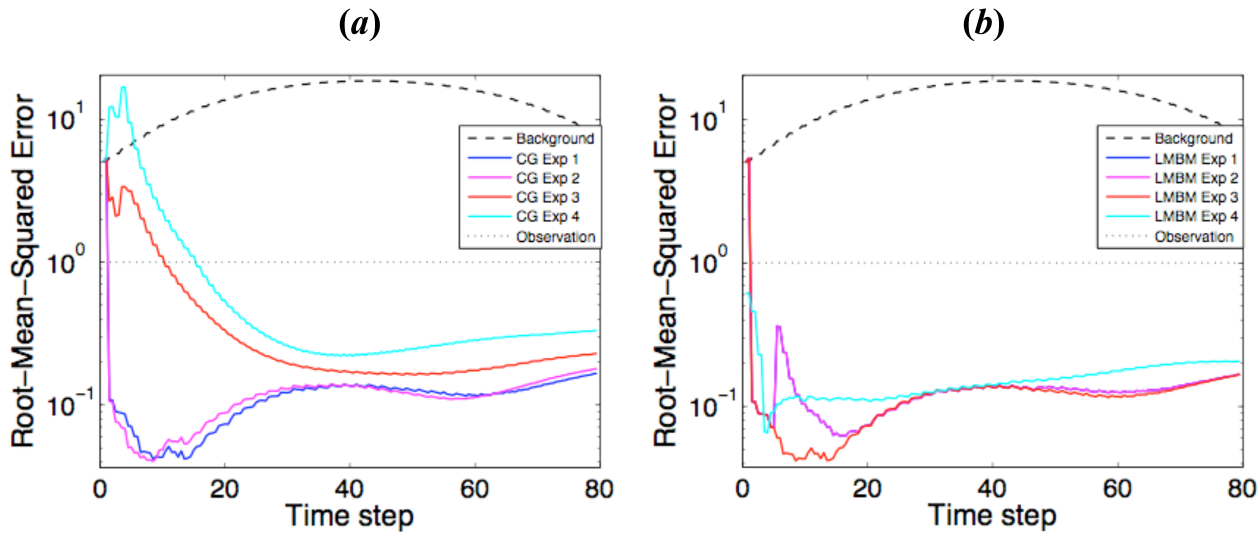

***Figure* 3**. The RMS error of the north-south wind component (ms^{-1}) in the MLEF ensemble data assimilation experiments with: (a) nonlinear CG, and (b) LMBM minimization methods. The dashed line represents the first guess RMS error, and the dotted line represents the observation error. In both experiments the final RMS error is generally smaller than either the background or the observation error. However, the LMBM algorithm eventually reaches a better fit, especially in the left side of the domain.

2) Develop computationally efficient code for the nonlinear/nondifferentiable/hybrid EnsDA system (*completed*)

Note that the LMBM algorithm includes the LBFGS for descent direction computation, with additional sub-gradient aggregation and modified limited memory SR1 update computations. This suggests that it may be possible to make an efficient use of the LMBM algorithm by allowing the LBFGS to be used in less difficult problems, and add more LMBM features as the minimization problem becomes more difficult. Therefore we adopt a new minimization setup shown in Fig. 4. As a default we use a derivative-free formulation of the generalized LBFGS (Zupanski et al. 2008), with optional addition of the LMBM for more difficult minimization problems (e.g., large Lipschitz constant). Since large-scale ensemble data assimilation requires error covariance localization, the minimization is typically performed locally, for smaller domains. The algorithmic setup presented in Fig. 4 allows an efficient use of the LBFGS-LMBM combination that could be done in parallel for all local domains.

Line search is critical for efficient and reliable performance of a nonlinear/nondifferentiable minimization algorithm. Reliable line search improves the efficiency of minimization algorithm. Therefore, in addition to the quadratic line search previously used in the generalized CG and LBFGS algorithms, an iterative cubic line search, followed by the check for Wolfe conditions, has been developed and implemented with MLEF. This development strengthens the default option of the LBFGS-LMBM algorithm, and implicitly the LMBM option as a preliminary line search.

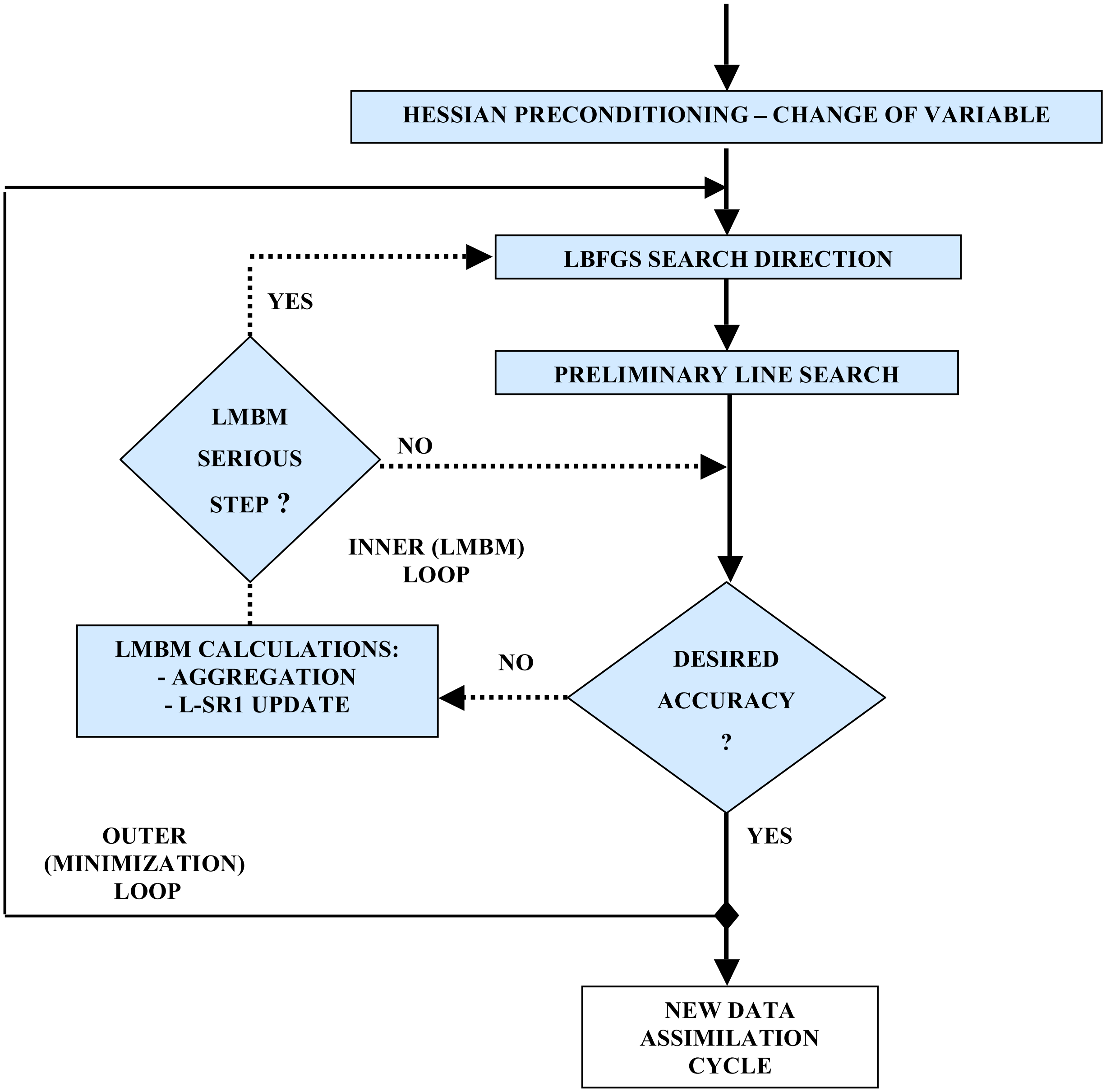

***Figure* 4**. Flow-chart of an efficient LBFGS-LMBM nonlinear/nondifferentiable minimization algorithm. For sufficient accuracy achieved in the initial LBFGS step the computationally more demanding LMBM step is avoided. When required, however, the LMBM will make use of the LBFGS code. This unconstrained minimization algorithm satisfies the efficiency requirement in computationally demanding data assimilation applications, and also relies on modular use of its components.

3) Perform extensive tests of the new EnsDA system. Make adjustments as necessary. (*in progress*)

Preliminary experiments with the new EnsDA system have been completed (Jardak et al. 2011; Steward et al. 2011). As indicated above, these experiments have produced a guidance regarding the choice of optimal nonlinear/differential minimization algorithms, as well as their impact in hybrid variational-ensemble data assimilation systems (Steward et al. 2011). In addition, our results suggest that some theoretically appealing methodologies, such as the particle filter, may require an order of magnitude larger number of ensembles than EnKF and MLEF, making it practically useless in realistic geosciences applications (Jardak et al. 2011). With respect to nonlinearity and nondifferentiability of the assimilation problem, however, the MLEF has shown an advantage over the EnKF. Our results imply that MLEF data assimilation with LBFGS-LMBM minimization represents the best available system for addressing high-dimensional nonlinear/nondifferentiable data assimilation problems found in geosciences.

We already started MLEF experiments using the WRF model with assimilation of conventional and satellite observations (originally planned for the Year 3). This will give us a head start for evaluating the LBFGS-LMBM minimization in realistic high-dimensional applications. In this system we access real data using the forward component of the community grid-point statistical interpolation (GSI) system. Bothe the WRF model and the GSI system are available from the Development Testbed Center (DTC) sponsored by National Center for Atmospheric Research (NCAR) and National Oceanic and Atmospheric Administration (NOAA) [at http://www.dtcenter.org/].

Since in planned experiments we address the nonlinear and nondifferentiable problems associated with assimilation of cloud-related observations, we include cloud microphysical variables as control variables in minimization. In order to illustrate the complexity of variable interactions in such system, we show the analysis increments from a single observation experiment. In particular, we assimilate a single observation of cloud snow at 500 hPa, which corresponds to high-frequency microwave satellite radiance channels, such as the 89.0 GHz band of the AMSR-E instrument. Note that such measurements of all-sky radiances are rarely assimilated even in an experimental mode, due to nonlinearity/nondifferentiability of all-sky radiative transfer operator, as well as due to complex forecast error covariance representation of microphysical variable uncertainty. The analysis increments (analysis minus first guess) are shown in Fig. 5. Note that

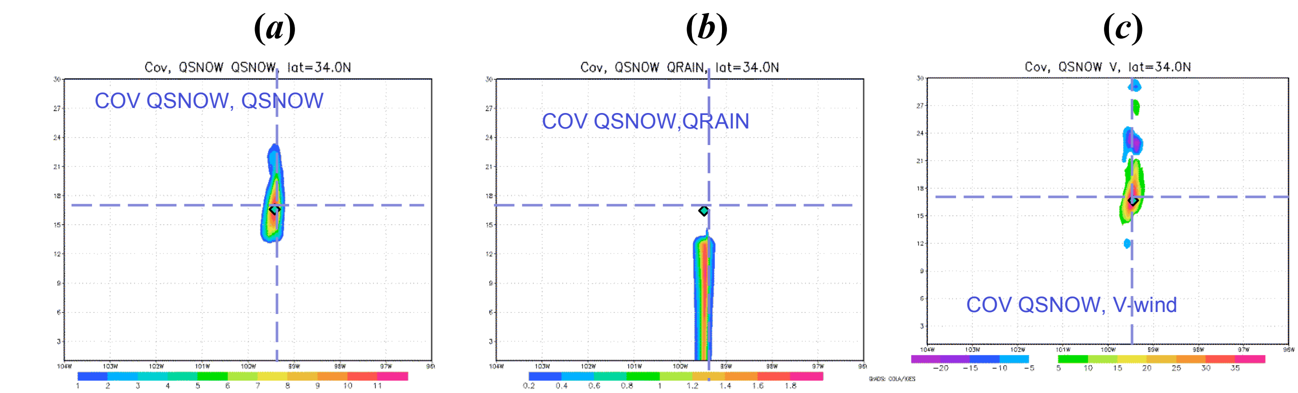

***Figure* 5**. Vertical cross-section of the analysis response of the MLEF-WRF system to a single observation of cloud snow at 500 hPa.: (*a*) snow, (*b*) rain, and (*c*) north-south component of wind. Note well defined, but asymmetric response of the cloud rain analysis.

in a single-observation experiment the analysis increments also reveal the structure of ensemble-based forecast error covariance. One can see that all analysis increments have localized responses, as expected. However, while the analysis increment of snow (Fig. 5a) is very much symmetric and centered around the observation, the cloud rain analysis increment is asymmetric, and also not centered around the observation. Although this feature is difficult to model in vaariational methods, it is a physically realistic response essentially implying that rain is not possible above the freezing level, but that it is connected to other levels through cloud dynamics. The wind analysis increment (Fig. 5c) shows that observation of cloud ice impacts not only the microphysical cloud variables but also the standard atmospheric variables, all linked through model dynamics.

4) Examine the quality of the Hessian preconditioning and the uncertainty estimation in nonlinear/nondifferentiable situation (*completed*)

Hessian matrix is of special importance not only for minimization and preconditioning, but also for data assimilation since its inverse provides an estimate of uncertainty. Using as an example the MLEF ensemble data assimilation with WRF model and satellite radiance assimilation, we show the ensemble-space Hessian spectrum at the initial first guess point, also used for preconditioning, and at the optimal point obtained as a result of minimization (Fig. 6). The particular storm was located over Texas and Oklahoma on 18-19 August 2007, and produced excessive precipitation and high winds.

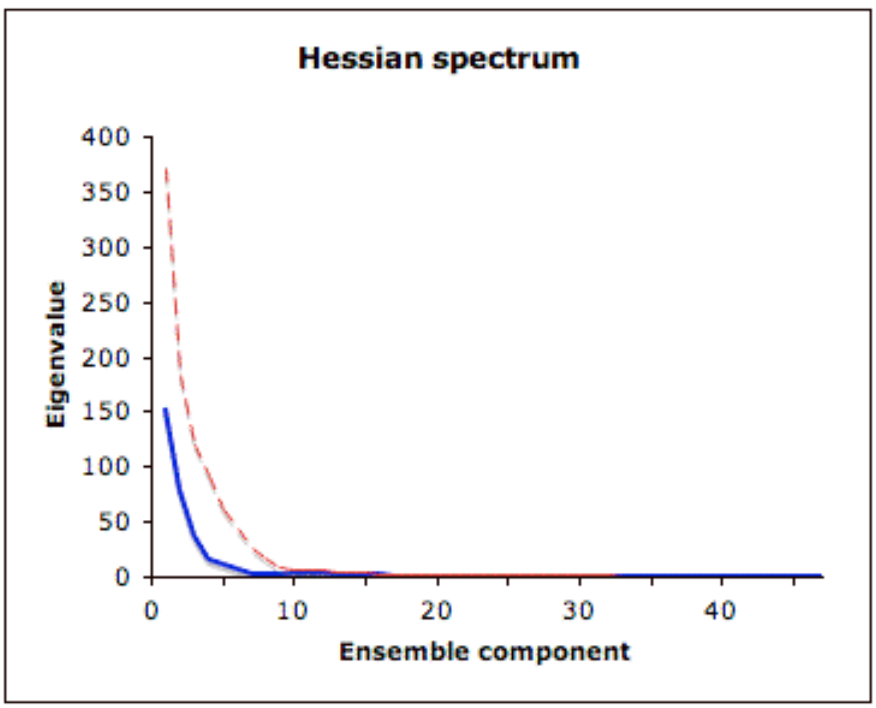

***Figure* 6**. Hessian spectrum for a local domain No. 131, corresponding to the location of the storm at 1800 UTC on August 18, 2007: (a) initial Hessian (solid blue line), and (b) final Hessian (dashed red line). Larger values of the final eigenvalues correspond to a reduction of uncertainty by data assimilation. Note that all eigenvalues are positive since the Hessian matrix is positive-definite.

One can note generally smaller eigenvalues at the initial point, considerably increasing at the optimal point. Since the inverse Hessian at the final point is a good approximation of the analysis error covariance, $P_a \approx H^{-1}$, this indicates that there is a reduction of uncertainty by data assimilation. Note that the eigenvalues are related by $\lambda(P_a) = \dfrac{1}{\lambda(H)}$. This uncertainty reduction can be quantified using the norm of the Hessian and its inverse. The Frobenious norms of the Hessian at the initial and final points are $\left\| H^{init} \right\| = 176$ and $\left\| H^{final} \right\| = 450$, respectively. The corresponding analysis error covariance norms are $\left\| P_a^{init} \right\| = 5.9$ and $\left\| P_a^{final} \right\| = 4.7$. Given that the uncertainty is proportional to the square root of analysis error covariance, one can find that the uncertainty reduction is 11% in this example. Similar behavior of the Hessian spectrum can be found for other local domains.

It is also of interest to quantify the variability of Hessian spectra over all local domains, since it indicates the uncertainty associated with ensemble components due to error covariance localization. In Fig.7 we show the variability of the Hessian spectra defined by the spread of spectra over all local domains. Specifically, we calculate the differences between (i) the maximum and mean spectra to define upper limit, and (ii) the minimum and the mean spectra to define the lower limit. The extremes are determined based on the Frobenious norm values. One

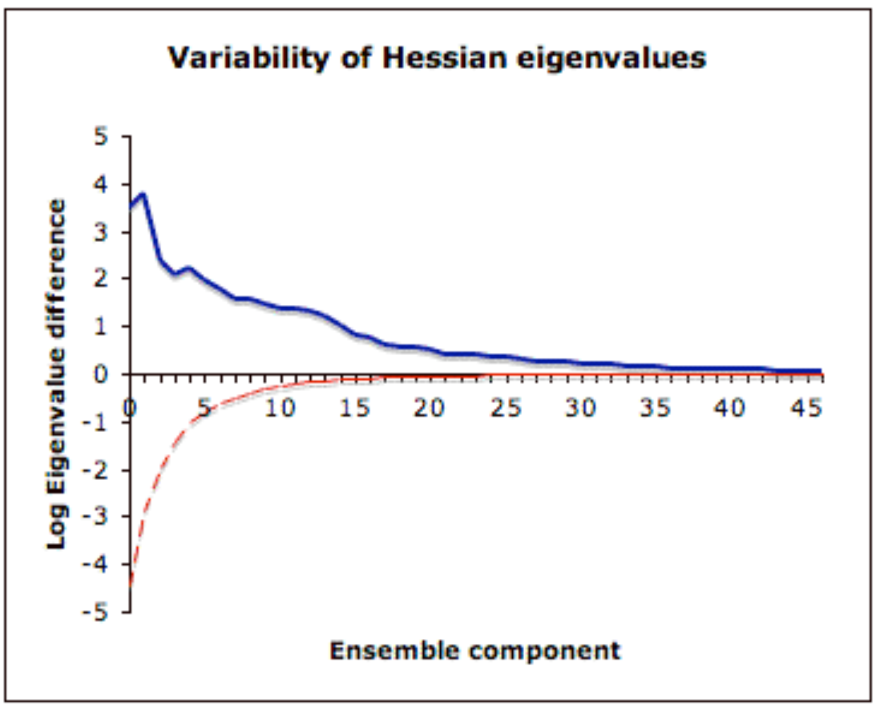

***Figure* 7**. Hessian spectra variability over all local domains, valid 1800 UTC on August 18, 2007: (a) upper limit (solid blue line), and (b) lower limit (dashed red line). Larger variability spreads over large and middle part of the spectrum, suggesting the need for good estimates of Hessian spectra throughout all ensemble components.

can note that most variability is associated with leading eigenvalues, but also that variability is noticeable for middle and smaller eigenvalues as well. This suggests that a good estimate of the Hessian spectra is required for most ensemble components. By the impact of local minimizations, this implies that the success of global minimization is impacted mostly by the lower and middle part of the Hessian spectra.

5) Submit manuscript(s) for publication in peer-reviewed scientific journals. Present the results on the web page, and at the conference/workshop on this subject (*completed*)

Two manuscripts are currently in review, both accepted subject to revisions:

(i) Jardak et al. (2011), submitted to the Journal of Geophysical Research, and
(ii) Steward et al. (2011), submitted to the Quarterly Journal of the Royal Meteorological Society.

Jeff Steward is preparing another manuscript that will employ the Weather Research and Forecting (WRF) model and satellite radiance assimilation, and Milija Zupanski is preparing two manuscripts, one related to nonlinearity and nondifferentiability of the carbon transport (with graduate student Biljana Orescanin), and another using WRF model and all-sky radiance assimilation.

As an invited guest speaker Dr. Milija Zupanski presented research related to this project at:

(i) Japan Meteorological Research Institute (MRI) (Zupanski 2011a), and
(ii) Japan Aerospace Exploration Agency (JAXA) (Zupanski 2011b),

and will give an invited talk at the upcoming meeting

(iii) The 6th International EnKF Workshop (Zupanski 2011c).

References:

Jardak, M., I. M. Navon, and M. Zupanski, 2011: Comparison of ensemble data assimilation for the shallow water equations model in the presence of nonlinear observation operator. *Q. J. Roy. Meteorol. Soc., accepted subject to revisions*.

Karmitsa, N., 2007: LMBM — FORTRAN Subroutines for Large-Scale Nonsmooth Minimization: User's Manual. *TUCS Technical report* No. 856, 32 pp.

Steward, J., I. M. Navon, M. Zupanski, and N. Karmitsa, 2011: Impact of non-smooth observation operators on variational and sequential data assimilation for a limited-area shallow water equations model. *Q. J. Roy. Meteorol. Soc.*, *accepted subject to revisions*.

Zupanski, M., I. M. Navon, and D. Zupanski, 2008: The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. *Q. J. Roy. Meteorol. Soc.*, **134**, 1039-1050.

Zupanski, M., 2011a: Cloud-Resolving Ensemble Data Assimilation for Predicting Local Heavy Rainfall. 2011 Meeting on the Study of Advanced Data Assimilation and Cloud Resolving Ensemble Technique for Prediction of Local Heavy Rainfall, February 28, 2011, Tsukuba, Japan.

Zupanski, M., 2011b: Satellite Data Assimilation: Ensemble data assimilation perspective. Invited presentation at JAXA, March 2, 2011, Tsukuba, Japan.

Zupanski, M., 2011c: A control theory approach to nonlinearity and non-differentiability in ensemble data assimilation. *The 6th International EnKF Workshop*, June 20-22, 2011, Ulvik, Norway.

Additional activities leveraged:

Our CMG research is closely related to several other projects lead by Milija Zupanski, that focus on hurricane, all-sky satellite radiance and precipitation assimilation using the MLEF algorithm. These projects are supported by the National Oceanographic and Atmospheric Administration (NOAA) and National Aeronautics Space Administration (NASA). The projects are:

i) NASA Global Precipitation Mission (GPM),
ii) Joint Centers for Satellite Data Assimilation (JCSDA) Utilization of GOES-R for hurricane prediction and analysis,
iii) NOAA NESDIS Extracting Maximum information from GOES-R data
iv) NOAA Hurricane Forecasting Improvement Program (HFIP).