A Strategically New Forecast Process Based on a Local Ensemble Kalman Filter

Application for the program THORPEX: A Global Atmospheric Research Program

PI: Dr. Istvan Szunyogh, Co-PI: Dr. Brian Hunt

September 5, 2003

Abstract

We propose a three year project to develop a strategically new forecast process, which is integrated, adaptive, and user controllable in accordance with the objectives of THORPEX. We will achieve this goal, through further developing our existing Local Ensemble Kalman Filter (LEKF) data assimilation scheme for the NCEP GFS and by developing a new adaptive observation technique applicable in the presence of strong nonlinearities. This system will integrate the sub-components of weather forecasting by: (1) enhancing the assimilation of weather observations, (2) generating ensemble initial conditions consistent with the spatio-temporally varying uncertainty in the analysis, (3) adaptively controlling the collection of observations, and (4) providing probabilistic forecasts for economic or societal applications. The adaptive nature of the Kalman Filter and the new observation techniques we are planning to test would provide a basis for adaptive observation, assimilation, and application procedures that would vary depending on the case and forecast situation. The proposed system is ideal for user controllable applications since it is computationally efficient, largely model independent, and easily portable between different computational platforms.

1 Introduction

Operational numerical weather prediction systems perform data assimilation by combining a short-term forecast (background) with observations to create an analysis. This procedure is a statistical interpolation based on the assumptions that uncertainties in the background and the observations are normally distributed with zero expected values and known covariance between the model variables and between observations. In reality, however, the covariance matrices cannot be directly computed since the true state of the atmosphere is unknown. Thus the implementation of a data assimilation system requires the development of statistical models that can provide estimates of the covariance matrices. The quality of a data assimilation system is primarily determined by the accuracy of these estimates (e.g., Daley, 1991; Kalnay, 2003).

Uncertainties in the background have a large variability in both space and time. Due to computational constraints, current operational data assimilation systems cannot take this variability into account. One of the most promising approaches to obtain a computationally cost efficient estimate of the background covariance matrix is the effort to develop operationally attainable ensemble Kalman filter schemes. The usefulness of this approach has been demonstrated for both simplified and more realistic models of the atmosphere (e.g. Houtekamer and Mitchell, 1998; Evensen, 1994; Anderson and Anderson, 2001; Hamill and Snyder, 2000; Whitaker et al., 2003).

Our proposed LEKF data assimilation system is an ensemble Kalman filter. The most important differences between our scheme and those proposed in the aforementioned papers is that (1) our scheme updates the analysis *concurrently at each grid point* (locally in model space), while the others update the analysis *sequentially at each observational location* (locally in observation space); (2) in our scheme the background covariance information is *localized by defining smoothly connected local regions* in the neigborhood of each grid point, while in the sequential schemes the background covariance information is *localized by the Gaussian filter of Gaspari and Cohn (1999) using a prescribed correlation length* in the neighborhood of each observation. We believe that our unique formulation of the ensemble Kalman filter is advantageous when many observations, especially those with correlated observational errors, are assimilated:

- The LEKF allows for significantly reducing the dimension of the analysis problem by using the dynamically most important local directions of the state space as the basis for the matrix operations. This reduced dimension depends only on the complexity of the dynamics, i.e., it is independent of the number of observations. This leads to an efficient filtering of redundant information in the observed data, which is present due to an oversampling by the observing instruments (a typical problem with high resolution remote sensors). In contrast, the sequential schemes require statistical techniques (e.g., superobing) to reduce the number of observations prior to doing the data assimilation.

- While the sequential data assimilation schemes can be extended to the case, in which the observations have correlated error, their efficiency quickly degrades as the number of correlated observations increases. This is due to the fact that all correlated observations have to be assimilated at the same time in this formulation. Hence the update step, which is a scalar equation for the case of uncorrelated observations, becomes a matrix equation of ever growing dimension as the number of correlated observations increases. In our scheme, the dimension of the matrix equations is fixed and, except for the inversion of the observational error covariance matrix, it is solely determined by the complexity of the dynamics.

- Although the computational cost of our scheme increases linearly with the number of grid points, the computation can be performed concurrently for each grid point. Thus, the wall clock computational time can be efficiently reduced in a parallel computational environment.

- We anticipate that a 4-dimensional extension of the algorithm, that has already been shown to work for a low-order analogue of the atmospheric dynamics, can be implemented on the NCEP GFS. We also anticipate that an efficient direct minimization of the cost function in the low dimensional local regions is also possible.

A detailed mathematical derivation and discussion of the new scheme can be found in Ott (2003a,b). Here, we first present an outline of the scheme providing no more detail than necessary to explain our unique concept of localization and to propose further enhancements to the system. We also show some results obtained by an implementation on the T62, 28-level, 2001 version of the full operational NCEP GFS.

2 Prior Research

2.1 Local vectors

A model state of the atmosphere is given by a vector field $\mathbf{x}(\mathbf{r}, t)$ where $\mathbf{r}$ is two dimensional and runs over discrete values $\mathbf{r}_{mn}$ (the grid in the physical space used in the numerical computations). Typically, the two components of $\mathbf{r}$ are the geographical longitude and latitude, and $\mathbf{x}$ at a fixed $\mathbf{r}$ is a vector of all relevant physical state variables of the model (e.g., wind velocity components, temperature, surface pressure, humidity, etc., at all height levels included in the model). Let u denote the dimensionality of $\mathbf{x}(\mathbf{r}, t)$ (at fixed $\mathbf{r}$); e.g., when five independent state variables are defined at 28 vertical levels, $u = 140$.

We do our analysis locally in model space. To explain this local procedure, we first introduce our local coordinate system and the approximations we make to the local probability distribution of $\mathbf{x}(\mathbf{r}, t)$. Since all the analysis operations take place at a fixed time t, we will suppress the t dependence of all vectors and matrices introduced henceforth.

We introduce at each point *local vectors* $\mathbf{x}_{mn}$ of the information $\mathbf{x}(\mathbf{r}_{m+m',n+n'}, t)$ for $-l \leq m', n' \leq l$. That is, $\mathbf{x}_{mn}$ specifies the model atmospheric state within a $(2l+1)$ by $(2l+1)$ patch of grid points centered at $\mathbf{r}_{mn}$. (This particular shape of the local region was chosen to keep the notations as simple as possible, but different (e.g., circular) shape regions and localization in the vertical direction can also be considered.) The dimensionality of $\mathbf{x}_{mn}$ is $(2l + 1)^2u$. The *local background error covariance matrix* and most probable state are are denoted by $\mathbf{P}^b_{mn}$ and $\bar{\mathbf{x}}^b_{mn}$, respectively.

We assume that, the rank of the $(2l + 1)^2u$ by $(2l + 1)^2u$ covariance matrix is much less than $(2l + 1)^2u$. Let

$$k = \text{rank}(\mathbf{P}^b_{mn}). \tag{1}$$

Thus $\mathbf{P}^b_{mn}$ has a $(2l + 1)^2u - k$ dimensional null space $\bar{\mathbb{S}}_{mn}$ and the inverse $(\mathbf{P}^b_{mn})^{-1}$ is defined for the component of the vectors $(\mathbf{x}^b_{mn} - \bar{\mathbf{x}}^b_{mn})$ lying in the k dimensional subspace $\mathbb{S}_{mn}$ orthogonal to $\bar{\mathbb{S}}_{mn}$

In the data assimilation procedure we describe in this paper, the background error covariance matrix $\mathbf{P}^b_{mn}$ and the most probable background state $\bar{\mathbf{x}}^b_{mn}$ are derived from a $k' + 1$ member ensemble of global state field vectors $\{\mathbf{x}^{b(i)}(\mathbf{r}, t)\}$, $i = 1, 2, \cdots, k' + 1$; $k' \geq k \geq 1$. The data assimilation algorithm is formulated in a k-dimensional space $\mathbb{S}_{mn}$ defined by the k principal components of the ensemble based estimate of $\mathbf{P}^b_{mn}$. The coordinates between the k-dimensional $\mathbb{S}_{mn}$ and the $(2l + 1)^2u$ dimensional local regions are changed by a $(2l + 1)^2u$ by k matrix, $\mathbf{Q}_{mn}$. We denote the projection of vectors into $\mathbb{S}_{mn}$ and the restriction of matrices to $\mathbb{S}_{mn}$ by a superscribed circumflex (hat). Thus for a $(2l + 1)^2u$ dimensional column vector $\mathbf{w}$, the vector $\hat{\mathbf{w}}$ is a k dimensional column vector given by

$$\hat{\mathbf{w}} = \mathbf{Q}^T_{mn}\mathbf{w}. \tag{2}$$

Let $\mathbf{x}^a_{mn}$ be the random variable at the current analysis time t representing the local vector after knowledge of the observations and background mean are taken into account. For simplicity, we assume that all observations collected for the current analysis were taken at the same time t. Let $\mathbf{y}^o_{mn}$ be the vector of current observations within the local region, and that the errors are normally distributed with covariance matrix $\mathbf{R}_{mn}$. An ideal (i.e., noiseless) measurement is a function of the true atmospheric state. Considering measurements within the local region (m, n), we denote this function $\mathcal{H}_{mn}(\cdot)$. That is, if the true local state is $\mathbf{x}^a_{mn}$, then the error in the observation is $\mathbf{y}^o_{mn} - \mathcal{H}_{mn}(\mathbf{x}^a_{mn})$. Assuming that the true state is near the mean background state $\bar{\mathbf{x}}^b_{mn}$, we approximate $\mathcal{H}_{mn}(\mathbf{x}^a_{mn})$ by linearizing about $\bar{\mathbf{x}}^b_{mn}$,

$$\mathcal{H}_{mn}(\mathbf{x}^a_{mn}) \approx \mathcal{H}_{mn}(\bar{\mathbf{x}}^b_{mn}) + \mathbf{H}_{mn}\Delta\mathbf{x}^a_{mn}, \tag{3}$$

where

$$\Delta\mathbf{x}^a_{mn} = \mathbf{x}^a_{mn} - \bar{\mathbf{x}}^b_{mn}, \tag{4}$$

and the matrix $\mathbf{H}_{mn}$ is the Jacobian matrix of partial derivatives of $\mathcal{H}_{mn}$ evaluated at $\bar{\mathbf{x}}^b_{mn}$. (If there are s scalar observations in the local $(2l + 1)$ by $(2l + 1)$ region at analysis time t, then $\bar{\mathbf{y}}^o_{mn}$ is s dimensional and the rectangular matrix $\mathbf{H}_{mn}$ is s by $(2l + 1)^2u$). The data assimilation step determines $\bar{\mathbf{x}}^a_{mn}$ (the *local analysis*) and $\mathbf{P}^a_{mn}$ (the *local analysis covariance matrix*) by minimizing the quadratic form,

$$\begin{aligned} J(\Delta\hat{\mathbf{x}}^a_{mn}) &= (\Delta\hat{\mathbf{x}}^a_{mn})^T(\hat{\mathbf{P}}^b_{mn})^{-1}\Delta\hat{\mathbf{x}}^a_{mn} \\ &+ (\hat{\mathbf{H}}_{mn}\Delta\hat{\mathbf{x}}^a_{mn} + \mathcal{H}_{mn}(\bar{\mathbf{x}}^b_{mn}) - \mathbf{y}^o_{mn})^T\mathbf{R}^{-1}_{mn} \times \\ &\quad (\hat{\mathbf{H}}_{mn}\Delta\hat{\mathbf{x}}^a_{mn} + \mathcal{H}_{mn}(\bar{\mathbf{x}}^b_{mn}) - \mathbf{y}^o_{mn}). \end{aligned} \tag{5}$$

Here $\hat{\mathbf{H}}_{mn} = \mathbf{H}_{mn}\mathbf{Q}_{mn}$ maps $\mathbb{S}_{mn}$ to the observation space.

For a fixed set of parameters, k, k', and l, the probability distribution of $\mathbf{x}^a$ (i.e., the ensemble $\mathbf{x}^{a(i)}(\mathbf{r}, t)$ can be estimated by the following serial algorithm (Figure 2.1):

1. The global analysis field is evolved to the next analysis time, thus obtaining a new background ensemble of global atmospheric states.

2. The global (in physical space) variables are partitioned to form the local input variables $\mathbf{y}^o_{mn}$, $\mathbf{R}_{mn}$, and $\mathbf{x}^{b(i)}_{mn}$ and the $(2l + 1)^2u \times (2l + 1)^2u$ outer product $\mathbf{P}^{b'}_{mn}$ is computed for each region.

3. The projection operator $\mathbf{Q}_{mn}$ is obtained and the local vectors are projected onto the k -dimensional subspace.

4. The local analysis covariance matrix is computed as

$$\hat{\mathbf{P}}^a_{mn} = \hat{\mathbf{P}}^b_{mn}\left[\mathbf{I} + \hat{\mathbf{H}}^T_{mn}\mathbf{R}^{-1}_{mn}\hat{\mathbf{H}}_{mn}\hat{\mathbf{P}}^b_{mn}\right]^{-1}. \tag{6}$$

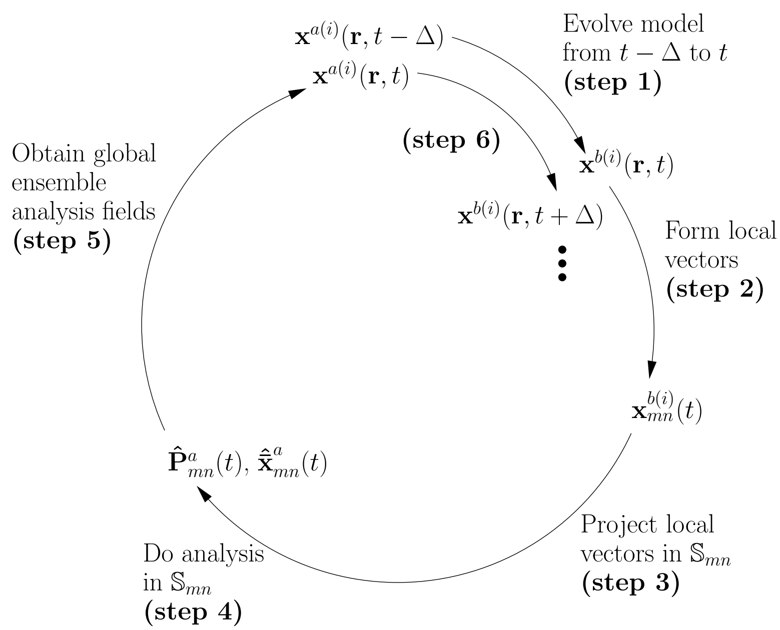

Figure 1: Illustration of the Local Ensemble Kalman Filter scheme as given by the six steps listed in the text

The local analysis (most probable state)

$$\bar{\mathbf{x}}^a_{mn} = \mathbf{Q}_{mn}\Delta\hat{\bar{\mathbf{x}}}^a_{mn} + \bar{\mathbf{x}}^b_{mn}, \tag{7}$$

is obtained from

$$\Delta\hat{\bar{\mathbf{x}}}^a_{mn} = \hat{\mathbf{P}}^a_{mn}\hat{\mathbf{H}}^T_{mn}\mathbf{R}^{-1}_{mn}\left[\mathbf{y}^o_{mn} - \mathcal{H}_{mn}(\bar{\mathbf{x}}^b_{mn})\right]. \tag{8}$$

and the local analysis ensemble, $\mathbf{x}^{a(i)}_{mn}i;\ i = 1, 2, \cdots, k' + 1$, is generated.

5. The $k'+1$ global fields $\left\{\mathbf{x}^{a(i)}(\mathbf{r}, t)\right\}$ are constructed from the $k'+1$ local analyses $\mathbf{x}^{a(i)}_{mn}$, which are given at each point $\mathbf{r}_{mn}$.

6. The procedure is repeated for the next analysis time.

This algorithm was first validated by an implementation on the Lorenz-96 model (Lorenz, 1996), which mimics the evolution of a scalar meteorological quantity along a latitude circle. We found that a full (non-localized) Kalman filter required an increasing number of ensemble members to reach a minimum rms error (about 20% of the observational noise for our experimental design), beyond which the analysis could not be reduced. The number of ensemble members needed was about 30, 50, and 80 for the 40, 80, and 120 grid point systems, respectively. The LEKF reached the same limit at 8 ensemble members independently of the system size (Ott, 2003a). We also found that the LEKF was able to handle data voids as efficiently as the full Kalman filter, and much more efficiently than a traditional data assimilation system using a static (constant) background matrix. In a later study (Ott, 2003b), we demonstrated that the LEKF scheme was robust both to errors in the Lorenz-96 model and to dynamical noise in the true state evolution.

2.2 Computational Implementation

The data assimilation begins with the construction of a vector containing suitably-scaled components of the dynamical variables at the grid points. Because we are not concerned with how they are generated, our algorithms are amenable to object-oriented programming techniques. The initial implementation, which we have used in our preliminary feasibility studies, contains a software layer called "Grid Manager" that handles the transport of data from the model grid to the data assimilation step. Grid Manager provides a consistent interface for the rest of the code. It comprises less than a third of the total lines of code, and it is the only portion of the software that needs to be altered when it is ported to a different model.

As is customary with high-performance modeling of this sort, the software is written in Fortran 95, which combines most of the necessary data-hiding and data-abstraction capabilities with simple, clean, high-performance language capabilities for numerical linear algebra, including parameterized floating-point types and two-dimensional arrays with column-oriented storage and variable dimensions, which are lacking in languages like C++. Our software also takes advantage of the code reuse that is possible in using well-tested, public-domain libraries like LAPACK.

Our system that can be easily adapted to upgraded or new versions of the forecast model, as neither our mathematical theory, nor the computer code we have been developing, depends on the details of any particular forecast model. This flexibility is a major advantage of the scheme over traditional 4D-Var schemes that require the modification of the tangent linear and adjoint model codes, whenever the model is modified. This is a critical advantage since numerous model upgrades can be expected during the long time span of THORPEX.

2.3 Results with the NCEP GFS

We carried out idealized data assimilation experiments with the T62 horizontal resolution, 28-level, 2001 version of the operational NCEP GFS. In our experiment, the "true" state was generated by a 30-day integration of the T62 GFS model, starting from the operational NCEP analysis at 0000 UTC on 1 January 2000. The observed data were generated by adding zero-mean, Gaussian random noise (observational error) to the true state. The variances of the errors are 1 K for the temperature, 1.1 m/s for the two horizontal components of the wind, and 1 hPa for the surface pressure.

As we expected, the scheme proved to be highly efficient. The assimilation of 1.5×10^6 simulated observations (wind, temperature, and surface pressure) takes about 12 minutes of wall-clock time on a Beowulf cluster of fourty 2.8-GHz Intel Xeon processors, a remarkably fast result for a reduced Kalman filter of this resolution. This result was achieved with local regions of about 750×750 km and $k = 79$ (an 80 member ensemble).

The timing results were obtained by observing all four variables at all grid points, in order to demonstrate the computational efficiency of our data assimilation system. The real strength of an ensemble Kalman filter, also demonstrated by our results for the Lorenz-96 model, is that it can efficiently extrapolate information to unobserved locations. To test the capability of the LEKF scheme to extract information from a reduced number of observations, we performed data assimilation experiments by gradually removing observations at random locations. Two examples of the results are shown in Figures 2 and 3, which demonstrate that the number of observational location can be reduced to 2000 (about 10% of the original number of locations) without losing the stability of the scheme and without considerably degrading the quality of the analyses. We found, as it also can be seen in our examples, that the analysis error is much smaller than the observational uncertainty for all variables at almost all atmospheric levels. The only exceptions are the temperature in the boundary layer and the wind components in the jet layer. These are the vertical layers, where the two variables have the largest gradient. To eliminate this problem, we plan to implement a vertical localization on the LEKF. We note that other teams developing ensemble Kalman filters also found the vertical localization of the covariances beneficial, but while they use the Gaussian filter of Gaspari and Cohn, we will use smoothly connected vertical layers in the true spirit of the LEKF.

3 Proposed Research

While our preliminary results with simulated observations are very promising, further significant theoretical and experimental research is needed before our methodology can be considered for operational implementation.

We are planning to test a number of enhancements to our system as part of the proposed research. The four most important of these are:

- Combination of the Ensemble Kalman Filter and the 4D-Var data assimilation

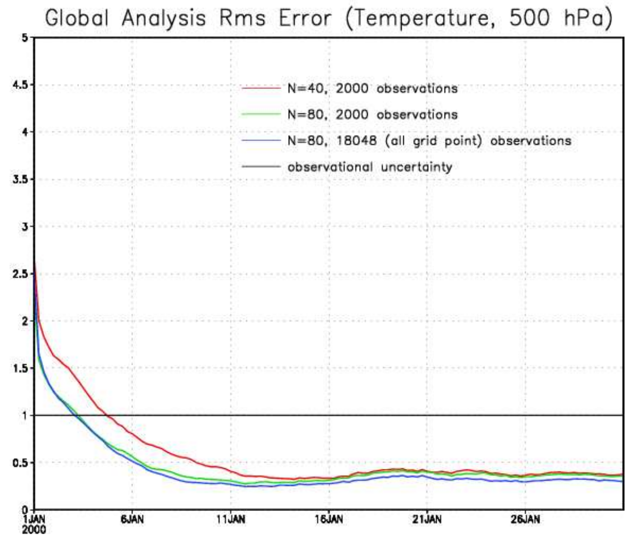

Figure 2: Time evolution of the global rms error in the analysis of the temperature at the 500 hPa pressure level. The areal average is taken over the entire globe.

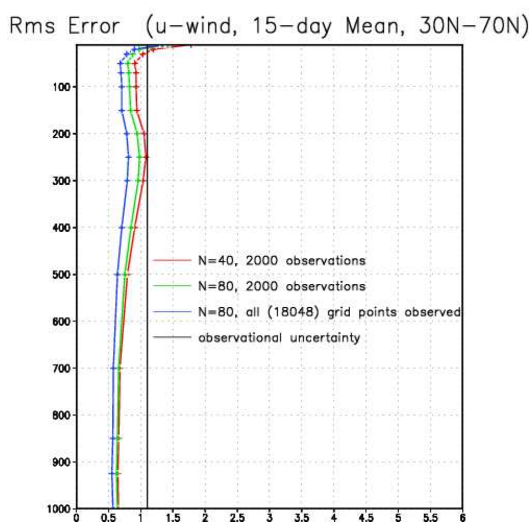

Figure 3: Vertical cross section of the rms error in the analysis of the zonal wind component at the 500 hPa level. The areal average is taken over the 30^0N-70^0N latitude belt.

approach. We propose to enhance the LEKF in such a way that asynchronous observations are assimilated at the correct time. This is essentially a 4D-Var extension of the LEKF, in which the linear model dynamics is inferred from the ensemble and not from the tangent-linear map, as it is done in a conventional 4D-Var scheme. Preliminary results with a low-order analogue of the atmospheric circulation (the 40-variable L96 model) show that this approach allows data assimilation every several model time steps to perform essentially as well as the original data assimilation every time step (Sauer et al., 2003).

- Direct minimization of the cost function. Ensemble Kalman Filters typically assume that the observational operator (which interpolates the background to the observational locations) is linear. For remotely-sensed data, however, the observational operator is often nonlinear. This issue is usually resolved by first linearizing the observational operator and then using it in the linear formulation of the problem. Our current implementation of the scheme follows this approach (see Eq. 8), though in our formulation of the Kalman Filter, the observational operator is allowed to be nonlinear in the cost function (see Eq. 5), and linearization of the operator is needed only to obtain the estimate of the analysis error covariance matrix (see Eq. 6), which is used for the generation of analysis ensemble perturbations. This allows for a direct minimization of the cost function including nonlinearity in the observational operator, which is expected to be computationally highly efficient for the small patches used in the LEKF. This may provide an important computational advantage over the current operational schemes, which seeks for the minimizer of the global cost function.

- Development of an additive variance inflation scheme to represent the effects of model errors on the analysis error covariance matrix in the NCEP GFS. A propotype of this scheme can be found in Ott (2003a), which was shown to be very effective in the L96 model, but an implementation on the NCEP GFS requires intense numerical experimentation.

- Development of an adaptive observation technique that can help improve forecasts into the medium and long forecast ranges (3-14 days). The main goal of the operational Winter Storm Reconnaissance (WSR) program is to improve the short range (1-3-day) prediction of extratropical cyclones. A detailed analysis of the forecast impact of the targeted observations, collected in the Winter Storm Reconnaissance programs, revealed that the added observations had an overall positive effect out to 7-day forecast lead time. It was found that this improvement was mainly due to an improvement in the prediction of the slowly varying, large-scale features of the atmospheric circulation (Szunyogh et al. , 2002), a finding also supported by the result of Miguez-Macho and Paegle (2001). We would like to capitalize on this observation by using diagnostic tools (e.g., Ensemble-dimension, local energetics) that can help maximize the beneficial effects of the adaptive observations into forecast ranges, where the error growth is non-linear. (Our most recent results on this subject can be found in Oczkowski, Szunyogh and Patil (2003)).

4 Outreach and Education

The University of Maryland Chaos Group has become one of the world's premier programs for training researchers in applied dynamical systems. In 1999, our graduate program in nonlinear dynamics was rated first in the country by *U. S. News and World Report* (since than nonlinear dynamics programs have not been rated). Our Meteorology Department and the Earth System Science Interdisciplinary Center (ESSIC) provide a unique environment for students working in earth sciences. Our Mathematics and Physics departments have proven records in promoting interdisciplinary interactions among students.

Currently, the U. S. has a lack of well-trained specialists in data assimilation at operational centers. In order to address this problem we have invested heavily in infrastructure to train students for careers in academia, industry, and government.

We strongly encourage our graduate students to use our large-scale computers (two Beowulf clusters, and an IBM SP2). This allows them to work on models of much greater sophistication and relevance than are typically available in graduate programs.

We will motivate interest from undergraduates in our proposed area through a mathematical modeling class taught by Dr. Hunt. The class is project based, and one area of focus will be interdisciplinary problems involving mathematics and the earth sciences.

5 Collaboration with NOAA scientists

We are planning to conduct our research in close collaboration with NCEP scientists. The close proximity of NCEP and the University of Maryland will allow us to have regular meetings with them. In fact, we have had several such meetings while developing the experimental version of the LEKF for the NCEP GFS. Our codes are built by using widely available standard mathematical packages and program libraries developed at NCEP. Since our codes input and output data in formats routinely used at NCEP, the transfer of our codes to NCEP for further testing and/or operational implementations, should be straightforward

6 Work Plan and Project Management

Our proposed research will consist of three main steps of enhancing the implementation of the LEKF on the NCEP GFS. We will also generate ensemble forecasts as part of each step, and the forecasts will be verified using a wide range of deterministic and probabilistic forecast scores. Our progress will also be tested by assessing the analysis and forecast impact of observations collected during THORPEX Observing-Systems Test (TOSTs) and THORPEX Regional Campaigns (TReCs) always using the most advanced version of our system.

Year 1 Our effort will be focused on implementing the 4D extension of the scheme and introducing the direct minimization of the cost function. Both new components will initially be validated by assimilating simulated observations. We will start assimilating real radiosonde and surface observations. This step will also include retrospective analysis/forecast experiments with dropsonde observations collected during operational Winter Storm Reconnaissance (WSR) field programs and THORPEX TOST and TReC field campaigns.

Year 2 Our efforts will be focused on improving the representation of model errors in the LEKF. While in the first year we are planning to use simple multiplicative and additive variancea inflation schemes to represent the effects of model uncertainties and dynamical noise on the analysis error covariance matrix; in the second year we expect to make several refinements to our variance inflation scheme. We expect that our final scheme will be an additive scheme that will inflate the variance in an adaptive way depending on the geographical location and the importance of the different phase space locations.

Year 3 Our efforts will be focused on the assimilation of remotely sensed observations. This would include generating locally high resolution "true" atmospheric states by integrating the NCEP Regional Spectral Model (RSM), implementing the latest version of the LEKF on the NCEP RSM, and running OSSE experiments to asses the analysis/forecast impacts of locally high resolution observational data sets. Intensive testing of the data assimilation system, hopefully including parallel runs with operational systems, would also take part in this year.

Project Management. The PI, Istvan Szunyogh, and co-PI Brian Hunt, will oversee the research project. They will be responsible for talking with the appropriate members of the research team on a regular basis, providing advice and direction.

NCEP is a 30-minute drive from the University of Maryland campus. The close proximity of the two institutions will allow for regular meetings between the research team and NCEP scientists to discuss the latest results and to help us maximize the value of our research products in supporting NOAA's mission within the THORPEX program . We will also be ready to help the transfer of our codes to NCEP's computer, once NCEP scientists feel that our product can be useful in supporting their mission.

7 Budget Justification

Our aforementioned progress has been made possible by one-time support from the W. M. Keck and McDonnell Foundations. From these grants, we have invested nearly $1.5 million into building a computational infrastructure that can support research with the NCEP GFS model, and into developing the prototype of the LEKF for the NCEP GFS. The support of the W. M. Keck Foundation has already ended and is nonrenewable. The support of the McDonnell Foundation will end in October 2005 and is much less than needed to support our interdisciplinary team. While we feel that the amount we request is small compared to what we have invested from non-federal funding sources, *the support of NOAA is absolutely necessary for us to be able to continue our work on further developing the LEKF for the NCEP GFS.*

The research team will be headed by **Istvan Szunyogh**, who is one of the main architects of LEKF data assimilation system for the NCEP GFS. He has extensive experience in operational numerical weather forecasting, especially in analyzing the impact of observations collected adaptively in field programs. He is one of the lead authors of several widely cited papers on the effect of targeted weather observations (Szunyogh et al., 1999a,b, 2000; Szunyogh et al. , 2002; Langland et al., 1999; Majumdar et al., 2001; Toth et al., 2000) and his research helped the operational implementation of the Winter Storm Reconnaissance program. He will devote approximately 33% of his time (4 months per year) to the project. The co-PI, **Brian Hunt**, is a mathematician and also a well-established expert on dynamical systems theory. He played a key role in working out the mathematical foundation of the LEKF scheme. Since he is a tenured associate professor, he will devote more than the 1 month per year to this project, for which we request funding. **Edward Ott** is a Distinguished University Professor, who is the top cited physicist in nonlinear dynamics and chaos. He developed the theoretical framework and worked out the mathematical details (with Hunt) of the LEKF algorithm. He is the lead author of our papers on the LEKF (Ott, 2003a,b). Since he is a tenured professor, he will devote more than the 0.5 month per year to this project, for which we request funding. **Eric Kostelich** is a Professor of Mathematics at the Arizona State University, who is a well known expert on dynamical systems theory and scientific computing. He is the main architect (with Szunyogh) of the LEKF code. He will play a key role in our further code development and numerical experimentation. He will spend spend one full month each year at the University of Maryland as a visiting professor.

We request funding for 1.5 month salary of the person who will manage our computer clusters used for the proposed research. We also request funding for the full-year support of a graduate student. Travel funds are requested to participate and present results at THORPEX meetings; to support one 2-week visit per year for Kostelich at the University of Maryland (in addition to his one month stay as a visiting professor) and a two-week visit by the PI at the Arizona State University. These visits will allow for intense discussions needed for planning code development and research papers. A modest fund is also requested for publication fees. Indirect cost is computed according to the rules of the University of Maryland.

References

J. L. Anderson, 1999: An Ensemble Adjustment Kalman Filter for Data Assimilation, *Mon. Wea. Rev.* **129**, 2884–2903.

C. H. Bishop and Z. Toth, 1999: Ensemble Transformation and Adaptive Observations, *J. Atmos. Sci.* **56**, 1748–1765.

R. Daley, 1999: *Atmospheric Data Analysis*. Cambridge University Press.

G. Evensen, 1994: Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics,

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Q. J. Roy. Meteor. Soc.,* **125,** 723-757.

T. M. Hamill and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme, *Mon. Wea. Rev.* **128**, 2905–2919.

P. L. Houtekamer and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Wea. Rev.* **126**, 796–811.

E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2003.

Langland, R. H., Z. Toth, R. Gelaro, I. Szunyogh, M. A. Shapiro, S. J. Majumdar, R. E. Morss, G. D. Rohaly, C. Velden, N. Bond, and C. H. Bishop, 1999: The North Pacific experiment (NORPEX-98): Targeted observations for improved North American weather forecasts.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability,* Vol. 1, ECMWF, Reading, Berkshire, UK, 1-18. *Bull. Am. Met. Soc.,***80,** 1363-1383.

Majumdar, S. J., C. H. Bishop, B. J. Etherton, Z. Toth, 2001a: Adaptive sampling with the Ensemble Transform Kalman Filter. Part II: Field program implementation. *Mon. WeaṘev.,* **130,** 1144-1165.

Majumdar, S. J., C. H. Bishop, I. Szunyogh, and Z. Toth, 20001b: Can an Ensemble Transform Kalman Filter predict reduction in forecast error variance produced by targeted observations? *Q. J. Roy. Meteor. Soc.,* (in press).

Miguez-Macho, G., and J. Paegle, 2001: Sensitivity of North American numerical weather prediction to initial state uncertainty in selected upstream subdomains. *Mon. Wea. Rev.,* **129,** 2005-2022.

Oczkowski, M., I. Szunyogh, and D. J. Patil, 2003: Scenarios for the development of locally low dimensional atmospheric dynamics. To be submitted. The latest version of the paper can be found online at `http://keck2.umd.edu/weather/weather_publications.htm`.

E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2003: A Local Ensemble Kalman Filter for Atmospheric Data Assimilation, submitted to *Monthly Weather Review*. The latest version of the paper can be found online at `http://keck2.umd.edu/weather/weather_publications.htm`.

E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2003: Estimating the state of large spatio-temporally chaotic systems" submitted to *Phys. Rev. Lett.* The latest version of the paper can be found online at `http://keck2.umd.edu/weather/weather_publications.htm`.

D. J. Patil, B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, "Local Low Dimensionality of Atmospheric Dynamics, *Phys. Rev. Lett.* **86**, 5878–5881.

Sauer, T., B. R. Hunt, J. A. Yorke......, 2003: Four-Dimensional Ensemble Kalman Filtering. The latest version of the paper is available at `http://math.gmu.edu/ tsauer/fourdvar`.

Szunyogh, I., Toth, Z., Majumdar, S. J., and Persson, A., 2002: On the propagation of the effect of targeted observations: The 2000 Winter Storm Reconnaissance Program. *Mon. Wea. Rev.* **130**, 1144-1165.

Szunyogh, I., Z. Toth, R. E. Morss, S. J. Majumdar, B. J. Etherton, and C. H. Bishop, 2000: The effect of targeted dropsonde observations during the 1999 Winter Storm Reconnaissance Program. *Mon. Wea. Rev.,* **128,** 3520-3537.

Szunyogh, I., Z. Toth, K. A. Emanuel, C. H. Bishop, C. Snyder, R. E. Morss, J. Woolen, and T. Marchok, 1999a: Ensemble-based targeting experiments during FASTEX: the effect of dropsonde data from the Lear jet. *Quart. J. Roy. Meteor. Soc.*, **125,**, 3189-3218.

Szunyogh, I., Z. Toth, S. J. Majumdar, R. E. Morss, C. H. Bishop and S. Lord, 1999b: Ensemble-based targeted observations during NORPEX. *Third Symposium on Integrating Observing Systems,* January, 10-15, 1999, Dallas, TX, AMS, 74-77.

Toth, Z., I. Szunyogh, S. J. Majumdar, R. E. Morss, B. J. Etherton, C. H. Bishop, S. J. Lord, M. Ralph, O. Persson and Z.-X. Pu, 2000: Targeted observations at NCEP: Toward an operational implementation. *Fourth Symposium on Integrating Observing Systems,* January, 10-14, 2000, Long Beach, CA, AMS, 186-193.

Whitaker, J. S., G. P. Compo, X. Wei, and T. H. Hamill, 2003: Reanalysis without radiosondes using ensemble data assimilation. Submitted to *Mon. Wea. Rev.*