DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Multi-Model Ensemble Approaches to Data Assimilation Using the 4D-Local Ensemble Transform Kalman Filter

Kayo Ide
University of Maryland
phone: (301) 405-0591 fax: (301) 314-9482 email: ide@umd.edu

Award Number: N000141010557
http://www.atmos.umd.edu/~ide

LONG-TERM GOALS

Uncertainties in the numerical prediction using a computational model of a physical system arise from two primary sources: i) errors within the model itself; and ii) imperfect knowledge of (a) the initial conditions to start the model and (b) boundary conditions and the forcing that is required to run the model. One way to examine these uncertainties is the multi-model approach, i.e., to compare results from multiple models. However, the multi-model approach cannot completely address either (i) or (ii) due to lack of knowledge of the real state. Another way is to compare the results with the "observations" that sample the real state. However, the observations introduce another source of uncertainties, i.e., iii) imperfect knowledge and/or improper assumptions within the observations including sampling error.

The ultimate objective of this project is to develop a framework for two purposes: one is the maximum reduction of the reducible uncertainties and the other is the diagnosis of the irreducible uncertainties in the numerical prediction. We will use a data-assimilation approach, which is ideal for this problem. Data assimilation is a method that was developed to primarily address the issues related to (ii-a) above by merging the observations into the numerical prediction. It attempts to optimally combine the "background"(or "forecast") information obtained by a short-term forecast using a numerical model with the observations taken within the forecast time window. The resulting state is the so-called "analysis", whose uncertainties are expected to be smaller than both the background and the observations. Some of these uncertainties are reducible by improving the data assimilation method. Other uncertainties are irreducible.

OBJECTIVES

To pursue our objectives, we integrate data assimilation into the multi-model approach. The Local Ensemble Transform Kalman Filter (LETKF) is our choice of the data assimilation method. Because it uses an ensemble to ESTIMATE the state uncertainties, it offers a perfect vehicle for the multi-model approach. In addition, a number of advantageous algorithms have been developed for the quantification and the reduction of uncertainties of all three types (i) - (iii), including both model-bias correction and observation-bias correction. Bias corrections, in a sense, transform part of the irreducible uncertainties (by other methods) into the reducible uncertainties. By integrating it into the multi-model approach, the LETKF will gain a powerful additional advantage: the combination of the ensemble weights and the calibration of the model will lead to improved performance over a single model LETKF. The resulting uncertainties are irreducible by the multi-model LETKF. We will extend the sensitivity diagnostics to examine the impact the observations and the background (forecast) in the uncertainty reduction using the multi-model LETKF. For improved and robust performance, we will also integrate the LETKF system into Variational approach.

APPROACH

To build the framework, we take hierarchical approach: first enhance individual elements and then integrate them consistently. The individual elements include improvement of existing methods, development of new methods, proof of concept of these methods using Observing System Simulation Experiments (OSSEs), and implementation to the real physical systems. The optimal framework for the ocean prediction system use sophisticated models such as Regional Ocean Model System (ROMS) and Geophysical Fluid Dynamics Laboratory Modular Ocean Model (MOM). For the development of the hybrid systems, we use the Simplified Parameterizations, primitivE-Equation Dynamics (SPEEDY) model.

WORK PURSUED

a. Development of new assimilation method without model

With the recognition of the importance of coastal oceans where the models have plenty of rooms to improve, we developed a new method for the detection of nonlinear instability in time series data without any physical model based on the concepts of time embedding and machine learning.

b. Coastal Ocean Data Assimilation

The LETKF has been interfaced with a ROMS implementation on the Chesapeake Bay (ChesROMS) as a first step towards a reanalysis and improved forecast system for the Chesapeake Bay. To account for forcing errors, a forcing ensemble is used to drive the ensemble states for the year 2003.

c. Global Ocean Data Assimilation

The LETKF has been implemented into the global MOM, along with two algorithms for enhanced performance, Incremental Analysis Update (IAU) and Running-in-Place (RIP). To evaluate the LETKF performance, Optimal Interpolation (OI) was also implemented. Results of the LETKF-IAU and LETKF–RIP are compared with the OI.

d. Hybrid data assimilation system

Hybrid data assimilation is the method of the choice for the operational numerical weather prediction at NOAA for its robust performance. To enhance the performance, a series of schemes are tested in the OSSEs setting using the SPEEDY. Particular topics are the dynamic constraints and variable localization (multivariate data assimilation) applied to LETKF.

RESULTS

a. Development of new data assimilation method without model

Applied in the reconstructed phase space based on time embedding, bred vectors have shown to correctly identify instabilities corresponding to those in the real, physical space (Fig.1). An effort is on going to develop a new data assimilation method based on the time-delay embedding without any physical model. Preliminary results are excellent, even outperforms the data assimilation in the physical space from the time series (Fig.2). Investigation is on going to better understand and further improve the new data assimilation method.

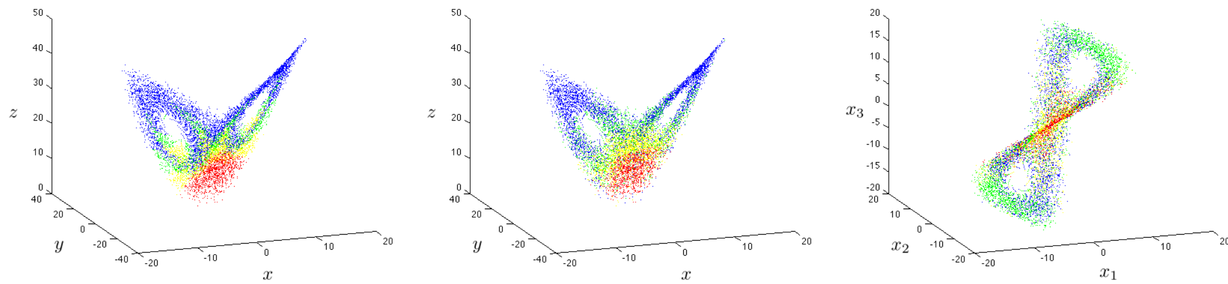

FIG. 1. (Color online) Growth rates of bred vectors in the Lorenz system using three different methods: (a) standard breeding in the phase space (x; y; z); (b) nearest-neighbor breeding in the phase space (x; y; z); and (c) nearest-neighbor breeding in the reconstructed phase space (x1; x2; x3). The colored points correspond to negative (blue), low (green), medium (yellow) and high (red) growth; see text for the value of the thresholds.

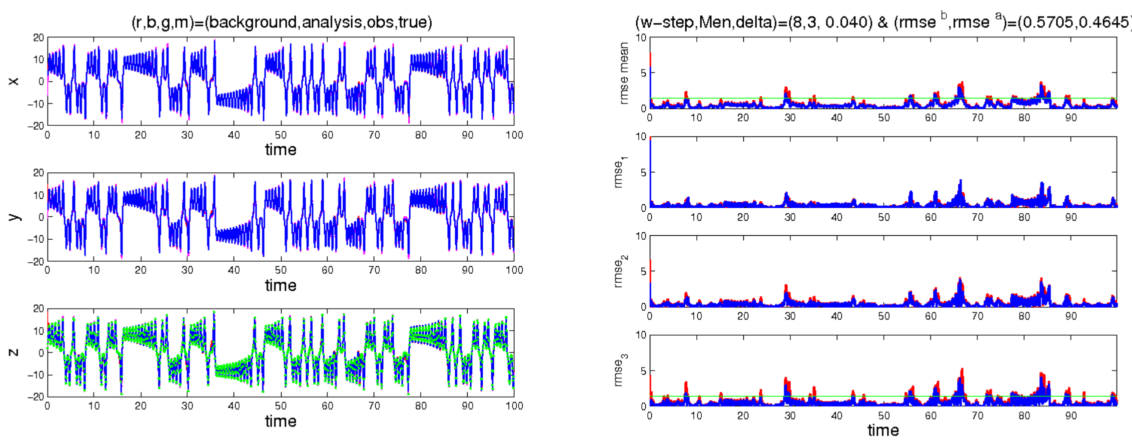

FIG.2. (Left) Forecast (red) and analysis (blue) using the most recent observed value (green) in the embedded space using the three-dimensional embedded space. (Right) corresponding root-mean square error.

b. Coastal Ocean Data Assimilation

In the observing system simulation experiments (OSSEs) using the ChesROMS-LETKF, the filter converges quickly and greatly reduces the analysis and subsequent forecast errors in the temperature, salinity, and current fields in the presence of errors in wind forcing. Errors in the Chesapeake Bay system are due more to errors in forcing than errors in initial conditions. To account or forcing errors, a forcing ensemble is used to drive the ensemble states for other years. A year-long model run for 2003 shows that the ChesROMS model captures the seasonal cycle of temperature and salinity in the Bay as well as many features of the Chesapeake circulation, but also contains significant bias and errors that can be corrected though assimilation. To investigate the contributions of errors from wind forcing and initial conditions, both of these fields were modified to produce free runs in the absence of data assimilation. When the only difference between the free run and the nature run is the initial condition, the free run converges to the nature run in approximately 2 - 4 weeks. This timescale gives some insight into the memory of the system for changes in initial condition and provides a crude bound for the limit for forecast improvements.

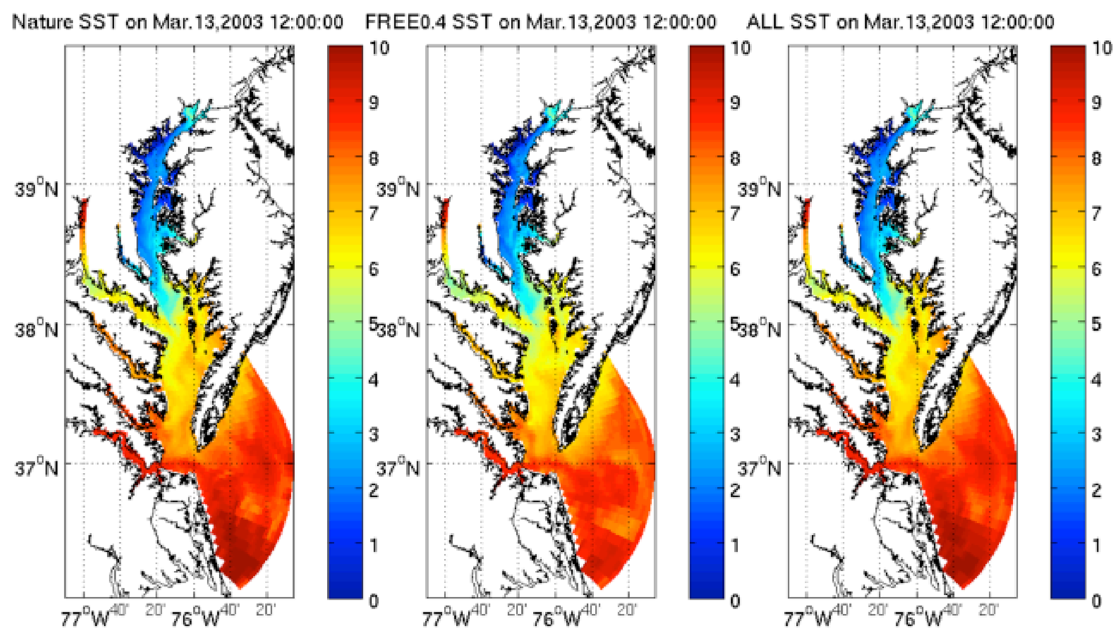

Figure 3. Comparison of Seat Surface Temperature for nature=truth (lef), without data assimlation (middle), and with data assimilation (right)

c. Global Ocean Data Assimilation

Examination of observation-minus-background differences shows a substantial reduction of errors as the data assimilation becomes more sophisticated (and computationally expensive) using LETKF-Running In Place (RIP). Reduction in the background error strongly implies a corresponding increase in the accuracy of the analysis. Root mean square difference (RMSD) is much smaller for RIP than for either Simple Ocean Data Assimilation or Incremental Analysis Update globally for temperature as well as salinity. Regionally the same results were found, with only one exception in which the salinity RMSD is slightly higher in the equatorial Pacific for RIP versus the other methods.

d. Global Atmospheric Data Assimilation

Hybrid assimilation methods have been proposed and developed to combine the advantages of the ensemble and variational methods. To enhance the ensemble effect, we developed a LETKF system with geostrophic constraints that eliminates the spurious correlation between uncorrelated variables, i.e., variable localization, even when only a small-size ensemble can be afforded because of the computational cost. Preliminary results clearly show the impact of both dynamic constraints and variable localization (Fig. 4).

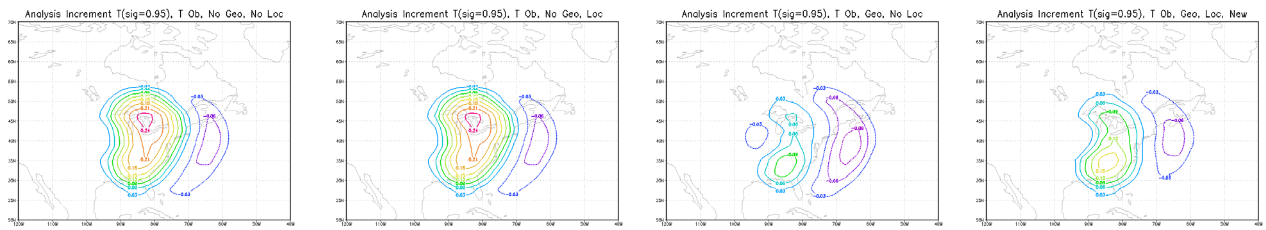

Figure43. Increment by single observation: left to right for (dynamic constraints, variable localization): (no, no), (no, yes), (yes, no), (no, no)

IMPACT/APPLICATIONS

The results from the LETKBF are encouraging and warrant further exploration of these assimilation techniques. EnSRF for complex system does not suffer from transient EC, and hence can be applied to realistic ocean data assimilation system. Development of data-based assimilation method will be advanced further to formulate a new type of hybrid data assimilation systems. This provides a new and novel way to model and predict sudden transitions in systems represented by time series data alone.

For the global ocean data assimilation, LETKF shows remarkable performance in comparison to the OI that have been used to produce the ocean reanalysis data sets. For coastal ocean data assimilation systems, the results from the OSSEs of the ChesROMS-LETKF system are promising for assimilating real data in the future. The LETKF is promising for the estimation of the uncertainty in the form of analysis error.

RELATED PROJECTS

N000140910418. Uncovering the Geometry of Ocean Flows and the Assimilation of Lagrangian Type Data

PUBLICATIONS

- Amezcua J., K. Ide, C. Bishop, E. Kalnay, 2012: Ensemble clustering in deterministic ensemble Kalman filter, Tellus A., 64, 18039, http://dx.doi.org/10.3402/tellusa.v64i0.18039
- de la Canara, A., A. M. Mancho, K. Ide, E. Serrano, C. R. Mechoso, 2012: Routes of Transport across the Antarctic Polar Vortex in the Southern Spring. *J. Atmos. Sci.*, **69**, 741-752. doi:DOI: 10.1175/JAS-D-11-0142.1

- Hoffman, M. J., T. Miyoshi, T. W. N. Haine, K. Ide, C. W. Brown, and R. Murtugudde, 2012: An advanced data assimilation system for the Chesapeake Bay: Performance Evaluation. *J. Atmos. Oceanic Tech.*, 29, 1542-1447.
- Norwoord, A., E. Kalnay, S.C. Yang, C. Wolfe, 2013: Lyapunov, singular and bred vectors in a multi-scale system: an empirical exploration of vectors related to instabilities. *J. Phys. A: Math. Theor.* 46 254021 doi:10.1088/1751-8113/46/25/254021
- Lynch, E., D. Kaufman, A.S. Sharma, K. Kalnay, and K. Ide, 2012: Breeding Vectors in the Phase Space Reconstructed from Time Series Data, Phys. Rev. submitted.
- Penny, S.G., E. Kalnay, J.A. Carton, B.R. Hunt, K. Ide, T. Miyoshi, G. Chepurin, 2013: The Running-in-Place algorithm applied to a Global Ocean General Circulation Model, Non. Process. In Geophys, sub judice.
- Santitissadeekorn, N., E. Spiller, C. Jones, R. Rutarindwa, L. Liu, K. Ide, 2013: Observing System Simulation Experiments of Cross-Layer Lagrangian Data Assimilation, *Dyn. Atmos. Ocean*