16. Data Assimilation Research Team

16.1. Team members

Takemasa Miyoshi (Team Leader)

Shigenori Otsuka (Postdoctoral Researcher)

Juan J. Ruiz (Visiting Researcher)

Keiichi Kondo (Student Trainee)

Yukiko Hayakawa (Assistant)

16.2. Research Activities

Data Assimilation Research Team (DA Team) was launched on 1 October 2012 and is composed of four research staff as of March 2013. Data assimilation is a cross-disciplinary science to synergize numerical simulations and observational data, using statistical methods and applied mathematics. As computers become more powerful and enable more precise simulations, it will become more important to compare the simulation with actual observations. DA Team performs cutting-edge research and development on advanced data assimilation methods and their wide applications, aiming at integrating computer simulations and observational data in the wisest way. Particularly, DA Team will tackle challenging problems of developing efficient and accurate data assimilation systems for high-dimensional simulations with large amount of data. The specific areas include 1) research on parallel-efficient algorithms for data assimilation with the super-parallel K computer, 2) research on data assimilation methods and applications by taking advantage of the world-leading K computer, and 3) development of most advanced data assimilation software optimized for the K computer.

In FY2012, we focused on 1) theoretical research on challenging problems in data assimilation, and 2) cutting-edge data assimilation research on meteorological applications. We have made substantial progress as follows:

1. An objective approach to model parameter estimation using data assimilation was investigated. (1 paper accepted)
2. Theories on discrete filtering to deal with model imperfections in ensemble-based data assimilation were explored.
3. A new approach to multi-scale covariance localization was invented and investigated.
4. The Local Ensemble Transform Kalman Filter (LETKF) system with a mesoscale numerical weather prediction model known as the Weather Research and Forecasting (WRF) model was ported to the K computer.

Main achievements are highlighted in the next section.

16.3. Research Results and Achievements

16.3.1. Model parameter estimation

In general, numerical simulation models contain a number of tuning parameters, and usually they are optimized manually and subjectively. Efforts have been made to make the tuning process objective and automatic, so that we find optimal parameters and make the simulation fit better to the observations. We have explored an approach to analyzing an augmented state vector consisting of both prognostic variables and model parameters through ensemble-based data assimilation. This allows estimating both prognostic state variables and model parameters simultaneously through data assimilation. Ruiz et al. (2013) explored the approach with an intermediate atmospheric general circulation model known as the SPEEDY model. Figure 1 shows the time series of three convective parameterization model parameters. Bad initial values converge to the right values (dashed lines) shortly after data assimilation both for temporally-fixed parameters (left panel) and for temporally-varying parameters (right panel).

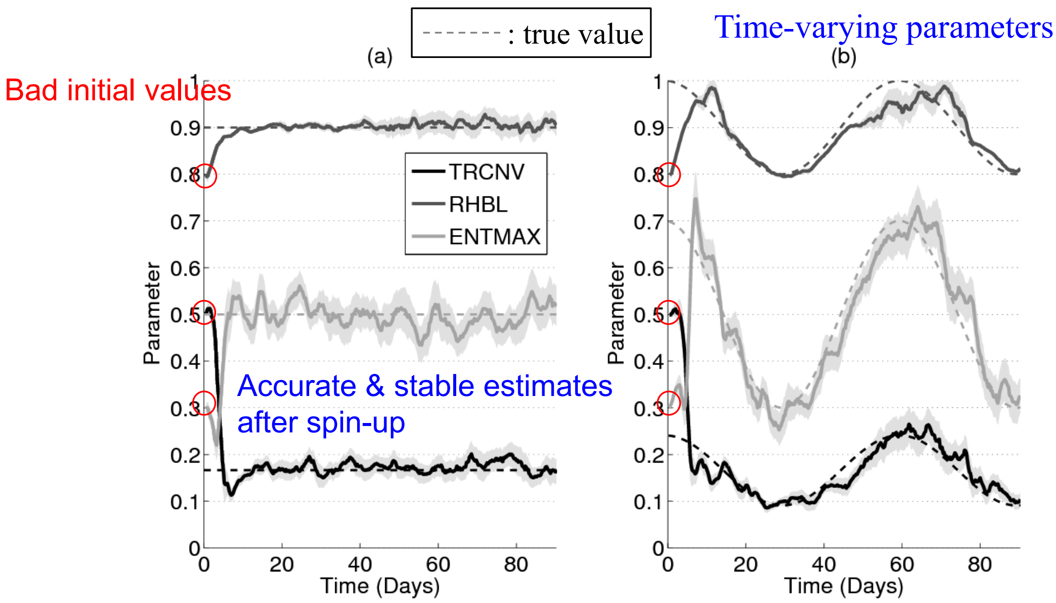

Fig. 1.

With the success from the idealized experiments with the SPEEDY model, we applied this approach to a real Typhoon case in 2008. Using the WRF-LETKF system (Miyoshi and Kunii 2012), air-sea exchange coefficients are estimated as a two-dimensional field. Figure 2 indicates estimated moisture exchange coefficients. The value 1 is the default; 1.5 and 0.5 correspond to 50% more and less effective air-sea exchange of the moisture fluxes, respectively. Over the Pacific, the estimated parameters suggest reducing moisture fluxes from the sea surface. As a result, biases are reduced, and Typhoon Sinlaku's forecast was improved.

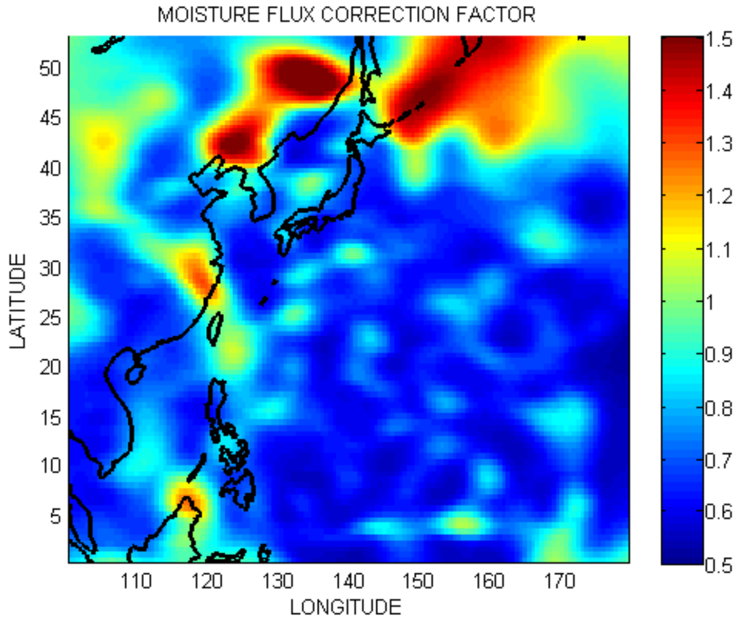

Fig. 2.

16.3.2 Theoretical development on multi-scale covariance localization

Ensemble-based data assimilation methods have been improved consistently and have become a viable choice in operational numerical weather prediction. A number of issues for further improvements have been explored, including flow-adaptive covariance localization and advanced covariance inflation methods. Dealing with multi-scale error covariance and model errors is among the unresolved issues that would play essential roles in analysis performance. With higher resolution models, generally narrower localization is required to reduce sampling errors in ensemble-based covariance between distant locations. However, such narrow localization limits the use of observations that would have larger-scale information. Previous attempts include successive covariance localization by F. Zhang et al. who proposed applying different localization scales to different subsets of observations. The method aims at using sparse radiosonde observations at a larger scale, while using dense Doppler radar observations at a small scale simultaneously. This study aims at separating scales of the analysis increments, independent of observing systems. Inspired by M. Buehner, we applied two different localization scales to find analysis increments at the two separate scales, and obtained astonishing improvements at all scales in simulation experiments using the SPEEDY model.

Figure 3 illustrates analysis increments from a single station observation at the star point. δx_h on the top right indicates the analysis increment from the raw ensemble perturbations with a narrower 500-km localization scale. This shows a small-scale feature near the observed location. δx_l on the bottom right indicates the analysis increment from a smoothed ensemble perturbations with a wider 1000-km localization scale. This shows smoother structure in the longer distance, although lacks structure in the shorter range. Merging δx_h and δx_l, although not the simple sum, gives the left panel, in which both smaller-scale structure in the shorter range and smoother structure in the longer range are preserved. With this new multi-scale localization method, we obtained large improvement mostly in the moisture and precipitation analyses.

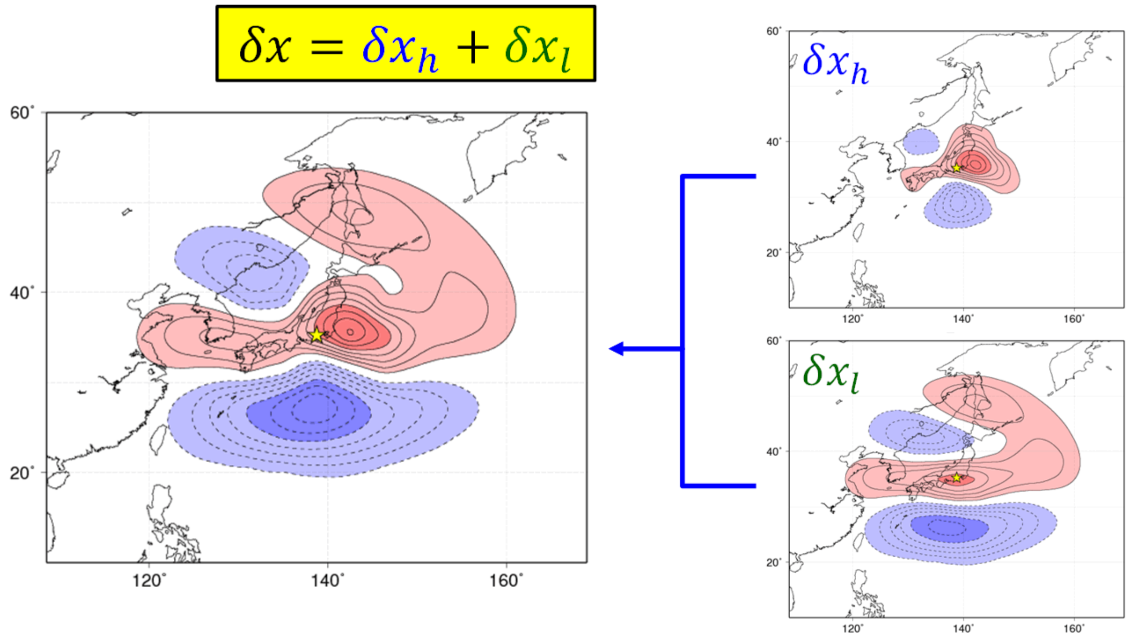

Fig. 3.

16.3.3 Porting the WRF-LETKF system to the K computer

For leading meteorological research using the K computer and for improving the computational efficiency of the widely-used LETKF code, we use the WRF-LETKF system (Miyoshi and Kunii 2012) as a testbed. SPIRE (Strategic Programs for Innovative Research) Field 3 performs research on super-high-resolution mesoscale numerical weather prediction using the LETKF, and DA Team aims at aiding their research through close collaboration on this LETKF code. As the first step, we ported the WRF-LETKF system on the K computer, and measured the current computational performance. Table 1 shows the computational time for a single-member, 6-hour WRF forecast for the default 60-km WRF-LETKF system, and its higher-resolution 20- and 5-km versions. Table 2 shows a similar table for the LETKF computations. These two tables suggest that the parallel efficiency is about 10% of the optimal for the 5-km experiment compared with the 60-km experiment. Our future work includes optimizing the parallel efficiency for the K computer, especially with larger ensemble sizes.

To confirm that the LETKF works properly on the K computer, 6-hour forecast fields from the LETKF analyses were investigated. Figure 4 shows 6-hour accumulated rain (shading, mm/6h) and mean-sea-level pressure (contours) for a low pressure system on 9 August 2008, after about a week spin-up of the LETKF data assimilation. The low pressure system is analyzed reasonably well for the three experiments at different resolutions. It is apparent that the 5-km experiment produces very fine structure of the rain bands associated with the low.

Table 1. Timing for the WRF forecasts.

Resolution	Number of nodes (CPU)	Max memory per node	Wall time (sec.)
60 km	13 (104)	0.914 GB	107
20 km	15 (120)	1.4 GB	452
5 km	50 (400)	6.3 GB	6320

Table 2. Timing for the LETKF analysis.

Resolution	Number of nodes (CPU)	Max memory per node	Wall time (sec.)
60 km	10 (80)	3.2 GB	130
20 km	40 (320)	4.4 GB	440
5 km	500 (4000)	8.4 GB	3000

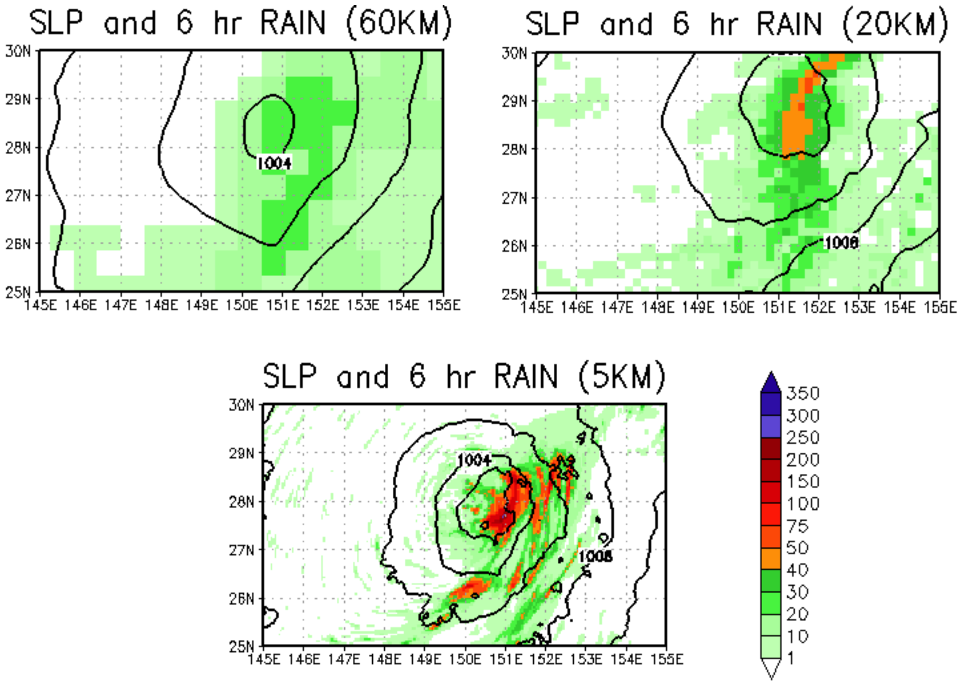

Fig. 4.

16.4. Schedule and Future Plan

Based on the achievements in FY2012, it is apparent that we need to optimize the LETKF code for the K computer, particularly for massively parallel computations. The current WRF-LETKF system does not perform well enough with the K computer, and algorithmic development will be necessary. Also, we will keep working on the fundamental theoretical problems of multi-scale and model-error treatments and of model parameter estimation. Besides, we will explore more theoretical aspects including non-linear and non-Gaussian treatments. These theoretical studies will improve the capability of ensemble-based data assimilation in wide applications. Moreover, we plan on seeking wider applications of data assimilation beyond geophysical applications. We will start working on exploratory investigations in FY2013.

16.5. Publication, Presentation and Deliverables

(1) Journal Papers

1. Greybush, S. J., R. J. Wilson, R. N. Hoffman, M. J. Hoffman, **T. Miyoshi**, K. Ide, T. McConnochie, and E. Kalnay, 2012: Ensemble Kalman Filter Data Assimilation of Thermal Emission Spectrometer (TES) Temperature Retrievals into a Mars GCM. *J. Geophys. Res.*, **117**, E11008. doi:10.1029/2012JE004097
2. Hoffman, M. J., **T. Miyoshi**, T. 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**, 1542-1557. doi:10.1175/JTECH-D-11-00126.1

3. Kunii, M. and **T. Miyoshi**, 2012: Including uncertainties of sea surface temperature in an ensemble Kalman filter: a case study of Typhoon Sinlaku (2008). *Weather and Forecasting*, **27**, 1586-1597. doi:10.1175/WAF-D-11-00136.1
4. Kang, J.-S., E. Kalnay, **T. Miyoshi**, J. Liu, and I. Fung, 2012: Estimation of surface carbon fluxes with an advanced data assimilation methodology. *J. Geophys. Res.*, **117**, D24101. doi:10.1029/2012JD018259
5. **Miyoshi, T.**, E. Kalnay, and H. Li, 2013: Estimating and including observation error correlations in data assimilation. *Inverse Problems in Science and Engineering*, **21**, 387-398. doi:10.1080/17415977.2012.712527
6. **Otsuka, S.**, S. Nishizawa, T. Horinouchi, and S. Yoden, 2013: An experimental data handling system for ensemble numerical weather predictions using a web-based data server and analysis tool "Gfdnavi". *J. Disaster Research*, **8**, 48-56.
7. **Ruiz, J. J.**, M. Pulido, and **T. Miyoshi**, 2013: Estimating model parameters with ensemble-based data assimilation: A review. *J. Meteorol. Soc. Japan*, **91**, in press.doi:10.2151/jmsj.2013-201

(2) Conference Papers

-None

(3) Invited Talks

1. **March 2013** Takemasa Miyoshi, "Challenges of Ensemble-based Data Assimilation for Large-Scale Simulations", Invited Presentation, The 3rd AICS International Symposium, RIKEN/AICS, Kobe, Japan.
2. **January 2013** Takemasa Miyoshi, "Advances and Challenges in Ensemble-based Data Assimilation in Meteorology", Invited Presentation, Third Data Assimilation Workshop, Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan.
3. **October 2012** Takemasa Miyoshi, "Observation Impact Estimates with an Ensemble-based Approach", Invited Presentation, International Symposium on Data Assimilation, German Weather Service (DWD), Offenbach, Germany.
4. **October 2012** Juan Ruiz, "Data Assimilation-based parameter estimation schemes", Exploring the Use of Data Assimilation Methods for the Detection and Attribution of Climate Change, Buenos Aires, Argentina.
5. **November 2012** Juan Ruiz, "Estimating model error with the ensemble Kalman filter", Invited Presentation, WCRP-SPARC Workshop, Buenos Aires, Argentina.

(4) Posters and presentations

1. **Miyoshi, T.***, **K. Kondo**, S.-C. Yang, and E. Kalnay: Spatial Structure of the LETKF Weights and Multi-scale Treatment in an EnKF. American Meteorological Society Annual Meeting, Austin, TX, USA, January 8, 2013.
2. **Ruiz, J. J.**, **T. Miyoshi***, M. Kunii, and M. Pulido: Self-optimization of Model Parameters with the LETKF: from Idealized Experiments to a Real-world Application. American Meteorological Society Annual Meeting, Austin, TX, USA, January 8, 2013.
3. **Miyoshi, T.*** and **K. Kondo**: Multi-scale Treatment in Ensemble Data Assimilation. Meteorological Research Institute, Tsukuba, Japan, February 13, 2013.
4. **Miyoshi, T.*** and **K. Kondo**: An approach to multi-scale localization. Nichii-gakkan, Kobe, Japan, March 21, 2013.
5. **Kondo, K.***, **T. Miyoshi**, and H. L. Tanaka: Multiscale localization in ensemble-based data assimilation. AICS International Workshop on Data Assimilation, RIKEN/AICS, Kobe, Japan, February 26-27, 2013.
6. **Ruiz, J. J.***, **T. Miyoshi**, and M. Kunii. Self-optimization of Model Parameters with the LETKF: a Real-world Application. AICS International Workshop on Data Assimilation, RIKEN/AICS, Kobe, Japan, February 26-27, 2013.
7. **Otsuka, S.***, N. J. Trilaksono, and S. Yoden: Statistics on Convections during the Jakarta Flood Event in 2007 Simulated by JMA-NHM. The 3rd AICS International Symposium, RIKEN/AICS, Kobe, Japan, February 28-March 1, 2013.

(5) Patents and Deliverables

The LETKF code is updated as needed and available at https://code.google.com/p/miyoshi/.