The computational aspect of the SCLAE-LETKF data assimilation system for rapid-update-cycle, high-resolution radar data assimilation

Guo-Yuan Lien, Seiya Nishizawa, Ryuji Yoshida, Hisashi Yashiro, Tatiana Martsinkevich, Takumi Honda, Shigenori Otsuka, Takemasa Miyoshi, Hirofumi Tomita, and Yutaka Ishikawa
RIKEN Center for Computational Science, Kobe, Japan

K computer
Computer simulations create the future

Introduction

- We have developed the **SCALE-LETKF** (Lien et al. 2017) regional data assimilation (DA) system for research.
  - It utilizes the Scalable Computing for Advanced Library and Environment-Regional Model (SCALE-RM; Nishizawa et al. 2015; Sato et al. 2015) model and the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al. 2007).
  - https://github.com/gylien/scale-letkf

Comparison of community DA code for research

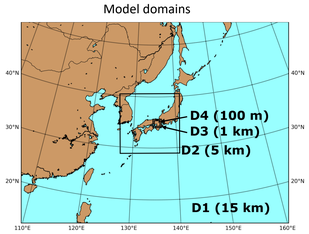
		WRFDA	Community GSI/EnKF	DART	SCALE-LETKF
Developer team		Mesoscale and Microscale Meteorology (MMM) Laboratory/ NCAR (U.S.)	Developmental Testbed Center (DTC)/ NCEP (U.S.)	Data Assimilation Research Section (DAReS)/ NCAR (U.S.)	RIKEN AICS
Assimilation algorithm		3DVar, 4DVar, Hybrid 3D/4DEnVar	3DVar, EnSRF, Hybrid 3D/4DEnVar	EAKF, EnSRF, Particle filter, … etc.	**LETKF**
Supported models	Global	--	GFS	MPAS, CESM, … etc.	--
	Regional	WRF	WRF	WRF, … etc.	**SCALE**
Supported observations		Conventional, satellite radiance, radar, GPSRO, … etc.	Conventional, satellite radiance, radar, GPSRO, … etc.	Conventional, radar, GPSRO, … etc.	Conventional, radar, satellite radiance (partial)

- We aim to use the SCALE-LETKF to conduct real-time, rapid-update-cycle, high-resolution assimilation of Phased Array Weather Radar (PAWR) data.
  - This requires very careful design in every part of the code to achieve high computational efficiency.

Targeted problem

- 30-second update cycle, 100-m model resolution, 100-member radar DA covering the radar observing range.

Model domains



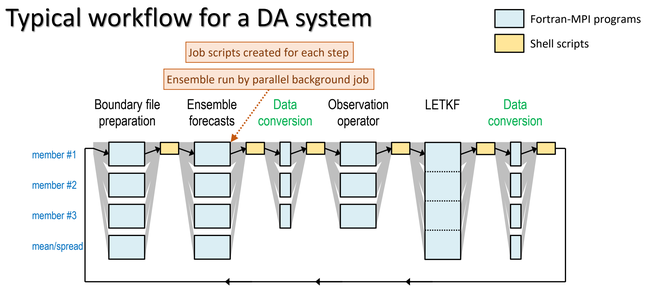
Highest resolution domain (D4)	
Model resolution	100 meters
Cycle length	30 seconds
Grid numbers	1200 x 1200 x 100
Meteorological variables	11
Ensemble members	100
Total data dimension (number of variables)	**1.6 x 10^{11}**
Total data size (Assume double-precision variables)	**1.3 TB**

K computer	
Total nodes	82,944
Cores / node	8
Memory / node	16 GB
Maximum disk space / node	29 GB

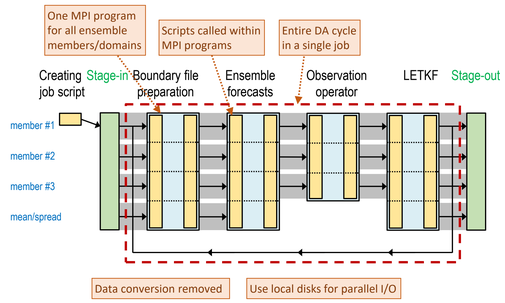
Observation	
Platform	Phased-Array Weather Radar (PAWR)
Highest resolution	100 meters
Frequency	30 seconds

Improvements in the workflow

Typical workflow for a DA system

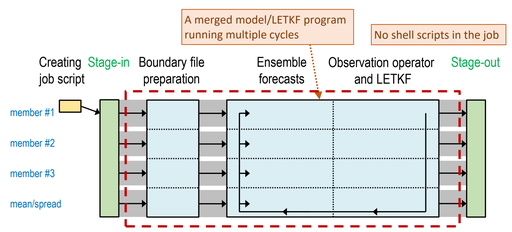


Improved workflow in the SCALE-LETKF

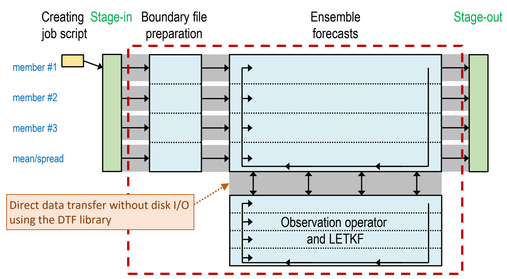


Specialized versions for real-time radar DA

A) Merged SCALE/LETKF program

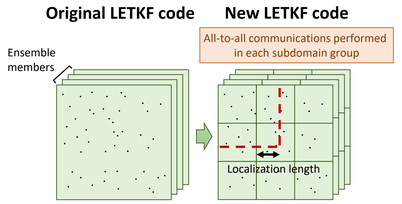


B) Direct data transfer between SCALE and LETKF using the DTF library *(in progress)*



Improvements in the LETKF program

- Employ domain splitting in the LETKF (similar to that usually done in the model):
  - Reduce the amount of all-to-all communications by maintaining the data locality, and also reduce the memory usage.
  - Similar to that done in Yashiro et al. (2016) for NICAM-LETKF.



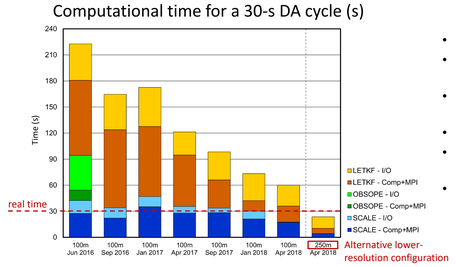
(One process is in charge of one subdomain for one or more member)

- Optimize the use of OpenMP.
- Implement MPI parallelization also in the observation operator calculation.
- Improve observation sorting and implement a fast observation number limit (Hamrud et al. 2015) scheme.
- Implement online, fast superobing in the LETKF. *(in progress)*

Results: Computational time

Progress we've made for the targeted problem

Computational time for a 30-s DA cycle (s)

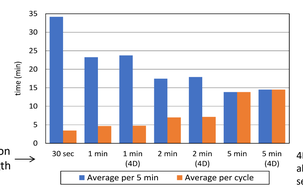


- Jun 2016 → Sep 2016: Merge OBSOPE and LETKF into a single program.
- Sep 2016 → Jan 2017: In LETKF, use QUICKSELECT algorithm for observation selection for the number limit.
- Jan 2017 → Apr 2017: Implement deterministic run (Schraff et al. 2016) and distribute the I/O of mean/spread files in a more parallel way.
- Apr 2017 → Sep 2017: Optimize the OpenMP directives.
- Sep 2017 → Jan 2018: Remove most run-time shell scripts and improve the observation processing code
- Jan 2018 → April 2018: Run multiple cycle using a merged SCALE/LETKF program and thus use direct memory copy to exchange data between the SCALE model and LETKF.

Computational time and forecast skill with respect to different cycle lengths

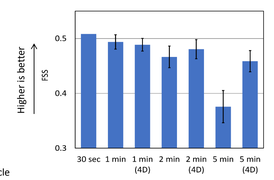
(1-km resolution; 208 nodes on the K computer)

Computational time (model + DA)



4D: 4D-LETKF assimilating all asynchronous (every-30-second) observations in a cycle

Forecast skill: Mean FSS in the 10- to 30-min forecasts



(18-fcst average)
Error bars: 95% confidence interval compared to 30-sec

Summary

- We have made remarkable progress in the code development to improve the computational efficiency and speed towards the goal of real-time experiments.
  - The 100-m resolution, 30-second-update configuration is still slower than the real time, while the 250-m resolution configuration can meet the goal.
- We believe that the SCALE-LEKTF system has become a useful research tool for broad mesoscale data assimilation studies.

Data Assimilation Team
RIKEN R-CCS 理化学研究所 計算科学研究センター RIKEN Center for Computational Science

This study was supported by CREST, JST and used computational resources of the K computer provided by the RIKEN Center for Computational Science (R-CCS) (Project ID: hp150019, hp160162, hp170178, hp180062, and ra000015).
It was also supported by FOCUS Establishing Supercomputing Center of Excellence, the FLAGSHIP2020, MEXT within the priority study4 (Advancement of meteorological and global environmental predictions utilizing observational "Big Data"), and JSPS KAKENHI Grant Number 17K14396.