Local Ensemble Transform Kalman Filter Experiments with the Nonhydrostatic Icosahedral Atmospheric Model NICAM

Koji Terasaki1, Masahiro Sawada2, and Takemasa Miyoshi1,2,3
1RIKEN Advanced Institute for Computational Science, Kobe, Japan
2University of Maryland, College Park, Maryland, USA
3Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
4Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan

Abstract

The Local Ensemble Transform Kalman Filter (LETKF) is implemented with the Non-hydrostatic Icosahedral Atmospheric Model (NICAM) to assimilate the real-world observation data. First, the NICAM-LETKF system was developed using grid conversions between the NICAM’s icosahedral grid and LETKF’s uniform longitude-latitude grid to take advantage of the existing codes of Miyoshi. The grid conversions require additional computations and may cause additional interpolation error. Therefore, the LETKF code is modified, so that the LETKF reads and writes the NICAM’s icosahedral grid data directly. We call this new version ICO-LETKF. In this study, the two systems are tested and compared using real conventional observations. The results show that the ICO-LETKF successfully accelerates the computations and improves the analyses.

(Citation: Terasaki, K., M. Sawada, and T. Miyoshi, 2015: Local ensemble transform Kalman filter experiments with the non-hydrostatic icosahedral atmospheric model NICAM. SOLA, 11, 23–26, doi:10.2151/sola.2015-006.)

1. Introduction

Data assimilation plays a key role in numerical weather prediction (NWP). Hunt et al. (2007) proposed the local ensemble transform Kalman filter (LETKF) as an advanced data assimilation method, and a number of studies have applied the LETKF to NWP and other geophysical systems (e.g., Miyoshi and Yamane 2007; Szunyogh et al. 2007; Hoffman et al. 2010; Miyoshi and Kunii 2012). The LETKF is originated from the local ensemble Kalman filter (LEKF; Ott et al. 2004), but applies the ensemble update of the ensemble transform Kalman filter (ETKF; Bishop et al. 2001) for more computational efficiency.

The Non-hydrostatic Icosahedral Atmospheric Model (NICAM) has been developed by the University of Tokyo, Japan Agency for Marine-earth Science and Technology (JAMSTEC), and RIKEN Advanced Institute for Computational Science (Tomita and Satoh, 2004; Satoh et al. 2008; Satoh et al. 2014). The NICAM is based on an icosahedral grid system. An icosahedron has 20 triangles, corresponding to 20 grids of the NICAM’s 0th grid division level. To increase the resolution, each triangular grid is divided into four triangles (cf. Fig. 1 of Satoh et al. 2005). Kondo et al. (2009) pioneered to implement the LETKF with NICAM and performed simulation experiments under the perfect model scenario. They simulated a synthetic, idealized observing system and did not use a realistic observation operator which converts model variables to an observed variable. Since the observations are simulated on the NICAM’s icosahedral grid, it was not necessary to use a longitude-latitude grid in their system. This study aims to develop a new NICAM-LETKF system including realistic observation operators and to assimilate the real observation data such as radiosondes. To the best of the authors’ knowledge, this is the first attempt to assimilate real observations with NICAM.

This study takes advantage of the existing LETKF code by Miyoshi (2005) and follow-on studies (e.g., Miyoshi and Yamane 2007; Miyoshi et al. 2007; Miyoshi 2011). The existing code includes realistic observation operators and has been tested with several different models such as the operational Japan Meteorological Agency (JMA) global model (Miyoshi et al. 2010), Atmospheric General Circulation Model (AGCM) for the Earth Simulator (AFES, Miyoshi and Yamane 2007; Miyoshi et al. 2007a), and Weather Research and Forecasting (WRF) model (Miyoshi and Kunii 2012), all of which are based on the longitude-latitude grid. The JMA global model and AFES are based on the Gaussian grid, and Miyoshi and Kunii (2012) employed the longitude-latitude grid for the WRF model while the model itself has several choices for map projection. However, the NICAM is based on the icosahedral grid. This study develops two versions of NICAM-LETKF: one with the grid conversions between the icosahedral and longitude-latitude grids to take advantage of the existing LETKF code (LL-LETKF, Fig. 1a), and the other without the grid conversions (ICO-LETKF, Fig. 1b). To remove the grid conversions, the existing LETKF code is modified so that it reads and writes the NICAM’s icosahedral grid directly. By removing the grid conversions, we would expect two advantages: saving the computer time and analysis accuracy for both systems and aims to address the scientific question of whether or not ICO-LETKF actually outperforms LL-LETKF in terms of computer time and analysis accuracy using the real conventional observations.
2. NICAM-LETKF

In this study, we implement new NICAM-LETKF systems with realistic observation operators based on the existing LETKF code by Miyoshi (2005) and follow-on updates. The existing LETKF code is written in FORTRAN90 and Message Passing Interface (MPI) and assumes the longitude-latitude grid for the global atmosphere. The LETKF computations are independent at each grid point. Therefore, the same LETKF core part is shared in the two NICAM-LETKF systems; the only difference is the I/O interface.

In LL-LETKF (Fig. 1a), the first guess from the NICAM output should be converted from the icosahedral grid to the uniform longitude-latitude grid at a 1.125° × 1.125° resolution and LETKF assimilation is performed in the longitude-latitude grid. The resulting analysis ensemble output is converted from the longitude-latitude grid to the icosahedral grid. Here, the grid-conversion tool included in the NICAM source code package is adopted. The grid conversion is performed based on the simple bilinear interpolation at each vertical level separately. Surrounding grid points are used for converting from the icosahedral grid to the longitude-latitude grid. From the longitude-latitude grid to the icosahedral grid, 4 surrounding grid points are used.

The observation operator is also modified for ICO-LETKF. The observation operator converts the model gridded data to the observed variable by variable conversions and spatial interpolation. The LL-LETKF system treats the longitude-latitude grid, so that we can keep using the existing observation operator. Here, the simple trilinear interpolation in 3-dimensional space using 8 surrounding grid points is applied. By contrast, the ICO-LETKF observation operator reads the icosahedral gridded data directly and uses the interpolation code included in the NICAM source code package. Here, the piecewise linear interpolation over triangles in 3-dimensional space using 6 surrounding grid points is applied.

The neighboring search algorithm within the LETKF also needs to be considered. When the LETKF analyzes a grid point, only neighboring observations are assimilated by applying a nearest-neighbor search based on a uniform longitude-latitude grid (Fig. 2a). The search area (green circle in Fig. 2a) has the radius of 2√N/3 times the localization scale σ, and the observations only within the search area are assimilated. Here, the observation data are sorted horizontally from the southwest corner to the northeast, and the accumulated numbers of the observations at each grid box are counted (cf. Fig. 1 of Miyoshi et al. 2007b). These counts serve as the pointers to the starting address of the observation data array. When we analyze a grid point (purple circle in Fig. 2a), we first select the observations around the grid point within the red box defined as the minimum square containing the search area (green circle) in Fig. 2a. Using the pointers accelerates to find the corresponding memory space within the red box. Next, the distances of all observations in the red box from the analyzing grid point (purple circle) are computed for LETKF computations. We can simply extend this search algorithm for the icosahedral grid system (Fig. 2b). To find the observations within the green circle in Fig. 2b, we can apply the virtual uniform longitude-latitude grid (orange circles) on top of the icosahedral grid (red circles). Figure 2c highlights the virtual grid for the search purpose, and we can simply apply the same neighboring search algorithm as in Fig. 2a. In this study, the resolution of the virtual uniform longitude-latitude grid is chosen to be 2.25° × 2.25°, which covers the whole globe. The virtual grid is used only for the neighboring search; the resolution affects only computational efficiency, not the search results or the eventual analysis.

3. Experiments

3.1 Experimental settings

The ensemble size is fixed at 20 in this study. The 20 initial conditions are chosen from the National Centers for Environmental Prediction (NCEP) Final Analysis (FNL) at 0000 UTC 1 November every year in the 20-year period from 1992 to 2001. These 20 states are spun-up for 2 days using NICAM with the grid division level 6 (approximately 112-km resolution) and 40 vertical levels. The model top is chosen to be 40 km (about 3 hPa). The data assimilation experiments are performed for a month in November 2011. Covariance localization is applied with the Gaussian function with σ = 400 km in the horizontal and 0.2 ln p in the vertical, but replaced by zero beyond 2√N/3 σ, where p is atmospheric pressure. The covariance inflation parameters are estimated adaptively at each grid point (Miyoshi 2011).

The NCEP PREPBUFR (available at http://rda.ucar.edu/data sets/ds337.0) observation data including radiosondes, wind profilers, aircraft reports, surface pressure, atmospheric motion vectors (AMVs) and sea surface winds including scatterometers are assimilated every 6 hours. The observation data within a 1-hour window from 30 minutes before to 30 minutes after the analysis time are assimilated. The observation error standard deviations included in the NCEP PREPBUFR data are used. Observations are rejected if the difference from the first guess is more than ten times of the observation error standard deviation. Figure 3 shows the observation coverage map at 1200 UTC 1 November 2011.

3.2 Timing results

ICO-LETKF removes the grid conversions and is expected to accelerate the computation. The timing for the NICAM ensemble forecasts is about 40 seconds on average using 800 nodes of the K computer and is completely identical for both systems. For data assimilation, 200 nodes of the K computer are used; the peak performance is 22.7 TFLOPS. Table 1 shows the computation time for LL-LETKF and ICO-LETKF. The timings for the observation operator and LETKF are similar in LL-LETKF and ICO-LETKF, although we find slight improvements in ICO-LETKF. The largest impact is found in the grid conversions. ICO-LETKF saves the entire 28.7 seconds for the grid conversions. Overall, ICO-LETKF accelerates the computations by about 40% compared to LL-LETKF.

Fig. 2. Schematic showing the observation neighboring-search problem based on (a) the longitude-latitude grid, (b) and (c) the icosahedral grid. Blue and red circles represent model grid points. Yellow stars represent observations. The purple circles are the analyzing grid point, and the surrounding green open circles represent the virtual uniform longitude-latitude grid points. The red boxes are defined as the minimum squares containing the search area.
3.3 Analysis accuracy

To find the exact impact of the grid conversions on the analyses, the differences of the analysis ensemble mean and spread after the first assimilation cycle between LL-LETKF and ICO-LETKF are shown in Fig. 4. The input first guess ensemble fields and adaptive inflation fields are exactly identical for the first assimilation step, so that the resulting differences are purely due to the grid conversions. The largest difference of the analysis ensemble mean appears over the Tibetan plateau (Fig. 4a). Generally fewer observations over the ocean may cause smaller differences of the analysis fields. We find more differences in the mid latitudes. As for the ensemble spread (Fig. 4b), we find generally smaller differences mostly in warm colors, which correspond to the larger spread of ICO-LETKF than LL-LETKF. Relatively large differences appear around the southern Greenland and near Alaska. The large differences may be caused by different quality control results. If LL-LETKF assimilates an observation but if ICO-LETKF rejects it, ICO-LETKF would have the larger ensemble spread. The LETKF systems include a commonly-used gross error check, in which observations are rejected when the observation-minus-first-guess value is greater than a prescribed threshold, in this study, 10 times the observation error standard deviation. The initial conditions for the first assimilation step are chosen from the previous 20-year atmospheric analyses and do not correspond to the atmospheric state of the day. This makes generally large observation-minus-first-guess values at the first assimilation step, and more observations would be closer to the gross-error threshold. Figure 4b shows the 500-hPa level, but we find generally small differences of the ensemble spread at different vertical levels. We also find occasional large differences but not necessarily in warm colors. The large differences would likely be related to the different gross-error check results between ICO-LETKF and LL-LETKF. In fact, the number of assimilated observation are 28,719 and 29,577 in LL-LETKF and ICO-LETKF, respectively, out of total 30,741 observations for the first assimilation step.

Figure 5 shows the time series of the root mean square differences (RMSD) for temperature (K) relative to the ERA-Interim. The RMSDs in both systems drop quickly for the first week. After the first week, ICO-LETKF shows the lower RMSD than LL-LETKF by about 0.2 to 0.4 K in the troposphere (700–200 hPa, Fig. 5c), about 10% improvement of analysis accuracy. In the upper levels beyond 50 hPa, the RMSDs decrease a little during the first week but remain at much higher values than in the troposphere in both systems. This is mainly because the conventional observations are very sparse at the upper levels where satellite radiances provide much information. Also, the model top height of NICAM is 40 km or about 3 hPa, much lower than 0.1 hPa of the ERA-Interim. Although only temperature is shown in Fig. 5, other variables (zonal and meridional winds and water vapor) also show general improvements in the troposphere (not shown). These improvements would be caused by reducing the interpolation error due to the grid conversions in LL-LETKF.

Table 1. Computational time (seconds) averaged over 10 assimilation cycles for each part of the LL-LETKF and ICO-LETKF systems using 200 nodes of the K computer.

<table>
<thead>
<tr>
<th></th>
<th>LL-LETKF</th>
<th>ICO-LETKF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid conversions</td>
<td>28.7</td>
<td>–</td>
</tr>
<tr>
<td>Observation operator</td>
<td>17.1</td>
<td>18.0</td>
</tr>
<tr>
<td>LETKF</td>
<td>35.1</td>
<td>31.2</td>
</tr>
</tbody>
</table>

Fig. 4. Differences of (a) the analysis mean field and (b) the analysis spread of 500-hPa temperature between LL-LETKF and ICO-LETKF after the first assimilation cycle. Warm colors correspond to the larger values of ICO-LETKF than LL-LETKF.

Fig. 5. Time-series of the global-mean root mean square differences (RMSD) for temperature (K) relative to the ERA-Interim. (a) LL-LETKF, (b) ICO-LETKF, and (c) the difference between LL-LETKF and ICO-LETKF. The horizontal and vertical axes represent the date and pressure, respectively. Negative values corresponds to ICO-LETKF’s advantage.
writes the NICAM’s icosahedral grid data directly (ICO-LETKF). The grid conversions include the interpolation and require additional computations. Also, the interpolation error would be an additional error source. The results with the LL-LETKF and ICO-LETKF systems showed overall computational acceleration by 40% and the analysis error reduction by 10% due to removing the grid conversions. Therefore, ICO-LETKF is the choice for the future developments. This study used only the NCEP PREPBUFR data. It is an important future direction to include satellite radiances and other satellite data such as the recently-launched GPM (Global Precipitation Measurement) satellite precipitation products.

Acknowledgements

The authors thank Masaki Satoh of the University of Tokyo, Hisashi Yashiro and Hirofumi Tomita of RIKEN AICS for fruitful discussions. This study was partly supported by JAXA/PMM.

References


4. Summary

This study developed the two versions of the LETKF with NICAM and tested them with the real NCEP PREPBUFR observations. One system includes the grid conversions between the NICAM’s icosahedral grid and the longitude-latitude grid (LL­LETKF). The other removes the grid conversions and reads and