Using Observations near the Poles in the AFES-LETKF Data Assimilation System

Akira Yamazaki¹, Takeshi Enomoto¹,², Takemasa Miyoshi¹,³,⁴, Akira Kuwano-Yoshida¹, and Nobumasa Komori¹

¹Application Laboratory, JAMSTEC, Yokohama, Japan
²Disaster Prevention Research Institute, Kyoto University, Uji, Japan
³RIKEN Advanced Institute for Computational Science, Kobe, Japan
⁴Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA

Abstract

The observation operators in the local ensemble transform Kalman filter (LETKF) were improved to enable use of observations in the vicinity of the poles in the data assimilation system composed of the atmospheric general circulation model for the Earth Simulator (AFES) and the LETKF. The improved observation operators allow to assimilate the observations located south (north) of northernmost (southernmost) Gaussian grid latitudes. An algorithm for searching the nearest observations from an analyzed grid for error covariance localization was also modified to efficiently assimilate observations near the poles.

The new algorithms were incorporated into the LETKF, and the impacts of routine radiosonde observations at the South Pole during the periods of July 2012 and January 2013 were assessed. The radiosonde observations suppressed an artificial expansion of the analysis ensemble spread which occasionally caused numerical instability in the upper troposphere and the lower stratosphere over the Antarctic regions. The analysis was also improved in the Antarctic regions.

(Citation: Yamazaki, A., T. Enomoto, T. Miyoshi, A. Kuwano-Yoshida, and N. Komori, 2017: Using observations near the poles in the AFES-LETKF data assimilation system. SOLA, 13, 41–46, doi:10.2151/sola.2017-008.)

1. Introduction

The local ensemble transform Kalman filter (LETKF, Hunt et al. 2007; Miyoshi and Yamane 2007) is a kind of the ensemble Kalman filter (EnKF), an advanced data assimilation (DA) technique. The LETKF can be implemented independently of the model, is suitable for ensemble forecasting and is efficient for parallel computing. Recently, the LETKF has been implemented with various models such as the global and regional atmosphere (e.g., Miyoshi and Aranami 2006; Miyoshi et al. 2010; Miyoshi and Kunii 2012; Terasaki et al. 2015), global and coastal ocean (Hoffman et al. 2008; Penny et al. 2013) and Martian atmosphere (Hoffman et al. 2010; Grebush et al. 2012).

The LETKF and the atmospheric general circulation model (AGCM) for the Earth Simulator (AFES) compose the AFES-LETKF ensemble DA systems named ALEDAS and ALEDAS2 (ALEDAS version 2) which have produced the ensemble atmospheric reanalysis datasets, ALERA (AFES-LETKF experimental ensemble reanalysis, Miyoshi et al. 2007a) and ALERA2 (Enomoto et al. 2013), respectively. Hereafter we do not distinguish their ensemble reanalysis, Miyoshi et al. 2007a) and ALERA2 (Enomoto et al. 2013). Miyoshi et al. (2007b, hereafter MYE07) updated the localization algorithm so that localization is more natural with the equal-distance circle; the current version searches nearby observations in circular areas with a constant radius defined by physical distances. However, even in the current algorithm, the search areas are slightly distorted near the poles (see Section 2.2 for more details). Therefore, this study aims to revisit the nearest-neighbor localization algorithm to avoid the distortion.

Another motivation to assimilate observations near the poles is to avoid an anomalous expansion of the analysis ensemble spread near the South Pole in the upper troposphere and the lower stratosphere. Since ALEDAS employs a spatio-temporally uniform multiplicative spread inflation, the analysis ensemble spread can be extremely large where the observing density is low. The large ensemble spread leads to unrealistic temperature or wind speeds and causes numerical instability during a forecast computation, i.e., AFES integration. In fact, ALEDAS occasionally ended abnormally during a forecast computation, especially in austral summer, if its time step was not set to short enough. The large ensemble spread near the South Pole can be suppressed if upper-atmospheric observations are effectively used.

Just located at the South Pole, daily radiosonde observations are routinely taken from the Amundsen-Scott South Pole Station.
Thanks to the continuous observations, we can assess the impact of the observations on the ALERA fields by comparing analyzed fields generated by ALEDAS with the current and modified algorithms. In this study, we will introduce the modified algorithms to use the observations at the South Pole and assess the impact of the observations on ALERA. The algorithms are described in Section 2, and the impact is assessed in Section 3. Conclusions are provided in Section 4.

2. The algorithms

Two parts in the pre-analysis algorithms of ALEDAS are modified to optimally use observations in the vicinity of the poles; one is for the interpolation in the observation operator (Section 2.1) and the other for the nearest-neighbor search used for identifying local observations for error covariance localization (Section 2.2).

2.1 Interpolation algorithm for observations near the poles

As mentioned in Section 1, the observation operator in the current ALEDAS corresponds to an interpolation to an observation point by using the surrounding 8 model gridded values. For this interpolation, because an observation point needs to be surrounded by model grids, polar observations located north (south) of the northernmost (southernmost) model grids are discarded. Therefore, we define model gridded values at the North (South) Pole as the interpolated ones from northernmost (southernmost) model grids. This way, the observation operators for observations at the poles are obtained.

Model scalar values such as temperature and moisture at the North (South) Pole are defined as the simple zonal-mean of the northernmost (southernmost) model gridded values at each model vertical level. The wind vector at the Pole is similarly defined as the zonal-mean values of the zonal and meridional components at the northern (southern) model grids, but the interpolation is done in the Cartesian (or absolute) coordinates after transformation from the spherical (or local) coordinates (Fig 1). The interpolated vector in the Cartesian coordinates is then transformed to that in the spherical coordinates at the Pole. Note that the vertical wind components are ignored in the transformation since they are small enough relative to the horizontal components and are not used in ALEDAS.

2.2 Nearest-neighbor search for error covariance localization

Localizing the error covariance in ALEDAS requires choosing observations within a circular search area with a constant radius in the physical space for each model grid (MYE07). To determine the area for the model gridded values analyzed at \((i, j)\) where \(i\) and \(j\) are longitudinal and latitudinal model grid indexes, respectively, the following steps are executed in the current algorithms:

1. applying a nearest-neighbor search based on a uniform longitude-latitude grid for each analyzed grid point. Namely, roughly selecting observations within a square region near the analyzed point,
2. computing physical distances between the analyzed grid point and all the observations selected via step 1,
3. discarding the distant observations beyond the localization radius to select the ones within the localization radius, and
4. giving distance-dependent localization weights (cf. Eq. (4) of MYE07) to the selected observations in step 3.

In step 1, the nearest-neighbor search is done by selecting only observations within the square region whose corners are defined as \(\pm \Delta i, \pm \Delta j\) for an analyzed grid point \((i, j)\), where \(\Delta i\) is the number of longitudinal grids equivalent to the physical distance of the localization radius at the latitude \(j\) and \(\Delta j\) is that of latitudinal grids, respectively. Since the current nearest-neighbor search processes two-dimensional sorting of the observational data and store the address (array index) of the observational data arrays at each model grid point (Fig. 1 of MYE07), observations within the square region can be rapidly selected in step 1. Next, in steps 2 and 3, the observations only within the search area are selected. Step 1 substantially reduces the computational cost, rather than computing the physical distances for all observations over the whole globe.

However, step 1 does not select observations within the search area when the observations are close to the poles, since observations within the search area must be surrounded by the square region. Figure 2a shows an example of the relationship between the search area and the square region for an analyzed model grid at (about) 75°N. Theoretically, all the observations within the search area (gray dots) should be included in the square region. However, because the grid number \(\Delta i\) is defined at the latitude of the analyzed grid, the longitudinal width of the square region at its polar side becomes shorter than that of the search area. Therefore, the circular shapes of the search areas slightly distort as close to the poles as illustrated in Figs. 2b, 2c, and 2d.

Therefore, we have slightly modified the step-1 algorithm to create wider square regions that entirely include the search areas. To do this, we adopt the spherical triangular method that can convert a physical distance into difference (degrees) between two grid points (see Supplement). This enables to find the longitudinal model grid numbers at the edge of the search area between the latitudes \(j - \Delta j\) and \(j + \Delta j\), and obtain the maximum/minimum longitudinal grid index \(i'\) where the number of longitudinal grids between \(i\) takes maximum of all the longitudinal model grids at the fringe: \([i - i']\) is defined as \(\Delta i'\). Here \(i'\) corresponds to the longitude of triangles in Fig. 2a or circles in Figs. 2b, 2c, and 2d. Once \(\Delta i'\) is obtained, we can create the wider square region with the corners \([i + \Delta i', j + \Delta j]\) entirely circumscribing the search area. Although the computational cost in the modified algorithm will slightly increase by using the wider regions, the modification would be reasonable to take precise circular localization near the poles while keeping a similar computational efficiency. It should be noted that although we have improved our search algorithm based on MYE07, other sophisticated search algorithms such as the one based on the K-d tree (Szunyogh et al. 2008) could further improve computational efficiency.
3. DA experiments using ALEDAS

We conduct DA experiments using ALEDAS with and without the new implementation developed in this study. With the new implementation, daily radiosonde observations at the South Pole can be assimilated; we can estimate the impact of using the observations by comparing the two experiments in terms of analysis ensemble spread and mean distributions and computational times. One experiment with the original implementation that MYE07 developed was performed as control experiment and the other with the new implementation was performed as test experiment. The ALEDAS descriptions are same as ALEDAS2 (Enomoto et al. 2013). Briefly, we used the AFES with a T119L48 resolution and 63 ensemble members. The localization scale $\sigma$ and the radius of the circular search area were set to 400 km and $2\sigma_{\sqrt{10/3}}$ (Gaspari and Cohn 1999), respectively. The observational data prepared from NCEP PREPBUFR datasets was used and assimilated every 6 hours with hourly slots. The initial ensemble members were chosen from ALERA2 datasets. Multiplicative 10% spread inflation was applied for background error covariance. The experimental periods are (i) July 2012 (from 00UTC 1 July to 00UTC 1 August 2012) and (ii) January 2013 (from 00UTC 1 January to 00UTC 1 February 2013) as representatives of austral winter and summer. With the new implementation 937 (2390) observations near the poles were additionally assimilated during the period of July 2012 (January 2013).

Figure 3 shows the difference of the analysis ensemble spread fields in the zonal wind and temperature averaged during the periods. It is found that the spread in the test experiment was substantially reduced over the Antarctic regions from the upper troposphere to the lower stratosphere for both periods. This reduction in turn implies that the spread there in the control experiment was extremely large potentially to cause numerical instability: in fact, forecast cycles during January 2013 in the control experiment experienced abnormal termination and thus needed a shorter time step than that during the other period. The termination would stem from the large ensemble spread distributions over the Antarctic regions because the cycles in the test experiment did not need the shorter time step. This is mainly associated with the low observing density over there as found in Fig. 4. As a result, daily radiosonde observations at the South Pole substantially stabilized the DA-forecast cycle in ALEDAS by suppressing artificial expansion of the analysis ensemble spread in the upper atmosphere over the Antarctic regions.

The radiosonde observations additionally improved the analysis ensemble mean fields over the Antarctic regions. Figure 5 shows zonal wind and temperature differences between the analysis ensemble mean fields in the experiments and the analysis in ERA Interim (Dee et al. 2011). We can find that the differences became smaller in the test experiment than those in the control especially at the upper troposphere and the lower stratosphere over the Antarctic regions during both winter and summer. In particular, the temperature field during the austral summer (January 2013) substantially improved when the ensemble spread became anomalously large (Figs. 5l and 3d), and the DA-forecast cycle in
the control experiment abnormally ended. Moreover, the analysis ensemble mean fields at the upper troposphere in middle latitudes and over the Arctic regions were slightly improved during the winter and summer, which probably owes to the modification of the nearest-neighbor search. Therefore, the new ALEDAS totally improved the analysis fields in ALERA.

Finally, we checked how the computational cost in the DA-analysis cycles increased by using the modified algorithms. Table 1 shows the computational times for the DA-analysis cycles averaged during the periods in the test and control experiments. As expected, the computational time in the test experiment became longer, but their difference was only a few percent of the total time of the analysis cycles. Thus, as about the computational cost, it is enough reasonable to use the LETKF with the new implementation for ALEDAS.

4. Concluding remarks

The observation operators and error covariance localization of ALEDAS have been modified to allow using observations near the poles. In the modified algorithms, the observation operator can be defined for the observations south (north) of the southernmost (northernmost) model grids. The nearest-neighbor search of observations as a part of error covariance localization was slightly updated for better treatment near the poles while keeping a similar computational efficiency. These modifications enable to assimilate all observations over the whole globe in a more appropriate manner.

An experiment using ALEDAS with the new implementation was performed and compared with the experiment with the original ALEDAS for austral summer and winter months. The main difference was that the new ALEDAS assimilated daily radiosonde observations at the South Pole, while the original ALEDAS.
The results showed that the experiment with the new implementation suppressed the large analysis ensemble spread in the upper troposphere and the lower stratosphere over the Antarctic regions and stabilize its DA-forecast cycle, and improved the analysis ensemble mean fields. Computational cost became only slightly larger in the DA-analysis part of the new ALEDAS, even though its DA-forecast part did not need the shorter time step during the austral summer. The new algorithms would be useful for ALEDAS-like atmospheric DA systems based on the longitude-latitude grid.

Another approach to suppress artificial expansion of the ensemble spread near the poles is to use more advanced cova-
riance inflation methods that can consider spatial and temporal distributions of observing density, such as the adaptive inflation methods proposed by Anderson (2009) and Miyoshi (2011). Since using other inflation methods in the LETKF does not conflict with the approach in this study, they are complementary to each other to improve the quality of ALEDAS.

Ideally, error covariance localization in EnKF should be removed, because localization destroys meaningful long-range error correlations. It has recently been elucidated that the LETKF with massive ensemble members, as large as 10000, does not require localization with an AGCM of intermediate complexity at T30L7 resolution (see Miyoshi et al. 2014; Kondo and Miyoshi 2016). But a realistic choice of the ensemble size in the state-of-the-art DA systems would be limited to O(100), and error covariance localization would be needed.

Acknowledgments

We are grateful to the reviewers for their helpful comments. The present work was partly supported by the Grants-in-Aids for Scientific Research (KAKENHI) numbers 26282111, 25800267, and 25400474. ALEDAS integrations were performed on the Earth Simulator with the support of JAMSTEC. The ALERA and ALERA2 datasets have been archived for about 2 and 5 years, respectively, which are available from https://www.jamstec.go.jp/. For data visualization, we used the GFD-Dennou Ruby libraries, GraDS, and MjoGraph.

Edited by: H. Kamahori

Supplement

Supplement includes mathematical explanation for the modified algorithms in Section 2.

References

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. Tellus, 61A, 72–83.

Dec, D. P., and co-authors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quart. J. Roy. Meteor. Soc., 137, 553–597.

Enomoto, T., M. Hattori, T. Miyoshi, and S. Yamane, 2010: Precursory signals in analysis ensemble spread. Geophys. Res. Lett., 37, L08804.

Enomoto, T., T. Miyoshi, Q. Moteki, J. Inoue, M. Hattori, A. Kuwano-Yoshida, N. Komori, and S. Yamane, 2013: Observing-system research and ensemble data assimilation at JAMSTEC. Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II), S.K. Park and L. Xu, Eds., Springer, 509–526.

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

Greybush, S. J., R. J. Wilson, R. N. Hoffman, M. J. Hoffmann, T. Miyoshi, K. Ide, T. McConnachie, and E. Kalnay, 2012: Ensemble Kalman filter data assimilation of Thermal Emission Spectrometer temperature retrievals into a Mars GCM. J. Geophys. Res., 117, E11008.

Hattori, M., J. Matsumoto, S. Ogino, T. Enomoto, and T. Miyoshi, 2016: The impact of additional radiosonde observations on the analysis of disturbances in the South China Sea during VPREX2010. SOLA, 12, 75–79.

Hoffman, R. N., R. M. Ponte, E. J. Kostelich, A. Blumberg, I. Szunyogh, S. V. Vinogradov, and J. M. Henderson, 2008: A simulation study using a local ensemble transform Kalman filter for data assimilation in New York Harbor. J. Atmos. Oceanic Technol., 25, 1638–1656.

Hoffman, M. J., and co-authors, 2010: An ensemble Kalman filter data assimilation system for the martian atmosphere: Implementation and simulation experiments. Icarus, 209, 470–481.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Physica D, 230, 112–126.

Inoue, J., T. Enomoto, and M. E. Hori, 2013: The impact of radiosonde data over the ice-free Arctic Ocean on the atmospheric circulation in the Northern Hemisphere. Geophys. Res. Lett., 40, 864–869.

Kuwano-Yoshida, A., and T. Enomoto, 2013: Predictability of explosive cyclogenesis over the northwestern Pacific region using ensemble reanalysis. Mon. Wea. Rev., 141, 3769–3785.

Kondo, K., and T. Miyoshi, 2016: Impact of removing covariance localization in an ensemble Kalman filter: Experiments with 10 240 members using an intermediate AGCM. Mon. Wea. Rev., 144, 4849–4865.

Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. Mon. Wea. Rev., 139, 1519–1535.

Miyoshi, T., and K. Aranami, 2006: Applying a Four-Dimensional Local Ensemble Transform Kalman Filter (4D-LETKF) to the JMA Nonhydrostatic Model (NHM). SOLA, 2, 128–131.

Miyoshi, T., and S. Yamane, 2007: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. Mon. Wea. Rev., 135, 3841–3861.

Miyoshi, T., S. Yamane, and T. Enomoto, 2007a: The AFES-LETKF experimental ensemble reanalysis: ALERA. SOLA, 3, 45–48.

Miyoshi, T., S. Yamane, and T. Enomoto, 2007b: Localizing the error covariance by physical distances within a Local Ensemble Transform Kalman Filter (LETKF). SOLA, 3, 89–92.

Miyoshi, T., Y. Sato, and T. Kadowaki, 2010: Ensemble Kalman filter and 4D-Var intercomparison with the Japanese operational global analysis and prediction system. Mon. Wea. Rev., 138, 2846–2866.

Miyoshi, T., and M. Kuni, 2012: The local ensemble transform Kalman filter with the Weather Research and Forecasting model: Experiments with real observations. Pure and Appl. Geophys., 169, 321–333.

Miyoshi, T., K. Kondo, and T. Imamura, 2014: The 10,240-member ensemble Kalman filtering with an intermediate AGCM. Geophys. Res. Lett., 41, 5264–5271.

Moteki, Q., and co-authors, 2011: The influence of observations propagated by convectively coupled equatorial waves. Quart. J. Roy. Meteor. Soc., 137, 641–655.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. Tellus, 60A, 113–130.

Penny, S. G., E. Kalnay, J. A. Carton, B. R. Hunt, K. Ide, T. Miyoshi, and G. A. Chepurin, 2013: The local ensemble transform Kalman filter and the running-in-place algorithm applied to a global ocean general circulation model. Nonlin. Processes Geophys., 20, 1031–1046.

Terasaki, K., M. Sawada, and T. Miyoshi, 2015: Local ensemble transform Kalman filter experiments with the Nonhydrostatic Icosahedral Atmospheric Model NICAM. SOLA, 11, 23–26.

Yamazaki, A., J. Inoue, K. Dethloff, M. Maturilli, and G. König-Langlo, 2015: Impact of radiosonde observations on forecasting summertime Arctic cyclone formation. J. Geophys. Res., 120, 3249–3273.

Manuscript received 27 November 2016, accepted 17 February 2017 SOLA: https://www.jstage.jst.go.jp/browse/sola/