Impacts of Frequent Assimilation of Surface Pressure Observations on Atmospheric Analyses

LILI LEI AND JEFFREY L. ANDERSON
National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 27 March 2014, in final form 16 August 2014)

ABSTRACT

To investigate the impacts of frequently assimilating only surface pressure (PS) observations, the Data Assimilation Research Testbed and the Community Atmosphere Model (DART/CAM) are used for observing system simulation experiments with the ensemble Kalman filter. An empirical localization function (ELF) is used to effectively spread the information from PS in the vertical. The ELF minimizes the root-mean-square difference between the truth and the posterior ensemble mean for state variables. The temporal frequency of the observations is increased from 6 to 3 h, and then 1 h. By observing only PS, the uncertainty throughout the entire depth of the troposphere can be constrained. The analysis error over the entire depth of the troposphere, especially the middle troposphere, is reduced with increased assimilation frequency. The ELF is similar to the vertical localization function used in the Twentieth-Century Reanalysis (20CR); thus, it demonstrates that the current vertical localization in the 20CR is close to the optimal localization function.

1. Introduction

Surface pressure (PS) observations provide significant information about the entire depth of the atmosphere. On synoptic scales, PS observations define the location and intensity of cyclones and anticyclones. Through geostrophy, PS observations yield an approximation to the barotropic component of the total flow. The PS tendency, related to the vertically integrated mass flux divergence, provides further information on the tropospheric circulation. On the mesoscale, PS observations define the location and intensity of convectively induced mesohighs and mesolows while PS tendency provides information on the evolution of mesoscale features.

Given the significant information provided by PS observations, the impact of assimilating PS observations has been investigated in both mesoscale and global simulations. The PS observations were assimilated using an ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Mitchell 1998) in a mesoscale model and yielded accurate depictions of mesoscale pressure patterns associated with two mesoscale convective systems (Wheatley and Stensrud 2010).

Whitaker et al. (2004) assimilated only PS observations in a simulation of the 1915 network using an EnKF and produced a useful Northern Hemisphere analysis of the middle and lower troposphere. The expected Northern Hemisphere 500-hPa analysis error is similar to the present-day 2–3-day forecast error. Compo et al. (2006) used an ensemble assimilation system to reanalyze both the near-surface and the upper-air circulation using only PS observations. The surface pressure observations could be used to reanalyze the entire extratropical troposphere circulation. Analysis error for Northern Hemisphere winter was on the order of the current 1–2-day forecast error in the lower troposphere and the 2–3-day forecast error in the middle and upper troposphere. Given the demonstrated advantages of using PS observations, the Twentieth-Century Reanalysis (20CR) spanning 1871 to the present was generated with an EnKF assimilating only PS observations (Compo et al. 2011).

These previous studies used 6-hourly assimilation; additional improvement to their results may be possible. Besides achieving similar results to Whitaker et al. (2004), Anderson et al. (2005) examined the impact of

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Lili Lei, National Center for Atmospheric Research, 1850 Table Mesa Dr., Boulder, CO 80307-3000, E-mail: lililei@ucar.edu

DOI: 10.1175/MWR-D-14-00097.1

© 2014 American Meteorological Society
varying the temporal frequency of PS observations in a simple atmospheric general circulation model. Increasing the assimilation frequency of PS observations resulted in a monotonic decrease of prior error for PS and all other state variables. Bengtsson (1980) with a linear quasigeostrophic system and Jarvinen et al. (1999) with a four-dimensional variational data assimilation (4DVAR) application also demonstrated that hourly assimilation of PS tendencies could improve upper-level analysis.

Motivated by these findings, the impact of frequently assimilating only PS observations on the troposphere is investigated here using the Community Atmosphere Model version 5 (CAM5; Neale et al. 2012), a more realistic model than that used in Anderson et al. (2005). An EnKF is used since flow-dependent background error information is crucial in assimilating PS observations and the EnKF provides estimates of analysis error not directly available from variational methods (Whitaker et al. 2004, 2009).

Localization is used in the EnKF to limit the impact of spurious correlations resulting from limited ensemble size. However, there are few theoretical studies of the vertical localization for PS observations. The simple atmospheric general circulation model used by Anderson et al. (2005) does not require vertical localization. Whitaker et al. (2004) used vertical localization of 1 below $\sigma = 0.2$, 0 above $\sigma = 0.5$, and linearly decreasing between these two $\sigma$ levels. The 20CR adopted the Gaspari–Cohn (GC; Gaspari and Cohn 1999) localization with half-width of two scale heights (Compo et al. 2011).

Anderson and Lei (2013) proposed an empirical localization function (ELF) that can automatically provide an estimate of localization functions for any observation type and state variable kind without empirically tuning the localization scale. The ELF was applied in a simple atmospheric general circulation model (Lei and Anderson 2014a) and a more realistic atmospheric general circulation model, CAM5 (Lei and Anderson 2014b, hereafter LA14b), and produced a smaller error than the best GC function. Thus, the ELF is used here to provide an estimate of vertical localization for PS observations.

Section 2 describes the experimental design. Section 3 discusses the ELF and uses it to estimate vertical localization for PS observations. Results are discussed in section 4. Section 5 gives the conclusions.

2. Experimental design

a. The DART/CAM system

Observing system simulation experiments (OSSEs) are conducted in the Data Assimilation Research Testbed and the Community Atmosphere Model (DART/CAM) system (Raeder et al. 2012). CAM5 uses a finite-volume grid with approximately 2° resolution ($96 \times 144$), and 30 vertical levels. At the surface, the model is forced by temporally interpolated observed monthly mean sea surface temperature using the default configuration from the Atmospheric Model Intercomparison Project (AMIP; Gates 1992). The land surface and sea ice models are fully active and coupled to the atmosphere model, but the ocean is specified.

The data assimilation system is DART (Anderson et al. 2009) with the ensemble adjustment Kalman filter (EAKF; Anderson 2001) used to combine synthetic observations with an ensemble of forecasts to produce an ensemble of analyses. To maintain ensemble spread, spatially and temporally varying state space adaptive inflation (Anderson 2009) is applied to the prior state. Sampling error correction (Anderson 2012) is used to maintain ensemble spread and reduce sampling errors due to small ensembles. The GC localization with half-width 0.2 rad (1274 km) is used for horizontal localization; vertical localization is discussed below.

b. Experimental design of the OSSE

Two sets of experiments are completed: one for Northern Hemisphere summer (July–August 2008) and one for winter (December 2008–January 2009). Summer experiments are described below. Winter experiments are similar except that simulation time is from 0000 UTC 1 December 2008 to 0000 UTC 1 February 2009.

A true initial condition (IC) is generated by advancing the model from an analysis at 0000 UTC 1 January 2008 (input data from CAM5) to 0000 UTC 1 July 2008. A nature run is obtained by advancing the model for two months from the true IC. Synthetic observations of PS are generated by adding random draws from a specified observational error distribution to spatially interpolated values of the true state. Observation error distributions are normal with mean 0 and variance 10 000 Pa². Observations are nearly uniformly distributed in the horizontal, and there are 7200 sites on the sphere as shown in Fig. 1. For each assimilation frequency, a separate nature run and corresponding synthetic observations are generated by stopping the forecast model at each assimilation time.

To generate the ICs for $N = 80$ ensemble members, small random perturbations from a normal distribution with a mean of 0 and a standard deviation of $10^{-11}$ are added to the temperature field of the true IC. These $N$ perturbed ensemble members are then integrated from 0000 UTC 1 July to 0000 UTC 1 September 2008 with the synthetic observations assimilated at the specified frequency for each case.
The root-mean-square error (RMSE) of the ensemble mean from the truth for PS, temperature, zonal and meridional winds, and specific humidity are evaluated. Data from the second month are used to compute the spatially and temporally averaged RMSE, while the first month is discarded to eliminate transient spinup. Time series of vertical profiles of the horizontally averaged RMSE and spread are also examined. Results are shown for the posterior since experiments with different assimilation frequencies have different forecast lead times. However, qualitatively similar results are obtained for the prior.

3. Vertical localization

As an integral observation, PS should require broader vertical localization than temperature and winds. However, there is limited research about appropriate vertical localization for PS. The ELF (Anderson and Lei 2013) uses the output of an OSSE and minimizes the RMS difference between the true value and the posterior ensemble mean. The ELF was demonstrated to integrate over all the possibilities of the synthetic observations follows Eq. (12) of Anderson and Lei (2013) as follows:

\[
\alpha = \sum_{k=1}^{K} (x_k^o - \overline{x}_k^o)\hat{b}_k(A_k + B_ky_k) / \sum_{k=1}^{K} [\hat{b}_k(A_k + B_ky_k)]^2 + (\hat{b}_k B_k)^2 \sigma_{o,k}^2, \tag{1}
\]

where

\[
A_k = \left( \frac{\sigma_{u,k}^2}{\sigma_{p,k}^2} - 1 \right) \overline{y}_k^p \quad \text{and} \quad B_k = \frac{\sigma_{u,k}^2}{\sigma_{o,k}^2} \tag{2}
\]

Superscripts \( t \) and \( p \) denote the true value and prior, the overbar denotes ensemble mean, \( \hat{b} \) is the sample regression coefficient, \( \sigma_{p,k}^2 \) and \( \sigma_{o,k}^2 \) are the prior and posterior ensemble sample variances for the estimated observation, respectively, and \( \sigma_{o,k}^2 \) is the observation error variance; \( k \) indexes the pair in the subset and \( K \) is the total number of pairs in the subset.

An OSSE with 6-hourly assimilation without vertical localization is conducted for the month of December. The first 11 days are discarded to eliminate transient effects, and the last 20 days are used to compute the ELFs. ELFs are computed for PS observations with temperature, zonal and meridional winds, and specific humidity. Following Lei and Anderson (2014a; LA14b), the true gridded PS values rather than the synthetic observations are used in order to increase the sample size for each subset and also lead to a smoother ELF by reducing the sampling error of the observations.

For the vertical ELF, the set of all pairs is divided into subsets \( SV(l) \), where \( S \) denotes subset and \( V \) stands for state variables of temperature, zonal and meridional winds, and specific humidity, and \( l \) indexes the 30 model levels. The subset \( SV(l) \) contains all pairs \((y, x)\) where \( y \) is a PS observation at any time \( t \) from the OSSE and \( x \) is a state variable \( V \) in the same vertical column as \( y \) at time \( t \) with model level of \( l \). Following Lei and Anderson (2014a; LA14b), the true gridded PS values rather than the synthetic observations are used in order to increase the sample size for each subset and also lead to a smoother ELF by reducing the sampling error of the observations.
outside the critical region for a 95% confidence, and then the localization value is set to 0.

The colored solid lines in Fig. 2 show the ELFs. The ELFs are close to 1 below model level 15 (~300 hPa), and then decrease to 0 at the model top. However, as discussed in LA14b, these ELFs are too noisy to be applied directly in a subsequent OSSE. Thus, a smooth localization function is obtained by fitting a GC function to these ELFs. The GC function that has the smallest RMS difference from these ELFs has half-width $4 \log p$ units, shown by the black solid line in Fig. 2. This GC function can be seen as an estimate of the optimal localization function without empirically tuning the localization scale (LA14b), and it is used in subsequent assimilation experiments.

Using the CAM5 pressure coordinate and assuming surface pressure is 1000.0 hPa, the GC function for the 20CR is computed and shown by the black dashed line in Fig. 2. The GC function fitted from the ELFs is very similar to the GC function used in the 20CR, except that the former has slightly larger localization value than the latter. Previous studies (Lei and Anderson 2014a; LA14b) suggest that the ELF can provide an estimate of the optimal localization function, thus, the ELF here demonstrates that the vertical localization for PS observations in the 20CR is very good.

4. Results

The temporally and spatially averaged RMSEs are shown in Fig. 3. For PS, RMSE in summer is 23.7 Pa when the assimilation frequency is 6 h, 20.4 Pa for 3 h (~14% reduction), and 16.1 Pa for 1 h (~32% reduction).

By only assimilating PS observations every 6 h, the temperature RMSE in summer is constrained to 0.51 K. This decreases to 0.43 (0.36) K when the assimilation frequency is increased to 3 (1) h, a 16% (29%) error reduction. Similar approximately linear relationships between RMSE and the assimilation frequency are obtained for zonal wind and specific humidity in summer, although specific values of error reduction vary. Results for meridional wind are similar to zonal wind (not shown).

RMSEs of PS, temperature, zonal and meridional winds, and specific humidity in winter also decrease with increasing assimilation frequency. The percentage of error reduction for PS is similar to summer, but the percentages of error reduction for other state variables are different from summer. For instance, temperature RMSE decreases from 0.48 to 0.47 (0.39) K when the assimilation frequency increases from 6 to 3 (1) h, which corresponds to a 2% (19%) error reduction. Therefore, for both summer and winter, the RMSEs of directly and nondirectly assimilated state variables can be constrained by only assimilating PS observations and the RMSEs are reduced with more frequent assimilation of PS observations.

For each state variable, the RMSE statistics are different in summer and winter, and the differences appear mostly near the surface. This is possibly due to unreasonable error growth of temperature and wind for grid points over land with ice cover in the winter experiments, which is independent of whether any assimilation is done and appears with more frequent restarting of the model. A Student’s $t$ test is used to assess the significance of the RMSE difference for each state variable given an assimilation frequency between the summer and the winter case. The differences for all state variables and assimilation frequencies are significant given a 95% confidence.

To examine the detailed impacts of assimilating PS observations on nondirectly observed state variables, time series of vertical profiles of horizontally averaged temperature RMSE for the winter are shown in Fig. 4. Vertical profiles before and after equilibration are shown, because the former shows how quickly the PS information spreads upward through the atmosphere and the latter presents vertical structures of saturated RMSE. By assimilating the 6-hourly PS observations, RMSEs in upper levels become saturated (achieve equilibrium) after about 1 month. For a time before the error saturated (e.g., 21 December), RMSE with 3-hourly assimilation is smaller than that with 6-hourly assimilation at nearly every vertical level. Similarly, the RMSE with 1-hourly assimilation is further reduced. Thus, with
more frequent assimilation, the impact of PS observations on temperature is more effectively and efficiently retained.

After the error equilibrates, RMSE in the whole column for 1-hourly assimilation is smaller than for 3-hourly assimilation, and both are smaller than for 6-hourly assimilation. This is consistent with the temporally and spatially averaged RMSE shown in Fig. 3. With 6- and 3-hourly assimilation, the largest RMSE after equilibration are concentrated aloft (between 500 and 70 hPa). By observing only the PS every 6 and 3 h, the temperature analysis of the middle and upper troposphere still has substantial errors. But these temperature errors are significantly improved by more frequent (1 h) assimilation, and the temperature error reduction between 1- and 3-hourly assimilation is 21%.

It is interesting that a relative maximum RMSE appears near 850 hPa for the three assimilation frequencies. The maximum RMSE near 850 hPa is located over the eastern subtropical Pacific and Atlantic Oceans (not shown) where the model has large areas of stratocumulus cloud; thus, it is possible that the relative maximum RMSE near 850 hPa is related to the shallow convection scheme in CAM (Park and Bretherton 2009). There is also a relative maximum RMSE around 100 hPa, and it is greatly reduced when assimilation frequency increases to 1 h; however, it is not clear why it appears around 100 hPa.

Figure 4 also shows that there is an increase in RMSE at the top of the model as assimilation frequency increases. The top three levels of CAM include a horizontal diffusion that acts as a simple sponge layer to absorb vertically propagating wave energy and reduce the strength of the wintertime jets (Neale et al. 2012). Data assimilation in CAM introduces some numerical imbalance that leads to vertically propagating waves. The strength of the horizontal diffusion is tuned for free runs of CAM that do not have this extra source of waves so these are not entirely damped in the assimilation cases. The more frequent assimilation cases lead to increased wave energy in the upper layers of the model that are reflected as increased RMSE.

Time series of vertical profiles of horizontally averaged RMSE of specific humidity (not shown) have similar
patterns to temperature except above 200 hPa where model humidity is nearly 0. With more frequent (~1 h) assimilation of PS observations, the analysis of specific humidity for the middle troposphere is significantly improved.

The time series of vertical profiles of the horizontally averaged RMSE of zonal wind for the assimilation experiments in winter are shown in Fig. 5. Similar to temperature, the RMSE of zonal wind before the error equilibrates is smaller when PS observations are more frequently assimilated. After the error equilibrates, more frequent assimilation reduces the zonal wind RMSE in the whole column, especially around 200 hPa. The maximum RMSE of zonal wind is around 200 hPa, and the large RMSEs around 200 hPa are mainly concentrated in the tropics (not shown). It is not immediately clear why the maximum RMSE of zonal wind appears near 200 hPa, but it is possible that this is related to the cumulus parameterization scheme in CAM (Zhang and McFarlane 1995). Results for meridional wind (not shown) are similar to zonal wind.

For each state variable, the ensemble spread has similar patterns to the RMSE except with smaller magnitude. The average adaptive inflation is slightly larger than 1.0, and inflation values larger than 1.0 are distributed approximately uniformly over the sphere (LA14b). Similar results to winter assimilation experiments are obtained for summer assimilation experiments (not shown).

5. Conclusions

The impacts of frequently observing only PS are examined here using DART/CAM in a perfect model context. The EAKF in DART is used since the flow-dependent background error covariances are essential to effectively assimilate only PS observations (Anderson et al. 2005; Whitaker et al. 2004, 2009). To effectively spread the PS information in the vertical, the ELF is used to provide an estimate of vertical localization for PS observations. The localization for PS observations is close to unity up to around 300 hPa, thus, it extends broadly in the vertical. The GC localization function fitted from the ELFs is very similar to the GC function used in the 20CR, which provide evidence that the vertical localization for PS observations in the 20CR is very good.

For both summer and winter, more frequent assimilation of PS observations (e.g., 1 and/or 3 h vs 6 h) can better constrain both directly and nondirectly observed state variables, and reduce the error and uncertainty through the entire depth of the troposphere, especially in the middle troposphere. The impacts of PS observations on state variable are more effectively and efficiently retained with more frequent assimilation of PS observations.
The ELF was demonstrated to outperform the best GC localization in CAM (LA14b) and the vertical localization for PS provided by the ELF produces promising results, thus, applying this vertical localization to frequent assimilation of real PS observations will be the next step. Also, the frequent assimilation of PS observation with the ELF that might vary with geographic regions and have temporal variation could improve future versions of the 20CR (Compo et al. 2011).

Acknowledgments. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. Thanks to Kevin Raeder, Tim Hoar, and Nancy Collins for technical support and helpful discussions. Jeffrey Whitaker provided valuable discussion on the vertical localization in the 20CR.

REFERENCES


