A data assimilation case-study using a limited-area ensemble Kalman filter

Sébastien Dirren¹, Ryan D. Torn and Gregory J. Hakim

University of Washington, Seattle, WA

Submitted to Monthly Weather Review

January 17, 2005

¹ Corresponding Author:
Dr. Sébastien Dirren
Department of Atmospheric Sciences, Box 351640
University of Washington
Seattle, WA 98195-1640
E-mail: sebastien.dirren@alumni.ethz.ch
Abstract

Ensemble Kalman filter (EnKF) data assimilation experiments are conducted on a limited-area domain over the Pacific Northwest region of the United States, using the Weather Research and Forecasting model. Idealized surface pressure, radiosoundings and aircraft observations are assimilated every 6 hours for a seven-day period in January 2004. The objectives here are to study the performance of the filter in constraining analysis errors with a relatively inhomogeneous, sparse observation network and to explore the potential for such a network to serve as the basis for a real-time EnKF system dedicated to the Pacific Northwest region of the United States.

When only a single observation type is assimilated results show that the ensemble-mean analysis error and ensemble spread (standard deviation) are significantly reduced compared to a control ensemble without assimilation for both observed and unobserved variables. Analysis errors are smaller than background errors over nearly the entire domain when averaged over the seven-day period. Moreover, comparisons of background errors and observation increments at each assimilation step suggest that the flow-dependent filter corrections are accurate in both scale and amplitude. An illustrative example concerns a mis-specified mesoscale 500-hPa shortwave trough moving along the British Columbia coast, which is corrected by surface pressure observations alone. The relative impact of each observation type upon different variables and vertical levels is also discussed.
1. Introduction

Research studies on the ensemble Kalman filter (EnKF) suggest that it is a potentially useful data assimilation technique for atmospheric and oceanic applications (Evensen 1994, Houtekamer and Mitchell 1998; Hamill 2005). Many studies have addressed basic properties of the EnKF in simplified models (e.g. Houtekamer and Mitchell 2001, Bishop 2001, Whitaker and Hamill 2002, Anderson 2001). Recently this work has expanded to applications in more complex NWP models using either synthetic observations (e.g. Mitchell et al. 2002, Snyder and Zhang 2003, Zhang et al. 2004a) and real observations (e.g. Keppenne and Rienecker 2002, Houtekamer et al. 2004, Dowell et al. 2004). The goal of the present study is to assess the potential for an operational EnKF to constrain errors in a region such as the Pacific Northwest region of the United States. This location presents a significant test for the robustness of data assimilation systems due to the presence of complex topography, land–ocean contrast, and a relatively inhomogeneous, sparse observing network.

The EnKF extends the traditional Kalman Filter (Kalman and Bucy 1961) using an ensemble of fully non-linear forecasts to estimate a flow-dependent background-error covariance matrix. This matrix determines (i) the relative weighting of an observation relative to the background estimate (e.g. a six-hour forecast), (ii) the effect of an observation on nearby locations, and (iii) the relationships between different fields, for example the appropriate effect of a height observation on the wind field. In principle, the use of a flow-dependent covariance matrix in the EnKF technique should yield analyses with smaller errors than for methods that employ fixed covariance matrices, such as 3DVAR. The EnKF algorithm might therefore be particularly suitable for regions of sparse observations, complex topography and wide range of spatial scales, such as considered here.

Studies on EnKF techniques in complex models have addressed mainly planetary (e.g. Houtekamer et al. 2004, Keppenne and Rienecker 2002, Whitaker et al. 2004) and convective scales (e.g. Snyder and Zhang 2003, Zhang et al. 2004b, Dowell and al. 2004), with fewer studies at the synoptic scale and mesoscale, which are of practical importance for both forecasting and analysis (verification) purposes. An important issue on mesoscale domains in particular is the need for properly posed ensemble boundary conditions on limited-area domains. Whereas convective-scale case-studies may avoid the use of ensemble boundary conditions over short time periods (Snyder and Zhang 2003, Dowell et al. 2004), this is not possible for longer time periods, since significant flow normal to the domain boundaries affects the domain interior. Torn et al. (2005) proposed and tested several boundary condition methods that do not require a pre-existing global ensemble and that offer new opportunities for limited-area data assimilation using EnKF techniques. Here we test one such method in EnKF experiments using the Weather Research and Forecasting model (WRF) over a seven-day period that is characterized by strong flow across the domain boundaries and significant weather features.

The perfect model assumption is made here in order to rigorously quantify the performance of the filter. Observations are assimilated for an idealized, but realistic, network of surface pressure observations, radiosoundings and aircraft measurements (ACARS). Filter performance is evaluated over the entire time period and at individual analysis times. The relative impact of each observation type upon different variables and vertical levels is also considered.
The outline of the paper is as follows. Section 2 describes the methodology of the study. Section 3 is devoted to a synoptic overview of the major flow characteristics during the seven-day period of interest. The average performance of the filter is then evaluated in Section 4 for surface pressure observations, whereas the analysis increments at selected individual assimilation times are analyzed in Section 5. Results for different observation types are presented in Section 6. The last section provides a summary and a concluding discussion on the potential relevance of these results to the assimilation of real observations.

2. Methodology

We implement the EnKF with WRF, a non-hydrostatic primitive equation, mesoscale model (Michalakes et al. 2001). The limited-area domain applies to the eastern North Pacific Ocean and adjacent region of western North America, and consists of a 102 x 88 horizontal grid with 45-km grid spacing and 32 vertical levels. All model runs employ the Mellor-Yamada-Janjic (Eta) TKE scheme, Betts-Miller-Janjic convective scheme and WRF Single Moment (WSM) 3-class (water vapor, liquid water, and “simple” ice) microphysics scheme. The lateral boundary zone is five grid points wide with the outer boundary points specified by interpolated GFS values and the inner four boundary points by a linear combination of interpolated GFS values and WRF dynamics.

An ensemble square-root filter (EnSRF) is employed, which does not require perturbed observations (Whitaker and Hamill, 2002). To overcome ensemble undersampling, we localize the covariances using eq. (4.10) of Gaspari and Cohn (1997) where all covariances go to zero 2000 km from an observation (e.g. Hamill et al. 2001). Ensemble perturbations are inflated by 6% at each assimilation time to prevent an overconfident ensemble estimate of the state (i.e. filter divergence).

All experiments are performed under the perfect-model assumption for a 7-day period in January 2004 with a 6-hour data-assimilation cycle. Truth is taken to be a solution of the WRF model starting with the GFS analysis at 12 UTC 20 January 2004, which is integrated forward for 7-days with boundary conditions taken from 6-hourly GFS analyses. Having a truth run gives us the ability to quantify the performance of the system over the entire domain without having to consider model error or observation representativeness.

Simulated observations consists of surface pressure (technically dry-air mass), radiosondes (u, v and temperature, each 12 hours) and Aircraft Communications Addressing and Reporting System [ACARS] observations (u, v and temperature). A total of four experiments are conducted, with the first three corresponding to assimilating only one observation type, and the fourth experiment corresponding to assimilating all observations. Observation values are taken from the truth run with added uncorrelated Gaussian error; the error variance for each observation type is provided in Table 1. Surface pressure observation locations represent a thinned version of the existing Automated Surface Observing System (ASOS) and fixed buoy network. The actual surface stations are placed on the nearest grid point with a minimum observation spacing of 4 grid points. Radiosonde stations are also re-located to the nearest model grid point, and only significant-level data is assimilated (Table 1). ACARS observation locations are generated by taking the actual position of all such observations within 30 minutes of an assimilation time, and thinning such that the observations are at least 2 grid points and 25 hPa apart. These procedures leads to 108 surface pressure stations, 25 vertical soundings and an average of 115 ACARS locations per assimilation cycle.
Ensemble boundary perturbations are derived using the “climatology time series” tech-
nique described in Torn et al. (2005). Archived GFS analyses are sampled for seven days
starting at arbitrary times for each of the 90 ensemble members. The ensemble mean of
these time series is replaced by the 6-hour GFS forecasts valid at each assimilation time,
and the perturbations are scaled by a fixed factor (0.15). This factor corresponds to the
average RMS difference between GFS analyses and 6-hour forecasts. Six-hour GFS forecasts
are used as the ensemble mean to allow for realistic boundary condition errors and to mimick
the global forecasts that would be available in a real-time system. The initial ensemble is
generated by choosing random GFS analyses, removing the mean, scaling the perturbations
by 0.15 and centering the ensemble around a 48-hour GFS forecast valid 12 UTC 20 January
2004. A relatively inaccurate estimate for the mean initial state (i.e. 48-hour GFS forecast)
is used so that the observations attract the state during the first few assimilation steps,
and so that the initial ensemble-mean error is similar to the ensemble spread defined by the
scaled perturbations.

3. Overview of the development

Fig. 1 shows the 7-day time evolution of the 500-hPa geopotential height and sea-level
pressure fields in the truth run. The evolution of the 500 hPa pattern is characterized
by a transition from a high-amplitude ridge over the eastern Pacific to more zonal flow.
During the first two days, a large-scale ridge exists over the West Coast of North America.
Simultaneously, a cyclone (denoted by closed surface pressure isobars and corresponding
upper-level trough) moves northward along the west side of the ridge toward western Alaska
(Fig. 1a-b). At 12 UTC 22 January, anticyclogenesis over the Pacific elongates the upper-
level trough located over western Alaska near the top of the ridge. A portion of this feature
breaks free, creating a new short-wave trough that subsequently propagates southeastward
along the West Coast of North America (Fig. 1c-d). By 12 UTC 24 January, the shortwave
trough moves over southern British Columbia in north-westerly upper-level flow. During the
next few days, the ridge breaks down as zonal flow becomes established over the eastern
Pacific.

4. Average Performance of the Filter

In order to evaluate the filter’s ability to constrain errors, we compare experiments where
observations are assimilated with a control experiment without any assimilation. In this
section we evaluate the filter performance in an averaged sense, by calculating both time-
and space-averaged RMS error of the ensemble mean, as well as the time-averaged ensemble
standard deviation. Comparisons between background errors and observation increments for
individual times are discussed in the next section.

Here we define two different error diagnostics: the space-averaged RMS error of a variable
$X$ (evaluated at a fixed vertical level) as a function of time,

$$ E_k^X = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} \left[ X_{k,i} - X_{tr,k,i} \right]^2}, $$

4
and the time-averaged RMS error as a function of space,

\[
E^X_i = \sqrt{\frac{1}{N_t} \sum_{k=1}^{N_t} \left[X_{k,i} - X_{tr,k,i}^r\right]^2},
\]

where \(X_{k,i}\) is the ensemble mean at time \(k\) and grid point \(i\), \(X_{tr,k,i}^r\) is the corresponding truth value and \(N_t\) and \(N_i\) correspond to the number of times and number of horizontal grid points, respectively.

Finally, the time-averaged analysis ensemble standard deviation (hereafter also called ensemble spread), \(\sigma^X_i\), is defined by

\[
\sigma^X_i = \sqrt{\frac{1}{N_t} \sum_{k=1}^{N_t} \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left[X_{k,i}^j - X_{k,i}\right]^2},
\]

where \(X_{k,i}^j\) is the value for ensemble member \(j\), and \(N_e\) represents the number of ensemble members.

a. Space-averaged diagnostics

Fig. 2 shows the space-averaged analysis RMS errors as a function of time for each ensemble member and the ensemble mean. All fields in the control experiment (Fig. 2, left panels) exhibit increasing ensemble spread with time due to the lack of observational constraint. Interestingly, the ensemble-mean error actually decreases with time. This result is due to the fact that the initial error is relatively large and the ensemble-mean boundary condition is set to the GFS six-hour forecast, which generally has small errors.

When surface pressure observations are assimilated, the ensemble-mean analysis errors and ensemble spread decrease toward stochastic equilibrium after two days at smaller values relative to the control case (Fig. 2, right panels). This result suggests that the 108 inhomogeneously distributed surface pressure observations efficiently constrain errors for this domain, even though no observations are assimilated over half of the domain. In particular the domain-average RMS analysis error for surface pressure is similar to or smaller than the 1 hPa observation error after two days, whereas corrections to non-observed variables also lead to significant reductions of both error and ensemble spread (Fig. 2c-h).

b. Time-averaged diagnostics

Fig. 3 gives the spatial distribution of the time-averaged ensemble-mean RMS analysis error for dry-air mass and the time-averaged ensemble spread. In the control experiment (left panels), persistent errors with large uncertainty are located along the West coast of Canada and over the eastern Pacific ocean, coincident with the areas of highest variance about the 7-day time mean (not shown).

In contrast, Fig. 3 (right panels) shows how the assimilation of surface pressure observations decreases both the error and ensemble variance in places near observations. Error and spread are smaller than observation error (i.e. \(\leq 1\) hPa) where the observation network is relatively dense, but are similar to the control experiment south of the Gulf of Alaska where
surface pressure observations are not available. Domain-averaged RMS analysis error and ensemble spread decrease by 50% and 64%, respectively, relative to the control case.

The assimilation of surface pressure observations also reduces the error in unobserved variables. Fig. 4 focuses on both 500-hPa geopotential height and surface wind speed, but similar results are obtained for both horizontal components of the wind at these levels (not shown). Similar to Fig. 3, errors are reduced in areas of dense observations. RMS errors (over space and time) and ensemble spread in the 500-hPa height field are reduced by 32% and 42%, respectively, compared to the control experiment. These results suggest that surface pressure observations drawn from the existing inhomogeneous ASOS network may constrain analysis errors not only at the surface but also in the middle and lower troposphere, consistent with results found by Whitaker et al. (2004). We note that the region of persistent errors in the northeast corner of the domain probably results from the fact that the flow during the period of interest is from the northwest, leaving little time for appropriate flow-dependent covariances to develop in this location.

For the surface wind speed field (Fig. 4e-h), the structure of time-averaged RMS errors and ensemble spread differ from those for the mass field in that the wind speed field has errors with smaller spatial scales, although errors are still generally larger over the ocean. The increased power in smaller spatial scales is attributed to the fact that the wind is approximately the derivative of the pressure field, which weights smaller scales. Despite the presence of several areas of persistent errors, mainly in data-sparse regions, the domain-averaged RMS ensemble-mean analysis errors and ensemble spread are reduced by 26% and 35%, respectively, relative to the control experiment.

c. Background-Analysis

In order to evaluate the performance of the data assimilation step, we consider the difference between the time-averaged RMS error of the ensemble-mean background and analysis fields,

\[
\Delta E^X_i = (E^X_i)_{\text{anal}} - (E^X_i)_{\text{back}}.
\]

Similarly, define the difference in the background and analysis ensemble spread by

\[
\Delta \sigma^X_i = (\sigma^X_i)_{\text{anal}} - (\sigma^X_i)_{\text{back}}.
\]

Positive values in these quantities indicate that, on average, the ensemble-mean analysis error and ensemble spread are smaller for the analysis than for the background. Fig. 5 shows that this is in fact the case over nearly the entire domain for the 500-hPa height and surface pressure fields. Note that the analysis errors are actually slightly larger than the background for the 500-hPa height field over small areas near the Gulf of Alaska and British Columbia. The largest average error reductions are near the West Coast of North America, co-located with much of the significant weather during this period, suggesting that observations along the coast are important in correcting short-term forecast errors in this region.

Domain averages of \(\Delta E^X_i\) and \(\Delta \sigma^X_i\) reveal the largest analysis-error reduction for the surface pressure fields (≈40%), with smaller reductions for unobserved fields (15% for 500-hPa height and 5% for surface wind speed). An important consideration in this comparison regards the fact that observation assimilation leads not only to better analyses but also to
better forecasts, which is apparent from $(E^X_i)^{\text{back}} < (E^X_i)^{\text{control}}$ for all fields (Figs. 3 and 4). For example, while surface wind-speed averaged RMS analysis errors are only 5.5% smaller than the averaged background errors, the RMS background errors are 22% smaller than in the control experiment.

5. Instantaneous Performance of the Filter

The previous section analysed the averaged performance of the limited-area EnKF in both space and time. Here, we consider instantaneous filter corrections by comparing background errors and analysis increments at individual assimilation times to reveal how the assimilation of surface pressure observations adjusts the background fields. As an illustrative example, Fig. 6 shows background errors (colors) and the analysis increment defined as the difference between analysis and background fields (negative values dashed) of four different quantities at 18 UTC 24 January. Ideally, the analysis increments would be the negative value of the background error at each grid point.

Analysis increments for surface pressure are accurate in both scale and amplitude when compared with background errors (Fig. 6a). More precisely, the large positive background errors centered over British Columbia are co-located with large negative analysis increments, and negative background errors near Alaska are co-located with positive analysis increments. Two localized regions have analysis increments of the wrong sign at this time: one over the Aleutian Islands and another over the ocean, near the center of the domain. The first region can be related to the prescribed background-error covariances on the lateral boundary, and the second is due to the lack of observations. Domain-averaged RMS pressure errors are reduced by 26% during this analysis cycle. As an alternative metric, we consider RMS errors averaged near each observation location. Surface stations with analysis errors smaller (larger) than background errors averaged within a 450 km radius of the stations are shown in green (red) circles. This reveals that 94% of the stations have analysis errors smaller than background errors (Fig. 6).

Analysis increments are also accurate for unobserved fields, as shown in Fig. 6(b-d). In general, regions characterized by large (in scale and amplitude) background errors typically experience accurate corrections. This result is particularly evident for the 500-hPa geopotential height field. In contrast, for 500-hPa temperature, although there exist two regions where large background errors are corrected, on a domain-averaged basis there is a slight increase in RMS analysis errors compared to background errors. This is a persistent issue during the entire period, which probably reflects weak correlations between surface pressure and mid-tropospheric temperature (e.g. Fig. 8 in Rabier et al. 1998; Fig 8 in Derber et al. 1999).

Recall from the synoptic discussion in section 3 that a shortwave trough moves southeastward along the West Coast of North America (Fig.1). This trough is associated with large errors in the control experiment (Fig. 7). Errors associated with this feature are reduced by the assimilation of only surface pressure observations in this experiment (Fig. 7). Specifically, after three six-hour analysis cycles, both the scale and magnitude of the errors are greatly reduced compared to both the control simulation (Fig. 7, left panels), and compared to the background (Fig. 7, right panels).

Finally, Fig 8 shows a scatterplot of the background errors versus the negative of the analysis increment at each of the observation locations for all assimilation times. The high
correlation coefficients for surface pressure and 500-hPa geopotential height indicate that the analysis increment plots presented in Fig. 6 and 7 are systematic during the entire period. The smaller correlation coefficient for surface wind speed can be related to the structure of the cross-variable covariances between surface pressure and wind speed, which may be expected to be maximized a few grid points from a surface pressure observation location since these variables are related approximately by a horizontal derivative. The small correlation coefficient in Fig. 8c indicates that the correlations between surface pressure and mid-tropospheric temperature are weak, as discussed above.

6. Different types of observations

The results of the previous sections suggest that only 108 surface pressure observations are sufficient to reduce both the ensemble-mean analysis error and the ensemble spread of different variables and through a deep layer of the troposphere. On the other hand, we have also shown that the mid-tropospheric temperature field is not as strongly affected by surface pressure observations due to small covariances. We consider now observations of winds and temperature at various vertical levels as would come from radiosoundings and ACARS data. The aim here is first to simulate aspects of a real-time data assimilation system based on conventional observations for the region of interest, and second to discuss the relative impact of each type of observations. We will also investigate the corrections of background errors in mid-tropospheric temperature.

Fig. 9 shows the difference between the time-averaged RMS errors of the ensemble mean background and analysis 500-hPa potential temperature (Eq. 1) and the difference in the ensemble spread (Eq. 2), based upon the assimilation of surface pressure (Fig. 9a,b), vertical soundings (Fig. 9c,d) and ACARS observations (Fig. 9e,f). Recall from above that surface pressure observations have relatively small impact on analysis errors and ensemble spread for mid-tropospheric temperature (Fig. 9a-b); however, errors in this field decrease considerably when radiosonde (Fig. 9c-d) and ACARS (Fig. 9e-f) observations are assimilated. Even though the locations of ACARS observations are time-dependent, a direct comparison of error reduction (Fig. 9e-f) at ACARS observation locations (Fig. 10) indicates a good correlation between analysis increments and dense regions of ACARS observations (predominantly distributed near major airports).

A comparison of the RMS errors associated with assimilating each of the different observation types supports and extends the previous results. Tables 2 and 3 show respectively the mean (over time and space) RMS analysis error and analysis ensemble spread for the assimilation of each observation type. It is apparent that the assimilation of a single observation type results in smaller analysis errors and ensemble spread as compared to the control experiment for both observed and unobserved variables. Furthermore, the assimilation of surface pressure observations leads to smaller errors and variance in the lower troposphere, whereas soundings and ACARS are more useful for reducing errors in the mid-troposphere. Interestingly, these results also suggest that errors in mid-tropospheric geopotential height are effectively reduced by surface pressure observations alone. A quantitative inter-comparison between different observation types is difficult since each observation type resolves different variables, levels and locations. In order to partly overcome this inhomogeneity, RMS errors and standard deviation have been averaged near the observation sites alone by the method described above (values in parenthesis in Tables 2 and 3). As might be expected, the smallest RMS errors are
realized when assimilating all observation types, whereas the relative influence of individual observation types depends on the field and on the vertical level considered.

7. Discussion and outlook

In this study we have tested the performance of a limited-area EnKF data assimilation system that resolves sub-synoptic scales on a domain with rich topography, inhomogenous observation density, and substantial flow normal to the domain boundaries. Three different types of synthetic observations have been assimilated over a seven-day period characterized by significant weather features. Single observation-type experiments show that both ensemble-mean error and ensemble spread are reduced for both observed and unobserved variables when compared to an ensemble control integration where no observations are assimilated.

Analysis errors are systematically smaller than background errors when averaged over the entire domain, over the entire time period, and near observation location; that is, the assimilation step reduces errors. A comparison of background errors and analysis increments at individual assimilation steps suggests that the analysis increments are accurate in both scale and amplitude, especially for the observed fields. Interestingly, we find that the assimilation of surface pressure observations is effective at constraining errors from the surface up through the mid-troposphere despite the use of an inhomogeneous observing network and the relatively small number of observations. Finally, whereas a quantitative comparison of the relative impact of each observation type strongly depends on the spatial distribution of observations, the results suggest that the largest error reduction is achieved when all observation types considered here are assimilated simultaneously.

These promising results offer a glimpse into how a data assimilation system at these sub-synoptic scales may perform given the current conventional observation network of surface stations, rawindsondes and aircraft (ACARS) data. However, the above results represent an upper bound on the EnKF performance when assimilating real observations since these experiments are based on the perfect model assumption. When assimilating real data, issues concerning model error and observation representativeness may affect the performance of the filter. With the goal of assimilating real observations in mind, we offer comments on each observation type below.

Surface pressure observations appear to be effective in constraining both lower tropospheric fields and geopotential height up through at least the middle troposphere. The existing surface pressure network is relatively dense over land and, although sparse, extends over the ocean in the form of buoys and ship reports. A significant issue with surface observations concerns representativeness due to differences between model topography and the actual station elevation, especially near complex terrain where stations are often located within unresolved valleys. To use these observations, quality control measures will be necessary so that observations are only assimilated for stations where model topography and station elevation are similar.

Although vertical soundings provide good vertical resolution at one location, the existing rawindsonde network is particularly sparse in the horizontal (e.g. only 25 stations over the domain considered here). Unlike surface stations, observation representativeness is less of an issue, especially above the surface.

ACARS observations may be a particularly useful source of information in regions where
rawindsone and surface pressure observations are sparse, especially over the Pacific Ocean (from transcontinental flights). However, ACARS observations are frequently over land in the form of airport approaches and takeoffs, and may therefore be subject to representativeness error, especially for valley airports. Finally, whereas observations of previous types are available only at prescribed time intervals, ACARS data is available continuously. For fixed assimilation intervals, a similar thinning procedure may be necessary to filter the observations. We note that, in general, the EnKF is not constrained to fixed assimilation intervals (synchronous observations), although high performance in the asynchronous case requires tighter coupling between the model and the EnKF assimilation algorithm.\footnote{The weak coupling between model and assimilation algorithm is usually regarded as a positive attribute of the EnKF. Tighter coupling in the case of asynchronous observations is required so that model fields are accessible in random access memory (fast), avoiding the need to frequently read and write ensemble data to disk (slow).}

Although our results show that analysis errors are effectively constrained over land, additional observation types are necessary to achieve similar performance over the ocean. One candidate is cloud-track wind observations. Although these winds have larger observation error compared to soundings and ACARS, the observation density is typically much higher. Satellite radiances are heavily used by many operational centers at synoptic scales, and therefore it may be useful to consider these observations in future experiments.

A final consideration for the application of EnKF techniques to limited-area domains concerns the specification of ensemble boundary conditions. Although the method used here to generate ensemble boundary conditions around global model analyses works well at 45 km resolution, it may not perform as well at higher resolution. Moreover, higher spatial resolution may require more frequent boundary updates, which may be lacking from global model fields (here GFS). One potential solution to these issues involves a nesting approach, whereby the boundary condition methods of Torn et al. (2005) are applied on a coarse outer grid each six hours. These coarse grids then provide natural ensemble boundary conditions for use on the boundaries of finer inner grids.

Acknowledgments.

This research was sponsored by National Science Foundation grant ITR-0205648 and National Oceanic and Atmospheric Administration C-STAR grant NA17RJ1232. SD was also supported by a postdoctoral fellowship from the Swiss National Foundation for Sciences.
References


Figure Captions

Figure 1. Time evolution of the 500-hPa geopotential height (solid lines every 50 m) and mean sea level pressure every 5 hPa (dashed lines every 50 m) in the truth run.

Figure 2. Time evolution of space-averaged RMS errors $[E^X_k]$ of the ensemble-mean analysis (solid lines) and of all individual ensemble members (dots) for the no-assimilation control experiment (left) and for the experiment with surface pressure observations (right). Corresponding time-mean RMS values are indicated in each panel. The fields considered are: (a,b) surface pressure (units: hPa); (c,d) 500 hPa geopotential height (units: m); (e,f) 500 hPa potential temperature (units: K); and (g,h) surface wind speed (m s$^{-1}$). The thin solid line in panels a,b represents observation error, and the dashed lines are drawn in other panels for reference.

Figure 3. Horizontal distribution of time-averaged RMS-errors $[E^p_{surf}]$ of the ensemble mean analysis (top) and of the time-averaged standard deviation $[\sigma^p_{surf}]$ of the analysis ensemble (bottom) for surface pressure over the 7-day period in the control experiment (left) and in the experiment where surface pressure observations are assimilated (right). Domain-averaged RMS values are displayed below each figure (units: hPa).

Figure 4. As in Fig. 3 except for 500-hPa geopotential height (a–d; unit: m) and for the surface wind speed (e–h; units: m s$^{-1}$).

Figure 5. Time-averaged RMS errors of the difference between the assimilation and no-assimilation experiments. (Eq. 1; left panels) and averaged standard deviations (Eq. 2; right panels) between background and analysis fields. Positive values indicate that observation assimilation systematically reduces errors, or ensemble spread. Also shown are the corresponding averages in RMS errors and the reduction between the analysis and the background (in %).

Figure 6. Instantaneous background errors (colored, $X^i_{back} - X^{tr}_i$) and analysis increments (contours, $X^i_{anal} - X^i_{back}$) for (a) surface pressure (units: hPa), (b) 500-hPa geopotential height (units: m), (c) surface wind speed (units: m/s), and (d) 500-hPa potential temperature (units: K) at 18 UTC 24 January 2004. Domain-averaged RMS errors and the associated reduction in RMS error (in %) are shown below each panel. Stations characterized by smaller RMS error in the analysis as compared to the background within 450 km of the observation location are colored green, and red otherwise.

Figure 7. Background error ($X^i_{back} - X^{truth}, y$-axis) versus the negative of analysis increments ($X^i_{back} - X^i_{anal}, x$-axis) at the observation locations for (a) surface pressure (units: hPa), (b) 500-hPa geopotential height (units: m), (c) 500-hPa potential temperature (units: °K), and (d) surface wind speed (units: m/s) when assimilating surface pressure observations. The linear regression and the ideal correction are given by the solid line dashed lines, respectively. Corresponding correlation coefficients are indicated below each figure.
Figure 8. As in Fig. 6, except for 500-hPa geopotential height for the control experiment (left) and the experiment with surface pressure observations (right).

Figure 9. Horizontal distribution of ACARS observations during the 7-day period. Low-, mid- and upper-level observations are defined by the 800-1050, 400-800 and 50-400 hPa layers, respectively.

Figure 10. Difference of time-averaged RMS errors (Eq. 1; left panels) and standard deviations (Eq. 2; right panels) between background and analysis 500-hPa potential temperature based upon the assimilation of surface pressure (top), vertical soundings (middle) and ACARS observations (bottom). Positive values denote areas where the filter systematically reduce errors, or spread, from the background values. The corresponding domain-averaged RMS error and the reduction in RMS error in the analysis relative to the background (in %) are shown below each panel.

Table 1. Characteristics of each observation type. Note that vertical soundings sample observations from the following ten levels: 925, 850, 700, 600, 500, 400, 300, 200, 150 and 100 hPa.

Table 2. Ensemble-mean RMS errors over time and space for each observation-type experiment. Values in parenthesis indicate averaged RMS errors averaged within 450 km of the observation locations.

Table 3. As in Table 2 except for the average standard deviation of the analysis ensemble $[\sigma_q(i)]$. 
Table 1: Characteristics of each observation type. Note that vertical soundings sample observations from the following ten levels: 925, 850, 700, 600, 500, 400, 300, 200, 150 and 100 hPa.

<table>
<thead>
<tr>
<th>Obs. type</th>
<th>stations</th>
<th>obs.</th>
<th>frequency</th>
<th>min. separation</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surf. pres. obs</td>
<td>108</td>
<td>108</td>
<td>6-hours</td>
<td>4 Δx</td>
<td>(p_{surf}: 1) mb</td>
</tr>
<tr>
<td>Vert. soundings</td>
<td>25</td>
<td>750</td>
<td>12-hours</td>
<td>4 Δx</td>
<td>(u, v: 1.5) m/s ; (θ: 1^\circ) K</td>
</tr>
<tr>
<td>ACARS</td>
<td>∼115</td>
<td>∼345</td>
<td>6-hours</td>
<td>2 Δx, 25 hPa</td>
<td>(u, v: 1.5) m/s ; (θ: 1^\circ) K</td>
</tr>
</tbody>
</table>

Table 2: Ensemble-mean RMS errors over time and space for each observation-type experiment. Values in parenthesis indicate averaged RMS errors averaged within 450 km of the observation locations.

<table>
<thead>
<tr>
<th>mean (E^q(i))</th>
<th>(p_{surf}) (hPa)</th>
<th>(z_{500}) (m)</th>
<th>(θ_{500}) (°K)</th>
<th>(u_{500}) (m/s)</th>
<th>(vel_{surf}) (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>2.02</td>
<td>16.68</td>
<td>1.25</td>
<td>4.12</td>
<td>2.11</td>
</tr>
<tr>
<td>Surf pres obs.</td>
<td>1.02 (0.58)</td>
<td>11.30 (10.97)</td>
<td>1.03 (1.17)</td>
<td>3.20 (3.41)</td>
<td>1.56 (1.22)</td>
</tr>
<tr>
<td>Soundings obs.</td>
<td>1.51 (1.26)</td>
<td>12.37 (11.72)</td>
<td>0.85 (0.82)</td>
<td>2.96 (2.25)</td>
<td>1.69 (1.14)</td>
</tr>
<tr>
<td>ACARS obs.</td>
<td>1.65 (1.55)</td>
<td>13.69 (13.79)</td>
<td>0.86 (0.69)</td>
<td>2.86 (1.99)</td>
<td>1.76 (0.97)</td>
</tr>
<tr>
<td>All obs.</td>
<td>1.03 (0.64)</td>
<td>9.63 (7.95)</td>
<td>0.78 (0.73)</td>
<td>2.61 (1.94)</td>
<td>1.50 (0.98)</td>
</tr>
</tbody>
</table>

Table 3: As in Table 2 except for the average standard deviation of the analysis ensemble [\(σ^q(i)\)].

<table>
<thead>
<tr>
<th>mean (σ^q(i))</th>
<th>(p_{surf}) (hPa)</th>
<th>(z_{500}) (m)</th>
<th>(θ_{500}) (°K)</th>
<th>(u_{500}) (m/s)</th>
<th>(vel_{surf}) (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>2.42</td>
<td>26.39</td>
<td>1.47</td>
<td>4.07</td>
<td>1.77</td>
</tr>
<tr>
<td>Surf pres obs.</td>
<td>0.88 (0.46)</td>
<td>15.30 (12.79)</td>
<td>1.13 (1.09)</td>
<td>2.73 (2.63)</td>
<td>1.14 (0.85)</td>
</tr>
<tr>
<td>Soundings obs.</td>
<td>1.28 (1.05)</td>
<td>13.64 (10.41)</td>
<td>0.71 (0.42)</td>
<td>2.02 (1.14)</td>
<td>1.03 (0.54)</td>
</tr>
<tr>
<td>ACARS obs.</td>
<td>1.47 (0.74)</td>
<td>15.12 (7.11)</td>
<td>0.88 (0.36)</td>
<td>2.30 (0.96)</td>
<td>1.26 (0.51)</td>
</tr>
<tr>
<td>All obs.</td>
<td>0.75 (0.33)</td>
<td>8.83 (4.06)</td>
<td>0.61 (0.32)</td>
<td>1.75 (0.89)</td>
<td>0.92 (0.42)</td>
</tr>
</tbody>
</table>
Figure 1: Time evolution of the 500-hPa geopotential height (solid lines every 50 m) and mean sea level pressure every 5 hPa (dashed lines every 50 m) in the truth run.
Figure 2: Time evolution of space-averaged RMS errors $[E_X^Y]$ of the ensemble-mean analysis (solid lines) and of all individual ensemble members (dots) for the no-assimilation control experiment (left) and for the experiment with surface pressure observations (right). Corresponding time-mean RMS values are indicated in each panel. The fields considered are: (a,b) surface pressure (units: hPa); (c,d) 500 hPa geopotential height (units: m); (e,f) 500 hPa potential temperature (units: K); and (g,h) surface wind speed (m s$^{-1}$). The thin solid line in panels a,b represents observation error, and the dashed lines are drawn in other panels for reference.
Figure 3: Horizontal distribution of time-averaged RMS-errors $|E_{i}^{\text{surf}}|$ of the ensemble mean analysis (top) and of the time-averaged standard deviation $\sigma_{i}^{\text{surf}}$ of the analysis ensemble (bottom) for surface pressure over the 7-day period in the control experiment (left) and in the experiment where surface pressure observations are assimilated (right). Domain-averaged RMS values are displayed below each figure (units: hPa).
Figure 4: As in Fig. 3 except for 500-hPa geopotential height (a–d; unit: m) and for the surface wind speed (e–h; units: m s$^{-1}$).
Figure 5: Time-averaged RMS errors of the difference between the assimilation and no-assimilation experiments. (Eq. 1; left panels) and averaged standard deviations (Eq. 2; right panels) between background and analysis fields. Positive values indicate that observation assimilation systematically reduces errors, or ensemble spread. Also shown are the corresponding averages in RMS errors and the reduction between the analysis and the background (in %).
Figure 6: Instantaneous background errors (colored, $X_i^{\text{back}} - X_i^{\text{tr}}$) and analysis increments (contours, $X_i^{\text{anal}} - X_i^{\text{back}}$) for (a) surface pressure (units: hPa), (b) 500-hPa geopotential height (units: m), (c) surface wind speed (units: m/s), and (d) 500-hPa potential temperature (units: K) at 18 UTC 24 January 2004. Domain-averaged RMS errors and the associated reduction in RMS error (in %) are shown below each panel. Stations characterized by smaller RMS error in the analysis as compared to the background within 450 km of the observation location are colored green, and red otherwise.
Figure 7: As in Fig. 6, except for 500-hPa geopotential height for the control experiment (left) and the experiment with surface pressure observations (right).
Figure 8: Background error ($X_{\text{back}} - X_{\text{truth}}$, y-axis) versus the negative of analysis increments ($X_{\text{back}} - X_{\text{anal}}$, x-axis) at the observation locations for (a) surface pressure (units: hPa), (b) 500-hPa geopotential height (units: m), (c) 500-hPa potential temperature (units: °K), and (d) surface wind speed (units: m/s) when assimilating surface pressure observations. The linear regression and the ideal correction are given by the solid line dashed lines, respectively. Corresponding correlation coefficients are indicated below each figure.
Figure 9: Difference of time-averaged RMS errors (Eq. 1; left panels) and standard deviations (Eq. 2; right panels) between background and analysis 500-hPa potential temperature based upon the assimilation of surface pressure (top), vertical soundings (middle) and ACARS observations (bottom). Positive values denote areas where the filter systematically reduce errors, or spread, from the background values. The corresponding domain-averaged RMS error and the reduction in RMS error in the analysis relative to the background (in %) are shown below each panel.
Figure 10: Horizontal distribution of ACARS observations during the 7-day period. Low-, mid- and upper-level observations are defined by the 800-1050, 400-800 and 50-400 hPa layers, respectively.