Reanalysis without radiosondes using ensemble data assimilation

Jeffrey S. Whitaker*, **Gilbert P. Compo,**

Xue Wei and Thomas M. Hamill

NOAA-CIRES Climate Diagnostics Center, Boulder, CO

to appear *Mon. Wea. Rev.* May 2004 issue.

15th April 2004

* *Corresponding author address:* NOAA Climate Diagnostics Center, 325 Broadway R/CDC1, Boulder, CO 80305-3328, Jeffrey.S.Whitaker@noaa.gov

Abstract

Studies using idealized ensemble data assimilation systems have shown that flow-dependent background-error covariances are most beneficial when the observing network is sparse. The computational cost of recently proposed ensemble-data assimilation algorithms is directly proportional to the number of observations being assimilated. Therefore, ensemble-based data assimilation should both be more computationally feasible and provide the greatest benefit over current operational schemes in situations when observations are sparse. Reanalysis before the radiosonde era (pre-1931) is just such a situation.

The feasibility of reanalysis before radiosondes using an ensemble square-root filter (EnSRF) is examined. Real surface pressure observations for 2001 are used, sub-sampled to resemble the density of observations we estimate to be available for 1915. Analysis errors are defined relative to a three-dimensional variational (3DVar) analysis using several orders of magnitude more observations, both at the surface and aloft. We find that the EnSRF is computationally tractable and considerably more accurate than other candidate analysis schemes which use static background-error covariance estimates. We conclude that a Northern Hemisphere reanalysis of the middle and lower troposphere during the first half of the 20th century is feasible using only surface pressure observations. Expected Northern Hemisphere analysis errors at 500 hPa for the 1915 observation network are similar to current 2.5 day forecast errors.

1. Introduction

A primary goal of climate change research is to understand variations in the frequency and intensity of severe weather events on decadal and longer time-scales. An obvious prerequisite for achieving this goal is an accurate baseline estimate of the frequency and intensity of severe weather over the last century. Analyses of long-term changes in extra-tropical cyclone frequency and intensity have been hampered by the inadequacy of current datasets (IPCC 2001, p. 163). Since synoptic-scale weather systems have time-scales of less than a week, a century-long dataset of tropospheric circulation fields at daily resolution is required. The NCEP-NCAR 50-year reanalysis (Kistler and Coauthors 2001) provides four times daily gridded circulation fields beginning in 1948, when digital upper-air observations became widely available. The only daily tropospheric circulation dataset available that extends back before 1948 is derived from charts of sea-level pressure hand-drawn by U.S. Air Force meteorologists in the 1940's and 1950's (United States Weather Bureau 1944). Although a remarkable achievement for its time, this original reanalysis suffers from serious problems associated with incorrect assumptions made by the analysts in data sparse regions (Jones 1987; Trenberth and Paolino 1980) and does not provide estimates of the full three-dimensional tropospheric structure. Clearly, a better dataset is needed - but is it possible to create a more accurate daily tropospheric circulation dataset for the first half of the 20th century given the paucity of available observations?

In this study we examine whether advanced data assimilation systems have significant advantages over currently available systems for sparse networks of surface pressure obser-

vations representative of the early part of the 20th century. Specifically, we examine the performance of the ensemble-based data assimilation system described by Whitaker and Hamill (2002) applied to a simulated 1915 observing network, created by sub-sampling the surface pressure observations for 2001. We focus on surface pressure observations since they are the most widely available and reliable observations in the early 20th century, and provide more information about the state of the free troposphere than surface wind and temperature observations. In a companion report we will carefully examine the available data record for the last 150 years and assess the performance of several analysis schemes with surface-only observation networks representative of 1890 to 1940.

Previous studies using idealized ensemble data assimilation systems (e.g. Hamill and Snyder 2000) have shown that their flow-dependent background-error covariances are most beneficial when the observing network is sparse. When observations are very dense, the background-error covariances do not change as much from time to time, so static background-error covariance models (such as used in 3DVar) can be nearly as effective. In addition, the computational cost of computing analysis increments in recently proposed ensemble-data assimilation algorithms (Houtekamer and Mitchell 2001; Whitaker and Hamill 2002) is directly proportional to the number of observations being assimilated. Therefore, ensemble-based data assimilation should be more computationally feasible and provide the greatest benefit over current operational schemes in situations when observations are sparse. Reanalysis before the radiosonde era is just such a situation.

The paper is organized as follows: section 2 contains a description of the experimental de-

sign, including the simulation of the 1915 observing network, the ensemble data assimilation system and forecast model. Section 3 presents the results of the assimilation experiments, which show that the EnSRF can produce mid-tropospheric analyses given surface observations at 1915 densities which are as accurate as 2.5 day forecasts are today. Section 4 contains a summary of the results and a discussion of the unresolved issues that need to be addressed before these techniques can be applied to a real reanalysis of the first half of the 20th century.

2. Experimental Design

a. The Observations

To simulate how a modern data assimilation system can be expected to perform on a historical observational network, the 2001 observational network was reduced to only surface observations with a density typical of 1915. The number of synoptic observations potentially available for each month during the period 1913-1917 was determined in 5x5 degree boxes from a detailed inventory of the digital land surface data holdings of the National Center for Atmospheric Research (NCAR), the National Climatic Data Center (NCDC), the Waves and Storms dataset (Schmith et al. 1997), and manuscript data holdings of NCDC. The Global Historical Climate Network (GHCN) surface pressure station locations were used a proxy for synoptic reports currently available only in manuscript form, some of which are now being digitized by NCDC (S. Doty, personal communication), Environment Canada (V. Swail,

personal communication), the European Union (P. Jones, personal communication) as well as other international efforts (R. Jenne, personal communication). The marine observations available were also determined in 5x5 degree boxes from a detailed inventory of ICOADS Release 2.0 (Woodruff et al. 1998; Diaz et al. 2002), the German Marine Meteorological archive (courtesy of V. Wagner), and the Kobe Collection 2001 (Manabe 1999). The 1915 observation network was chosen for this study since it is representative of data availability during the earlier part of the 20th century (Fig. 1). The number of surface observations increases dramatically in the 1930's and 1940's.

The quality-controlled observations used as input to the NCEP-NCAR reanalysis (Kistler and Coauthors 2001) for 2001 were sub-sampled to simulate the 1915 network. The 2001 observations were first reduced by retaining only surface pressure observations from radiosonde and marine reports issued within 30 minutes of the analysis time. The location of the radiosonde stations gives an excellent approximation to the location of historically available land surface pressure stations. This reduced the total network from over 150,000 observations to less than 2000 per analysis. The simulated historical network was then constructed by randomly selecting from the reduced network in each five-degree box with a probability equal to the average number of daily historical surface pressure observations in the box normalized by the average number of daily surface pressure observations in the reduced 2001 network. To replicate the historical station temporal inhomogeneity, this random selection from the reduced network is performed for each analysis. The network reduction and sampling procedures together produce a simulated land station network with locations that are

nearly constant but whose temporal sampling is as variable as expected historically. The simulated marine network reproduces the expected ship tracks. Figure 2 shows a map of the probabilities assigned to each five-degree box. Surface temperature observations were included with each surface pressure observation (although surface temperature observations were not assimilated, they were used to reduce the surface pressure observation to the model orography as discussed in the next paragraph). A map illustrating a typical simulated surface pressure network at 00 and 12 UTC is shown in the right hand panel of Figure 5. In this example there are 204 surface pressure and temperature observations in the Northern Hemisphere poleward of 20^oN. At 06 and 18 UTC, the number of surface marine observations is nearly the same, but there are almost no observations over land areas.

Observational error standard deviations were the same as those used in the NCEP-NCAR reanalysis, 1.6 hPa for ship observations and 1 hPa for land stations. In situations where the absolute difference between the model orography and the real orography is less than 600 meters at the observation location, and a co-located temperature observation is available, the surface pressure observation is reduced to the model orography assuming the mean temperature in the intervening layer is $T_{ob} + \frac{1}{2}\Gamma\Delta z$, where T_{ob} is the temperature observation, Γ is $-6.5^o\mathrm{Kkm}^{-1}$ and Δz is the difference between the model and real orography. The observation error is adjusted accordingly, assuming that the error in the estimate of the lapse rate Γ is $3^o\mathrm{Kkm}^{-1}$. If $|\Delta z| > 600$ m then the surface pressure observation is not used. If a co-located temperature observation is not available, the surface pressure observation is used without modification if $|\Delta z| < 10$ m, otherwise it is not used.

To assess the benefit of flow dependent background-error covariances, the analyses produced by the ensemble data assimilation system are compared to those produced by two other simpler systems with static background-error covariance estimates. The 3DVar system used to produce the NCEP-NCAR reanalysis, a.k.a the Climate Data Assimilation System, or CDAS (Kistler and Coauthors 2001), was adapted to the 1915 observation network by multiplying the background-error covariances used in the reanalysis by a constant factor > 1, and turning off the divergence tendency constraint. We call this modified CDAS system CDAS-SFC.

Increasing the background-error covariance amplitude in the CDAS-SFC system was necessary since the background-error covariances used in CDAS were tuned to the modern observing network (Kistler and Coauthors 2001), with several orders of magnitude more observations. By trial and error we settled on a factor of 16. Values less than 16 produced an inferior analysis, while values greater than 16 did not result in a significantly better analysis. The spatial structure of the background-error covariances (described in Parrish and Derber (1992)) was not modified.

By performing single observation assimilation experiments we found that the divergence tendency constraint severely limited the size of the analysis increments when the observation increment (first-guess minus observation) was large. With the constraint turned off, the CDAS-SFC system produced a larger analysis increment. The divergence tendency constraint in the CDAS system was intended to control the excitation of large-amplitude gravity waves in the analysis. However we observed no significant increase in gravity wave noise after

turning off this constraint, but we did find a significant reduction in analysis error.

The CDAS-SFC and EnSRF assimilation cycles were started on 15 November 2001 and run through December 2001. The reanalysis fields for the same calendar day in the previous year were used to start the CDAS-SFC analysis cycle. The EnSRF analysis was initialized with a random sample of reanalysis states from Novembers 1971-2000. Only results for December 2001 (using observations sub-sampled at 1915 densities in the manner described above) are presented, the analyses for the 15-day spin-up period are discarded.

A simple statistical interpolation (SI) analysis was also performed, using the 1971-2000 reanalysis climatology as a first guess, and climatological anomaly covariances from the reanalysis as a model for the background-error covariances. The procedure for performing the SI analysis is exactly the same as the procedure used to perform the EnSRF analysis described in the following section, except that instead of an ensemble of model forecasts a random ensemble of NCEP-NCAR reanalysis states is used to compute the background-error covariances.

For all the analysis experiments, analysis error was estimated by computing the root-mean squared difference with the NCEP-NCAR reanalysis for 2001. Because the reanalysis used several orders of magnitude more observations, including radiosondes, aircraft and satellite soundings, we expect that this difference is significantly larger than the error in the reanalysis itself.

b. The Ensemble Data Assimilation System

Ensemble data assimilation systems transform a forecast ensemble into an analysis ensemble with appropriate statistics. This can be done stochastically, treating the observations as random variables, (e.g. Houtekamer and Mitchell 1998; Burgers et al. 1998), or deterministically, requiring that the covariance of the updated ensemble satisfy the Kalman filter analysis error covariance equation. Deterministic analysis ensemble updates are Monte-Carlo implementations of Kalman square-root filters, hence we call them ensemble square-root filters. Ensemble square-root filters are not unique (Tippett et al. 2003), since different ensembles can have the same covariance. This non-uniqueness has led to the development of several different algorithms for updating the analysis ensemble (Bishop et al. 2001; Anderson 2001; Whitaker and Hamill 2002). Here we implement the latter variant, which reduces to a particularly simple form when observations are assimilated serially (one after another).

Following the notation of Ide et al. (1997), let $\mathbf{x}^b$ be an m-dimensional background model forecast; let $\mathbf{y}^o$ be a p-dimensional set of observations; let $\mathbf{H}$ be the operator that converts the model state to the observation space; let $\mathbf{P}^b$ be the $m \times m$-dimensional background-error covariance matrix; and let $\mathbf{R}$ be the $p \times p$-dimensional observation-error covariance matrix. The minimum error-variance estimate of the analyzed state $\mathbf{x}^a$ is then given by the traditional Kalman filter update equation (Lorenc 1986),

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b), \tag{1}$$

where

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T+\mathbf{R})^{-1}. \tag{2}$$

The analysis-error covariance $\mathbf{P}^a$ is reduced by the introduction of observations by an amount

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b(\mathbf{I}\text{-}\mathbf{K}\mathbf{H})^T+\mathbf{K}\mathbf{R}\mathbf{K}^T = (\mathbf{I}\text{-}\mathbf{K}\mathbf{H})\mathbf{P}^b. \tag{3}$$

In ensemble data assimilation, $\mathbf{P}^b\mathbf{H}^T$ is approximated using the sample covariance estimated from an ensemble of model forecasts. For the rest of the paper, the symbol $\mathbf{P}$ is used to denote the sample covariance from an ensemble, and $\mathbf{K}$ is understood to be computed using sample covariances. Expressing the variables as an ensemble mean (denoted by an over-bar) and a deviation from the mean (denoted by a prime), the update equations for the EnKF may be written as

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{K}(\overline{\mathbf{y}}^o\text{-}\mathbf{H}\overline{\mathbf{x}}^b), \tag{4}$$

$$\mathbf{x}'^a = \mathbf{x}'^a + \widetilde{\mathbf{K}}(\mathbf{y}'^o - \mathbf{H}\mathbf{x}'^b), \tag{5}$$

where $\mathbf{P}^b\mathbf{H}^T = \overline{(\mathbf{H}\mathbf{x}'^b)\mathbf{x}'^{bT}} \equiv \frac{1}{n-1}\sum_{i=1}^{n}\mathbf{H}\mathbf{x}_i'^b\mathbf{x}_i'^{bT}$, n is the ensemble size (= 100 unless otherwise noted), $\mathbf{K}$ is the traditional Kalman gain given by (2) and $\widetilde{\mathbf{K}}$ is the gain used to update deviations from the ensemble mean. Note that an over-bar used in a covariance estimate implies a factor of $n-1$ instead of n in the denominator, so that the estimate is unbiased. In the EnKF, $\widetilde{\mathbf{K}} = \mathbf{K}$, and $\mathbf{y}'^o$ are randomly drawn from the probability distribution of observation

errors (Burgers et al. 1998). This choice of $\mathbf{y}^{'o}$ ensures that for an infinitely large ensemble, (3) will be satisfied exactly (Burgers et al. 1998). However, as pointed out by Whitaker and Hamill (2002), for a finite ensemble (3) will not be satisfied exactly, and the noise added to the observations acts as an extra source of sampling error, degrading the performance of the filter. In the EnSRF, $\mathbf{y}^{'o}= \mathbf{0}$ and $\widetilde{\mathbf{K}}$ is given by

$$\widetilde{\mathbf{K}}= \mathbf{P}^b\mathbf{H}^T \left[(\sqrt{\mathbf{H}\mathbf{P}^b\mathbf{H}^T+\mathbf{R}})^{-1}\right]^T (\sqrt{(\mathbf{H}\mathbf{P}^b\mathbf{H}^T+\mathbf{R})}+\sqrt{\mathbf{R}})^{-1} \tag{6}$$

(Andrews 1968). This choice guarantees that (3) is satisfied exactly. If $\mathbf{R}$ is diagonal, observations may be assimilated serially, the analysis after assimilation of the Nth observation becomes the background estimate for assimilating the $(N+1)th$ observation (Gelb et al. 1974), and the above expression simplifies to

$$\widetilde{\mathbf{K}}= \left(1+\sqrt{\frac{R}{\mathbf{H}\mathbf{P}^b\mathbf{H}^T+R}}\right)^{-1} \mathbf{K}, \tag{7}$$

where R and $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$ are scalars, while $\mathbf{K}$ and $\widetilde{\mathbf{K}}$ are vectors of the same dimension as the model state vector. This was first derived by J. Potter in 1964 (Maybeck 1979). Although (6) requires the computation of two matrix square-roots, the serial processing version (7) requires the computation of only a scalar factor to weight the traditional Kalman gain, and therefore is no more computationally expensive than the EnKF.

As discussed in Whitaker and Hamill (2002), sampling error can cause filter divergence in any ensemble data assimilation system, so some extra processing of the ensemble covari-

ances is usually necessary. The two techniques used here are distance-dependent covariance localization (Houtekamer and Mitchell 2001; Hamill et al. 2001) and covariance inflation (Anderson and Anderson 1999).

Covariance localization is a filter that forces the ensemble covariances to go to zero at some horizontal distance L from the observation being assimilated. It is intended to counter the tendency for ensemble variance to be excessively reduced by spurious long-range correlations between analysis and observations points. For all the experiments shown here, L is set to 5000 km and the horizontal structure of the filter is the same as used in Whitaker and Hamill (2002). We also use a covariance filter in the vertical which forces ensemble covariances to go to zero at $\sigma = 0.05$ (roughly 50 mb). The vertical covariance localization function has a value of 1 below $\sigma = 0.2$, zero above $\sigma = 0.05$, and decreases linearly in σ these levels. The same covariance filter is also applied to the SI analysis scheme, which uses a 100-member ensemble of randomly chosen December reanalysis states, instead of an ensemble of 6-h model forecasts.

Covariance inflation simply inflates the deviations from the ensemble mean first-guess by a factor $r > 1.0$ for each member of the ensemble, before the computation of the background-error covariances and before any observations are assimilated. We have found that different inflation factors were required in the Northern and Southern Hemisphere, due to the large differences in the density of the observing networks. In the limit that there are no observations influencing the analysis in a given region, it is easy to see how inflating the ensemble every analysis time can lead to unrealistically large ensemble variances, exceeding the clima-

tological variance. The simulated 1915 network has very few observations in the Southern Hemisphere extratropics, generally less than 20 per analysis time. Therefore, only a very small inflation factor is needed there. In the Northern Hemisphere, however, we found it necessary to use a significantly larger inflation factor to avoid filter divergence. For all the results presented here, we use an inflation factor that varies smoothly across the equator at $\sigma = 1$ from a value $r = 1.07$ in the Northern Hemisphere to a value of $r = 1.007$ in the Southern Hemisphere. Our use of vertical covariance localization means that no observations ever affect the analysis above $\sigma = 0.05$. Therefore, in order to avoid the development of excessive ensemble variance in the stratosphere we use a vertically varying r which is constant from $\sigma = 1$ to $\sigma = 0.2$, then decreases linearly to unity at $\sigma = 0.05$.

Observations of surface pressure are processed serially. The ensemble mean and ensemble deviations are updated using equations 4 and 5, with the Kalman gains given by equations 2 and 7. Covariance inflation is applied to the background ensemble deviations before assimilating any observations. Only those grid points within L km of the observation are updated, where $L = 5000$km is the covariance localization length scale. The analysis is performed on a 128x64 Gaussian grid, and the forward operator $\mathbf{H}$ represents bi-linear interpolation to the observation location (and, if necessary, a reduction to the model orography as described in section 2a). The analysis updates for each observation are parallelized by partitioning the model state vector by vertical level. For example, with our vertical covariance localization only the lowest 23 (out of 28) model levels are updated. To run the assimilation on 4 processors, 23 levels of zonal wind are updated on processor 1, 23 levels of meridional wind on

processor 2, 23 levels of temperature on processor 3, and 23 levels of specific humidity plus surface pressure on processor 4. The only communication between processors needed during the analysis update involves the propagation of $\mathbf{Hx}^b$ from the processor updating surface pressure (processor 4 in this example) to the other processors. The results shown here were run on 31 2.2 GHz Intel processors. Only 6 seconds wall-clock time were needed to process 417 surface pressure observations for a 100-member ensemble with $L = 5000$ km, or about 0.015 seconds per observation. Running the forecast ensemble to generate the background estimates for the next assimilation time is trivially parallel, since every ensemble member runs independently on a separate processor.

Although we do not employ an explicit treatment of model error, our implementation of covariance localization and inflation can be considered a crude parameterization of model error. Covariance inflation increases the magnitude of the background-error covariance estimate. This is necessary to deal with biases caused by sampling error (as discussed by Whitaker and Hamill (2002)), but it can also be thought of as accounting for unrepresented model-error covariance. This can only account for model errors that are in the same subspace as the background ensemble. Covariance localization forces the covariance between background-error estimates at two locations to go to zero as the distance between the two locations increases. This is necessary to deal with sampling error, even in the absence of model error. However, the effect of covariance localization is to increase the rank of the background-error covariance estimate (Hamill et al. 2001). The extra degrees of freedom introduced into the background error estimate can be thought of as representing model errors

that do not project onto the sub-space spanned by the background ensemble.

c. The Forecast Model

The forecast model used is a recent version of the NCEP global medium-range forecast model (MRF), which was operational until mid-1998. The model is spectral with a triangular truncation at wavenumber 62, with 28 sigma levels. A detailed description of the model physics can be found in Wu et al. (1997). Boundary conditions are taken from the NCEP-NCAR reanalysis and are the same for each ensemble member. No initialization is performed during the CDAS-SFC or EnSRF analysis cycles.

3. Assimilation Results

Figure 3 summarizes the assimilation results for the simulated December 1915 network. Shown are time series of the root-mean square (RMS) analysis error for (A) mean sea-level pressure (MSLP) and (B) 500 hPa geopotential height (Z500). The analysis error (defined relative to the NCEP-NCAR reanalysis and averaged over the Northern Hemisphere poleward of 20^oN) is shown for the EnSRF (black curve) , SI (red curve), and CDAS-SFC (blue curve) analyses, along with the spread of the EnSRF analysis (green curve). The CDAS-SFC analysis is significantly better than the SI analysis, indicating that using a six-hour model forecast as a first-guess in a 3DVar analysis is an improvement over a climatological background, even for this small number of surface pressure observations. However, the EnSRF

analysis is about 50% more accurate than the CDAS-SFC analysis. In fact, there is not a single case in the one-month period where the CDAS-SFC is as accurate as the EnSRF. We note that since we have made no attempt to tune the *structure* of the background-error covariances used in CDAS for this very sparse observation network, it is possible that a better specification of static background-error covariances in the CDAS-SFC analysis system would improve the CDAS-SFC analysis. The SI analysis, which uses climatology as a background, has a large diurnal cycle associated with the differences in the number of observations available at 00 (12) UTC and 06 (18) UTC. This occurs because there is no mechanism in the SI analysis to propagate information forward in time.

Averaged over the Northern Hemisphere extratropics, the spread in the EnSRF has nearly the same mean value as the RMS error, indicating that our implementation of covariance localization and inflation has successfully countered the loss of ensemble variance expected from sampling error and the lack of an explicit treatment of model error. The pattern of ensemble spread, averaged over the month-long assimilation period, is also quite similar to the pattern of the ensemble mean RMS error (Fig. 4). However, close inspection of Figure 4 reveals that in data-dense (data-sparse) regions, the ensemble mean spread is somewhat smaller (larger) than analysis error. This is in part due to our implementation of covariance inflation. Over data-sparse areas that are only weakly influenced by observations, covariance inflation increases ensemble variance too much. We have tuned the inflation factor so that the Northern Hemisphere average value of spread is similar to analysis error, consequently ensemble variance over data-dense areas must be deficient in order to balance the tendency

for ensemble spread to be too large over data-sparse areas. This behavior may also be in part a manifestation of spatially correlated observation error, perhaps associated with "representativeness error" (e.g. Liu and Rabier 2002). Our analysis algorithm assumes that the errors for each observation are independent. If observation errors are actually spatially correlated, the filter may reduce the variance of the ensemble too much. Further study is needed to assess the impact of mis-specification of observation-error covariances on the performance of ensemble data assimilation systems, particularly with dense observation networks.

The inhomogeneity of the ensemble variance shown in Figure 4 helps explain how the EnSRF is able to outperform the CDAS-SFC system, which assumes that the background-error variance is a function of latitude only (Parrish and Derber 1992). A 3DVar system could probably be tuned to the 1915 surface pressure network to produce a better estimate of the spatial varying nature of the error in the first guess, and hence a better analysis. However, the EnSRF is able to do this with little tuning (only the specification of the covariance localization and inflation parameters). We consider the ability of the EnSRF to adapt to large changes in the structure of the observing network to be a very desirable property for an analysis system to be used for a reanalysis of the entire 20th century.

Also included in Figure 4 is a map showing the RMS error of 60-hour forecasts run with same version of the MRF used in the assimilation experiments, but initialized from the 00 UTC reanalyses for all Decembers between 1979 and 2002[1]. The mean RMS error averaged over the Northern Hemisphere extratropics is 39 m, very close to the mean December EnSRF analysis error for the simulated 1915 network (Fig. 3). Therefore, we expect that a reanalysis

[1]These forecasts were run as part of a separate "re-forecast" project described in Hamill et al. (2003).

of the early 20th century using only surface pressure observations should be about as accurate at 500 hPa (in a Northern Hemisphere average sense) as 60-h forecasts are today. In fact, Fig. 4c shows that the EnSRF is actually more accurate than modern 60-h forecasts over most of the Northern Hemisphere, except over the polar region and Asia, where there are voids in the simulated 1915 observation network.

A sample 500 hPa EnSRF analysis is shown in Figure 5 for 00UTC December 14, along with the corresponding map from the NCEP-NCAR reanalysis (using all available observations). The EnSRF analysis, using only a few hundred surface pressure observations, is clearly able to reproduce most of the significant mid-tropospheric flow features present in the reanalysis, including the synoptic-scale short waves.

One source of error in the EnSRF scheme is associated with sampling error in the estimate of the background-error covariances. Assuming this source of error is significant compared to other possible sources of error (e.g. mis-representation of model-error or observation-error covariance) then increasing the ensemble size should improve the accuracy of the EnSRF analysis. We have performed additional assimilation experiments with 50 and 200 member ensembles to assess the impact of ensemble size on analysis quality. The 50 (200) member ensemble was run with a covariance localization length scale L of 4000 (6000) km (the 100 member ensemble experiment was run with $L = 5000$ km). Since covariance localization is intended primarily to remove spurious long-range correlations between analysis and observations points which arise from sampling error less localization should be needed for larger ensembles. The covariance inflation parameters for the 50 and 200 member experiments were

fixed to the values used in the 100 member experiment. Figure 6 shows time series of 500 hPa analysis error for these experiments. For the 16-day assimilation period, the error of the 50-member ensemble is about 8% more than the 100-member ensemble. Increasing the ensemble size from 100 to 200 members results in a very slight reduction in analysis error. From these results we conclude that increasing the ensemble size beyond 100 provides little benefit. This could either be because all sources of analysis error for this sparse observation network are well sampled with an ensemble of 100, or more likely, other sources of error, such as model-error, become relatively more important than sampling error when ensemble sizes exceed 100.

We have also performed experiments using observations for June 2001 sub-sampled to simulate the 1915 network (not shown). The 500 hPa geopotential height RMS Northern Hemisphere analysis errors for both the EnSRF and CDAS-SFC analyses are about the same for June as they are for December. However, since the climatological variance of 500 hPa geopotential height is smaller in June than December, the analysis errors expressed in terms of anomaly correlation decline in June relative to December (0.85 versus 0.95 for the EnSRF). When compared to forecasts initialized from the NCEP/NCAR reanalysis for all Junes from 1979-2001, the EnSRF analyses generated from the simulated June 1915 network have an RMS 500 hPa geopotential height error roughly equivalent to the 84-h forecast (as compared to 60-h forecasts for the simulated December 1915 network). The degradation in the June analysis errors relative to December, which occurs for all candidate analysis schemes, is likely due to the fact that covariances between surface pressure and other variables at other levels

in the troposphere is larger in winter, when coherent baroclinic systems are more prevalent.

4. Discussion and Conclusions

A Northern Hemisphere reanalysis of the middle and lower troposphere for the first half of the 20th century is feasible using only surface pressure observations. Ensemble data assimilation techniques are particularly well suited to the task, and can be expected to produce 500 hPa analyses with errors similar to current 2-3 day forecasts.

Before such a reanalysis can be undertaken, methods for quality control of the historical surface pressure observations will need to be developed. Background-error covariance estimates from the ensemble data assimilation system could provide a basis for a simple "background check" (Dee et al. 2001) which marks as suspect all observations whose deviation from the ensemble mean first guess is greater than some factor times the background-error variance estimate at the observation location. Typically, those observations flagged as suspect by the background check are then subjected to a "buddy check" (ibid), which compares suspect observations to nearby observations which passed the background check. Unfortunately, when observations are sparse, there may not be very many "buddies". Observation errors are likely to be larger than they are assumed to be in this study, particularly for ships. Currently we have no reliable estimates of surface pressure observation error for observations taken in the early 20th century.

Data assimilation is typically used to generate initial conditions for numerical weather forecasts. Therefore, each analysis is based upon only current and past observations. How-

ever, when producing a retrospective reanalysis, one is free to use all available observations, including those data collected after the analysis time. A Kalman smoother is a direct generalization of the Kalman filter which incorporates observations both before and after the analysis time. Whitaker and Compo (2002) introduced an ensemble square-root smoother (EnSRS), which is a generalization of the EnSRF based upon the fixed-lag Kalman smoother proposed by Cohn et al. (1994). We will be investigating the use of the EnSRS to see if using observations taken past the analysis time can improve upon the accuracy of the EnSRF analyses presented here.

Ensemble data assimilation systems are only now moving from the realm of perfect model, "identical twin" experiments to real world cases with actual observations. There is still research needed to fully realize the potential that ensemble data assimilation holds for improving analyses and forecasts. In particular, our results suggest that spatially correlated "errors of representation" (which are incorporated into the overall observation error in most data assimilation schemes) may adversely affect the performance of ensemble filters which assume spatially uncorrelated observation errors. In addition, parameterizations of model error that are more sophisticated than the combination of covariance localization and inflation employed here will almost certainly improve the performance of ensemble data assimilation systems.

Our results demonstrate that with some further development, advanced ensemble data assimilation systems and the available surface pressure observations could be used to create a reanalysis of the entire 20th century. Such a dataset would be useful for determining decadal

variations of synoptic-scale variability in the Northern Hemisphere extratropics.

Acknowledgments

Fruitful discussions with J. Barsugli, C. Bishop, E. Kalnay, P. Sardeshmukh, and C. Snyder are gratefully acknowledged. This project would not have been possible without the support of the University of Colorado's Cooperative Institute for Research in Environmental Science, which provided an Innovative Research grant, and NOAA's Forecast Systems Laboratory, which provided access to their High Performance Computing System. One of us (Hamill) was supported by NSF grant ATM0112715. Compo was supported by a grant from NOAA's Office of Global Programs.

References

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.

Andrews, A., 1968: A square root formulation of the Kalman covariance equations. *AIAA J.*, **6**, 1165–1168.

Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.*, **129**, 420–436.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Cohn, S. E., N. S. Sivakumaran, and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.*, **122**, 2838–2867.

Dee, D. P., L. Rukhovets, R. Todling, A. M. da Silva, and J. W. Larson, 2001: An adaptive buddy check for observational quality control. *Quart. J. Roy. Meteor. Soc.*, **127**, 2451–2471.

Diaz, H., C. Folland, T. Manabe, D. Parker, R. Reynolds, and S. Woodruff, 2002: Workshop on advances in the use of historical marine climate data. Bulletin 51, WMO.

Gelb, A., J. F. Kasper, R. A. Nash, C. F. Price, and A. A. Sutherland, 1974: *Applied Optimal Estimation*. M. I. T. Press, 374 pp.

Hamill, T. M. and C. Snyder, 2000: A hybrid ensemble Kalman filter-3d variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.

Hamill, T. M., J. S. Whitaker, and X. Wei, 2003: Ensemble re-forecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, submitted.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

— 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: operational, sequential, and variational. *J. Met. Soc. Japan*, **75 (1B)**, 181–189.

IPCC, 2001: *Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the International Panel on Climate Change.*, Cambridge University Press. 572 pp.

Jones, P. D., 1987: The early twentieth century Arctic High - fact or fiction? *Climate Dyn.*, **1**, 63–75.

Kistler, R. and Coauthors, 2001: The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–268.

Liu, Z.-Q. and F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Quart. J. Roy. Meteor. Soc.*, **128**, 1367–1386.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.

Manabe, T., 1999: The digitized Kobe collection, Phase I: Historical surface marine meteorological observations in the archive of the Japan Meteorological Agency. *Bull. Amer. Meteor. Soc.*, **80**, 2703–2715.

Maybeck, P. S., 1979: *Stochastic Models, Estimation and Control*, Academic Press, volume 1, chapter 7. 368–409.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's Spectral Satistical-Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1967–1988.

Schmith, T., H. Alexandersson, and H. Tuomenvirta, 1997: North Atlantic-European pressure observations 1868-1995 (WASA dataset version 1.0). tech. rep. 97-3, Danish Met. Inst.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble Square Root Filters. *Mon. Wea. Rev.*, **131**, 1485–1490.

Trenberth, K. E. and D. A. Paolino, 1980: The northern hemisphere sea-level pressure dataset: Trends, errors and discontinuties. *Mon. Wea. Rev.*, **108**, 855–872.

United States Weather Bureau, 1944: *Daily Synoptic Series, Historical Weather Maps, Northern Hemisphere Sea-Level, January 1899 to June 1939*. Cooperative project of U.S. Army, Air Force, and Weather Bureau Statistics Division.

Whitaker, J. S. and G. P. Compo, 2002: An ensemble Kalman smoother for Reanalysis. *Proceedings of the Symposium on observations, data assimilation and probabilistic prediction*, American Meteorological Society, Orlando, FL, 13-17 January 2002, 144–147.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Woodruff, S., H. Diaz, J. Elms, and S. Worley, 1998: COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth*, **23**, 517–527.

Wu, W., M. Iredell, S. Saha, and P. Caplan, 1997: Changes to the 1997 NCEP operational MRF model analysis/forecast system. Technical Procedures Bulletin 443, NCEP, National Weather Service, Office of Meteorology, Programs and Plans Division, Silver Spring, MD, 20910.

URL `http://www.nws.noaa.gov/om/tpb/indexb.htm`

Figure Captions

Figure 1: Number of surface pressure observations (per day) available in the Northern Hemisphere poleward of 20^oN each year from all available digital sources, including those currently being digitized.

Figure 2: Map of probabilities used to create simulated 1915 surface pressure network. Probabilities represent the average number of daily historical surface pressure observations in each 5 degree box normalized by the average number of daily surface pressure observations in the reduced 2001 network. See text for details.

Figure 3: Time series of RMS analysis error and ensemble spread averaged over the Northern Hemisphere poleward of 20^oN for (A) mean sea level pressure (hPa) and (B) 500 hPa geopotential height (m).

Figure 4: Maps of 500 hPa geopotential height (A) ensemble mean analysis error, and (B) ensemble spread averaged over the 124 December analyses for the simulated 1915 surface pressure network. (C) Mean 60-h 500 hPa geopotential height forecast error for forecasts initialized from all 744 00 UTC December NCEP/NCAR reanalyses for 1979-2002. Units are meters.

Figure 5: 500 hPa geopotential height analysis for 00UTC Dec 14 2001 (contour interval 50 m). CDAS analysis, using all available observations is shown on left. EnSRF analysis, using the simulated 1915 surface pressure observation network is shown on right. Black dots indicate locations of surface pressure observations used in the EnSRF analysis.

Figure 6: Time series of EnSRF RMS 500 hPa geopotential height analysis error averaged over the Northern Hemisphere poleward of 20°N for 50, 100 and 200 member ensembles.

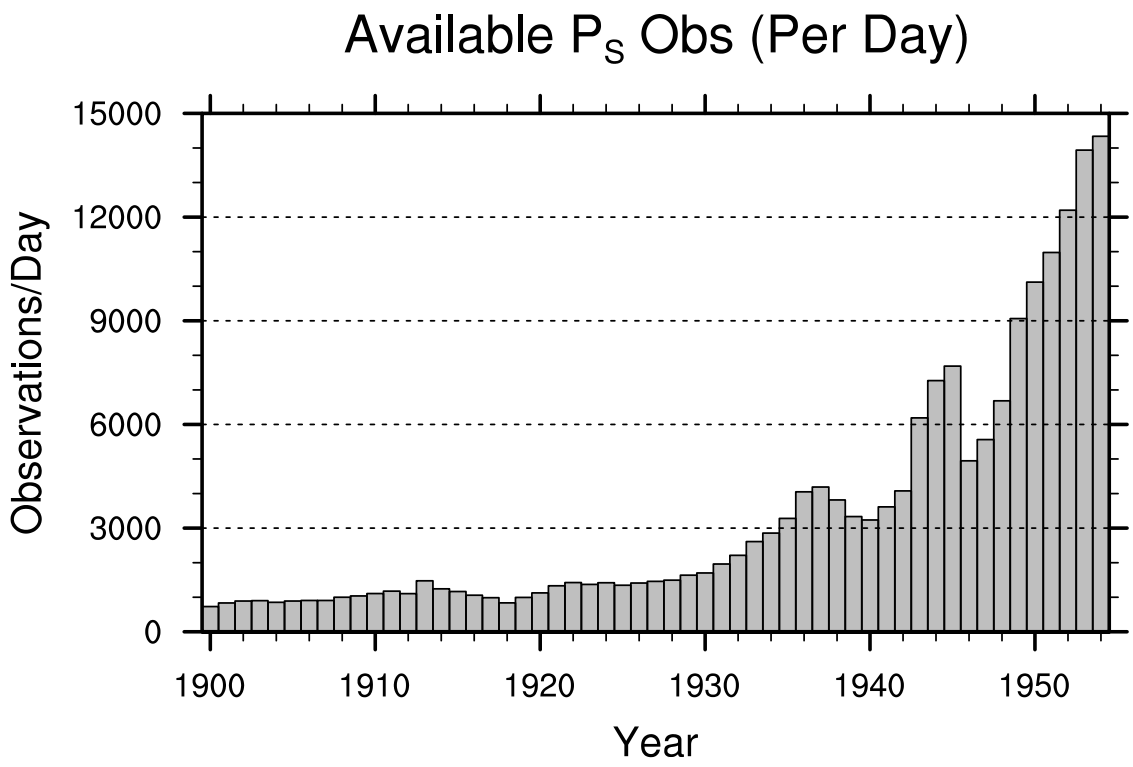

Figure 1: Number of surface pressure observations (per day) available in the Northern Hemisphere poleward of $20^o N$ each year from all available digital sources, including those currently being digitized.

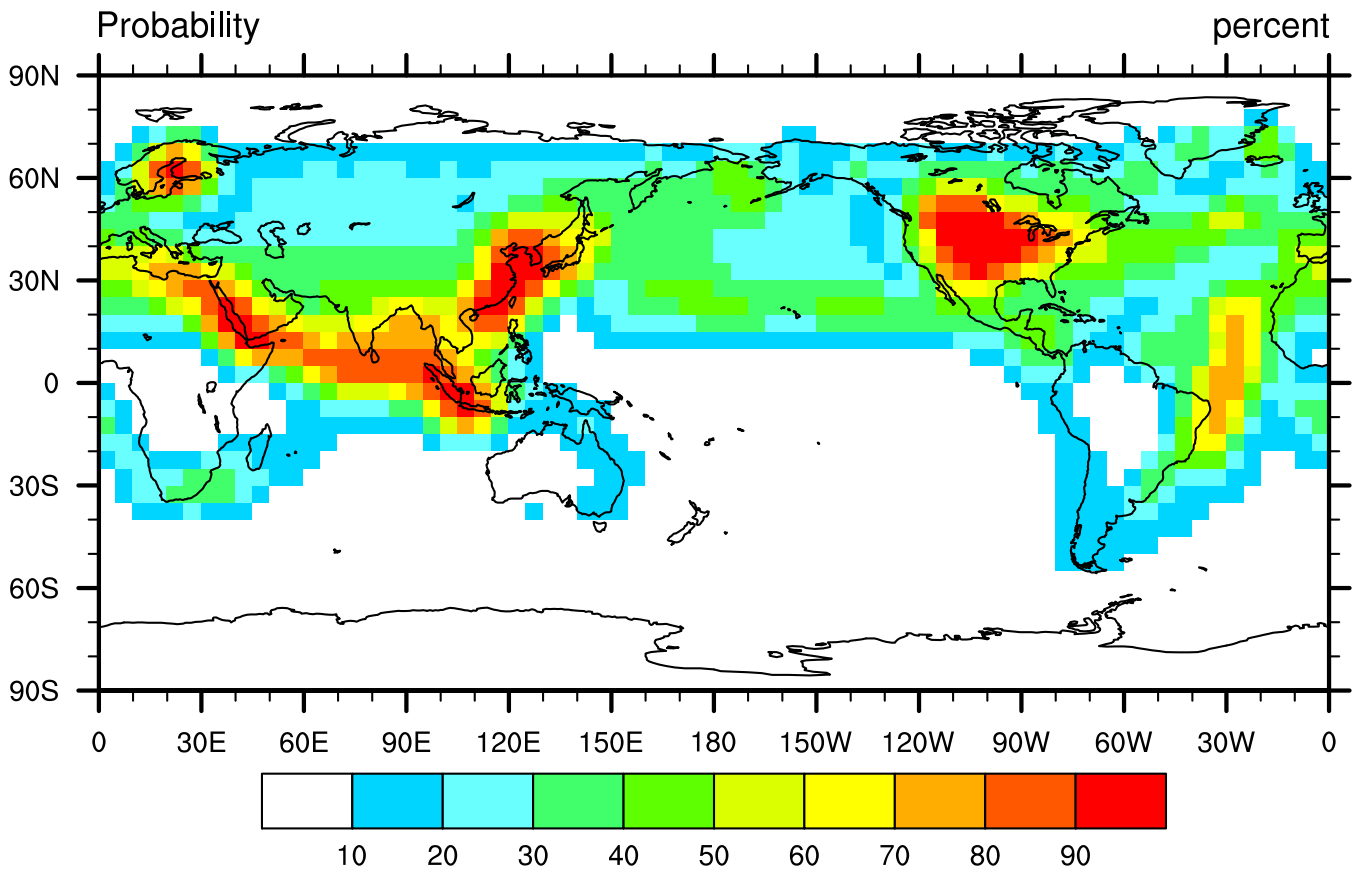

Figure 2: Map of probabilities used to create simulated 1915 surface pressure network. Probabilities represent the average number of daily historical surface pressure observations in each 5 degree box normalized by the average number of daily surface pressure observations in the reduced 2001 network. See text for details.

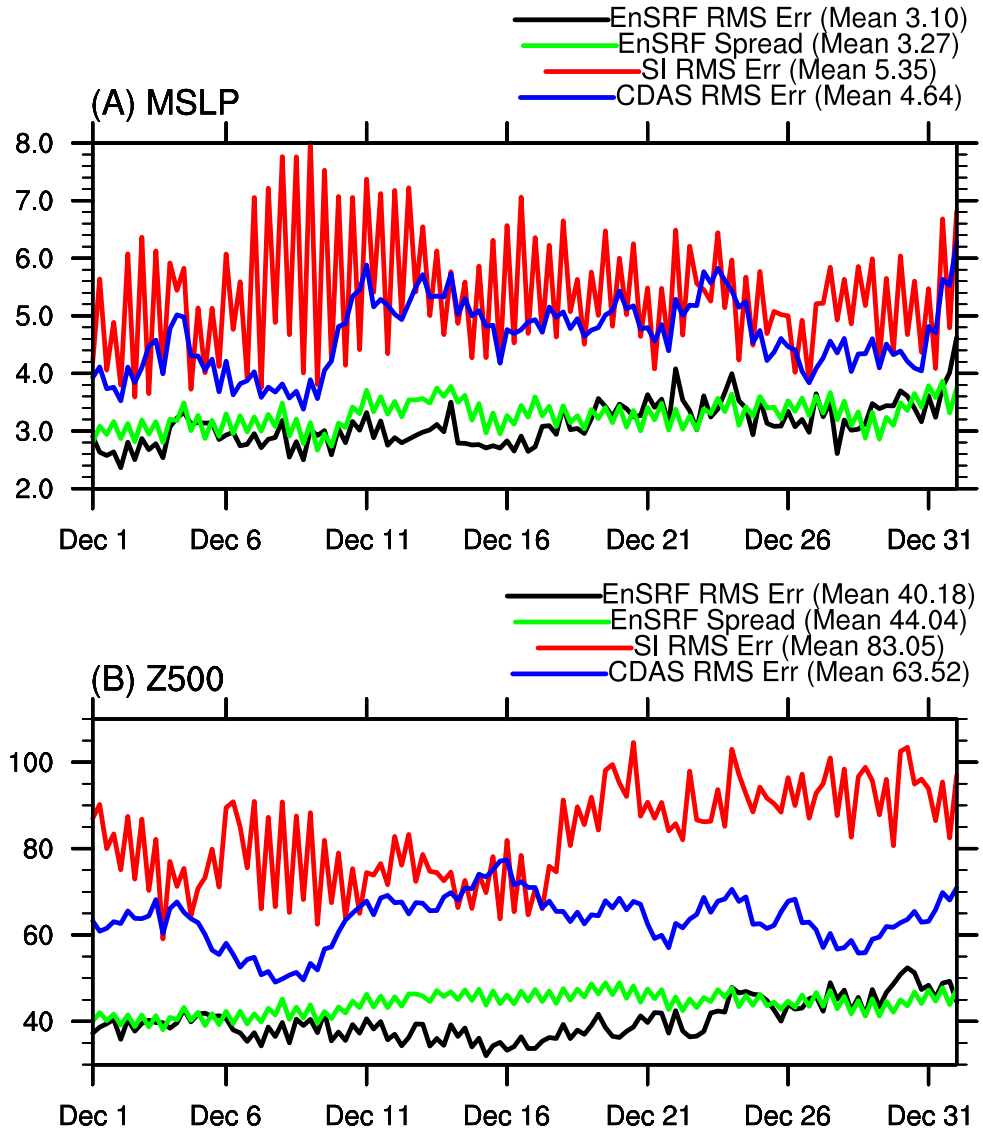

Figure 3: Time series of RMS analysis error and ensemble spread averaged over the Northern Hemisphere poleward of 20°N for (A) mean sea level pressure (hPa) and (B) 500 hPa geopotential height (m).

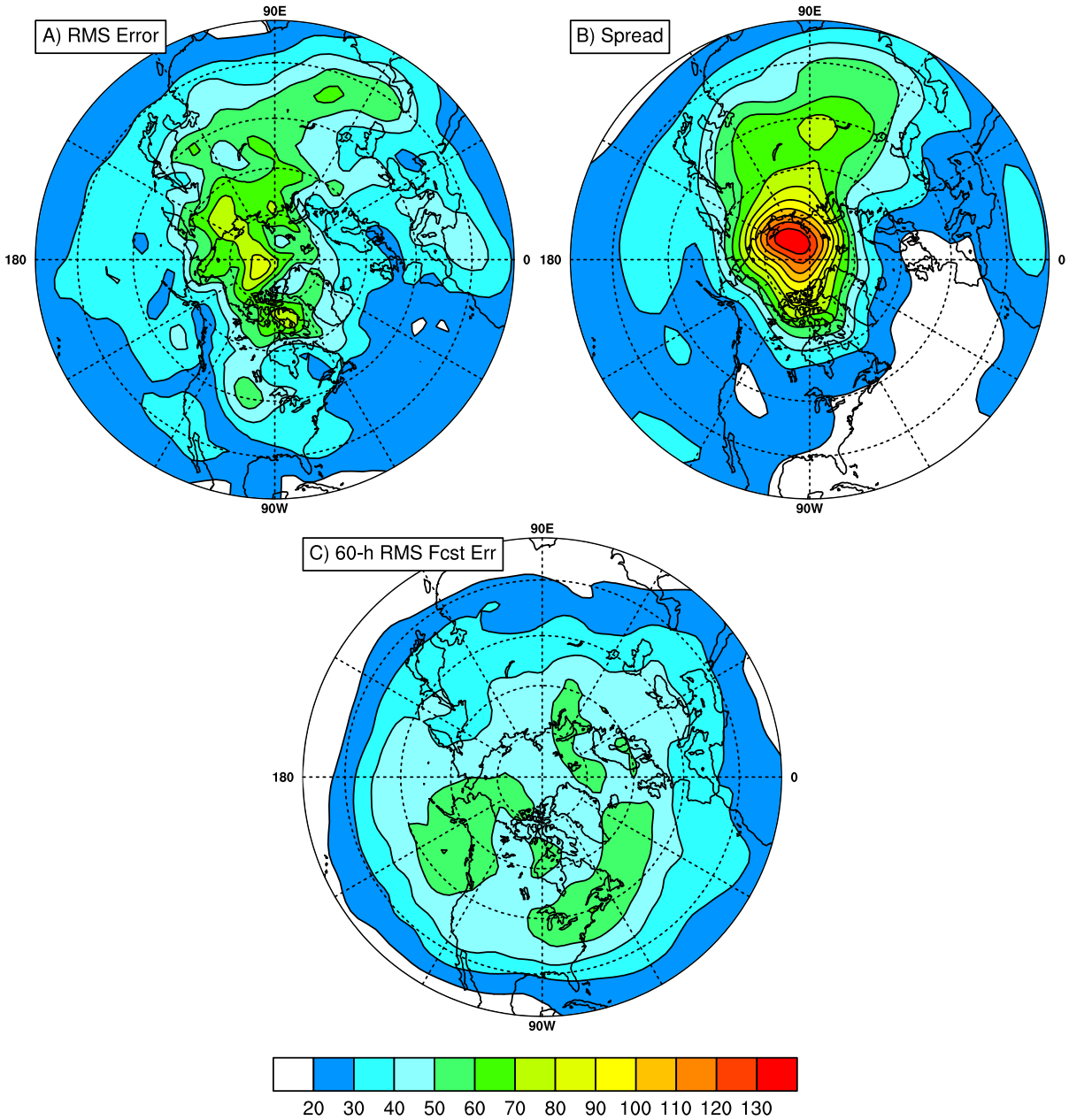

Figure 4: Maps of 500 hPa geopotential height (A) ensemble mean analysis error, and (B) ensemble spread averaged over the 124 December analyses for the simulated 1915 surface pressure network. (C) Mean 60-h 500 hPa geopotential height forecast error for forecasts initialized from all 744 00 UTC December NCEP/NCAR reanalyses for 1979-2002. Units are meters.

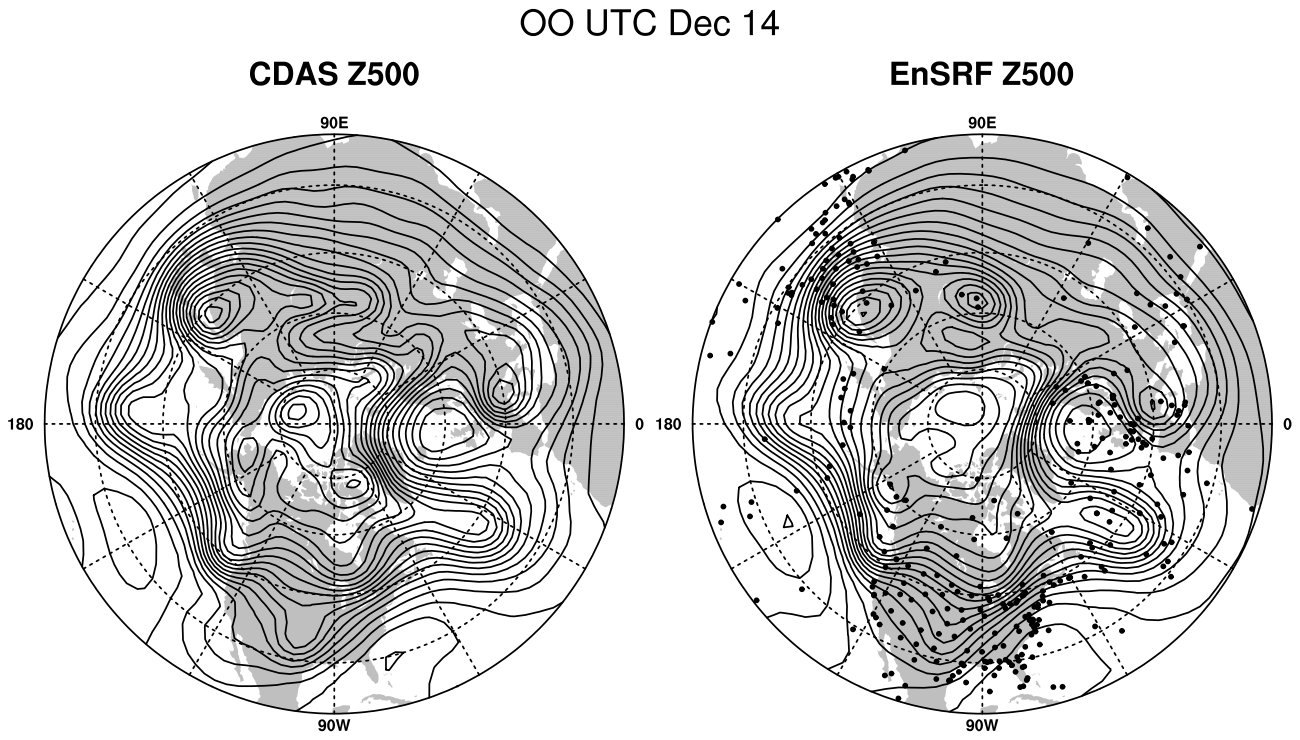

Figure 5: 500 hPa geopotential height analysis for 00UTC Dec 14 2001 (contour interval 50 m). CDAS analysis, using all available observations is shown on left. EnSRF analysis, using the simulated 1915 surface pressure observation network is shown on right. Black dots indicate locations of surface pressure observations used in the EnSRF analysis.

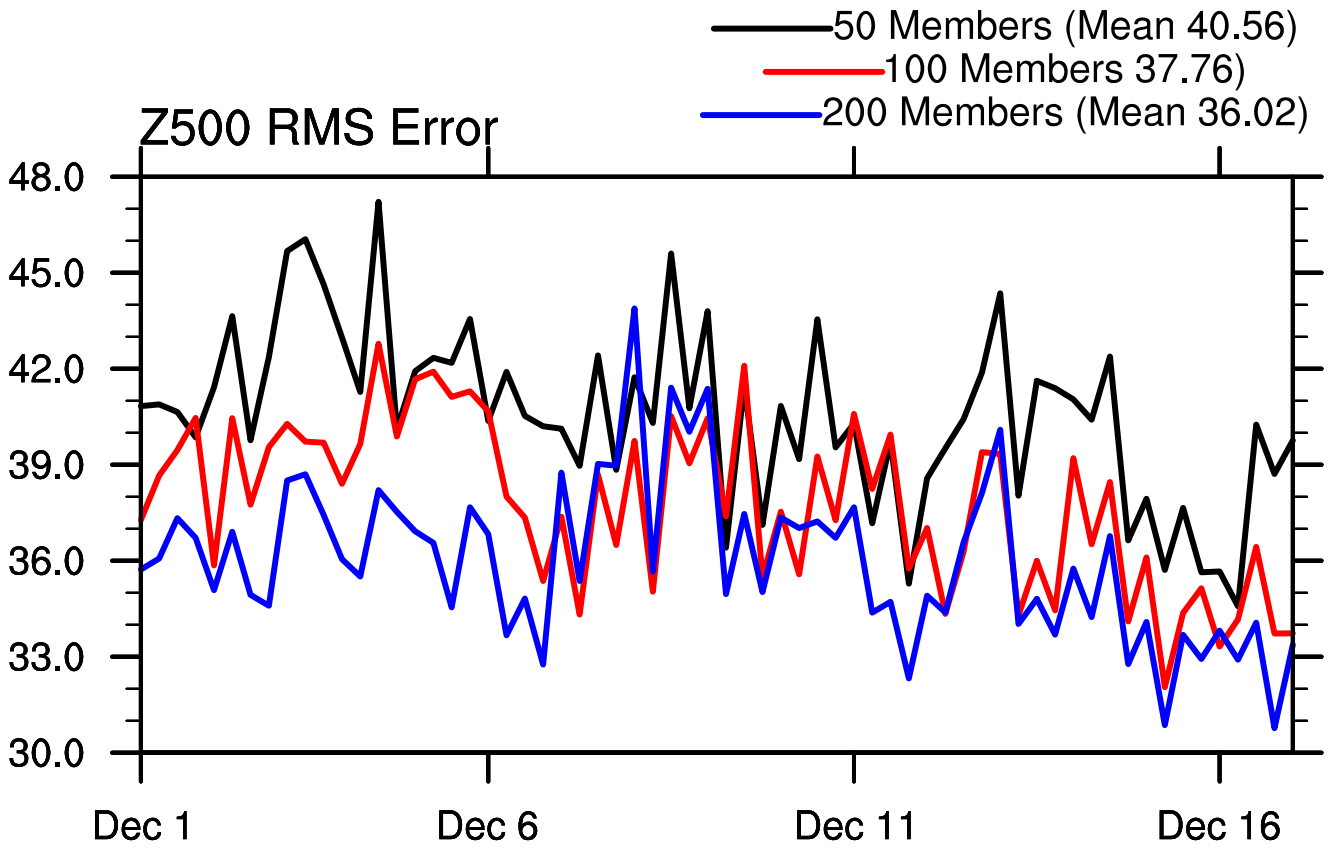

Figure 6: Time series of EnSRF RMS 500 hPa geopotential height analysis error averaged over the Northern Hemisphere poleward of 20°N for 50, 100 and 200 member ensembles.