Clim Dyn (2010) 35:995–1009
DOI 10.1007/s00382-009-0714-5

Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model

Helga S. Huntley · Gregory J. Hakim

Received: 23 February 2009 / Accepted: 24 November 2009 / Published online: 31 December 2009
© Springer-Verlag 2009

Abstract The problem of reconstructing past climates from a sparse network of noisy time-averaged observations is considered with a novel ensemble Kalman filter approach. Results for a sparse network of 100 idealized observations for a quasi-geostrophic model of a jet interacting with a mountain reveal that, for a wide range of observation averaging times, analysis errors are reduced by about 50% relative to the control case without assimilation. Results are robust to changes to observational error, the number of observations, and an imperfect model. Specifically, analysis errors are reduced relative to the control case for observations having errors up to three times the climatological variance for a fixed 100-station network, and for networks consisting of ten or more stations when observational errors are fixed at one-third the climatological variance. In the limit of small numbers of observations, station location becomes critically important, motivating an optimally determined network. A network of fifteen optimally determined observations reduces analysis errors by 30% relative to the control, as compared to 50% for a randomly chosen network of 100 observations.

H. S. Huntley
Department of Applied Mathematics, University of Washington,
Box 352420, Seattle, WA 98195-2420, USA

Present Address:
H. S. Huntley (✉)
School of Marine Science and Policy, University of Delaware,
106 Robinson Hall, Newark, DE 19716, USA
e-mail: helgah@udel.edu

G. J. Hakim
Department of Atmospheric Sciences, University of Washington,
Box 351640, Seattle, WA 98195-1640, USA
e-mail: hakim@atmos.washington.edu

Keywords Data assimilation · Paleoclimate · Ensemble Kalman filter · Atmospheric modeling

1 Introduction

Accurate estimates of past climate change provide important benchmarks for comparing and understanding future climate change. Climate reconstructions based on models alone lack a link to observations, which constrain the space of possible states of the climate. Given the short length of time for the instrumental record, approaches based on observations necessarily rely on measurements of proxy quantities. A major challenge with these measurements is the quantitative assessment of the relationship between the proxies and the physical fields, such as wind and temperature, near and far from the location of the proxy measurement. Empirical relationships are frequently used, but they lack the physical constraints that relate the proxies to dynamically consistent fields in space and time. The approach we advocate here attempts to overcome these shortcomings by blending information from observations and model estimates of the observations using ensemble-based state estimation.

One obstacle to progress on this problem using state estimation is that, unlike most weather observations, proxy measurements represent integrals over long periods of time; often a year or longer. A solution to this problem was proposed by Dirren and Hakim (2005) and tested on a Lorenz system with slow and fast modes. The crux of this method is that the time averaging and Kalman-filter-update operators are linear and commute. This property allows one to update the time-averaged state independently from the deviation from the time average, the latter being essentially a high-frequency signal unresolved by the observations.

The purpose of this paper is to test this method in an atmospheric model, which is idealized yet contains dynamically relevant variability at a range of frequencies as a result of a baroclinic storm track interacting with a mountain. Such a framework is realistic enough to assess the generality of the low-order results of Dirren and Hakim (2005), but simple enough to allow testing over a range of observation network and model configurations. This simplicity also comes at a price: Although our motivation is paleoclimate reconstruction, the model we employ is insufficiently complex to capture true paleoclimate variability, which necessarily involves additional components of the Earth climate system such as oceans and ice. Nevertheless, the simpler model chosen here provides a testbed for solution techniques to key aspects of the problem of paleo data assimilation, while supplying some of the rich dynamical interactions across different timescales representative of the complete problem. It serves, thus, as a stepping stone in the hierarchy of models, moving toward proper assimilation of paleo proxy data in more complicated models.

The remainder of the paper is organized as follows. The method is described in Sect. 2, and results from a control experiment without any data assimilation and a set of baseline experiments with a fixed, randomly distributed network of observation are described in Sect. 3. The sensitivity of the baseline experiment to changes in the observation averaging period are described in Sect. 4. Experiments exploring the sensitivity of the baseline results to the observational error and to the number and the location of stations are described in Sects. 5 and 6, respectively. A limited assessment of the impact of model error, in the form of degraded resolution and erroneous damping timescale, is described in Sect. 7. Section 8 provides a concluding summary.

2 Methodology

2.1 Notation

Throughout, ensemble averages will be denoted using angle brackets $\langle\cdot\rangle$, while averages over time will be indicated using an overbar $\bar{\cdot}$. Quantities from the “true state” simulation, or observations taken therefrom and modified with noise, will be denoted by a subscript o. The averaging time period is denoted by τ.

2.2 The model

With the idea of a model hierarchy in mind, we chose a quasi-geostrophic (QG) atmospheric model to test the assimilation method described below, building on earlier research with an idealized one-dimensional model (Dirren and Hakim 2005). The QG model provides some of the realistic phenomenology and chaotic behavior of the atmosphere, such as jets, baroclinic waves, cyclones and fronts, yet it is simple enough to run quickly and to permit a straight-forward analysis of the results. Specifically, the model describes a turbulent jet stream on a periodic domain for a uniform potential vorticity troposphere, bounded above and below by rigid surfaces. A convenient consequence of such a configuration is that the interior field (between ground and tropopause) is determined completely and without approximation by the potential temperature boundary conditions; therefore the dynamics collapse computationally to two dimensions. Throughout we will be analyzing the potential temperature field θ on the boundaries.

The model implementation employs a pseudo-spectral method and a third-order accurate Adams-Bashforth time integration scheme. The dynamics are damped by hyperviscosity and an Ekman layer on the lower boundary. Forcing is supplied by relaxation back to a prescribed jet (Hoskins and West 1979) on a timescale of about 20 days, which is meant to mimic radiative restoration of midlatitude baroclinicity.

The size of the domain is approximately 28,000 km in x, 11,000 km in y, and 10 km in z, with a de-aliased spectral resolution equivalent to 64×32 grid points (about 435 km × 345 km). Time and space are scaled and expressed in nondimensional units, where one unit corresponds to 9.26 h and 1,000 km, respectively. Moreover, potential temperature (the prognostic variable in this model, which we will analyze) is also non-dimensionalized, with one unit scaled to 9.18 K. Further details on the model and scaling are provided in Hakim (2000). A mountain is positioned at the center of the domain, with height roughly 6 km and of spherical Gaussian shape (with a damping length scale of 1,000 km). Compared to a flat bottom boundary, this topography results in a richer low-frequency spectrum in the potential temperature field and facilitates the study of different averaging times.

2.3 The assimilation procedure

The method we employ for assimilating time-averaged data is a variation of the (square-root) ensemble Kalman filter (EnKF), which is derived as the optimal unbiased state estimate under the assumptions of a linear model and Gaussian error statistics. This novel version of the algorithm was first suggested by Dirren and Hakim (2005). Rather than determining innovation increments to instantaneous quantities from averaged observations, the time averages themselves are updated and unchanged perturbations therefrom added to the analysis to recover the full state. See Appendix 1 for further details.

For the experiments presented here, an ensemble of fifty members is cycled over fifty averaging periods. Standard measures are employed to address sampling error, including spatial localization of covariances by a combination of a hierarchical filter (Anderson, 2007) with five sub-ensembles and a Gaspari-Cohn function (Gaspari and Cohn 1999).

For the EnKF, observations impact only those variables that covary with the model estimates of the observations. In theory, if the true statistics of the system were known precisely (and provided the observations are a linear function of the state vector), the resulting time averages of the algorithm used here would be equivalent to those achieved by applying the EnKF directly to the instantaneous variables, due to the commutation of the linear operators (time average and filter update; see Appendix 2). In practice, the statistics must be estimated from the ensemble, and we expect covariance relationships of time-averaged observations with time-averaged state variables to be less prone to noise than those with instantaneous state variables.

In addition, and more importantly, the Dirren and Hakim method has the advantage that far fewer covariance estimates have to be carried out at each assimilation step in order to update the time-averaged state than with a comparable traditional implementation of the EnKF. In experiments based on a dense observation network (Fig. 1), using the present method, where only time averages are updated during assimilation, and using another where instantaneous states are updated and then averaged a posteriori, we found both analysis and forecast estimates of the time averages to be essentially equally good from both methods. The estimates of instantaneous quantities are slightly better when they are directly updated, although only at short averaging times. However, the instantaneous-assimilation method becomes increasingly expensive for long time averages, since the time required for assimilation scales linearly with the number of assimilation times; the method advocated here requires only a single assimilation step per averaging interval. For the $\tau = 100$ case, the difference in computing time amounts to roughly a factor of 60.

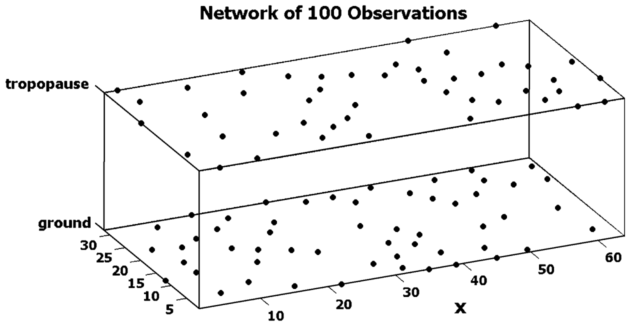


Fig. 1 Baseline experiment observation network

2.4 Statistical validation methods

An assumption of ensemble modeling is that the discrete statistics of the finite ensemble are a good representation of the continuous statistics of the real system. Even in the case of a perfect model experiment, as we have here, it is well known that sampling error can result in poorly calibrated ensemble data assimilation systems with, for example, an underestimate of the variance (see, e.g., Anderson and Anderson (1999)).

The validation method we employ here, based on Murphy (1988), compares the expected value of the squared error in the ensemble mean to the expected value either of the ensemble variance or of the squared error in a single run. Let the state variable be denoted by θ and a single ensemble member by θ_j. The "truth" is θ_o and the ensemble average $\langle\theta\rangle$. An expected value will be indicated by $[\![\cdot]\!]$. Let N be the number of ensemble members. The ensemble variance is then defined as

$$S = \frac{1}{N-1}\sum_{j=1}^{N}\left(\theta_j - \langle\theta\rangle\right)^2. \tag{1}$$

The squared errors in the mean and in a single run are, respectively,

$$E_M = \left(\langle\theta\rangle - \theta_o\right)^2 \quad \text{and} \tag{2}$$

$$E_j = \left(\theta_j - \theta_o\right)^2. \tag{3}$$

Murphy showed that, under the assumption that the "truth" is drawn from the same distribution as the ensemble members,

$$[\![E_M]\!] = \frac{N+1}{N}[\![S]\!] \tag{4}$$

and

$$[\![E_M]\!] = \frac{N+1}{2N}[\![E_j]\!]. \tag{5}$$

These results will be used to check the calibration of the ensemble. Note that the errors in these expressions are *squared*, not *root-mean-square*. While it is common in the literature to use the ratio of the root-mean-squared errors, $[\![\sqrt{E_M}/\langle\sqrt{E_j}\rangle]\!]$, Murphy's theory (which is often cited as justification) in fact does not extend to this expression, since $[\![\sqrt{X}]\!] \neq \sqrt{[\![X]\!]}$ and $[\![X/Y]\!] \neq [\![X]\!]/[\![Y]\!]$ in general.

2.5 Diagnostics

Ensemble data assimilation is designed to improve upon a background estimate by adjusting the ensemble mean closer to observations and by reducing the ensemble variance. We focus primarily on error reduction in the ensemble mean to evaluate the performance of our algorithm in the QG model. Specifically, we consider the root-mean-square (rms) error in the potential temperature (θ) field taken over the ground and tropopause, given by

$$R(t) = \sqrt{\frac{1}{K}\sum_x [\langle\theta(x)\rangle - \theta_o(x)]^2}, \tag{6}$$

where K is the number of points in space and either the analysis or the forecast is substituted. A similar expression is used to analyze the time averages relevant in the context of averaged observations.

A reference value is generated by carrying out these calculations on a control ensemble created without any assimilation. The resulting rms error samples are used to determine an rms error time mean,

$$R^{\text{avg}} = \frac{1}{T}\sum_t R(t), \tag{7}$$

and standard deviation,

$$\sigma_R = \sqrt{\frac{1}{T-1}\sum_t (R(t) - R^{\text{avg}})^2}. \tag{8}$$

Due to averaging, ensemble mean forecasts are expected to have smaller errors than forecasts from a single simulation (Eq. (5) above; see Murphy 1988). To distinguish the error reduction due to the use of an ensemble from that due to the data assimilation, it is important to derive the reference value from an ensemble mean rather than from a single simulation.

3 Baseline experiments

3.1 The setup

Many factors affect filter performance, including observation locations and error statistics and the number of ensemble members. Throughout the experiments we fix parameters associated with the assimilation technique, such as ensemble size, covariance localization method, and assimilation interval. All ensembles are initialized by fifty random draws from the model climate history. In the experiments that follow the observation averaging interval is varied to investigate the performance of the assimilation algorithm (Sect. 4). In subsequent experiments (Sects. 5 and 6) we consider the effect of observation error, location, and number on the results.

For the baseline experiments, 100 observation locations are chosen randomly but evenly spread across the two surfaces (Fig. 1). A single truth run is created for generating observations and for verification of analyses and forecasts. For each observation averaging period, τ, observations are drawn at the specified locations from the truth run, averaged, and modified by random noise (normally distributed, with mean 0 and standard deviation $\sigma_o = 0.27$, which is about half the climatological variability in the model).

3.2 Two examples: $\tau = 100$ and $\tau = 20$

Assimilation results are reviewed here for two cases: a relatively long averaging time $\tau = 100$ (i.e., about a month) and a relatively short averaging time $\tau = 20$ (i.e., about a week). For paleoclimate records, even monthly averages are unrealistically short, but the absolute timescale is not the main point here. The more general question addressed here is to what extent a low-frequency signal can be determined from time-averaged observations in the presence of large high-frequency variability, such as the baroclinic waves in the QG model, which dynamically interact with the low-frequency signal.

Due to Ekman damping, there is less variability on the lower surface of the model than at the tropopause. Similarly, there are north–south spatial differences in variability due to the chaotic motion of the jet. We will not address results associated with these spatial variations, but focus instead on summary results for both boundaries taken together. Our conclusions hold for each boundary taken separately as well.

For an averaging time of $\tau = 100$, time-averaged analyses ($\overline{\theta}$, red line in Fig. 2 (a)) improve upon the control (black line), reducing the rms error on average (over the 50 assimilation intervals) by over 18 % . The rms error is almost always less than the ensemble-adjusted control value, R_c^{avg}, despite the fact that the observational error (green line) actually exceeds the mean control rms error (i.e., climatology). The improved accuracy of the analyses does not, however, translate to improvements in the forecast (blue line). We conclude that the time-averaged observations constrain the time-averaged potential temperature field, but for long time averages, the error still saturates in the forecast cycle. In fact, for averaging times greater than $\tau = 30$, using an ensemble randomly drawn from the model climatology, as opposed to the ensemble advanced by the model from the previous analysis time, yields essentially identical analysis rms errors (not shown). In other words, in this situation, the model adds little information over that available from climatology, and the assimilation procedure primarily serves to interpolate spatially between the observations. For $\tau = 100$, the rms error in the instantaneous values exhibits little improvement over the control, since variability at shorter timescales dominates (Fig. 2b).

Results differ for shorter averaging times. For example, $\overline{\theta}$ analysis errors for $\tau = 20$ are larger in absolute terms,

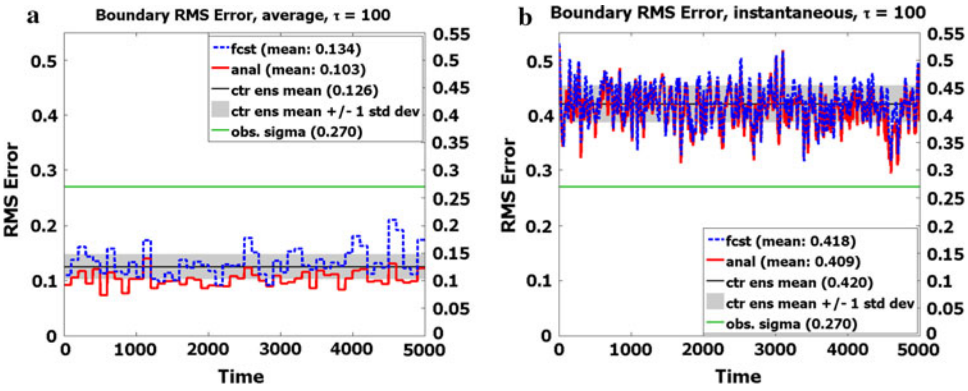
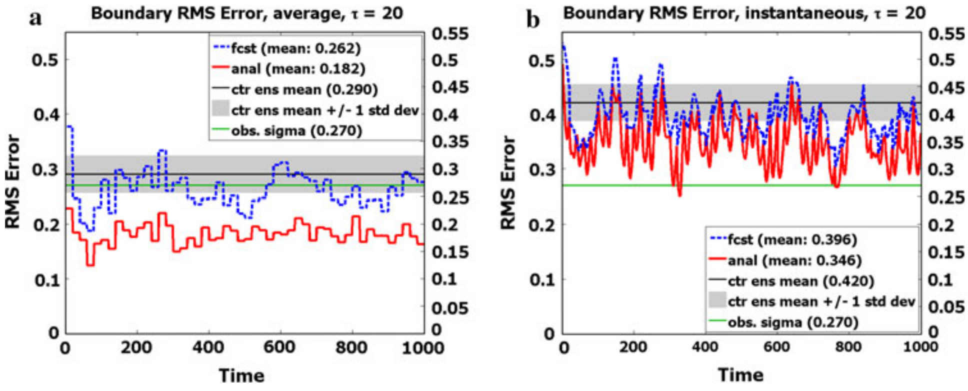


Fig. 2 Root-mean-square error at the boundaries of the model run with $\tau = 100$, compared to the "truth" run. The left panel (**a**) and the right panel (**b**) show results for averaged $\overline{\theta}$ and instantaneous θ, respectively. *Red* indicates analysis, *blue* forecast values. The *green lines* represent the standard deviation for the error added to the observations. The *solid black lines* show the mean rms error of the time-averaged and instantaneous quantity, respectively, for the control run ensemble, with the *shaded grey band* indicating one standard deviation from this mean

Fig. 3 Same as Fig. 2, but for $\tau = 20$



because of the larger variability on shorter timescales, but show greater improvement relative to the corresponding time-averaged control (Fig. 3a). Moreover, forecast errors are also substantially reduced, as are errors in both analyses and forecasts of the instantaneous values (Fig. 3b). Note that the observational error in this case is below the control (R_c^{avg}), which effectively gives the observation more weight; this issue will be revisited in Sect. 5.

We remark here that application of the filter alone, as can be seen in Figs. 2 and 3, will inevitably lead to some discontinuities at each assimilation step in a time-series of multiple averaging periods, which do not represent physical processes. A smoother transition can be attained through a post-analysis step of filtering for low-frequency variability. Whether such a smooth solution is any closer to the "true" physical state is unclear, but such a filter-smoother combination will be more appropriate for some types of analyses. As we do not focus here on the dynamics across multiple assimilations, our results are not impacted by the discontinuities, and the many issues surrounding this phenomenon, associated with any reanalysis application, are beyond the scope of this paper.

3.3 Statistical validation

As described above in Sect. 2.4, an ensemble data assimilation technique not only has to improve the estimate of the true state but should also reflect the probability distribution from which the true state is drawn. The main challenge is to maintain proper ensemble variance. To verify that our algorithm is properly calibrated, relationships (4) and (5) are evaluated by averaging over all available instances in time, and in the case of $\llbracket E_j \rrbracket$ also over all ensemble members. Since fifty ensemble members were used, the theoretically predicted values for the ratios are

$$\frac{\llbracket E_M \rrbracket}{\llbracket S \rrbracket} = \frac{N+1}{N} = \frac{51}{50} = 1.02 \tag{9}$$

$$\frac{\llbracket E_M \rrbracket}{\llbracket E_j \rrbracket} = \frac{N+1}{2N} = \frac{51}{100} = 0.51 \tag{10}$$

The values computed for the two example cases $\tau = 100$ and $\tau = 20$ are given in Table 1. These numbers indicate that our ensembles are generally well calibrated, although the spread tends to be smaller in the analysis of the averaged quantity.

Table 1 Statistical validation of the ensemble variance for the baseline cases using Murphy’s formulae

	Inst fcst.	Inst anal.	Avg fcst.	Avg anal.	Target
$\tau = 20$	0.97	0.97	1.05	1.16	$\frac{\llbracket E_M \rrbracket}{\llbracket S \rrbracket} = 1.02$
$\tau = 100$	1.03	1.03	1.06	1.13	
$\tau = 20$	0.50	0.50	0.52	0.54	$\frac{\llbracket E_M \rrbracket}{\llbracket E_j \rrbracket} = 0.51$
$\tau = 100$	0.51	0.51	0.52	0.54	

4 Effects of changing the averaging period τ

To construct a more complete analysis of the effects of varying the averaging time of the observations, experiments were performed for τ varying from 10 to 100 in steps of 10. In all cases, the rms errors of the analysis of averaged quantities, at both the tropopause and the ground, (red line, Fig. 4) are significantly less than in the control (black line). The analysis error is also smaller than the assumed observational error (green line) for all values of τ. For $\tau = 10$ and 20, forecasts (blue line) exhibit average rms error reduction of more than approximately one standard deviation σ_R, as derived from the unassimilating control; decreases for $\tau = 30$ are slight and disappear altogether for larger values of τ. Although the observation error becomes unrealistically large relative to intrinsic variability for $\tau > 30$, we show in Sect. 5 that this factor cannot account for the lack of improvement to forecasts for these time averages. This leads us to conclude that, for these timescales, the growth to saturation of errors at higher frequencies is shorter than the averaging time, which no reduction in analysis error can change; that is, these timescales are beyond the limit of predictability for this model assimilating such averaged observations with our method.

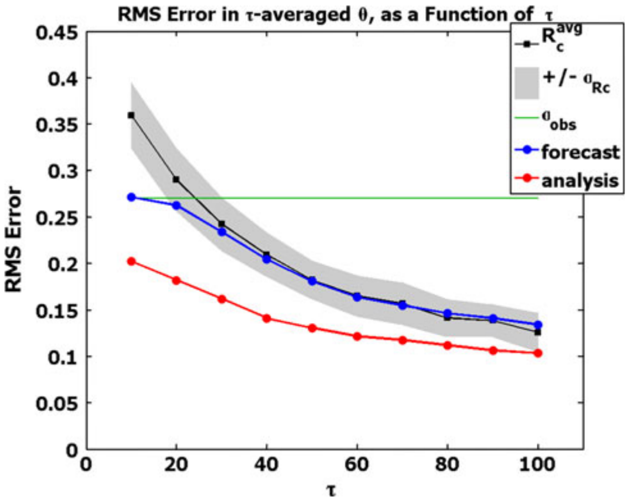


Fig. 4 Mean root-mean-square errors R^{avg} for $\overline{\theta}$ of the analyses (*red*) and forecasts (*blue*) as a function of averaging time τ. For comparison, the plot also shows σ_o (*green*, independent of τ) and the control R_c^{avg} (*black*, *squares*) for ensemble forecasts, along with the span of $\pm\ \sigma_R$ (*grey shading*)

Thus far we find that assimilating τ-averaged observations improves τ-averaged analyses. Here we generalize the assessment of reduction in R^{avg} to a range of output averaging times, τ^*, by considering for each observation averaging time τ the error reduction function

$$g(\tau, \tau^*) = \left(\frac{R_c^{\mathrm{avg}}(\tau^*) - R_{\mathrm{anal}}^{\mathrm{avg}}(\tau, \tau^*)}{R_c^{\mathrm{avg}}(\tau^*)} \right) \cdot 100\%. \tag{11}$$

Note that the control value R_c^{avg} depends on τ^* only, while the analysis value $R_{\mathrm{anal}}^{\mathrm{avg}}$ is a function of both observation and output averaging times. The results for a range of values of τ and τ^* are summarized in Fig. 5. The forecast used for the second panel in each case is run for a length of τ (the observation averaging period) between assimilations, as displayed in Figs. 2 and 3; consecutive sections of length τ^* (the output averaging period) enter the calculation.

For a given analysis state variable averaging time τ^*, assimilation with a shorter observation averaging time generally produces a greater improvement; i.e., along a horizontal line in Fig. 5 a, the error reduction decreases. Exceptions to this rule occur for most values of τ^* with $\tau = \tau^*$ giving, in fact, a better result than the pattern suggests, due to the in-phase relationship between assimilation period and output averaging period. Similarly, along a vertical line in Fig. 5a, for a given observation averaging time τ, a greater error reduction is realized for longer state variable averaging times τ^*. The exceptions here are somewhat more noticeable and, again, fall predominantly along the line $\tau = \tau^*$ but also on values of τ^* that are integer multiples of τ, as one would expect. Maybe most noteworthy is that the error reduction across almost the entire figure is quite substantial, with significant improvements ($\geq$10%) well below the $\tau = \tau^*$ line. In fact, the 10%-cut-off follows more closely $\tau^* = 0.5\ \tau$ for $\tau \geq 50$ and is even lower for smaller τ. In other words, the method is fairly successful at constraining some higher-frequency variability with low-frequency observations.

Interestingly, these patterns in the error reduction in the analysis are not reflected by the error reduction in the forecast (Fig. 5b). As seen in Fig. 4, the error in the τ-averaged forecast saturates at the control error for $\tau \gtrsim 30$. However, the forecast averaged over longer periods τ^* still exhibits some error reduction ($\geq$5%) for τ up to 70. On the other hand, there is virtually no error reduction for $\tau^* < \tau$, with the exception of $\tau = 10$ and 20. Thus, improving the longer timescales does not have a significant influence on shorter timescale dynamics.

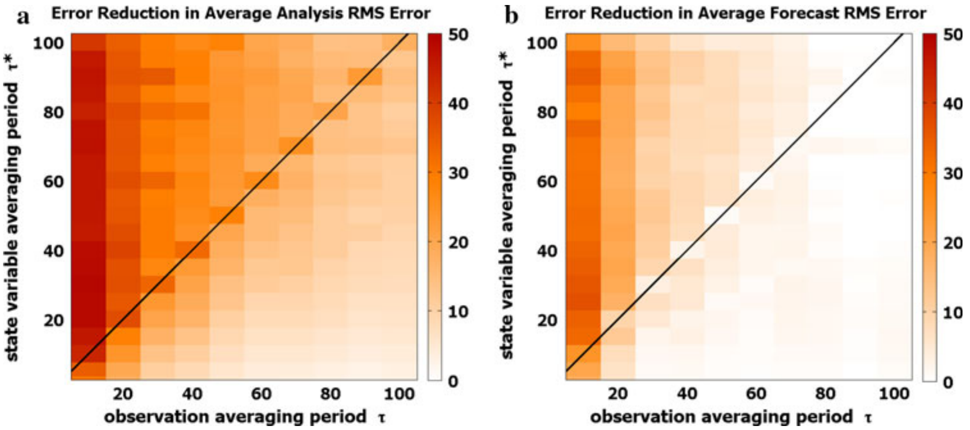


Fig. 5 Percentage reduction in the average rms error of (**a**) the analysis and (**b**) the forecast of the τ^*-averaged state variable for different observation averaging times τ (see Eq. (11)). The line $\tau = \tau^*$ is indicated in *black* and corresponds to the case of a τ-averaged state variable addressed in previous figures. *Each box* represents the τ and τ^* values at its center

5 Impact of observation error

In the baseline experiments, the assumed observational error is fixed for all averaging times. As a consequence, the observational error actually significantly exceeds climatological variance for larger τ. It is conceivable that the lack of improvement in the forecast could be ascribed to overly noisy observations. To test this idea, experiments are repeated with σ_o now taken consistently as a third of the climatological rms error R_c^{avg} for each value of τ. As Fig. 6 shows, the decreased observational noise (dashed green line, vs. the original noise as a solid green line) does significantly lower the error in the analysis of the τ-averaged potential temperature (dashed red line with triangles, vs. solid red line with circles). For $\tau \geq 30$, the rms error is cut by roughly a third or more (almost half for $\tau = 100$). The forecasts, however, are barely improved for $\tau = 10$ and 20 and not at all for any larger values of τ (dashed blue lines with triangles, vs. solid blue lines with circles). In other words, the averaging periods are long enough that the errors saturate in the model forecast, even if the initial conditions are markedly more accurate. As mentioned before, in these cases, the model serves to interpolate between the observations, and the assimilation can be performed on an ensemble drawn from climatology at each assimilation step with similar results. Instantaneous analyses improve by 2–6%, while forecasts have practically the same errors as in the baseline experiments (not shown).

To assess further the impact of observation error, we chose $\tau = 20$ as an example where both analysis and forecast of $\overline{\theta}$ are improved by reducing σ_o from 0.27 to 0.10. σ_o is then varied over a larger range of values (0, 0.1, 0.22, 0.5, 1, and 2; essentially a doubling each time after 0). Figure 7 shows the rms error in the averaged quantity as a function of observational error. The plot for instantaneous θ looks remarkably similar, except with larger values. (Greater variability at the shorter timescales leads to a control rms error of 0.42 vs. 0.29 for $\tau = 20$.) Clearly, when σ_o becomes too large, the assimilation has no noticeable effect. While this is hardly surprising, note that this threshold is well beyond the climatological uncertainty. In fact, even when the observations have an error three times that of climatology, i.e. much greater than what can be expected for observations from actual measurements, the analysis still shows a reduction in rms error by about one standard deviation (and >10%). With no observational error the error reduction in the analysis is 57%. Initially, as σ_o increases, so does the rms error, with a more or less linear relationship. Eventually, the analysis rms error curve levels off. The pattern in the forecast rms error is similar, although the reduction as σ_o tends to 0 is far smaller. These results depend to some extent on the properties of the observation network, the sensitivity to which is considered next.

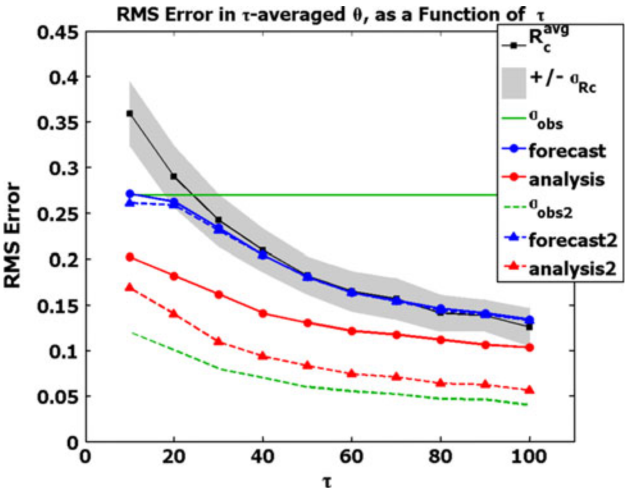


Fig. 6 Mean root-mean-square errors R^{avg} for $\overline{\theta}$ of the analyses (*red*) and forecasts (*blue*) as a function of averaging time τ. The *solid lines* and *circles* show the baseline results for constant σ_o; the *dashed lines with triangles* are for decreasing σ_o. Corresponding observational error is plotted in *green* lines. The control R_c^{avg} for ensemble forecasts is also shown (*black*, *squares*), along with the span of $\pm\sigma_R$ (*grey shading*)

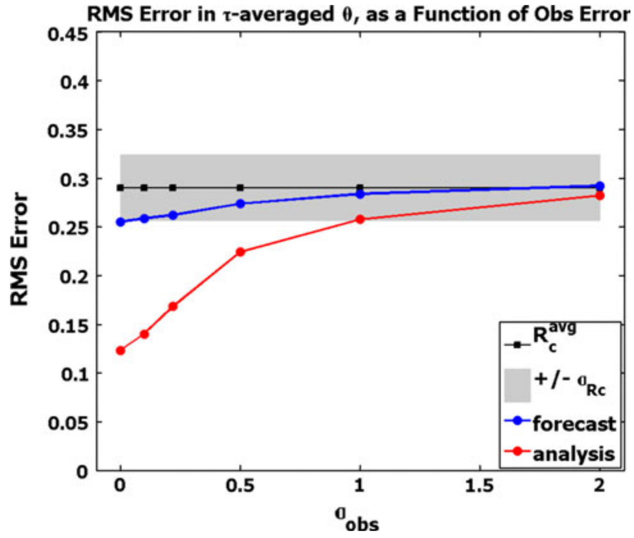


Fig. 7 Mean root-mean-square errors R^{avg} for $\overline{\theta}$ of the analyses (*red*) and forecasts (*blue*) as a function of observational error σ_o for averaging time $\tau = 20$. The control R_c^{avg} for ensemble forecasts is shown (*black*, *squares*), along with the span of $\pm\sigma_R$ (*grey shading*)

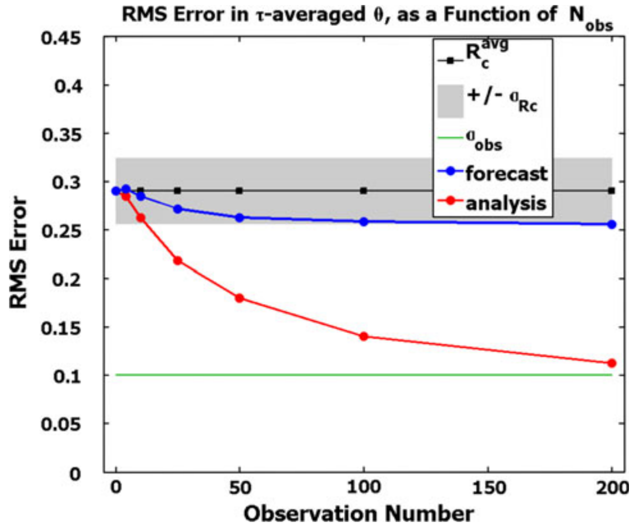


Fig. 8 Mean root-mean-square errors R^{avg} for $\overline{\theta}$ of the analyses (*red*) and forecasts (*blue*) as a function of number of observations for averaging time $\tau = 20$ and observational error $\sigma_o = 0.10$ (*green*). The control R_c^{avg} for ensemble forecasts is shown (*black*, *squares*), along with the span of $\pm\sigma_R$ (*grey shading*)

6 Impact of observation number and location

6.1 Observation number

In the experiments discussed so far, the observation network is fairly dense and evenly distributed (Fig. 1). We found that assimilating these 100 time-averaged observations significantly improves analyses for all averaging periods considered, as well as the forecasts for τ less than 30. Here we address the sensitivity of the results to the number of observations, an issue of particular relevance in the context of paleoclimate where observational networks tend to be relatively sparse.

As in the previous section, to be concise we focus on a specific case, with $\tau = 20$ and $\sigma_o = 0.10$, and consider the following range of observation number: 0, 4, 10, 25, 50, 100, and 200. For reference, note that the total number of model variables is 4,096, providing an upper bound on the degrees of freedom in the model. Observation locations are randomly chosen across the ground and tropopause subject to a minimum horizontal spacing. Because location becomes relevant for smaller numbers of observations, four experiments were run for $N = 4$ and two for $N = 10$, and the errors averaged. Results show little impact for four observations, with a sharp drop in error between 10 and 50 observations, which then levels off for larger numbers of observations (Fig. 8). Forecast errors exhibit a small improvement for 25 or more observations, but little change when the number is increased past 50.

6.2 Observation location

As alluded to above, results for small numbers of observations become sensitive to observation location, which motivates the notion of designing an optimal network for small numbers of observations. In order to objectively determine an optimal network, a metric must be defined to measure impact. Theoretically, it is possible in the context of a perfect model experiment such as this to choose the rms error in the forecast as the metric. However, such an approach has little relevance for any applications outside this narrow context, as the "true" forecast has to be known a priori. Here we choose as a metric the magnitude of the projection onto the first empirical orthogonal function (EOF) of the $\overline{\theta}$ field, which represents the pattern that captures the most variability. It also represents the large-scale oscillations that are likely constrainable by a sparse observation network.

The theory of ensemble-based optimal network design we follow is described in Ancell and Hakim (2007), and example applications are shown in Hakim and Torn (2008) and Torn and Hakim (2008). A similar methodology is also described and applied in Khare and Anderson (2006a) and Khare and Anderson (2006b). Ancell and Hakim (2007) derive a recursive formula to find the n'th most sensitive spot, conditional on the assimilation of the first $(n - 1)$ locations. We extend this work in the special case of choosing a single location at a grid point in each step, arriving at a formula that does not rely on matrix calculations; a summary is provided in Appendix 3. As in previous experiments, the averaging period is fixed at $\tau = 20$ and observation error at $\sigma_o = 0.10$.

The first EOF for $\overline{\theta}$ with $\tau = 20$ exhibits an east–west asymmetry due to the mountain near the center of the domain (Fig. 9). The similarity in structure at the ground and

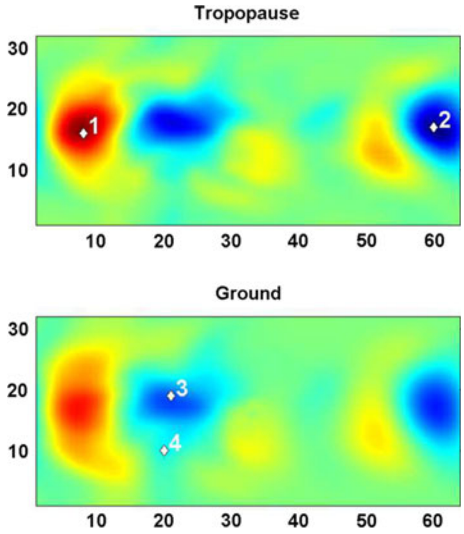


Fig. 9 First EOF for $\overline{\theta}$, $\tau = 20$, all boundary points; tropopause (*top*) and ground (*bottom*). The *marked locations* indicate the four observations to whose assimilation the coefficient of the first EOF is most sensitive, conditional on the assimilation of the earlier found locations. The assumed observation error variance is 0.10

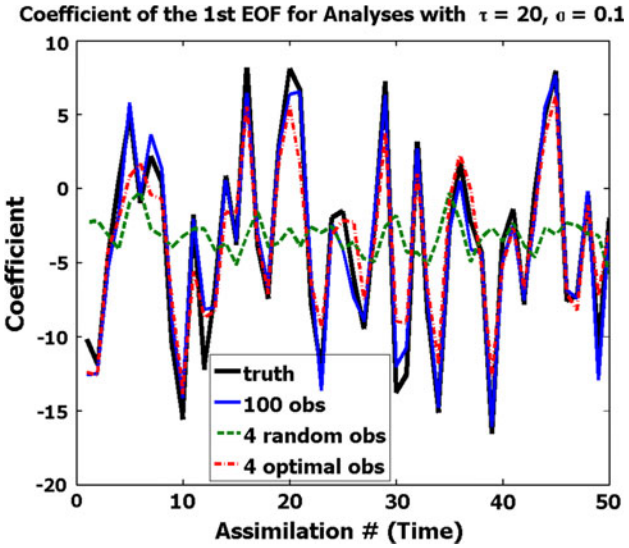


Fig. 10 The coefficients of the first EOF at each assimilation time over the course of each run are compared to the "truth" (*black*) for analyses from runs assimilating the 100 baseline observations (*blue*, *solid*), four randomly placed observations (*green*, *dashed*), and four optimal observations (*red*, *dot-dashed*). Remarkably, just four well-placed observations track the "truth" almost as well as 100 random ones

tropopause reflects an approximately barotropic flow. The first three locations that produce the greatest reduction in error variance for the leading EOF are located near extrema in the EOF, while the fourth is slightly to the south and west of the center of the domain. The fifth point (not shown) is located just to the east of point three, and the subsequent three points either coincide exactly with an earlier observation or are located at the same horizontal location but on the other surface. Moreover, the expected error-variance reduction drops by one order of magnitude from the first to the second observation, decreases for observations three and four, and is another order of magnitude less by the fifth observation. These results suggest that just four observations may significantly constrain the first EOF.

To determine the importance of observation location, we compare the coefficient of the first EOF of the "truth" run with those of the analysis and forecast from each of three experiments: (a) the 100 baseline observations (see Fig. 1), (b) four randomly placed observations (repeated four times to test a variety of configurations), and (c) the four optimal observations defined above (as marked in Fig. 9). As Fig. 10 indicates, assimilating four random observations has little influence on the analysis error of the EOF coefficient (dashed green line). (Note that the variability in this run is less than that in the "truth" run because an ensemble mean is taken.) On the other hand, for the four optimal observations, the analysis (red dot-dashed line) and forecast (not shown) curves track the true state (solid black line) almost as well as for the case with 100 observations (solid blue line). More quantitatively, 100 observations reduce the error in the EOF coefficient of the analysis by 81% from the control case, four random observations by $<1\%$, and four optimal observations by 62%.

The following strategies are now considered for extending the optimal network calculation beyond four observations: eight optimal observations for the first EOF, four optimal observations for the first EOF plus four for the second EOF, and an experiment with fifteen optimal observations for the first five EOFs (four each for the first and the second, two each for the third and the fourth, and three for the fifth). The specific numbers for the last experiment were chosen to include, for each EOF, only observations with large variance reduction relative to the next best observation. In all cases, observation locations are conditional on the assimilation of previously selected locations. Results are compared in Table 2 in terms of percentage improvement in rms errors over the control (no assimilation) case for the coefficients of the first two EOFs and the rms errors of $\overline{\theta}$ and θ.

By any measure, assimilating four randomly placed observations is not effective. However, the error reduction in the forecast produced by the baseline case (100 random observation locations) is approximated well by assimilating just fifteen optimally chosen observations. For the analysis, 100 random observations are better than the fifteen, which result in about two thirds as much error reduction. Note that using eight observations for the first EOF does not reduce rms errors as much as adding four additional observations for the second EOF, despite the fact that the second EOF accounts for 12% of the variance as compared to 20% for

Table 2 Error reduction due to data assimilation

Reduction in	Error in 1st EOF coeff		Error in 2nd EOF coeff		$\overline{\theta}$ rms error		θ rms error	
	fcst. (%)	anal. (%)	fcst. (%)	anal. (%)	fcst. (%)	anal. (%)	fcst. (%)	anal. (%)
100 random	33	81	39	84	11	52	6	23
4 random (avg)	0	0	1	5	0	0	0	0
4 for 1st EOF	10	62	19	21	3	12	2	7
8 for 1st EOF	13	65	26	36	4	16	3	8
4 (1st) + 4 (2nd EOF)	28	67	26	62	8	22	5	11
15 for 5 EOFs	36	66	34	73	10	30	6	15

Four types of error are evaluated: errors in the coefficient of the first and second EOF of $\overline{\theta}$ and rms errors in $\overline{\theta}$ and in θ. Both analysis and forecast data from six experiments are compared, as described in the text

the first EOF. Moreover, the reduction in forecast errors in the coefficient of the first EOF is less than half when all eight observations are chosen to constrain this EOF versus when four of the eight observations are chosen to constrain the second EOF. A better estimate of the projection onto the second EOF helps produce a better forecast of that onto the first, suggesting that the different patterns are dynamically coupled. Collectively, the first five EOFs account for 54% of the variance, and the experiment with 15 optimal observation locations yields a 30% rms error reduction for $\overline{\theta}$ as compared to a 52% reduction for the baseline case with 100 randomly chosen observations.

7 Accounting for model error

So far we have discussed perfect model experiments, where the true state is taken to be an independent ensemble member. The methods described here prove successful in this idealized context, but the real problem involves an imperfect model. While a comprehensive study of model error is beyond the scope of this study, we briefly discuss two additional experiments that provide some indication of the robustness of the results to the relaxation of the perfect model assumption. The first experiment addresses model resolution, which is especially important for paloeclimate simulation since the computational expense scales roughly with the cube of the horizontal grid spacing. The second experiment addresses a parametric error in the timescale for the damping of the solution to an assumed background temperature profile, which may be thought of as an error in radiative parameterization or solar forcing. In both cases, observations are drawn from the same truth simulation employed previously, by the method described in Sect. 3.1.

7.1 Reduced resolution

For the resolution test case, the baseline assimilation case with 100 random observations is repeated with doubled horizontal grid spacing. Averaging times considered include $\tau = 10$, 20, 30, and 40, and the assumed observational error is one third of the control rms error. The effect of the model error on the rms error in the time-averaged control variable $\overline{\theta}$ increases with averaging time τ.

The results show, as expected, that the rms errors for the reduced-resolution case (red line, Fig. 11) increase relative to the baseline case (blue line) by about 20–50% for the analyses. Note, however, that the rms errors for the control (no-assimilation) experiments differ only by 1–5 % (black lines with colored squares). While the data assimilation still produces significant error reduction in the analyses, its effectiveness is reduced noticeably when this type of resolution error is introduced. Forecast errors also increase in the reduced-resolution run compared to the identical model case (not shown), leading to insignificant improvement over the control for $\tau \geq 30$.

7.2 Reduced relaxation time

Here we consider an experiment where the time for damping back to a prescribed temperature field is reduced by one third. Effectively, the model variability is reduced, as the state is forced more strongly towards the background state. This is reflected in the larger errors in the control experiments without assimilation when damping is strengthened (black line with red squares, vs. black line with blue squares, Fig. 12). Nevertheless, the rms errors with assimilation compare very closely with the case having the correct damping timescale (red line, vs. blue line). Results are similar for forecast errors (not shown) and, interestingly, the rms errors in the instantaneous θ analyses and forecasts are essentially unchanged from the perfect model case (not shown). The assimilation counteracts the parameterization error more effectively than the resolution deficiency. We hypothesize that although the damping adversely affects low-frequency features, this effect is offset by a reduction in amplitude of high-frequency features that are poorly resolved by observations.

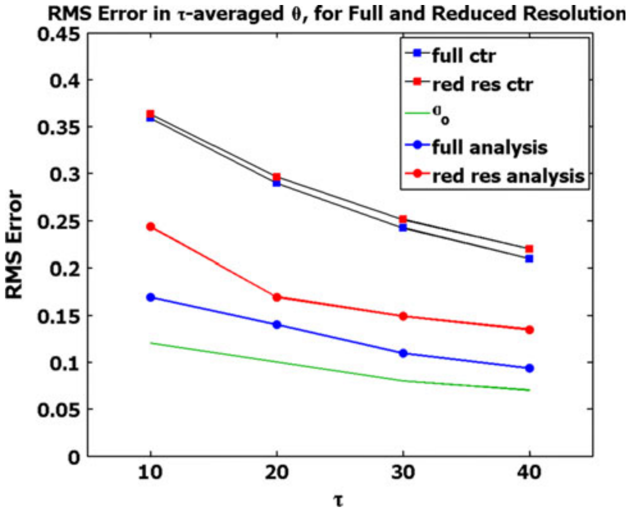


Fig. 11 Mean root-mean-square errors R^{avg} for $\overline{\theta}$ of the original control ensemble (*blue squares on black line*), the reduced resolution control ensemble (*red squares on black line*), the original model analyses (*blue lines with circles*), and the reduced resolution model analyses (*red lines with circles*). As a reminder the imposed observational errors are also shown (*green*)

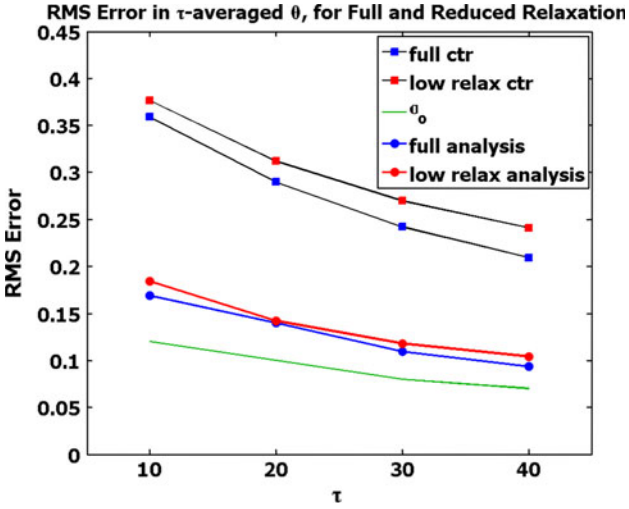


Fig. 12 Mean root-mean-square errors R^{avg} for $\overline{\theta}$ of the original control ensemble (*blue squares on black line*), the changed background state control ensemble (*red squares on black line*), the original model analyses (*blue lines with circles*), and the changed background state model analyses (*red lines with circles*). As a reminder the imposed observational errors are also shown (*green*)

8 Summary and Conclusions

The performance of a data assimilation method for time-averaged observations is evaluated for a quasi-geostrophic model of a midlatitude atmospheric jet interacting with a mountain. While the model is simple enough to allow a large number of simulations, it resolves realistic atmospheric climate phenomena, such as jets, baroclinic waves, cyclones and fronts, and their nonlinear interaction with a mountain, which produces lower frequency signals in the field. This makes the model well suited for addressing the problem of assimilating time-averaged observations that are affected by high-frequency transients, even though it lacks aspects necessary to fully simulate paleoclimate variability. Assimilation experiments reveal that, over a wide range of observation averaging times τ, significant reduction in analysis rms error is found relative to an ensemble simulation without assimilation (18% for $\tau = 100$ to 44% for $\tau = 10$ in the averaged variable). Moreover, errors in the analyses averaged over times shorter than the observations are also significantly improved. For observation averaging times shorter than the model's predictability timescale (i.e., the time period after which a model forecast becomes indistinguishable from climatology, approximately 30 time units), errors in the instantaneous state are also reduced.

These results suggest that assimilation of paleoclimate observations, which are typically averages over relatively long time periods, may successfully constrain a wide range of timescales. Since averaged quantities tend to have a spatially more coherent covariance pattern than instantaneous ones as well, relatively localized observations typical of the paleoclimate proxies may suffice to constrain a wide geographical region. An interesting result is that errors are reduced for averaging times shorter than the observations, which derives from the fact that the model physics spread information from observations in space and time.

Forecast errors show less improvement with observation assimilation than do analysis errors, particularly at longer averaging times. Beyond averaging times of about 30 time units, forecasts have no skill over climatology. In fact, for these timescales, assimilation using random draws from climatology to represent the background errors yields comparable results to the case where ensemble forecasts are used to estimate background errors. This outcome is potentially advantageous, since the majority of the EnKF computational expense is in the forecast step, and since very large ensembles can be drawn from climatology to deal more consistently with sampling error as compared to conventional techniques of inflation and localization. However, this approach basically reduces to the statistical reconstruction techniques we wish to improve upon, suffering from stationary statistics and model bias. Future research will be devoted to addressing this problem.

Experiments indicate limited sensitivity to observation error and number. Significant rms error reduction is found even for observations with errors much larger than climatological uncertainty, especially for analyses. Errors decrease steeply as a function of the number of randomly located observations from about ten to about fifty, and then decrease more gradually.

Optimal observation network design has also been considered, and we find that, for this model and an observation averaging time of $\tau = 20$, 15 optimally chosen observations recover about half of the error reduction of a random, evenly spaced network of 100 observations. These results are compelling, and suggest the possibility of targeting future proxy measurement campaigns to optimally reduce reconstruction errors. Clearly, further work with more sophisticated models is required before such ideas can be practically pursued, as well as to test the analysis and forecast results presented here. In particular, in order to extend the predictability timescale of the model beyond the annual averaging typical of many proxy measurements, a coupled ocean–atmosphere model seems essential. Finally, as many proxy measurements are chemical or biological, research is needed to construct methods for robust, adaptive forward models for the proxies.

Acknowledgments We thank Ryan Torn for discussions on ensemble filtering, Jeff Anderson for helpful comments on an earlier version of the manuscript, and Angie Pendergrass for discussions on literature related to paleoclimate state estimation and for reviewing an earlier version of the paper. We also thank two anonymous reviewers, whose comments have made this a better paper. The first author was supported in part by a National Science Foundation VIGRE grant and the second author was supported by the following grants: National Oceanic and Atmospheric Administration CSTAR Grant NA17RJ1232, Office of Naval Research Grant N00014-06-1-0510, and National Science Foundation grants 0552004 and 0902500, awarded to the University of Washington.

Appendix 1: Details of the assimilation procedure

We use a version of the ensemble Kalman filter for data assimilation. Its primary advantage over other methods is that the model statistics are derived with a Monte Carlo technique rather than being specified a priori or calculated explicitly, which allows them to be flow-dependent. The twist on the method introduced by Dirren and Hakim (2005) is the application of the filter to time-averaged state variables instead of the instantaneous ones. Below is an outline of the resulting procedure.

Let the model estimate of the state of the system at time t be given by the vector $X(t)$. For each assimilation step, a background forecast of the state of the system $X^b(t)$ is generated. The analysis after the assimilation will be represented as $X^a(t)$. Individual members of an ensemble will be denoted as $X_n(t)$ for $n = 1, 2, \ldots, N$. Each ensemble member can be decomposed into an ensemble mean $\langle X \rangle(t)$ and a perturbation from this mean $X_n{}^*(t)$. Similarly, the state can be decomposed into a time average $\overline{X}$ and a perturbation from this average $X'(t)$.

The observations y_o are averages over a time interval $[t_o, t_o + \tau]$. The model estimate of these observations is given by

$$y^e = \overline{H(X(t))}, \tag{12}$$

for some function H, possibly nonlinear. In our application, H chooses the potential temperature at a particular location, but such a simple form is not required by the theory. Note that the averaging is explicitly excluded from H. This differentiates our method from a traditional application of the EnKF to data collected as averages, which subsumes the averaging into H.

The assimilation algorithm proceeds as follows:

1. The state is advanced from t_o to $t_o + \tau$ to create an ensemble of N background state estimates
$$X_n^b(t), \quad n = 1, 2, \ldots, N.$$
2. For each ensemble member, the model estimate of the observations is computed,
$$y_n^e = \overline{H(X_n^b(t))}, \quad n = 1, 2, \ldots, N. \tag{13}$$
3. Each ensemble forecast is decomposed into its time average and a perturbation:
$$X_n^b(t) = \overline{X_n^b} + X_n'(t), \quad n = 1, 2, \ldots, N. \tag{14}$$
4. The time averages $\overline{X_n^b}$ are updated using established EnKF methods; here we assimilate observations sequentially, with the square-root filter described by Whitaker and Hamill (2002). We split time averages and model estimates of observations into an ensemble mean and perturbations from this mean:
$$\overline{X_n^b} = \left\langle \overline{X^b} \right\rangle + \overline{X_n^b}^* \tag{15}$$
$$y_n^e = \langle y^e \rangle + y_n^{e*}, \quad n = 1, 2, \ldots, N. \tag{16}$$
Updates are performed for each observation by
$$\left\langle \overline{X}^a \right\rangle = \left\langle \overline{X^b} \right\rangle + K(y_o - \langle y^e \rangle) \tag{17}$$
$$\overline{X}_n^{a*} = \overline{X_n^b}^* + \beta K\left(-y_n^{e*}\right), \quad n = 1, 2, \ldots, N, \tag{18}$$
where the Kalman gain is given by
$$K = \frac{\mathrm{cov}\left(\overline{X^b}^*, y^{e*}\right)}{\mathrm{var}(y^{e*}) + \mathrm{var}(y_o)} \tag{19}$$
and the reduced Kalman gain factor by
$$\beta = \left(1 + \sqrt{\frac{\mathrm{var}(y^{e*})}{\mathrm{var}(y^{e*}) + \mathrm{var}(y_o)}}\right)^{-1}. \tag{20}$$
The observations variance, $\mathrm{var}(y_o)$, is assumed known and is fixed in our experiments; other variances and covariances are estimated from the ensemble.

The analysis ensemble of time averages is then reconstructed:

$$\overline{X}_n^a = \langle \overline{X}^a \rangle + \overline{X}_n^{a*}, \quad n = 1, 2, \ldots, N. \tag{21}$$

5. An instantaneous analysis is derived by adding the perturbations from the average over time:

$$X_n^a(t) = \overline{X}_n^a + X_n'(t), \quad n = 1, 2, \ldots, N. \tag{22}$$

Returning to step 1, these analyses are used as initial conditions for advancing the state over the next time period τ.

Appendix 2: Equivalence of the Kalman filter applied to time averages and instantaneous quantities

The first part of the proposition we prove here shows equivalence of the two Kalman filter implementation under the assumption that averaged quantities and perturbations from these averages do not significantly co-vary. In the application at hand, this is a reasonable assumption (not shown). The second part concerns the equivalence in the averaged quantity, even when the assumption is relaxed.

We will continue to use overbars to indicate time averages and primes for perturbations from time averages.

Proposition 1 *Let H be the linear map from the state to the unaveraged observations. Let y denote the averaged observations.*

(a) *Assume $\mathrm{cov}(X, y) = \mathrm{cov}(\overline{X}, y)$, or equivalently $\mathrm{cov}(X', y) = 0$. Then applying the Kalman filter to instantaneous quantities or to averaged quantities results in the same analysis (within numerical accuracy).*

(b) *If the assumption does not hold, then the time average of the analyses will be identical, but the perturbations therefrom may differ.*

Proof Since the result is not dependent specifically on the ensemble implementation of the Kalman filter, we will keep the proof general and only stipulate that variances and covariances are derived consistently.

A background state $X^b(t)$ is given for $t \in [t_o, t_o + \tau]$, along with observations y_o. Let A denote the time-averaging operator on the state space, and let $P = I - A$ be the perturbation operator. We define $\hat{H} = HA$. Note that, by linearity of H, the estimates of the observations from the background state y^e become

$$y^e = \overline{H(X^b(t))} = H\left(\overline{X^b(t)}\right) = \hat{H}\left(X^b(t)\right). \tag{23}$$

Also provided are the background covariance matrix B for the augmented state vector containing variables for all $t \in [t_o, t_o + \tau]$ and the observation covariance matrix R. Note that the covariance matrix of the averaged state is given by $B^A = A\ B\ A^T$.

The traditional application of the Kalman filter to instantaneous quantities results in the analysis

$$X_1^a = X^b + K(y_o - y^e), \tag{24}$$

where

$$K = B\hat{H}^T\left[\hat{H}B\hat{H}^T + R\right]^{-1}. \tag{25}$$

Equation (24) can be separated in the update of the average and the update of the perturbation: $X_1^a = \overline{X_1^a} + X_1^{a\prime} = AX_1^a + PX_1^a$ with

$$\overline{X_1^a} = \overline{X^b} + AK(y_o - y^e), \tag{26}$$

$$X_1^{a\prime} = X^{b\prime} + PK(y_o - y^e), \tag{27}$$

With our approach, the time average alone is updated:

$$\overline{X_2^a} = \overline{X^b} + \tilde{K}(y_o - y^e), \tag{28}$$

where

$$\tilde{K} = B^A H^T\left[HB^A H^T + R\right]^{-1} = AK, \tag{29}$$

yielding the full analysis

$$X_2^a = \overline{X_2^a} + X^{b\prime}. \tag{30}$$

Substituting Eq. (29) into Eq. (28) and comparing to Eq. (26) shows that $\overline{X_1^a} = \overline{X_2^a}$, proving part (b). It follows that

$$X_1^a = X_2^a \Longleftrightarrow X_1^{a\prime} = X^{b\prime} \tag{31}$$

$$\Longleftrightarrow PK(y_o - y^e) = 0. \tag{32}$$

The assumption in part (a) is $\mathrm{cov}(X', y) = 0$, i.e., $PB\hat{H}^T = 0$, which gives us (32). ☐

Appendix 3: Observation sensitivity theory

The following is based on the sensitivity theory described in Ancell and Hakim (2007). We will briefly state those results needed here and then discuss the extension to a formula without matrix calculations for finding the n'th most sensitive observation location, conditional on the assimilation of the previous $(n - 1)$ observations. An alternative derivation for a similar approach can be found in Khare and Anderson (2006a).

Let $\mathbf{x}$ denote the state (column) vector and $\mathbf{X}$ the ensemble state matrix formed from N ensemble members. Further, let J be the scalar forecast metric of interest.

The sensitivity of J is measured in terms of the reduction of its variance σ^2 (or, equivalently, its error variance) that is expected from the assimilation of one or more new observations. Note that the variance is independent of the actual observed values, which determine the change in the mean of

J. Forming the ensemble of forecast metric values into the row vector $\mathbf{J}$, an estimate of the variance is given by

$$\sigma^2 = \frac{1}{N-1}\delta\mathbf{J}\delta\mathbf{J}^T. \tag{33}$$

Approximating the change in J from a leading-order Taylor-series truncation,

$$\delta\mathbf{J} = \left[\frac{\partial J}{\partial \mathbf{x}}\right]^T \delta\mathbf{X}, \tag{34}$$

it follows that

$$\sigma^2 = \left[\frac{\partial J}{\partial \mathbf{x}}\right]^T \mathbf{B}\left[\frac{\partial J}{\partial \mathbf{x}}\right], \tag{35}$$

where $\mathbf{B} = 1/(N-1)(\delta\mathbf{X})(\delta\mathbf{X})^T$ is the ensemble estimate of the state error covariance matrix.

For the Kalman filter, the relationship between the background error covariance matrix $\mathbf{B}_{i-1}$ and the analysis error covariance matrix $\mathbf{B}_i$ is given by

$$\mathbf{B}_i = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{B}_{i-1}, \tag{36}$$

where $\mathbf{H}_i$ is the map from the state to the new (i'th) observation and $\mathbf{K}_i$ is the Kalman gain associated with the new observation

$$\mathbf{K}_i = \mathbf{B}_{i-1}\mathbf{H}_i^T[\mathbf{H}_i\mathbf{B}_{i-1}\mathbf{H}_i^T + \mathbf{R}_i]^{-1}. \tag{37}$$

$\mathbf{R}_i$ here represents the i'th observation error covariance matrix. To simplify the notation, let us further define $\mathbf{E}_i$ as the total innovation error covariance matrix $\mathbf{H}_i\mathbf{B}_{i-1}\mathbf{H}_i^T + \mathbf{R}_i$, so that

$$\mathbf{B}_i = \mathbf{B}_{i-1} - \mathbf{B}_{i-1}\mathbf{H}_i^T\mathbf{E}_i^{-1}\mathbf{H}_i\mathbf{B}_{i-1} \tag{38}$$

The reduction in variance, which we seek to maximize, is then

$$\delta\sigma_i^2 = \left[\frac{\partial J}{\partial \mathbf{x}}\right]^T (\mathbf{B}_{i-1} - \mathbf{B}_i)\left[\frac{\partial J}{\partial \mathbf{x}}\right] \tag{39}$$

$$= \left[\frac{\partial J}{\partial \mathbf{x}}\right]^T \mathbf{B}_{i-1}\mathbf{H}_i^T\mathbf{E}_i^{-1}\mathbf{H}_i\mathbf{B}_{i-1}\left[\frac{\partial J}{\partial \mathbf{x}}\right]. \tag{40}$$

Note that the formula (40) can always be reduced to a function of the original covariance matrix $\mathbf{B}_o$, using successive substitutions from Eq. (38).

In particular,

$$\delta\sigma_1^2 = \left[\frac{\partial J}{\partial \mathbf{x}}\right]^T \mathbf{B}_o\mathbf{H}_1^T\mathbf{E}_1^{-1}\mathbf{H}_1\mathbf{B}_o\left[\frac{\partial J}{\partial \mathbf{x}}\right] \tag{41}$$

and

$$\delta\sigma_2^2 = \left[\frac{\partial J}{\partial \mathbf{x}}\right]^T \mathbf{B}_o(\mathbf{H}_2^T - \mathbf{H}_1^T\mathbf{E}_1^{-1}\mathbf{H}_1\mathbf{B}_o\mathbf{H}_2^T)\mathbf{E}_2^{-1}\ldots$$

$$(\mathbf{H}_2 - \mathbf{H}_2\mathbf{B}_o\mathbf{H}_1^T\mathbf{E}_1^{-1}\mathbf{H}_1)\mathbf{B}_o\left[\frac{\partial J}{\partial \mathbf{x}}\right], \tag{42}$$

where

$$\mathbf{E}_2 = \mathbf{H}_2\mathbf{B}_1\mathbf{H}_2^T + \mathbf{R}_2 \tag{43}$$

$$= \mathbf{H}_2\mathbf{B}_o\mathbf{H}_2^T - \mathbf{H}_2\mathbf{B}_o\mathbf{H}_1^T\mathbf{E}_1^{-1}\mathbf{H}_1\mathbf{B}_o\mathbf{H}_2^T + \mathbf{R}_2. \tag{44}$$

Formulae like these can, theoretically, be written down for any i, but they quickly get rather complicated. Instead, we consider how these calculations simplify in the case of each "set" of observations consisting of a single one taken directly from the state vector.

In this special case, $\mathbf{H}_j$ simplifies to a row vector of zeros with a single entry of 1, say the $i(j)$'th entry: $\mathbf{H}_j = [0\ldots010\ldots0]$. To simplify the notation, we will denote this element chosen as the j'th observation by x_j rather than $x_{i(j)}$, even though, of course, it does not necessarily occupy the j'th position in the state vector. Let R_j be the error variance of the observation of x_j, and define a function f such that

$$f_i(x_j, x_k) = \mathbf{H}_j\mathbf{B}_{i-1}\mathbf{H}_k^T \quad \text{and} \tag{45}$$

$$f_i(J, x_j) = \left[\frac{\partial J}{\partial \mathbf{x}}\right]^T \mathbf{B}_{i-1}\mathbf{H}_j^T. \tag{46}$$

Note that $f_1(a, b) = \mathrm{cov}(a, b)$ for any a, b.

At the i'th step, we want to maximize (rewriting Eq. 40)

$$\delta\sigma_i^2(x_i) = \frac{[f_i(J, x_i)]^2}{E_i(x_i)} \tag{47}$$

over all x_i, where

$$E_i(x_i) = f_i(x_i, x_i) + R_i. \tag{48}$$

Note that, at this stage, x_i is the independent variable, but x_1 through x_{i-1} are already chosen and fixed in previous steps.

For $i = 1$, (47) and (48) are trivial to determine, yielding

$$\delta\sigma_1^2 = \frac{[\mathrm{cov(J, x_i)}]^2}{\mathrm{var}(x_i) + R_1}. \tag{49}$$

For larger i, the following calculations have to be carried out, relying on the results from previous steps:

$$f_1(x_{i-1}, x_i) = \mathrm{cov}(x_{i-1}, x_i) \quad (M - i + 1 \text{ new covariance calculations}) \tag{50}$$

$$f_k(x_{i-1}, x_i) = f_{k-1}(x_{i-1}, x_i) - f_{k-1}(x_{k-1}, x_i)\frac{f_k(x_{i-1}, x_{k-1})}{E_{k-1}(x_{k-1})}$$

$$\text{for } k \in \{2, \ldots, i-1\} \quad \text{(2M + 1 operations/value of k)} \tag{51}$$

$$f_i(J, x_i) = f_{i-1}(J, x_i) - f_{i-1}(x_{i-1}, x_i)\frac{f_{i-1}(J, x_{i-1})}{E_{i-1}(x_{i-1})} \quad (2M + 1 \text{ operations}) \tag{52}$$

$$f_i(x_i, x_i) = f_{i-1}(x_i, x_i) - \frac{[f_{i-1}(x_{i-1}, x_i)]^2}{E_{i-1}(x_{i-1})} \quad (3M \text{ operations}) \tag{53}$$

$$E_i(x_i) = f_i(x_i, x_i) + R_i \qquad (M \text{ additions}) \tag{54}$$

$$\delta\sigma_i^2(x_i) = \frac{[f_i(J, x_i)]^2}{E_i(x_i)} \qquad (2M \text{ operations}) \tag{55}$$

The total computational cost consists of $M - k + 1$ covariances, $(4M - 1) + (2M + 1)k$ arithmetic operations, and the maximization.

References

Ancell B, Hakim GJ (2007) Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. Mon Weather Rev 135:4117–4134

Anderson J, Anderson S (1999) A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon Weather Rev 127(12):2741–2758

Anderson JL (2007) Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. Phys D 230:99–111

Dirren S, Hakim GJ (2005) Toward the assimilation of time-averaged observations. Geophys Res Lett 32(4):L04804

Gaspari G, Cohn SE (1999) Construction of correlation functions in two and three dimensions. QJR Meteorol Soc 125:723–757

Hakim GJ (2000) Role of nonmodal growth and nonlinearity in cyclogenesis initial-value problems. J Atmos Sci 57(17):2951–2967

Hakim GJ, Torn RD (2008) Ensemble synoptic analysis: synoptic-dynamic meteorology and weather analysis and forecasting—a tribute to Fred Sanders, p 36

Hoskins BJ, West NV (1979) Baroclinic waves and frontogenesis. Part II: uniform potential vorticity jet flows—cold and warm fronts. J Atmos Sci 36(9):1663–1680

Khare SP, Anderson JL (2006a) An examination of ensemble filter based adaptive observation methodologies. Tellus A 58:179–195

Khare SP, Anderson JL (2006b) A methodology for fixed observational network design: theory and application to a simulated global prediction system. Tellus A 58:523–537

Murphy JM (1988) The impact of ensemble forecasts on predictability. QJR Meteorol Soc 114:463–493

Torn RD, Hakim GJ (2008) Ensemble-based sensitivity analysis. Mon Weather Rev 136:663–677

Whitaker JS, Hamill TM (2002) Ensemble data assimilation without perturbed observations. Mon Weather Rev 130:1913–1924