The GEOS Retrospective Data Assimilation System: The 6-hour lag case*

Yanqiu Zhu[†‡], Ricardo Todling[†], Jing Guo[†], Stephen E. Cohn,
I. Michael Navon[§], and Yan Yang[§]
Data Assimilation Office, NASA/GSFC, Greenbelt, Maryland

To be submitted to the *Mon. Wea. Rev.*, 2002

*Preliminary results of this investigation were presented at the Second International Symposium on Frontiers of Time Series Modeling: Nonparametric Approach to Knowledge Discovery, Nara, Japan, December 2000.

[†]Additional affiliation: Science Applications International Corporation, Beltsville, Maryland.
[‡]Corresponding author address: Dr. Yanqiu Zhu, Data Assimilation Office, NASA/GSFC, Code 910.3, Greenbelt, MD 20771. *e-mail:* yzhu@dao.gsfc.nasa.gov.
[§]Affiliation: Department of Mathematics and C.S.I.T, Florida State University, Tallahassee, FL 32306

Abstract

The fixed-lag Kalman smoother (FLKS) has been proposed as the framework to construct data assimilation procedures capable of producing high-quality climate research datasets. The FLKS-based systems, referred to as retrospective data assimilation systems, are basically an extension to three-dimensional filtering procedures with the added capability of incorporating observations not only in the past and present time of the estimate, but also at future times. A variety of simplifications are necessary to render FLKS-based retrospective assimilation procedures practical.

In this article, we present an FLKS-based retrospective data assimilation system implementation for the Goddard Earth Observing System Data Assimilation System (GEOSDAS). The practicality of this implementation comes from the practicality of its underlying (filter) analysis system, i.e., the physical-space statistical analysis system (PSAS). The behavior of two schemes are studied here. The first retrospective analysis (RA) scheme is designed to simply update the regular PSAS analyses with observations available at times ahead of the regular analysis times. Although our GEOSDAS implementation is general, results are only presented for when observations 6-hours ahead of the analysis time are used to update the PSAS analyses and calculate the so-called lag-1 retrospective analyses. Consistency tests for the RA scheme show that the lag-1 retrospective analyses indeed have better 6-hour predictive skills than the predictions from the regular analyses. This motivates the introduction of the second scheme which, at each analysis time, uses the 6-hour retrospective analysis to replace the first-guess normally used in the PSAS analysis, and therefore allows the calculation of a revised (filter) PSAS analysis. Since in this scheme the lag-1 retrospective analyses influence the filter results, this procedure is referred to as the retrospective-based iterated analysis (RIA) scheme. Results from the RIA scheme indicate its potential for improving the overall features of the assimilation.

1 Introduction

The concept of retrospective data assimilation, as invoked in the present article, was introduced by Cohn et al. (1994) as the procedure to calculate analyses using observations in the past, present, and future of the times at which analyses are sought for. In estimation theory, estimates of the state of a system produced this way are known as smoother estimates, and the techniques behind deriving them are known as smoothing. In sequential data assimilation the most natural smoothing technique to employ is that of fixed-point smoothing. In this case, the usual filter estimate obtained at a fixed time using observations before and at the analysis time is sequentially updated as future observations became available. Future observations can be used for as long as experimentation show their impact to be useful. The idea of estimating the state of a system at a fixed time over and over again the more observations become available can be taken a step further by seeking fixed point estimates at a series of consecutive fixed times. This is what is accomplished by fixed-lag smoothing. Specifically, for linear systems under the typical assumption of unbiased Gaussian-distributed errors the fixed-lag Kalman smoother (FLKS) provides the best linear unbiased estimate of the state of the system at a given time using observations in the past, present, and at a time lag-ℓ ahead of the time of the estimate. The FLKS was proposed by Cohn et al. (1994; CST94 hereafter) as the strategy to follow when developing retrospective data assimilation schemes designed to produce the highest possible quality datasets for climate research.

The FLKS is composed of two major components: the Kalman filter (KF) portion and the fixed-lag smoother portion. The FLKS is fully dependent on the KF as it is formulated on the basis of the observation-minus-forecast residuals used by the KF. In general, when the filter is not the KF, but rather some suboptimal implementation of it, we can still think of suboptimal implementations of FLKS-based retrospective data assimilation schemes as consisting of a filter portion and a smoother (or retrospective) portion. Todling et al. (1998) used this explicit separation between filtering and smoothing to study the behavior of a variety of combinations of filter and smoother approximations to the linear FLKS. One particular approximation studied then, namely the adaptive CCF-based retrospective data assimilation scheme, could be seen as having the potential for being implemented in practice. The abbreviation CCF stands for *constant forecast error covariance filter* which in many respects is a good representation of some practical three-dimensional variational analysis systems operational today. The spectral statistical-interpolation analysis system of Parrish and Derber (1992) is an example of such a system; the U.S. Navy analysis system of Daley and Barker (2001) is another; and so is the physical-space statistical analysis system (PSAS) of Cohn et al. (1998), which is also central to the work in the present article.

To really take forward the idea of developing a practical retrospective data assimilation system the linear FLKS formulation of CST94 has to be extended to handle nonlinear dynamics. Since the retrospective portion of the algorithm relies completely on the filter, designing nonlinear filters immediately result in designing nonlinear smoothers. Todling and Cohn (1996; TC96 hereafter) derived a nonlinear FLKS algorithm based on the traditional extended Kalman filter (EKF). Similar derivations can be found elsewhere (e.g., Biswas and Mahalanabis 1973; Verlaan 1998). The way smoothers use future observations to calculate updates to state estimates is by propagating information back in time using the adjoint dynamical model. For nonlinear dynamics the adjoint of the tangent linear dynamics must be provided in principle, e.g., as when using an EKF-based FLKS. Four-dimensional variational (4D-var) procedures such as that of Rabier et al. (2000) also require the adjoint of the tangent linear dynamics. The need for the adjoint model can be avoided if the retrospective assimilation strategy is designed on the basis of ensemble techniques such as that of Evensen and van Leeuwen (2000).

In this article, we study the performance of a PSAS-based retrospective analysis (RA) system developed for the Goddard Earth Observing System (GEOS) data assimilation system (GEOS-

DAS). Since the current forecast error covariance matrix of PSAS varies very slowly in time we can identify the GEOS-based suboptimal RA procedure studied here with the CCF scheme of Todling et al. (1998). Our RA implementation in GEOSDAS is general and applicable to any number of time lags, but in the present article we concentrate on results for the 6-hour, i.e. lag-1, retrospective analysis. Motivated by some of the results obtained with this version, and by the ideas of constructing so-called iterated filters and smoothers common in the engineering literature, we also study here the performance of a retrospective-based iterated analysis (RIA) scheme. In the RIA, the lag-1 retrospective analysis at a given time t_{k-1} is used to produce a new first-guess at time t_k that is used to revise the filter analysis, i.e. PSAS, at the same time t_k. In the RIA the final analysis is the second (iterated) analysis calculated using the first-guess generated from the lag-1 retrospective analysis. This is a considerably different use of the "static" retrospective analyses proposed by CST94. Though a formal argument for the RIA procedure is not presented here, the procedure is found to beneficially impact the overall quality of the analyses. This lag-1 RIA scheme makes the retrospective procedure resemble a 4D-var cycle (e.g., Courtier et al. 1994; Rabier et al. 2000).

Indeed, the original FLKS-based retrospective analysis formulation of CST94 is meant as an alternative approach to 4D-var. The FLKS framework is a natural four-dimensional extension to three-dimensional procedures formulated sequentially rather than variationally. Four-dimensional variational procedures are an extension of 3D-var that take into account observations within a time interval. Ménard and Daley (1996) have showed the equivalence of 4D-var and fixed-interval smoothing. Similarly, for linear dynamics, the FLKS is algebraically equivalent to 4D-var and can be derived from the 4D-var cost function by solving a two-point boundary value problem (Zhu et al. 1999). The main distinction between 4D-var and the FLKS is in their computational approach. The former involves an iterative optimization procedure to get to the solution, whereas the latter deals directly with the analytical solution of the problem. One practical consequence of this distinction relates to how these procedures allow to account for model error. As pointed out by Todling et al. (1998), FLKS-based assimilation schemes directly enherit any model error covariance parameterization embbeded in the filter portion. Various techniques can be found in the literature that account for model error in 4D-var by using the dynamical model as a weak constraint to the optimization problem (e.g., Derber 1989; Bennett et al. 1996; and Zupaski 1997). However, until reasonable understading of model error is acquired, and the corresponding model error covariance parameterizations can be relied upon, this distinction between 4D-var and FLKS-based assimilation is rather mute. Another important point to make relates to what it now seems to be recognized (Fisher and Andersson 2001) as one of the main advantages of 4D-var over 3D-var-like procedures, namely that, the former uses the observations nearly at their proper times [as it is the case in the European Centre for Medium-Range Forecasts (ECMWF) 4D-var implementation of Rabier et al. 2000)], whereas in the latter it is more common to bundle the observations into 6-hour batches. This can be resolved, particularly in sequential, 3D-var assimilation procedures by using a rapid update cycle strategy. Though this is not explored in the present article, since in GEOSDAS observations are bundled into 6-hour batches, we should point out that there is no intrinsic difficulty in building an FLKS-based retrospective analysis system under the rapid update cycle filtering strategy.

In the sequel, we briefly review, in section 2, the theoretical framework behind retrospective analysis. The presentation is based on the EKF and the corresponding nonlinear extension of the FLKS. In section 3, we describe the framework of our practical implementation directed to adding a retrospective component to GEOS; here, both the RA and RIA schemes are presented. In section 4, results of a preliminary evaluation of these retrospective schemes are presented and discussed. Conclusions are drawn in section 5.

2 Theoretical framework: the fixed-lag Kalman smoother

In this section we briefly recapitulate the formulations of the fixed-lag Kalman smoother of CST94 and TC96. Following Todling et al. (1998) we separate the FLKS into a filter portion and a retrospective portion. The filter portion is based on the linear Kalman filter, or more generally in any nonlinear extension of the KF; the retrospective portion is based on the linear fixed-lag Kalman smoother, or any equivalent nonlinear extension compatible with the underlying filter. As in TC96, the discussion below is based on the EKF.

(a) The filter portion

Using the notation of CST94, the filter portion of the FLKS formulation of TC96 can be summarized by the usual EKF equations

$$\mathbf{w}^f_{k|k-1} = \mathcal{A}_{k,k-1}(\mathbf{w}^a_{k-1|k-1}) \ , \tag{1a}$$

$$\mathbf{w}^a_{k|k} = \mathbf{w}^f_{k|k-1} + \mathbf{K}_{k|k}\mathbf{v}_k \ , \tag{1b}$$

$$\mathbf{K}_{k|k} = \mathbf{P}^f_{k|k-1}\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k \ , \tag{1c}$$

$$\mathbf{P}^f_{k|k-1} = \mathbf{A}_{k,k-1}\mathbf{P}^a_{k-1|k-1}\mathbf{A}^T_{k,k-1} + \mathbf{Q}_k \ , \tag{1d}$$

$$\mathbf{P}^a_{k|k} = (\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_k)\mathbf{P}^f_{k|k-1} \ . \tag{1e}$$

The first two expressions refer to the state evolution, which depend on the last three expressions essentially related to error covariance evolution and update. At time t_k, the forecast n-vector $\mathbf{w}^f_{k|k-1}$ evolves through the nonlinear dynamical operator $\mathcal{A}_{k,k-1}$ from the analysis n-vector $\mathbf{w}^a_{k-1|k-1}$, according to (1a). The dynamical operator $\mathcal{A}_{k,k-1}$ stands for, say, a general circulation model, and possibly any transformations necessary to convert the model prognostic variables into the filter state vector, and vice-versa.

The main difference in the EKF equations written above and the way they more commonly appear in the atmospheric data assimilation literature (e.g., Miller et al. 1994) is in the time subscripts. Here, the subscripts follow standard engineering notation developed in linear estimation theory and mostly suitable to the development of smoothers. This subscript notation is also particularly helpful in reminding that for linear systems perturbed by Gaussian-distributed noise the forecast $\mathbf{w}^f_{k|k-1}$ and analysis $\mathbf{w}^a_{k|k}$ state vectors are actually the *conditional means* of the true state n-vector $\mathbf{w}^t_k$, that is,

$$\mathbf{w}^f_{k|k-1} = \mathcal{E}\{\mathbf{w}^t_k|\mathbf{w}^o_{k-1}, \cdots, \mathbf{w}^o_1\}\,, \tag{2a}$$

$$\mathbf{w}^a_{k|k} = \mathcal{E}\{\mathbf{w}^t_k|\mathbf{w}^o_k, \mathbf{w}^o_{k-1}, \cdots, \mathbf{w}^o_1\}\,, \tag{2b}$$

at time t_k. The conditioning, represented by the vertical bar in the ensemble mean operator $\mathcal{E}\{\bullet|\bullet\}$, is on the time series of observations $\mathbf{w}^o_k$. The forecast at time t_k is an estimate of the true state ensemble mean conditioned on all observations prior to time t_k; equivalently, the analysis at time t_k is an estimate of the true state ensemble mean conditioned on all observations up to and *including* those at time t_k.

The EKF, as the KF, is based on the residual p_k-vector $\mathbf{v}_k$ in (1b) formed of the difference between the p_k-vector of observations $\mathbf{w}^o_k$ and the model predicted "observations" $\mathcal{H}_k(\mathbf{w}^f_{k|k-1})$ at time t_k, that is,

$$\mathbf{v}_k \equiv \mathbf{w}^o_k - \mathcal{H}_k(\mathbf{w}^f_{k|k-1})\,. \tag{3}$$

The nonlinear operator $\mathcal{H}_k$ stands for the transformations involved in converting filter state vector quantities into observables. Optimality of the filter depends on the $n \times p_k$ weighting matrix $\mathbf{K}_{k|k}$ given to this observation-minus-forecast (OMF) residual vector $\mathbf{v}_k$ through (1b). Although the expression for the weighting matrix $\mathbf{K}_{k|k}$ in the EKF is very similar in form to its linear KF equivalent, contrary to the linear case, $\mathbf{K}_{k|k}$ in (1c) is now state-dependent since the $p_k \times n$ Jacobian matrix $\mathbf{H}_k$ of the observation operator $\mathcal{H}_k$ is linearized around the forecast state vector $\mathbf{w}^f_{k|k-1}$. State dependence of the EKF weighting matrix $\mathbf{K}_{k|k}$ also comes from its dependence on the OMF residuals covariance matrix $\mathbf{\Gamma}_k$, given by

$$\mathbf{\Gamma}_k = \mathbf{H}_k\mathbf{P}^f_{k|k-1}\mathbf{H}^T_k + \mathbf{R}_k\,, \tag{4}$$

for uncorrelated observation and forecast errors. Here, $\mathbf{R}_k$ is the $p_k \times p_k$ observation error covariance matrix and $\mathbf{P}^f_{k|k-1}$ is the state-dependent $n \times n$ forecast error covariance matrix. The dependence of the forecast error covariance matrix in (1d) on the state comes from the $n \times n$ Jacobian matrix $\mathbf{A}_{k,k-1}$ of the dynamics operator $\mathcal{A}_{k,k-1}$ and its linearization about model trajectory initialized from the analysis vector $\mathbf{w}^a_{k-1|k-1}$. The forecast error covariance matrix also depends on the model error covariance matrix $\mathbf{Q}_k$, which is normally assumed to be known. Since the forecast error covariance matrix $\mathbf{P}^f_{k|k-1}$ evolves from the $n \times n$ analysis error covariance matrix $\mathbf{P}^a_{k-1|k-1}$ it depends further on the accuracy of the previous estimate calculated by the filter, i.e., through (1e) applied at time t_{k-1}.

In the linear case, the dynamics and observation operators reduce $\mathcal{A}_{k,k-1} = \mathbf{A}_{k,k-1}$ and $\mathcal{H}_k = \mathbf{H}_k$, respectively, and (1) reduces to the linear KF for known model and observation error statistics. Moreover, as pointed out in TC96, in the linear Gaussian-distributed noise case, the forecast and analysis error covariance matrices are the *conditional mean* error covariances. It is when the observation errors are white, Gaussian, in time that the time series of residual vectors $\mathbf{v}_k$ can be identified with the innovation sequence (see for example, Anderson and Moore 1979, section 5.3)

(b) The retrospective portion

In the FLKS, the retrospective portion uses the OMF residual vector $\mathbf{v}_k$ at time t_k to calculate corrections to previous filter analyses and to previous retrospective analyses using an update equation similar to the state update expression (1b) of the filter portion. That is, the lag-ℓ retrospective analysis is calculated by

$$\mathbf{w}^a_{k-\ell|k} = \mathbf{w}^a_{k-\ell|k-1} + \mathbf{K}_{k-\ell|k}\mathbf{v}_k \ , \tag{5}$$

for $\ell = 1, 2, \ldots, \min(k, L)$, and a maximum desired lag $\ell = L$. Each retrospective analysis corresponds to an "incremental" correction to an estimate of the state calculated previously.

For example, when $k = \kappa$ and $\ell = 1$, the lag-1 retrospective analysis $\mathbf{w}^a_{\kappa-1|\kappa}$ corresponds to a correction to the most recently available state estimate at time $t_{\kappa-1}$, i.e., the filter analysis $\mathbf{w}^a_{\kappa-1|\kappa-1}$; when $k = \kappa + 1$ and $\ell = 2$, the lag-2 retrospective analysis $\mathbf{w}^a_{\kappa-1|\kappa+1}$ corresponds of a correction to the most recently available state estimate at time $t_{\kappa-1}$ which is now the lag-1 retrospective analysis $\mathbf{w}^a_{\kappa-1|\kappa}$; and so on up to the desired lag $\ell = L$ when the estimate at time $t_{\kappa-1}$ is given by the lag-L retrospective analysis $\mathbf{w}^a_{\kappa-1|\kappa+L-1}$.

This example to illustrate the mechanism for correcting consecutive state estimates at a given time with successive smoother calculations makes the FLKS algorithm resemble very much the fixed-point smoother. This is simply because in this example we chose to fix the time at which estimates are being sought, that is, time $t_{\kappa-1}$. The resemblance between the fixed-lag and fixed-point smoothers is no coincidence. The FLKS of CST94 and CT96 can be derived from a fixed-point smoother formulation using, for example, the approach of state augmentation (e.g., Biswas and Mahalanabis 1973). The point of matter here is that the "incremental" corrections to the state estimates at time $t_{\kappa-1}$ are calculated on the basis of the OMF residual vectors $\mathbf{v}_\kappa$, $\mathbf{v}_{\kappa+1}$, and so on up to $\mathbf{v}_{\kappa+L-1}$. That is, each lag of the algorithm introduces corrections to the state estimate by using observations at times ahead of the retrospective analysis time, up to the maximum desired lag L.

Because the retrospective analyses are based on the same OMF residual vectors used in the filter portion of the algorithm, the retrospective $n \times p_k$ weighting matrix $\mathbf{K}_{k-\ell|k}$ depends on the OMF residual covariance matrix $\mathbf{\Gamma}_k$ in (4). Furthermore, $\mathbf{K}_{k-\ell|k}$ also depends on the $n \times p_k$ matrix $\mathbf{H}_k^T$, the transpose of the Jacobian of the observation operator, and on the $n \times n$ forecast-analysis cross-covariance matrix $\mathbf{P}^{fa}_{k,k-\ell|k-1}$, through the EKF-based expression

$$\mathbf{K}_{k-\ell|k} = (\mathbf{P}^{fa}_{k,k-\ell|k-1})^T \mathbf{H}_k^T \mathbf{\Gamma}_k^{-1} \ , \tag{6}$$

as can be found in TC96. The forecast-analysis cross-covariance $\mathbf{P}^{fa}_{k,k-\ell|k-1}$ evolves from previously calculated analysis error covariances and analysis-analysis error cross-covariances through the Jacobian $\mathbf{A}_{k,k-1}$ of the dynamics operator. Its evolution equation and the update equations for the retrospective analysis error cross-covariances are

$$\mathbf{P}^a_{k-\ell|k} = \mathbf{P}^a_{k-\ell|k-1} - \mathbf{K}_{k-\ell|k}\mathbf{H}_k\mathbf{P}^{fa}_{k,k-\ell|k-1} \,, \tag{7a}$$

$$\mathbf{P}^{aa}_{k,k-\ell|k} = (\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_k)\mathbf{P}^{fa}_{k,k-\ell|k-1} \,, \tag{7b}$$

$$\mathbf{P}^{fa}_{k,k-\ell|k-1} = \mathbf{A}_{k,k-1}\mathbf{P}^{aa}_{k-1,k-\ell|k-1} \,, \tag{7c}$$

and the details of their derivation can also be found in TC96.

That retrospective analyses are built on the basis of future observations can be simply understood by recalling the meaning of the time subscript notation used here. In the linear Gaussian-distributed noise case the time subscript notation signifies that the retrospective analysis estimates are indeed estimates of the *conditional means*. In this case, the retrospective analysis at time $t_{k-\ell}$ is

$$\mathbf{w}^a_{k-\ell|k} = \mathcal{E}\{\mathbf{w}^t_{k-\ell}|\mathbf{w}^o_k, \mathbf{w}^o_{k-1}, \cdots, \mathbf{w}^o_1\} \,, \tag{8}$$

where now, in contrast to the filter estimates (2), the ensemble mean is conditioned on all observations before, during and after time $t_{k-\ell}$ and up to time t_k. As mentioned previously, in the linear optimal case, when the underlying filter is the KF and the sequence of OMF residual vectors is actually the innovation sequence, the retrospective portion just described reduces to the optimal FLKS. Independently of nonlinearities, in general, if the filter is suboptimal the corresponding retrospective analyses are suboptimal as well. This is simply because both the filter and the smoother are based on the same sequence of OMF residual vectors $\mathbf{v}_k$. Unfortunately, in the suboptimal case, there is no guarantee that consecutive retrospective lagged estimates will represent improvements over estimates with smaller lag(s) or even over the filter results (see Todling et al. 1998 for illustration).

As pointed out by Todling et al. (1998), one interesting feature of the FLKS that is directly related to it being formulated on the basis of an underlying filter is that it incorporates model error covariances naturally (see also appendix A, here). A variety of techniques exist to incorporate model error in 4D-var (e.g., Derber 1989; Bennett et al. 1996; and Zupanski 1997). Since 4D-var is equivalent to fixed-interval smoothing (see Ménard and Daley 1996; and Zhu et al. 1999) and for all practical purposes we can always choose a lag L in fixed-lag smoothing that accomplished the same benefit as fixed-interval smoothing (Moore 1973), FLKS-based assimilation procedures present a potential alternative to 4D-var. Since we currently lack the necessary knowledge to parameterize model error covariances this advantage of the FLKS over 4D-var is not very significant, but it may prove to be relevant in the future.

3 Practical framework: GEOSDAS considerations

When it comes to practical applications, the algorithm described in the previous section serves mainly as a guide to help design suitably feasible data assimilation procedures. Besides the overly stressed computational cost requirements argument made against practical implementations of assimilation procedures such as the one presented in the previous section, there is the more fundamental argument made about the nearly utter lack of knowledge of the underlying error statistics, specifically, of the model and the observation error statistics. Both these arguments have motivated the study of a number of simplifications to both filtering (e.g., Cohn and Todling 1996, and references therein) and smoothing (e.g., Todling et al. 1998, and references therein) procedures. In this section, we describe the details of the implementation of the FLKS-based retrospective procedure as we see fit to GEOSDAS. Before describing the GEOS retrospective analysis portion of the implementation we review the current GEOSDAS that is the practical equivalent of the filter portion of the algorithm.

(a) The GEOS analysis and data assimilation system

The DAO operational GEOS data assimilation system consists of three major components:

an atmospheric general circulation model (GCM); the physical-space statistical analysis system (PSAS); and the incremental analysis update (IAU) procedure. At the so called analysis times, the GCM provides a first-guess field to PSAS so it can process OMF residuals and generate the analysis state. Essentially, PSAS is the practical equivalent to the EKF equations (1b)-(1c), and obtains the analysis state as a correction to the model first-guess. For practical reasons, PSAS neglects the error covariance expressions (1d) and (1e) and therefore it is a suboptimal filter. The suboptimality of GEOS analyses is not unique to this system but indeed all operational systems to date are suboptimal implementations of their underlying mathematical framework. Each PSAS analysis is used in the IAU procedure of Bloom et al. (1996) to construct a tendency term that is used to force the GCM during a 6-hour period around the analysis time. The GCM trajectory obtained during the IAU integration is known as the *assimilated* trajectory.

In GEOSDAS the state-space of the GCM is different than the state-space of the analysis system and it is convenient to define a specific nomenclature for the purposes of the present article. In what follows, we refer to *background* as the state-vector provided by the GCM and to *forecast* or *first-guess* as the background field transformed to the analysis space. The model and analysis spaces are different because their state variables and grids are different. For instance, the GCM state variables are surface pressure, potential temperature, specific humidity and the zonal and meridional components of the wind, where all variables are defined on the Arakawa C-grid and on a vertical sigma coordinate system. On the other hand, the analysis state vector is composed of sea level pressure, zonal and meridional components of the sea level wind, zonal and meridional components of the upper-air wind, mixing ratio, and geopotential heights, where all variables are defined on the Arakawa A-grid and in pressure coordinates (see for example DAO 1996, for details).

We designate an m-dimensional GCM state vector by $\mathbf{y}(\sigma)$ and an n-dimensional analysis state vector by $\mathbf{w}(p)$, where the space distinction between the model and analysis is emphasized in the notation by showing explicitly the vertical coordinate system these states are defined on. For our purposes, we can represent a GCM integration during the time interval τ as

$$\frac{d\mathbf{y}(\sigma)}{dt} = \mathcal{M}[\mathbf{y}(\sigma)] + \alpha\,\delta\mathbf{y}^a_{k|k}(\sigma)\,. \tag{9}$$

Here, $\mathcal{M}$ is the nonlinear GCM operator and the second term in the rhs corresponds to the constant IAU forcing term applied to the GCM during the IAU integration period. The parameter α controls when and how the model-space analysis increment $\delta\mathbf{y}^a_{k|k}(\sigma)$ affects the integrations. During the IAU time interval $\tau = \tau^{iau} = t_{k+1/2} - t_{k-1/2}$ we set $\alpha = 1/\tau^{iau}$; and during the 3-hour GCM background integration time interval $\tau = t_{k+1} - t_{k+1/2}$ we set $\alpha = 0$. At an analysis time t_k, the GCM-provided background field $\mathbf{y}^b_{k|k-1}(\sigma)$ is converted into the analysis first-guess through the operation

$$\mathbf{w}^f_{k|k-1}(p) \quad = \quad \Pi[\mathbf{y}^b_{k|k-1}(\sigma)]\,, \tag{10}$$

where for convenience we use similar time subscript notation as that used in the previous section. The space conversion operator Π is nonlinear since it represents not only simple interpolation from one grid to another but also variable transformations such as conversion from potential temperature to geopotential heights.

The forecast vector $\mathbf{w}^f_{k|k-1}(p)$ is used to construct the OMF residual p-vector $\mathbf{v}_k$ in (3). Instead of calculating explicitly the weighting matrix (1c), PSAS splits the calculation of the last term in the analysis equation (1b) in two steps. The first step is to solve the following linear system of equations

$$\mathbf{\Gamma}_k\mathbf{x}_k = \mathbf{v}_k\,, \tag{11}$$

for the variable $\mathbf{x}_k$, so that in a second step the analysis $\mathbf{w}^a_{k|k}(p)$ can be calculated by

$$\mathbf{w}^a_{k|k}(p) = \mathbf{w}^f_{k|k-1}(p) + \mathbf{P}^f_{k|k-1}\mathbf{H}^T_k\mathbf{x}_k\,. \tag{12}$$

To keep notation simple, we denote the PSAS forecast error covariance with the same symbol $\mathbf{P}^f_{k|k-1}$ used in the previous section. However, as mentioned above, PSAS does not use the equivalent of (1d) to calculate the forecast error covariance matrix. Instead, the forecast error covariance in PSAS is parameterized using simple dynamical constraints. Only its variance fields vary (slowly) in time; its correlations are constant in time. A consequence of such simplification is that the forecast $\mathbf{w}^f_{k|k-1}(p)$ and the analysis $\mathbf{w}^a_{k|k}(p)$ vectors in (12) are also distinct from those of the previous section, even though they are designated with the same symbols as in the previous section. Notice also that currently the forecast error covariance formulation of PSAS is for the analysis variables and, in particular, in pressure coordinates. Moreover, the observation operator $\mathcal{H}_k$ in PSAS is linear, that is, $\mathcal{H}_k = \mathbf{H}_k$.

To proceed with the GEOS IAU assimilation, the analysis in (12) is converted back to the model space, through a conversion operator Π^+, as in

$$\mathbf{y}^a_{k|k}(\sigma) \;\; = \;\; \Pi^+[\mathbf{w}^a_{k|k}(p)]\,, \tag{13}$$

which is then used to finally construct the IAU $\delta\mathbf{y}^a_{k|k}(\sigma)$ increment to be used in (9),

$$\delta\mathbf{y}^a_{k|k}(\sigma) \;\; = \;\; \mathbf{y}^a_{k|k}(\sigma) - \mathbf{y}^b_{k|k-1}(\sigma)\,. \tag{14}$$

The actual implementation of Π^+ is such that it renders minimal the difference $||\mathbf{w}(p) - \Pi[\Pi^+[\mathbf{w}(p)]]||$. In other words, the error is minimal when transforming an analysis state vector into the model space and then transforming the resulting vector back to the analysis space. In a nonlinear sense, Π^+ is a pseudo-inverse of Π.

A schematic representation of the IAU assimilation procedure is shown in Fig. 1. In GEOS-DAS observations are processed at 6-hour intervals, which in the IAU framework implies that the GCM is integrated for 6 hours starting 3 hours before the analysis time. Going from the left to right in the diagram, at an analysis time, say $t = 6$Z, observations and a 3-hour model first-guess (represented by the north-eastward pointing dashed arrow) are combined in PSAS to calculate the filter analysis. This analysis is used to construct the IAU increment (14) and the model is integrated forward forced by the IAU tendency starting from $t = 3$Z up to $t = 9$Z.

Beyond this time, the IAU forcing is set to zero and the model runs "free" for the next 3 hours. At the end of this free 3-hour integration the GCM provides the background to be used in the PSAS analysis of the 12Z observations, and the cycle is repeated. The assimilated trajectory is represented in the figure by the thick-solid eastward-pointing arrows.

(b) The GEOS retrospective analysis

We have now the challenge of converting the retrospective portion of the FLKS as presented in the previous section into a practical algorithm. We have seen above that when building a practical filtering procedure such as PSAS one of the main approximations is to avoid dealing directly with the error covariance equations (1d)-(1e). Analogously, when building a practical implementation of the retrospective portion of the FLKS we want to calculate retrospective increments

$$\delta \mathbf{w}^a_{k-\ell|k}(p) \;\; \equiv \;\; \mathbf{w}^a_{k-\ell|k}(p) - \mathbf{w}^a_{k-\ell|k-1}(p) \; = \; \mathbf{K}_{k-\ell|k}\mathbf{v}_k \, , \tag{15}$$

for lags $\ell = 1, 2, \cdots, min(k, L)$, without having to calculate the smoother cross-covariances implicit in the retrospective gains $\mathbf{K}_{k-\ell|k}$ through (6) and (7). As it turns out, calculating these cross-covariances can be avoided since the retrospective gain matrices $\mathbf{K}_{k-\ell|k}$ can be written as

$$\mathbf{K}_{k-\ell|k} \;\; = \;\; \mathbf{P}^f_{k-\ell|k-\ell-1} \left[\prod_{j=k-\ell+1}^{k} (\mathbf{I} - \mathbf{K}_{j-1|j-1}\mathbf{H}_{j-1})^T \mathbf{A}^T_{j,j-1} \right] \mathbf{H}^T_k \mathbf{\Gamma}^{-1}_k \, , \tag{16}$$

(see appendix A), with the consequence that the retrospective increments in (15) become

$$\delta \mathbf{w}^a_{k-\ell|k}(p) \;\; = \;\; \mathbf{P}^f_{k-\ell|k-\ell-1} \left[\prod_{j=k-\ell+1}^{k} (\mathbf{I} - \mathbf{H}^T_{j-1}\mathbf{\Gamma}^{-1}_{j-1}\mathbf{H}_{j-1}\mathbf{P}^f_{j-1|j-2})\mathbf{A}^T_{j,j-1} \right] \mathbf{H}^T_k \mathbf{x}_k \, . \tag{17}$$

where we used (11) to replace $\mathbf{\Gamma}^{-1}_k \mathbf{v}_k$ with $\mathbf{x}_k$. We see from this expression that the lag-ℓ retrospective increment correspond is calculated as a linear combination of the columns of the forecast error covariances $\mathbf{P}^f_{k-\ell|k-\ell-1}$. The advantage of the expression above is that it refers only to quantities used by the filtering portion of the FLKS: the (filter) forecast error covariance matrix $\mathbf{P}^f_{j-1|j-2}$; the observation error covariance matrix $\mathbf{R}_{j-1}$; the linear (or linearized) observation operator $\mathbf{H}_{j-1}$ and its transpose (adjoint); and the adjoint of the Jacobian $\mathbf{A}_{j,j-1}$ of the dynamics operator. The smoother error cross-covariances $\mathbf{P}^{fa}_{k,k-\ell|k-1}$ and $\mathbf{P}^{aa}_{k,k-\ell|k}$, and smoother error covariance $\mathbf{P}^{a}_{k-\ell|k}$ never appear in (17).

At a given analysis time t_k, the retrospective increments can be calculated through a succession of operations similar to the two-step PSAS operations (11) and (12). Defining an n-vector $\mathbf{z}_k$ as

$$\mathbf{z}_k \equiv \mathbf{H}_k^T \mathbf{x}_k \,, \tag{18}$$

corresponding to the PSAS conjugate gradient solution $\mathbf{x}_k$ converted from the observation space to the analysis space by $\mathbf{H}_k^T$, the term in the square brackets of (17) can be calculated using the following algorithm:

$j = k$

while $j > 1$ and $j \geq \mathbf{max}(1, k - \ell + 1)$

$$\mathbf{z}^p_{j-1} = \mathbf{A}^T_{j,j-1}\mathbf{z}_j \tag{19a}$$

$$\mathbf{\Gamma}_{j-1}\mathbf{x}^r_{j-1} = \mathbf{H}_{j-1}\mathbf{P}^f_{j-1|j-2}\mathbf{z}^p_{j-1} \tag{19b}$$

$$\mathbf{z}_{j-1} = \mathbf{z}^p_{j-1} - \mathbf{H}^T_{j-1}\mathbf{x}^r_{j-1} \tag{19c}$$

$$\delta\mathbf{w}^a_{j-1|k}(p) = \mathbf{P}^f_{j-1|j-2}\mathbf{z}_{j-1} \tag{19d}$$

$j = j - 1$

endwhile

for a maximum number of time lags $\ell = L$. In this algorithm the n-vector $\mathbf{z}^p_{j-1}$ is the result of the adjoint dynamics evolution of the auxiliary n-vectors $\mathbf{z}_j$, for each backward integration j. This backward propagated vector $\mathbf{z}^p_{j-1}$ serves as the input to an equation similar to the first step (11) of the regular PSAS analysis, but with now a different rhs that is aimed at getting the retrospective solution $\mathbf{x}^r_{j-1}$ in (19b). The following step in the retrospective analysis loop is to update the n-vector $\mathbf{z}^p_{j-1}$ with the analysis-space projection of $\mathbf{x}^r_{j-1}$ as in (19c). Finally, the n-vector $\mathbf{z}_{j-1}$ in (19c) is used to calculate the retrospective analysis increment for each desired lag ℓ up to a maximum lag $\ell = L$ through application of the forecast error covariance operator as in (19d).

Notice that the entire retrospective analysis algorithm (18)-(19) works in the analysis space. In particular, the propagation operator $\mathbf{A}^T_{k,k-1} = \mathbf{A}^T_{k,k-1}(p)$ in (19a) is defined in pressure coordinates and it operates on geopotential heights, mixing ratio, zonal and meridional winds, etc, that is, the analysis variables. In fact, the linearized dynamical operator $\mathbf{A}_{k,k-1}(p)$ is given by

$$\mathbf{A}_{k,k-1}(p) \equiv \mathbf{\Pi}_k \mathbf{M}_{k,k-1}(\sigma) \mathbf{\Pi}^+_{k-1} \,. \tag{20}$$

where $\mathbf{M}_{k,k-1}(\sigma)$ is the $m \times m$ Jacobian matrix of the nonlinear operator $\mathcal{M}$ in (9),

$$\mathbf{M}(\sigma) \equiv \left. \frac{\partial \mathcal{M}[\mathbf{y}]}{\partial \mathbf{y}} \right|_{\mathbf{y}=\mathbf{y}(\sigma)} \tag{21}$$

and $\mathbf{\Pi}$ and $\mathbf{\Pi}^+$ are given by

$$\mathbf{\Pi} \equiv \left. \frac{\partial \Pi[\mathbf{y}]}{\partial \mathbf{y}} \right|_{\mathbf{y}=\mathbf{y}(\sigma)} , \tag{22a}$$

$$\mathbf{\Pi}^+ \equiv \left. \frac{\partial \Pi^+[\mathbf{w}]}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}(p)} , \tag{22b}$$

and correspond to the $n \times m$ and $m \times n$ Jacobian matrices of Π and Π^+, respectively. Similarly as for the nonlinear counterpart of these operators, we should require that $\mathbf{\Pi}^+$ be a pseudo-inverse of $\mathbf{\Pi}$. In practice, however, numerical inaccuracies and the character of the original nonlinearities are such that this can only be achieved to a certain extent. As we can see, the adjoint operator (20) needed in (19a) involves more than simply the adjoint of the tangent linear GCM. It involves the adjoint of the linearized transformations Π and Π^+ in (10) and (13), responsible for converting model state vectors and analysis state vectors back and forth between the two spaces.

A few remarks can be made at this point.

- Currently in PSAS the analysis error covariance matrix $\mathbf{P}^a_{k|k}$ is never referenced. Indeed, the current implementation of PSAS parameterizes the forecast error covariance matrix in a such a simple manner that none of the terms in the rhs of (1d) are taken into account. This is an approximation based partially on practical considerations that in the long line of further PSAS developments is bound to change. However, when the expressions (7) for the smoother error cross-covariances are bypassed and retrospective increments are calculated using the gains in (16) there are actually no approximations involved. The only consequence of not calculating the smoother error covariances is that we get no estimates for the accuracy of the retrospective analyses — which, in principle, can be extracted from $\mathbf{P}^a_{k-\ell|k}$. Expression (16) is exact for the linear FLKS and its nonlinear EKF-based extension.

- We see from (17) that an FLKS-based retrospective scheme allows future observations to be used to correct previous filter and retrospective analyses impaired by the lack of observations over a certain region earlier on in the assimilation. That is, when at, say, time t_{k-1} there are no observations over a certain region, the filter analysis at this time will essentially equal the first-guess over that region - aside from possible contributions by farther away regions through the forecast error correlations. If at, say, time t_k, observations become available over the region in question, or information from observations at downstream nearby regions get propagated through the adjoint of the tangent linear dynamics $\mathbf{A}^T_{k,k-1}$ into the region in question, this new information can be used to calculate a correction to the filter analysis at t_{k-1} as the lag-1 retrospective analysis represented in (17). In these cases, it is the the first term in the square-bracket of (17) that mostly contributes to the correction to the filter analysis. This argument can be taken beyond this simple lag-1 reasoning.

- Notice that the linear system (19b) solved within the retrospective analysis algorithm involves exactly the same operators required to calculate the sensitivity of forecasts to observations changes, as measured by some pre-specified cost function, as in the approach of Baker and Daley [2000; compare with their eq. (2.7a)]. Furthermore, (19c) involves exactly the operator required to examine forecast sensitivity with respect to changes in the background. It has been pointed out elsewhere that some of the operations in 4D-var, are closely related to operations required to study forecast sensitivity; the same is true of the operations in FLKS-based retrospective analysis schemes.

- When the forecast errors do not dynamically propagate the smoother follows an algorithm similar to (19) but with the adjoint operator in (19a) replaced by the identity. Since in the current implementation of PSAS the forecast error covariance is not dynamically determined, and even with its slowly varying forecast error variances it can be thought of having a time-independent forecast error covariance, it is conceivable that replacing the adjoint by the identity operator in (19a) may result in a reasonable retrospective analysis approximation consistent with the current underlying PSAS statistics. Todling (2000) has experimented with this idea using an identical-twin configuration setup for GEOS and has found a significant improvement in the mean error due to lag $\ell = 1$ and even to lag $\ell = 2$ retrospective analyses.

(c) The GEOS lag-1 retrospective-based iterated analysis

When the system is nonlinear, the idea to feedback the filter estimate into the analysis equation is particularly attractive, especially if we expect the filter analysis to be a better estimate of the state of the system than the first-guess provided by the model. Indeed, filtering strategies making use of such feedback procedures are commonly found in the literature. For instance, Jazwinski (1970, Theorem 8.2) introduces the so-called iterated EKF, which is suitable for nonlinear observation operators. Cohn (1997) proposes a similar procedure as an extension to PSAS so it can handle such types of operators. Iterated procedures aimed at dealing with nonlinearities of the observation operator are sometimes referred to as locally-iterated methods, since the iterations are performed at a single time. Jazwinski (1970, Theorem 8.3) also presents an iterative procedure that is aimed at correcting errors due to the dynamical linearizations required by the EKF. This latest procedure involves integrating the model with a newly estimated trajectory at each iteration and for this reason it resembles a smoother procedure referred to as the iterated linear filter-smoother algorithm. Combining ideas of filtering and smoothing leads to the possibility of developing globally-iterated procedures in which the filter analyses may be revised by a backward-filter integration within a certain time interval. Most these iterative procedures are inspired by Newton-type methods for solving systems of nonlinear equations (see Navon and Legler 1987, for a review of Newton-type methods).

Motivated by these methods we introduce here a procedure to use the retrospective analysis and try to improve the overall GEOS IAU-based assimilation. At first, the algorithm is based only on the lag-1 retrospective analyses. At any single time t_k, when a lag-1 retrospective analysis $\mathbf{w}^a_{k|k+1}(p)$ is available we can construct a model-space lag-1 IAU retrospective increment as

$$\delta\mathbf{y}^a_{k|k+1}(\sigma) \quad = \quad \Pi^+[\mathbf{w}^a_{k|k+1}(p)] - \mathbf{y}^b_{k|k-1}(\sigma) \ , \tag{23}$$

which is similar to (14), but is constructed using observations one step ahead of time t_k. This lag-1 retrospective increment can now be used to integrate the GCM over an IAU integration period already covered before. This is illustrated schematically in Fig. 2. The diagram resembles much the regular IAU procedure presented before in Fig. 1. In fact, the top part of the diagram, above the horizontal dotted line, is identical to the regular IAU procedure. However, now at, say, time $t = 12$Z we calculate a retrospective analysis by integrating the transformed PSAS solution vector in (18) back in time using the adjoint operation (19a); this is represented in the diagram by the south-westward pointing dashed arrow. A new PSAS-like linear system problem can be solved as in (19b) with the corresponding update (19c); which is represented in the diagram by the box tagged "Retro Ana". In the end, a lag-1 retrospective increment at $t = 6$Z is construct as in (23), and the GCM is integrated for 6 hours using this increment as the tendency term in (9). From this point on, the procedure follows the regular IAU schematics until it is time to process the observations at $t = 18$Z when the lag-1 retrospective analysis at $t = 12$Z can be calculated and the whole cycle repeated. The final assimilated trajectory is represented, as in Fig. 1, by the thick solid arrows. At a given analysis time, the relevant iterated PSAS analysis is represented in the diagram as the analysis from the lowest PSAS box in a column of the diagram (see thick vertical dashed lines).

Looking at the diagram in Fig. 2 we see that the retrospective-based iterated analysis amounts to a considerable increase in computational requirements when compared with the regular procedure in Fig. 1. Each iteration of the iterated analysis scheme requires one extra 9-hour GCM integration and two extra PSAS analyses. Such increased in the computational cost can only be justified if the procedure results in considerably improved analyses. If indeed this procedure proves to be worth, its computational burden can be reduced by calculating some of the steps in (19) at different resolutions. Similarly, to the strategy of incremental 4D-var of Courtier et al. (1994), we can for example integrate the adjoint of the tangent linear GCM in (19a) at lower resolution than the actual model integration (9). Also, the retrospective PSAS-like linear system (19b) can be solved at lower resolution than the regular linear system (11) solved in the first step of PSAS. For that matter, the calculations in (19a) and (19b) do not even have to be performed at the same resolution. Clearly, this type of approach to reduce computational cost involves the development of additional interpolation operators and their corresponding adjoints. Results obtained with full resolution tests will likely not translate equally to modifications of the algorithm involving resolution changes. Still, only experimental tests will demonstrate the feasibly and accuracy of such approach.

Independently of the IAU, in a linear optimal sense it can be shown that to feedback the lag-1 retrospective analysis at, say, t_{k-1} to calculate a revised filter analysis at time t_k does not amount to an improved filter analysis. In this iterated procedure, an optimal analysis can be calculated using the first-guess from the lag-1 retrospective analysis when the cross-covariance between the revised first-guess and the observations are properly taken into account. In fact, since the retrospective-based iterated analysis procedure here amounts to a modified filtering procedure the optimal gains in this case are similar to the usual modified filter gains when the forecast and observations are correlated (e.g., Jazwinski 1970, Example 7.5). Since in practice it would be quite difficult to account for this cross-covariance, we choose to neglect the cross-covariance terms all together.

4 GEOS experiments results

(a) Configuration and experimental setup

The retrospective analysis procedures of the previous section were implemented as an extension to GEOSDAS. The retrospective portion of the GEOS software is compatible with the first operational version of GOES, designed to support NASA's Earth Observing System mission and its Terra satellite. We refer to this earlier operational version as GEOS-3$'$ to avoid possible confusion with the considerably upgraded version of GEOS-3 operational at the time of this writing. The GEOS-3$'$ GCM operates at a resolution of 1^o latitude by 1^o longitude and it has 48 vertical sigma levels, with a dynamical core much similar to that of Suarez and Takacs (1995). At the synoptic hours, PSAS calculates analysis at a resolution of 2^o latitude by 2.5^o longitude on 20 pressure levels. Details on the implementation of PSAS can be found in da Silva et al. (1996), Guo et al. (1998), and Larson et al. (1998). And, as we have mentioned in the previous section, GEOS-3$'$ uses the IAU procedure of Bloom et al. (1996) to generate a time-continuous state trajectory referred to as the assimilation. To simplify matters, the experiments performed for the present article used both the GCM and PSAS at the coarse horizontal resolution of 4^o latitude by 5^o longitude; the GCM and PSAS vertical resolutions were kept unchanged. Except for sea-wind satellite observations, all observation data types used in GEOS-3′ are included in our experiments. Conventional observations from ships, environment and drifting buoys, surface stations, winds from pilot balloons, aircraft reports, and radiosonde stations are used. Cloud track wind retrievals and TOVS geopotential height retrievals are used as well. Furthermore, the Wentz (1997) SSM/I-derived total precipitable water retrievals are assimilated, though not through PSAS but rather by using the method of Hou et al. (2000).

Four new components are required to implement the retrospective capability in GEOSDAS: the adjoint of the tangent linear GCM; the additional PSAS-like operators involved in (19b); the adjoint of the tangent linear operator (10) taking model-space variables into analysis-space variables; and the adjoint of the tangent linear operator (13) taking analysis-space variables into model-space variables. Presently, the adjoint of the GCM includes the hydrodynamics adjoint and the adjoint of a simple diffusion scheme. No attempts to produce the adjoint of parts of the model physical processes has been made. Most modifications required to PSAS were quite cosmetics since they only required rearranging operators already available in the original PSAS software. Some effort was devoted to derive the proper tangent linear and adjoint operators for the transformations (10) and (13), particularly, because we striven to make sure that the back and forth operations would render minimal error. Some of this work was done by hand, some was done using the automatic differentiation tool of Giering and Kaminski (1998).

In the present article, only results for the lag-1 (6-hour) retrospective analysis are discussed. The behavior and impact of the simplest possible implementation of an FLKS-based assimilation scheme needs to be fully assessed and understood before engaging into experimentations with higher lags. Furthermore, we focus on the mean overall impact of the retrospective analysis. Therefore, leaving synoptic impact studies to future presentations. Since only the mean overall impact is of interest, we simplify the experimental configuration by updating the GCM trajectory needed during the adjoint integrations with the frequency of the analyses, that is, we update the adjoint basic state only every 6 hours.

We compare the results of three experiments conducted over the period of January 1998. To avoid possible differences due to spin-up issues, the experiments are actually started on 14 December 1997, but the results are ignored during this half-month period. Our first experiment is taken as the control and it uses the reduced resolution GEOS-3$'$ data assimilation system mentioned above. The control is referred to as the CTL experiment. The second experiment, referred to as the RA experiment, involves calculating lag-1 (6-hour) retrospective analyses for the entire month of January 1998 using the background fields and OMF time series of the control experiment. The third, and final, experiment is aimed at evaluating the lag-1 (6-hour) retrospective-based iterated analysis procedure introduced in the previous section, and it is referred to as the RIA experiment.

Mainly, we evaluate the RA and RIA experiments by examining the statistics of their corresponding residuals time series. That is, depending on the case, we calculate time root-mean-square (RMS) bias and standard deviation from the differences of the observations with either the forecast, or the analysis, or the retrospective analysis, or the retrospective forecast (see below). To ease comparisons, we grid the residuals over a 4^o latitude by 5^o longitude on the 20 pressure levels of the analysis space before calculating any statistics. Since during the course of one month observations, particularly from certain instruments, can still be sporadic within a grid-box we make sure that only grid-boxes containing 15 or more reports enter the calculation of the corresponding statistic. In the discussion that follows, we concentrate on results obtained in the troposphere.

(b) Evaluation of the 6-hour retrospective analysis

We start by comparing the results of the CTL and RA experiments using the set of observations assimilated in the CTL experiment. If everything works properly, and the 6-hour retrospective analyses are indeed an improvement over the regular control analyses we should see that in some mean sense the RA observation-minus-analysis (OMA) residuals are reduced in comparison to the OMA residuals of the control experiment. As a matter of fact, we can show that in the linear optimal case,

$$\mathcal{E}\{(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k+1}^a)(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k+1}^a)^T\} < \mathcal{E}\{(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k}^a)(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k}^a)^T\}\,. \tag{24}$$

Although there is no guarantee of this holding in general for the, suboptimal, nonlinear case under study, we would like to see the extent to which this is so. In practice, to assess this quantity we must take the usual ergodic assumption and replace the ensemble average by a time average. Examination of the time RMS biases and standard deviations for the OMA residuals when the analysis are either the regular filter analyses of the CTL experiment or the lag-1 retrospective analyses of the RA experiment has shown that they are virtually the same (not shown). Therefore, this point of view we might be led to think that there is no payoff in calculating lag-1 retrospective analyses.

Because of nonlinearities and suboptimalities in a practical system such as GEOS it is possible that we may see benefits from the RA scheme by looking at quantities other than the OMA residuals. If there is any value in the retrospective analyses, and indeed they correspond to a better analyses than the regular (filter) analyses, we should expect the former to have improved predictive skills over the latter. We cannot expect this to hold true for any forecast length, but it would be nice to see if it at least holds for predictions up to the time-lag period of the retrospective analyses; 6 hours in the lag-1 case here. Hence, we can compare the OMF residuals statistics when the forecasts are the regular filter forecasts of the CTL experiment and the so-called retrospective forecasts issued from the lag-1 retrospective analyses. Since the OMF residuals from a regular GEOSDAS run, such as the CTL experiment, involve 6-hour forecasts that are produced from partly integrating the GCM with the IAU forcing for 3 hours and partly integrating the GCM for another 3 hours without the influence from the IAU tendencies (see Fig. 1), we must carefully use the retrospective analyses when constructing OMF residuals from them.

To make a fair comparison, we calculate OMF residuals from the 6-hour retrospective analyses following a forecasting procedure based on IAU. For each available retrospective analysis for the entire month of January 1998 a retrospective forecast is issued following the the schematic representation shown in Fig. 3. As illustrated in the figure, the retrospective OMF residuals at, say, 12Z are calculated by converting the 6Z retrospective analysis to the model space and constructing the corresponding increment on the model space, following (23). This retrospective analysis increment is used as a tendency term during a 6-hour GCM integration, started at 3Z. At the end of the 6-hour integration the retrospective tendency term is turned off, by setting $\alpha = 0$ in (9), and the model is left to run free for another 3 hours when the OMF residuals at 12Z can be calculated using the observations at the same time.

Using these retrospective forecasts, Fig. 4 shows the time RMS bias (top panels) and standard deviation (bottom panels) for the geopotential height OMF radiosondes residuals for the CTL (solid curves) and RA (dashed curves) experiments averaged over the western (left) and eastern (right) quadrants of the Northern Hemisphere, for latitudes higher than 20N. These two domains are chosen because they represent the largest concentration of radiosondes in the globe. We see from the top panels that, in the RMS bias sense, the forecasts from the lag-1 retrospective analyses correspond to a considerable improvement over the regular GEOSDAS analyses. However, the bottom-left panel for the RMS standard deviations shows that the retrospective forecasts are considerably noisier than the regular forecasts over what is mostly North America; results are roughly neutral over most of Europe and Asia, as seen from the bottom-right panel. Figure 5 displays similar quantities but now for the zonal wind radiosondes OMF residuals. Except in the Northwestern region, between pressure levels 700 mb to 400 mb, the RMS bias for the zonal component of the zonal wind OMF radiosonde residuals are improved when forecasts are issued from the lag-1 retrospective analyses. In this same region, the zonal wind OMF standard deviations [panel (b.1)] shows a minor deterioration at levels below 400 mb, confirming the deterioration seen in the OMF heights in Fig. 4b.1; minor improvement in the standard deviations are seem above 400 mb. Over the Northeastern region a minor but consistent improvement is observed in both the RMS bias and standard deviation, as indicated from the panels on the right.

The statistics of OMF residuals for other variables and other observing systems can also be examined. Figure 6 shows the time RMS biases (top panels) and standard deviations (bottom panels) for the geopotential height TOVS retrievals OMF residuals. Since TOVS provides global coverage in the course of a single day, the spatial averages now cover the entire Northern Hemisphere (left panels) and the entire Southern Hemisphere (right panels). We see considerable improvement in the OMF biases and standard deviations from the retrospective forecast residuals. Interestingly, the standard deviation results over the Northern Hemisphere [panel (b.1)] contradict the deterioration observed in the geopotential heights of radiosondes OMF residuals [panel (b.1) of Fig. 4]. We attribute this contradiction over North American to contradictions between the geopotential height observations from the radiosondes and the TOVS retrievals themselves and not to the retrospective analysis procedure.

Another quantity we have studied is simply the spatial average of the residuals time mean. Though we expect considerable cancellation of errors in this quantity, it still serves as an indicator of the overall behavior of the residuals and of the underlying procedure used to produce them. Figure 7 shows the time mean OMF residuals for the CTL 6-hour first-guesses (forecasts) and the lag-1 retrospective forecasts. The globally-averaged time means for geopotential height TOVS retrievals and radiosondes OMF residuals are displayed in panels (a) and (b), respectively. We see mostly a reduction in the time mean residuals when the retrospective forecasts are used instead of the regular forecasts, with some overshooting at levels below 700 mb for the TOVS retrievals residuals. The zonal and meridional components of the radiosondes wind OMF residuals are displayed in panels (c) and (d), respectively, and again, we see an overall reduction when the retrospective forecasts are used, with some overshooting of the mean meridional wind around

150 mb.

It seems fair to say that, overall, in the nonlinear suboptimal case of the GEOS application, we see some benefit in producing the 6-hour forecasts issued from the lag-1 retrospective analyses over the regular GEOSDAS forecasts. This serves to indicate that there is some benefit from the RA scheme. This also serves as further motivation to use the iterated retrospective analysis procedure proposed in the previous section since it makes direct use of these retrospective forecasts (see Fig. 2).

(c) Evaluation of the 6-hour retrospective-based iterated analysis

We now evaluate the performance of the 6-hour (lag-1) retrospective-based iterated analysis scheme of Fig. 2. We start by comparing the OMA residuals between the CTL and the RIA experiments. Figure 8 shows the globally-averaged time RMS bias of geopotential heights from TOVS retrievals and radiosondes OMA residuals [panels (a) and (b), respectively], and the same quantity for the zonal and meridional components of the radiosondes winds [panels (c) and (d), respectively]. Although small, we actually see improvement in the OMA residuals due to the iterated analysis. To the extent that the ensemble mean can be replaced by the time RMS mean the inequality (24) holds when $\mathbf{w}^a_{k|k+1}$ corresponds to the iterated analysis, at least in a globally-averaged sense. Another proxy to (24) would be the time RMS standard deviations. We have examined this quantity for the OMA of both the CTL and RIA experiments and have found it to change very insignificantly, sometimes in favor of the RIA scheme, and other times with the CTL experiment results being slightly less noisy (not shown). The results in the time RMS standard deviations can however be made neutral or favor the RIA procedure by properly re-tuning the PSAS error statistics, but this is outside the scope of the present article.

Even though small, the improvement due to the RIA scheme seen in Fig. 8 is also visible directly from the time series of the globally-averaged OMA residuals RMS bias. Furthermore, this improvement is seen not only for TOVS retrievals and radiosondes OMA residuals, but for other instruments as well. An illustration is presented in Figure 9 by displaying the globally-averaged RMS bias of the zonal (top) and meridional (bottom) cloud-track winds OMA residuals at 200 mb. The thin curves correspond to the OMA residual times series from the CTL experiment and the thick curves are for the RIA experiment. The global reduction in the RMS bias can be as much as 1 m s^{-1} at certain times. This confirms the reduction in the globally-averaged time RMS bias of the radiosondes OMA residuals observed in panels (c) and (d) of Fig. 8 around the same pressure level.

Frequently, changes made to assimilation systems are evaluated and validated by making comparisons with independent observations, that is, observations which are not assimilated by the system. There are certainly plenty of reasons to take such comparisons and validations with caution as independent observations may contain errors of their own. Moreover, it is often difficult to find independent data sources that are not assimilated, particularly those providing global coverage. Data withholding experiments are commonly used to assess the impact of a particular observing system and can also be used to evaluate the impact of system changes (e.g., Bouttier and Kelly 2001, and references therein). Here we choose to verify the change in the 200 mb winds of Figs. 8 and 9 by using wind observations from the Global Aircraft Data Set (GADS) of the British Airways Boeing 747-400 flights, and by using further wind aircraft observations from ACARS (Aircraft Communications, Addressing, and Reporting System). Neither of these observation types are used in our assimilation experiments and therefore provide for legitimately independent checks.

The GADS wind observations have been shown by Rukhovets et al. (1998) to be of potential value to GEOSDAS if used regularly in the PSAS analyses. This suggests that any changes made to GEOS that show its analyses to draw more closely to these observations, even when they are not assimilated, should be considered an improvement. With that in mind, we used a dataset for the January 1998 GADS observations to construct OMA residuals for the analyses of both the CTL and RIA experiments. Figure 10 shows maps of the time RMS standard deviation of the zonal (top) and meridional (bottom) GADS winds OMA gridded residuals for the RIA experiment subtracted by the similar quantity for the CTL experiment. Essentially, with this differencing the color scheme in the figure indicates that blue (negative values) corresponds to improvements due to the RIA procedure. Though we see areas where the impact of RIA is neutral or negative, in most places it seems that the analyses of the RIA experiment are closer to the GADS observations than the analyses of the CTL experiment.

Similarly, Fig. 11 shows the differences shown in Fig. 10, but now for the ACARS wind OMA residuals. The maps are focused over North America since that is where the majority of the observations are concentrated in this case. Relatively neutral results are seen in the meridional component of the wind (bottom map), but undeniable improvement due to the RIA scheme is seen in the zonal component of the wind (top map).

Ultimately, as emphasized by CST94, one of the main motivations for performing retrospective analysis is to produce the best possible dataset for climate research. As such, it is important to examine the climatological impact of changes induced by the RIA procedure. Since the results of the experiments discussed here are still preliminary we do not want to duel too much on the significance of performing RA and RIA for the purposes of improving the climatological aspects of the assimilation strategy — recall that our experiments are for a very low resolution version of GEOS-3′. Still, we cannot avoid looking more closely to see what is the climatological impact of changes such as those observed in the wind field. In fact, the significance of the RIA scheme impact to the upper level winds can be seen more clearly by looking directly at the monthly-averaged winds. For instance, Fig. 12 shows the zonally-averaged, January 1998 monthly mean, meridional wind (top) for the RIA experiment and its difference from the CTL experiment (bottom). The bottom panel, shows a distinct tropical wind strengthening at the upper levels and a slight weakening at the mid- to lower levels when the RIA scheme is used.

This change in the tropical meridional wind is bound to affect the tropical circulation, namely, the Hadley circulation. To see the meridional circulation, we calculate the mass stream function ψ by integrating the zonally-averaged monthly mean meridional wind using the expression

$$\psi = \frac{2\pi R \cos\phi}{g} \int [\overline{v}]\, dp\,, \tag{25}$$

where v stands for the meridional wind, the operator $\overline{\bullet}$ represents the time mean, the operator $[\bullet]$ represents for the zonal average, R is the mean radius of the earth, g is the gravity constant, ϕ is the latitude, and the integral is taken downward from the pressure of 10 mb to the surface. Figure 13 shows the January 1998 mass stream function for both the CTL (top) and RIA (bottom) experiments. We see a clear enhancement of the Hadley circulation when the RIA procedure is used with the mass stream function peaking at about 16×10^{10} kg s^{-1} in contrast to the weaker peak of 12×10^{10} kg s^{-1} for the circulation of the CTL experiment. Although we do not expect the circulation pattern to be completely well represented at the coarse resolution we use in our experiments here, we must say that the circulation obtained with the RIA procedure looks very close to the circulation pattern of the full resolution, 1^o latitude by 1^o longitude, GEOSDAS (not shown), with its tropical circulation peaking at 18×10^{10} kg s^{-1}.

As we have stated above, the main motivation for, say, the RIA procedure is to obtain overall improved assimilated fields. Results such as the one in Fig. 13 seem to suggest that the RIA scheme has the potential for improving climatologically relevant features. Still, it would be hard to convince any operational group to apply an FLKS-based scheme without showing that it improves medium-range forecast skills or at least that it leaves the skills essentially unchanged. Therefore, we compared the skills of 5-day forecasts when the forecasts are issued from the CTL and the RIA analyzed fields. Notice that these are now regular forecasts, instead of the IAU-based forecasts obtained by following the schematic in Fig. 3. Since our experiments are confined to the single month of January 1998, there are so many independent samples we can obtained over this period. We issued 5-day forecasts starting from 2 January 1998 every 3 days until 26 January 1998, to have a small sample of 9 5-day forecasts. We verified that the overall conclusions and skills calculated from this small ensemble were not affected by the size of the sample by reducing the size of the sample further down to 5 members and performing cross-validation. As a measure of forecast skill we calculated anomaly correlations and RMS errors [e.g., see von Storch and Zwiers 1999, Eqs. (18.17) and (18.18)]. Both the CTL and RIA forecasts were evaluated against their own analyses. Furthermore, anomalies were calculated using a 10 year climatology obtained from the ECMWF operational analyses for the period of 1988 to 1997, and interpolated to the resolution of our experiments. We should say that the scores shown below are not representative of the actual scores of the operational GEOS-3 data assimilation system.

Figure 14 shows the anomaly correlations for the 500 mb geopotential height field calculated over four different regions for the 5-day forecasts issued from the CTL (solid curves) and RIA (dashed curves) analyses. We see that over the Northern Hemisphere extra tropics (top-left panel) forecasts from RIA are of similar skills as forecasts from the control analyses, at least up to day 4. In North America (bottom-left panel) the forecast skills from RIA analyses show some deterioration when compared against the skills of the regular forecasts. As when studying the OMF obtained from the retrospective forecast using the RA analyses, this deterioration over North America might be related to contradictions in the observing system over this area. In fact, this seems to be an issue confined to this region since, for example, over the Southern Hemisphere (top-right panel) and Europe (bottom-right panel) we actually see some improvement in skills when the 5-day forecasts are issued from the RIA analyses.

As a final illustration along the lines of comparing the skills of 5-day forecast from the CTL and RIA experiments we examine the RMS error of the wind fields at 850 mb and 200 mb. Figure 15 displays these quantities averaged over the tropics for both the zonal (right panel) and meridional (left panels) components of the wind. The RMS errors at 850 mb are virtually identical, while at 200 mb we see a slight improvement when using forecasts from the analyses of the RIA experiment. Although these are small improvements they serve as further confirmation of what we have seen previously when comparing the analyses of the CTL and the RIA experiments with independent observations.

5 Conclusions

The purpose of atmospheric data assimilation is to produce the best possible estimate of the the state of the atmosphere at any single time. In theory this can be accomplished by using smoothing techniques since they are aimed at maximizing data usage through inclusion of observations in the past, present, and future of the time an estimate is sought for. In the context of sequential data assimilation, the fixed-lag Kalman smoother (FLKS) provides a particularly attractive framework. The FLKS formulation is fully based on the underlying filtering strategy. Its standard formulation requires no error covariance information beyond what is require by the filtering approach. Indeed, the FLKS can be separated into a filter portion and a retrospective analysis (RA) portion and this separation renders practical implementation of FLKS-based procedures a relatively simple extension of an already existing (filter) analysis scheme.

Two different types of retrospective procedures are investigated in the present work. The first is the original FLKS-based formulation referred to simply as RA. The second is an iterated version of the original algorithm, referred to as RIA, in which lag-1 retrospective analyses are used to revise a previously calculated filter analysis. Both these procedure are implemented as an extension of the Goddard Earth Observing System (GEOS) data assimilation system. The new components required for implementing a retrospective capability in GEOS are the adjoint of the tangent linear of the GEOS general circulation model (GCM); the rearrangement of a few operators already available in the physical-space statistical analysis system (PSAS) of GEOS;

and the development of the tangent linear and adjoint operators responsible for transforming model-space variables into analysis-space variables back and forth as well as the adjoint of the operator transforming analysis-space variables into observables. The adjoint of the tangent linear GCM used in the present work includes the adjoint of the tangent linear hydrodynamics and the adjoint of a simple diffusion term; the adjoint of the physics is not included.

Only results for the 6-hour, lag-1, retrospective analysis are studied here. Although close examination of the observation-minus-analysis (OMA) residuals seem to suggest a rather neutral overall benefit from the lag-1 retrospective analysis, we see improved skills in the 6-hour forecasts issued from these lag-1 retrospective analyses. That is, the so-called retrospective forecasts are a closer match to the observations than the regular GEOSDAS forecasts. This is not to be taken as a surprising result but rather as verification that the RA scheme works. This improved 6-hour skill motivates the investigation of the RIA scheme since this scheme makes explicit use of the retrospective forecasts. Evaluation of the analyses from the RIA procedure indicates them to be closer to the observations than the usual PSAS analyses. The OMA residuals for independent observations not used during the assimilation further confirms some of the improvements due to the RIA scheme. More significant improvements are seen when examining climatologically relevant fields such as the mass stream function describing the meridional wind circulation. Lastly, anomaly correlations and root-mean-square errors from a small sample of 5-day forecasts indicate a mild improvement in skill scores when analyses from the RIA procedure are used for the 5-day forecasts instead of the regular GEOS analyses. Although the skill scores are not improved everywhere in the globe, there are improvements nonetheless.

Much work remains to be done to show that either RA or RIA are a fit extension to the usual PSAS analysis procedure of GEOS. The present work used only a reduced resolution version of GEOS and a study with a higher resolution version is necessary. Just as in four-dimensional variational procedures, one of the main feature of the retrospective analysis is its capability to incorporate backward-propagated information into the retrospective analyses via the model adjoint. This should, in particular, result in improved representation of synoptically relevant atmospheric events. This has not been explored in the present work and should serve as material for future studies. Furthermore, more and more it becomes evident that instead of windowing the observations in 6-hour batches, as it is commonly done in the 3D-var system, much can be gained by assimilating observations at their proper time, as it is done in 4D-var. For a PSAS-like system the rapid update cycle strategy is more readily implementable for this purpose and it will be important to investigate the impact of retrospective analysis in this context. Whether these studies will be done in the context of GEOS or in the context of the newly developed finite-volume data assimilation system in our group is still to be decided. In any case, we feel that the present work serves to justify the development of retrospective analysis capabilities in this new assimilation system.

Acknowledgments. We thank N. S. Sivakumaran for his contribution during the initial phase of this work. We are also thankful to A. M. da Silva for providing the program for gridding residuals; to J. Tenenbaum and L. Rukhovets for providing the January 1998 GADS dataset used in one of our comparisons; and to J. Ardizzone for providing the package to calculate forecast error skill scores. The numerical results in this work were obtained on an SGI 2000 through cooperation with the NASA Center for Computational Sciences at the Goddard Space Flight Center and the NASA Ames Research Center. This research was partially supported by the NASA EOS Interdisciplinary Project on Data Assimilation.

Appendix A

Retrospective gains as a function of filter variables only

The purpose of this appendix is to derive the alternative expression (16) for the retrospective gain matrix (see also Zhu et al. 1999). Using (6) and (7) with $\ell = 1, 2, ..., j$ we have

$$\begin{aligned}
\mathbf{K}_{k-1|k} &= \mathbf{P}^a_{k-1|k-1}\mathbf{A}^T_{k,k-1}\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k \\
&= \mathbf{P}^f_{k-1|k-2}\left[(\mathbf{I}-\mathbf{K}_{k-1|k-1}\mathbf{H}_{k-1})^T\mathbf{A}^T_{k,k-1}\right]\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k\,, \qquad \text{(A.1a)} \\
\mathbf{K}_{k-2|k} &= \left(\mathbf{P}^{aa}_{k-1,k-2|k-1}\right)^T\mathbf{A}^T_{k,k-1}\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k \\
&= \mathbf{P}^a_{k-2|k-2}\mathbf{A}^T_{k-1,k-2}\left[(\mathbf{I}-\mathbf{K}_{k-1|k-1}\mathbf{H}_{k-1})^T\mathbf{A}^T_{k,k-1}\right]\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k \\
&= \mathbf{P}^f_{k-2|k-3}\left[(\mathbf{I}-\mathbf{K}_{k-2|k-2}\mathbf{H}_{k-2})^T\mathbf{A}^T_{k-1,k-2}\right]\left[(\mathbf{I}-\mathbf{K}_{k-1|k-1}\mathbf{H}_{k-1})^T\mathbf{A}^T_{k,k-1}\right] \\
&\quad \times\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k\,, \qquad \text{(A.1b)} \\
\mathbf{K}_{k-j|k} &= \left(\mathbf{P}^{aa}_{k-1,k-j|k-1}\right)^T\mathbf{A}^T_{k,k-1}\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k \\
&= \mathbf{P}^f_{k-j|k-j-1}\left[(\mathbf{I}-\mathbf{K}_{k-j|k-j}\mathbf{H}_{k-j})^T\mathbf{A}^T_{k-j+1,k-j}\right]\cdots\left[(\mathbf{I}-\mathbf{K}_{k-1|k-1}\mathbf{H}_{k-1})^T\mathbf{A}^T_{k,k-1}\right] \\
&\quad \times\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k\,. \qquad \text{(A.1c)}
\end{aligned}$$

This can be written generally as in (16) or, making explicit use of (1c) for the filter gain matrix, we can also write

$$\mathbf{K}_{k-\ell|k} = \mathbf{P}^f_{k-\ell|k-\ell-1}\left[\prod_{j=k-\ell+1}^{k}(\mathbf{I}-\mathbf{H}^T_{j-1}\mathbf{\Gamma}^{-1}_{j-1}\mathbf{H}_{j-1}\mathbf{P}^f_{j-1|j-2})\mathbf{A}^T_{j,j-1}\right]\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k\,, \qquad \text{(A.2)}$$

which shows that, as pointed out in the main text, the retrospective gains only depend on filter quantities. Todling et al. (1998) have pointed out that the retrospective portion of the FLKS implicitly accounts for model error. The equation above serves to re-emphasize that remark as it shows that the retrospective gains depend directly on the forecast error covariance matrix, which is the filter quantity containing model error covariance information.

References

Anderson, B. D. O., and J. B. Moore, 1979: *Optimal Filtering*, Prentice-Hall, 357 pp.

Baker, N. L., and R. Daley, 2000: Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Q. J. R. Meteorol. Soc.*, **126**, 1431-1454.

Bennett, A. F., B. S. Chua, and L. M. Leslie, 1996: Generalized inversion of a global numerical weather prediction model. *Meteor. Atmos. Phys.*, **60**, 165-178.

Biswas, K. K., and A. K. Mahalanabis, 1973: Suboptimal algorithms for nonlinear smoothing. *IEEE Trans. Aerosp. Electron. Syst.*, **9**, 529-534.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256-1271.

Bouttier F., and G. Kelly, 2001: Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Quart. J. Roy. Meteor. Soc.* **127**, 1469-1488.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteorol. Soc. Japan*, **75**, No. 1B, 257-288.

Cohn, S. E., and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteorol. Soc. Japan*, **74**, 63-75.

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.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.*, **126**, 2913-2926.

Courtier, P., J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy operational implementation of 4-D VAR using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367-1387.

Daley, R., and E. Barker, 2001: NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.*, **129**, 869-883.

da Silva, A. M., and J. Guo, 1996: Documentation of the physical-space statistical analysis system (PSAS). Part I: The conjugate gradient solver version PSAS – 1.00. DAO Office Note 96-2, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

DAO, 1996: Algorithm Theoretical Basis Document Version 1.01. Data Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

Derber, J. C., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437-2446.

Evensen, G., and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852-1867.

Fisher, M., and E. Anderson, 2001: Developments in 4D-var and Kalman filtering. ECMWF Tech. Memo. 347, 36 pp.

Giering, R., and T. Kaminski (1998): Recipes for adjoint code construction. *ACM Trans. Math. Software*, **24**, 437-474.

Guo, J., J. W. Larson, G. Gaspari, A. da Silva, and P. M. Lyster, 1998: Documentation of the physical-Space statistical analysis system (PSAS). Part II: The factored-operator formulation of error covariances DAO Office Note 96-04, NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA.

Hou, A. Y., D. V. Ledvina, A. M. da Silva, S. Q. Zhang, J. Joiner, R. M. Atlas, G. J. Huffman, and C. D. Kummerow, 2000: Assimilation of SSM/I-derived surface rainfall and total precipitable water for improving the GEOS analysis for climate studies. *Mon. Wea. Rev.,* **128**, 509-537.

Jazwinski, A.H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 376 pp.

Larson, J. W., J. Guo, G. Gaspari, A. da Silva, and P. M. Lyster, 1998: Documentation of the physical-space statistical analysis system (PSAS). Part III: The software implementation. DAO Office Note 98-05, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

Ménard, R., and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*, **48A**, 221-237.

Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037-1056.

Moore, J. B., 1973: Discrete-time fixed-lag smoothing algorithms. *Automatica*, **9**, 163-173.

Navon, I. M., and D. M. Legler, 1987: Conjugate gradient methods for large scale minimization in meteorology. *Mon. Wea. Rev.*, **115**,1479-1502.

Parrish, D. F. and J. C. Derber, 1992: The national meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1764.

Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four dimensional variational assimilation. Part I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143-1170.

Rukhovets L., J. Tenenbaum, and M. Geller, 1998: The impact of additional aircraft data on the Goddard Earth Observing System analyses. *Mon. Wea. Rev.*, **126**, 2927-2941.

Suarez, M., and L. L. Takacs, 1995: Documentation of the ARIES/GEOS Dynamical Core, Version 2. NASA Tech. Memo. 104606, Vol 5, 45 pp. [Available from Data Assimilation Office, NASA, Code 910.3, Greenbelt, MD 20771.]

Todling, R. 2000: Retrospective data assimilation schemes: fixed-lag smoothing. *Proc. Second Intl. Symp. Frontiers of Time Series Modeling: Nonparametric approach to knowledge discovery*, Nara, Japan, December, 155-173.

Todling, R. and S. E. Cohn, 1996: Some strategies for Kalman filtering and smoothing. *Proc. ECMWF Seminar on Data Assimilation*, 91-111.

Todling, R., Cohn, S. E. and N. S. Sivakumaran, 1998: Suboptimal schemes for retrospective data assimilation based on the fixed-lag Kalman smoother. *Mon. Wea. Rev.*, **126**, 247-259.

Verlaan, M., 1998: Efficient Kalman filtering algorithms for hydrodynamics models. Ph.D. thesis, Technische Universiteit Delft, The Netherlands, 201 pp.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.

Wentz, F. J. (1997): A well calibrated ocean algorithm for SSM/I. *J. Geophys. Res.*, **120**, 8703-8718.

Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.*, **125**, 2274-2292.

Zhu, Y., R. Todling, and S. E. Cohn, 1999: Technical remarks on smoother algorithms. DAO Office Note 99-02, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

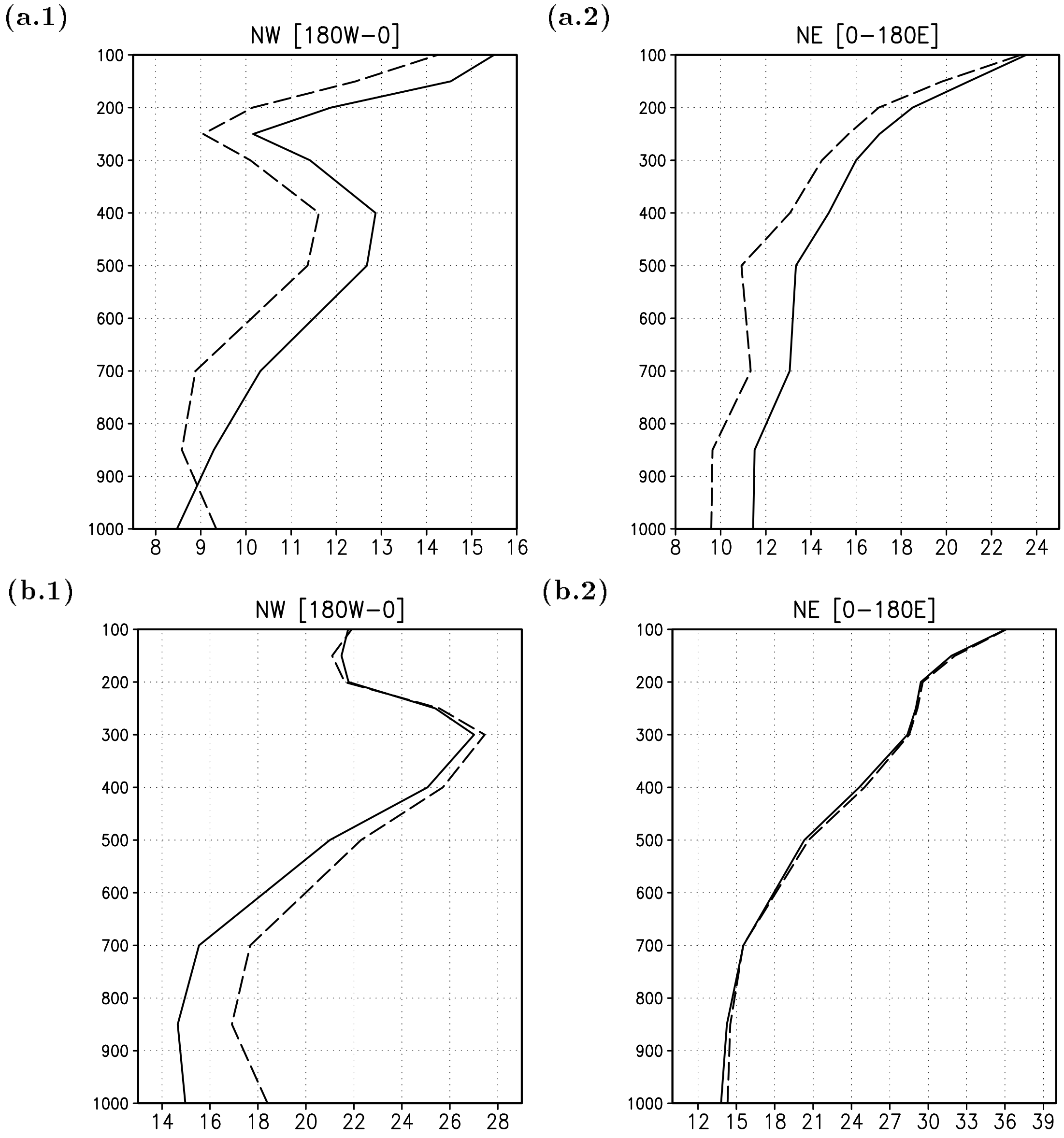


Figure 4: Time root-mean-square (RMS) bias (panels a) and standard deviation (panels b) for the geopotential height radiosondes gridded OMF residuals for the control experiment (solid curves) and for when forecasts are calculated from the lag-1 retrospective analyses from the RA experiment (dashed curves). Panels 1 on the left are for the Northwestern quadrant of the globe defined between longitudes 180W-0 and between latitudes 20N-90N; panels 2 on the right are for the Northeastern quadrant of the globe between longitudes 0-180E and latitudes 20N-90N. Units are in 1 m, and the scales in the abscissa is properly adjusted in each panel.

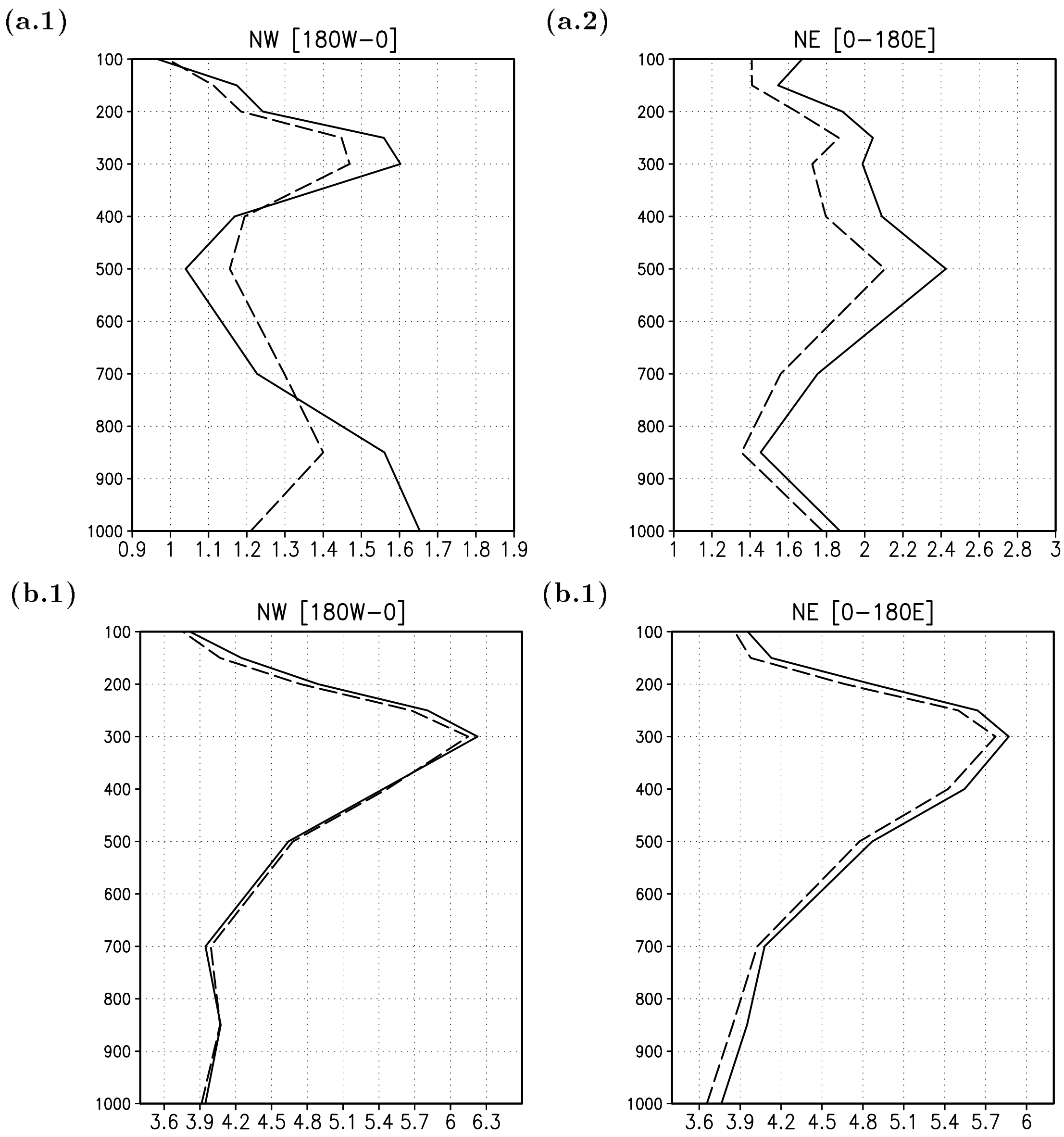
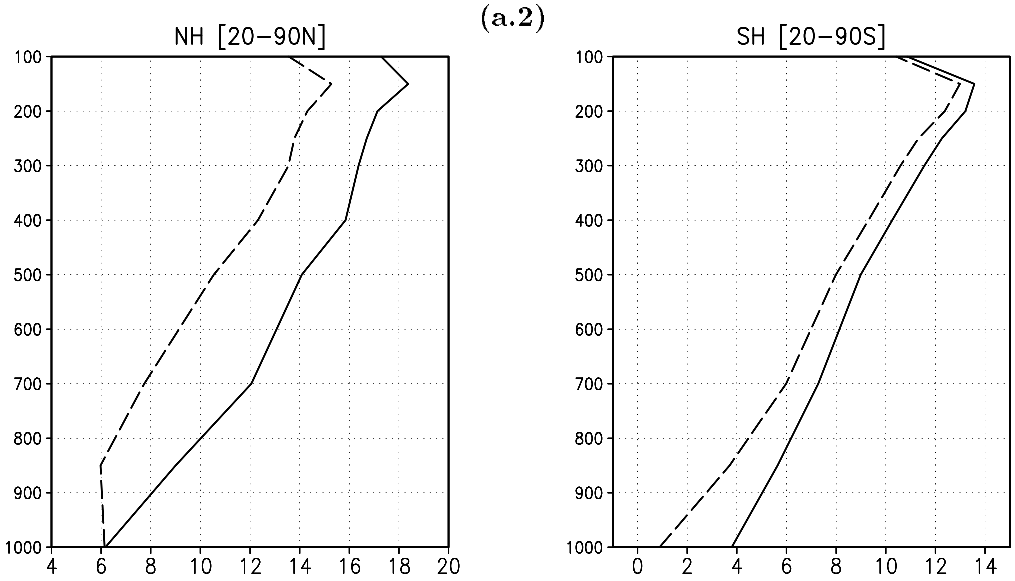


Figure 5: As in Fig. 4, but for zonal wind radiosondes OMF residuals. Units are now in 1 m s^{-1}.

(a.1)

(a.2)



(b.1)

(b.2)

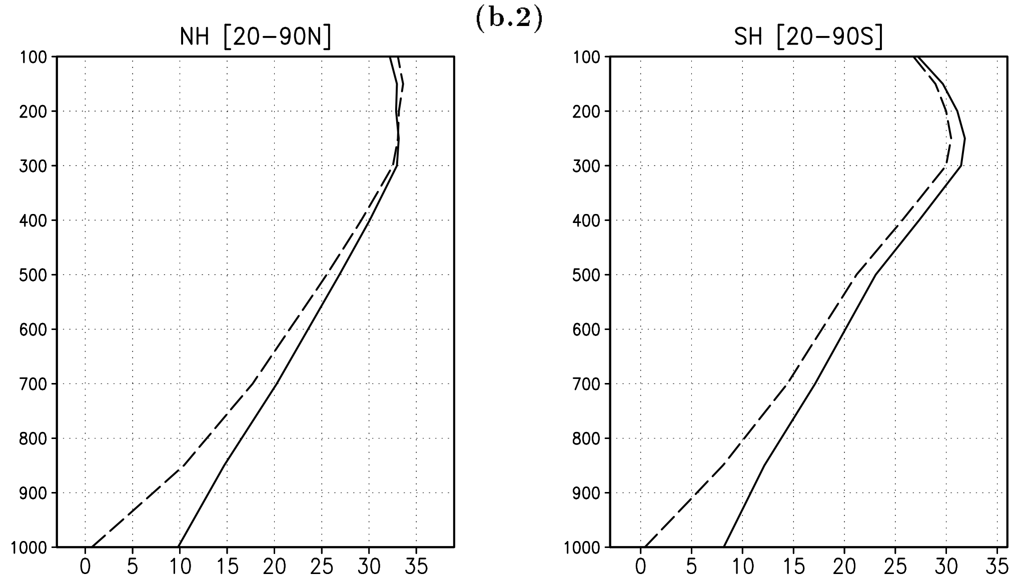


Figure 6: Similar to Fig. 4, but for TOVS geopotential heights retrievals OMF residuals and somewhat different regions: left panels are now for the Northern Hemisphere; right panels are now for the Southern Hemisphere.

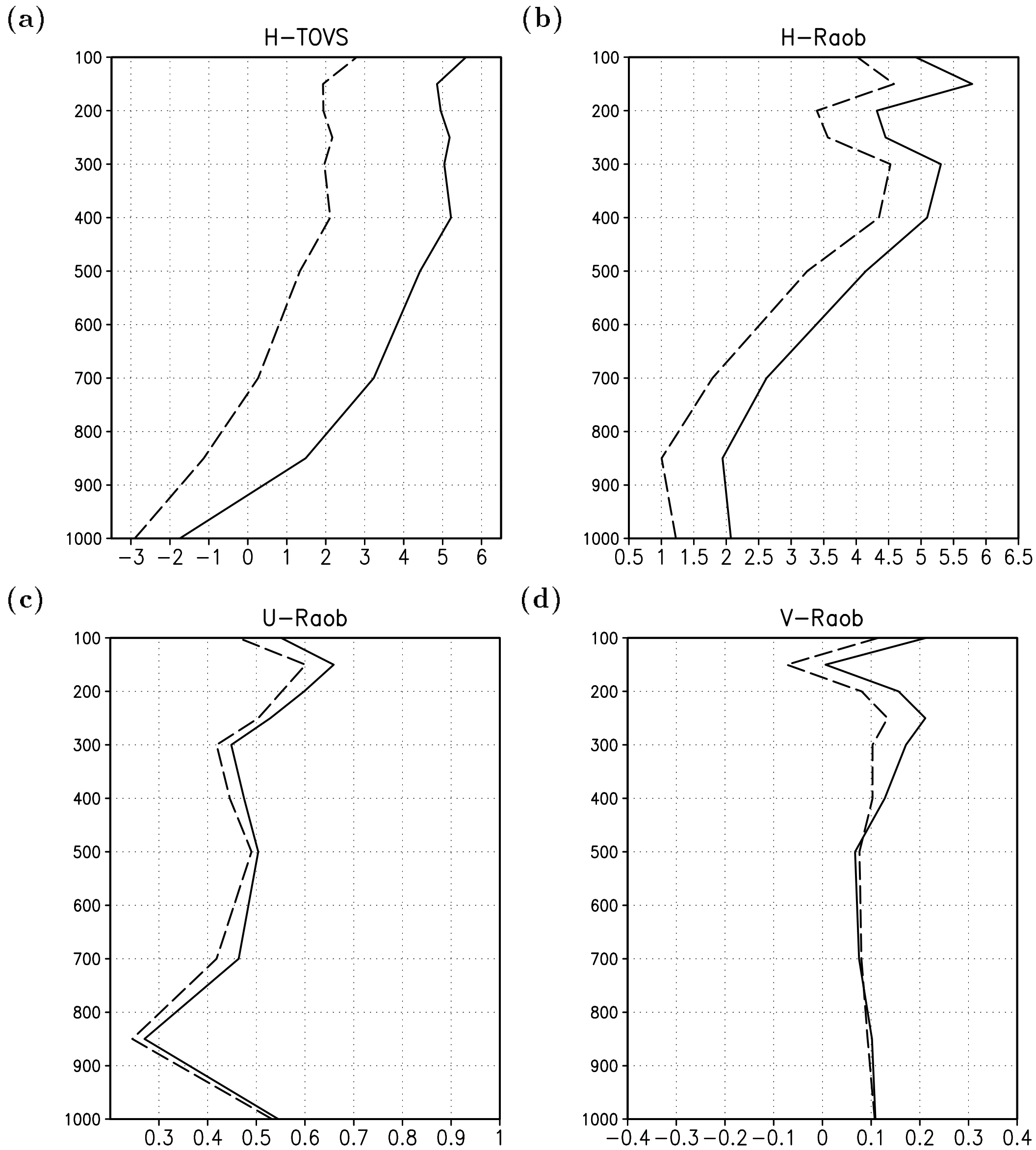


Figure 7: Globally-averaged time mean OMF residuals for the CTL experiment (solid curves) and for the retrospective forecasts (dashed curves). Panel (a) is for the geopotential height TOVS retrievals residuals; panels (b)-(d) are for the geopotential height, zonal wind, meridional wind radiosondes residuals, respectively.

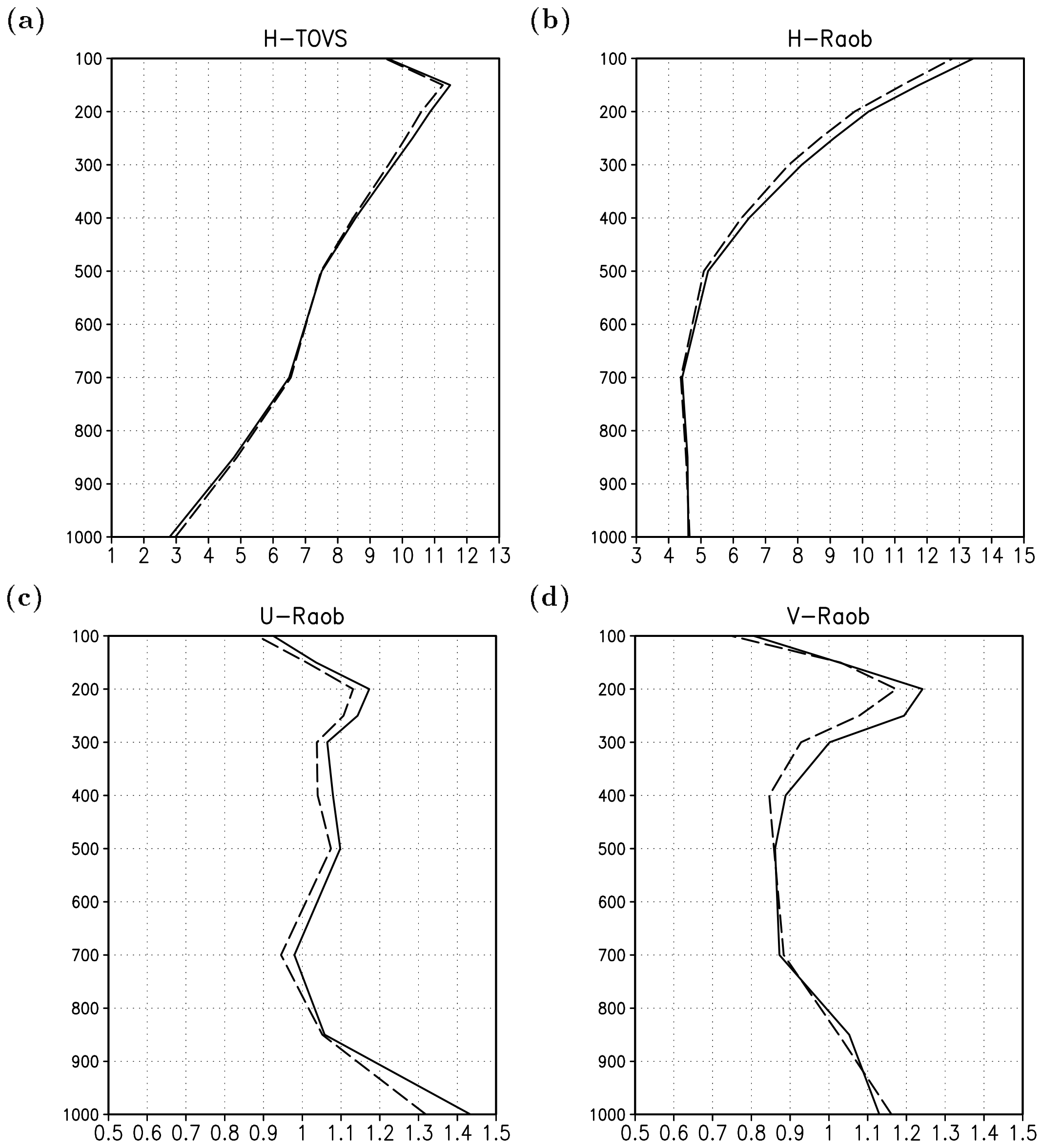


Figure 8: Globally-averaged time RMS bias of observation-minus-analysis (OMA) residuals for analysis of CTL experiment (solid curves) and retrospective assimilation from RIA experiment. Panels are arranged as in Fig. 7: panel (a) is for the geopotential height TOVS retrievals residuals; panels (b)-(d) are for the geopotential height, zonal wind, meridional wind radiosondes residuals, respectively.

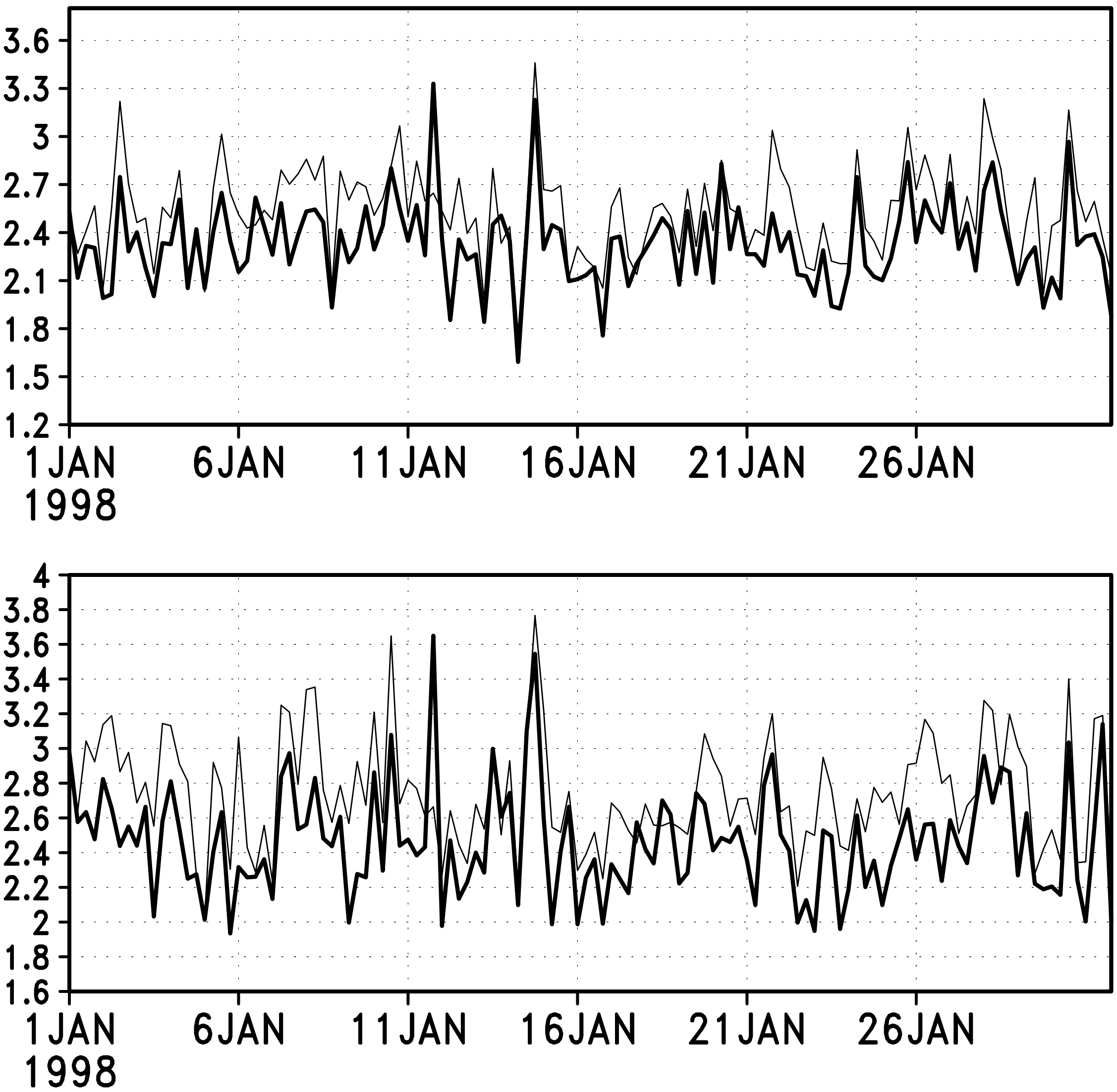


Figure 9: Time series of the glabally avaraged RMS bias for the zonal (top) and meridional (bottom) component of the cloud-track winds OMA residuals at 200 mb. The thin curves are for the CTL experiment, and the thick curves are for the RIA experiment. Units are in 1 m s^{-1}.

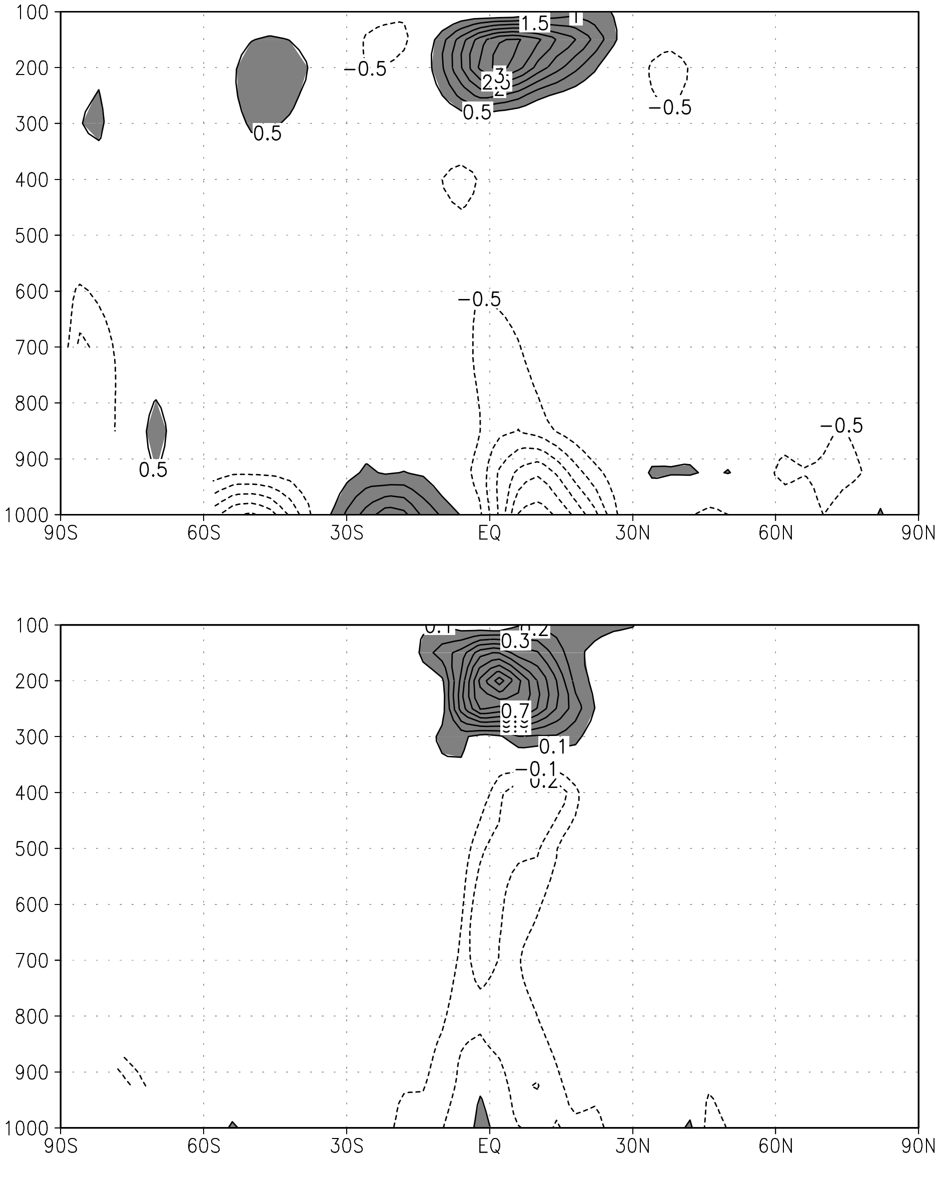


Figure 12: January 1998 zonally-averaged meridional wind for the RIA experiment (top) and its difference from the zonally-averaged meridional wind of the CTL experiment (bottom). Unit are in 1 m s^{-1}

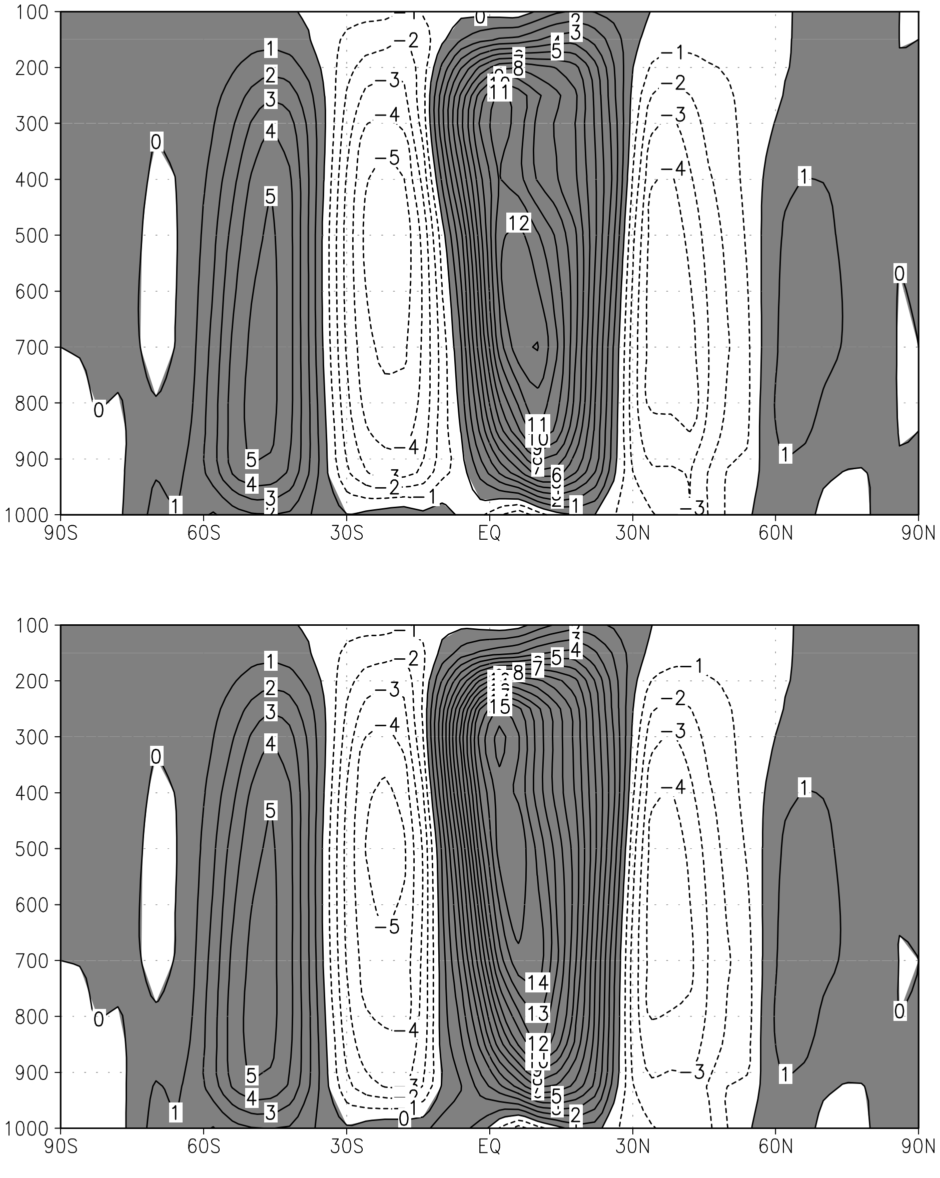


Figure 13: Mass stream function for CTL (top) and RIA (bottom) experiments. Units are in 10^{10} kg s^{-1}.

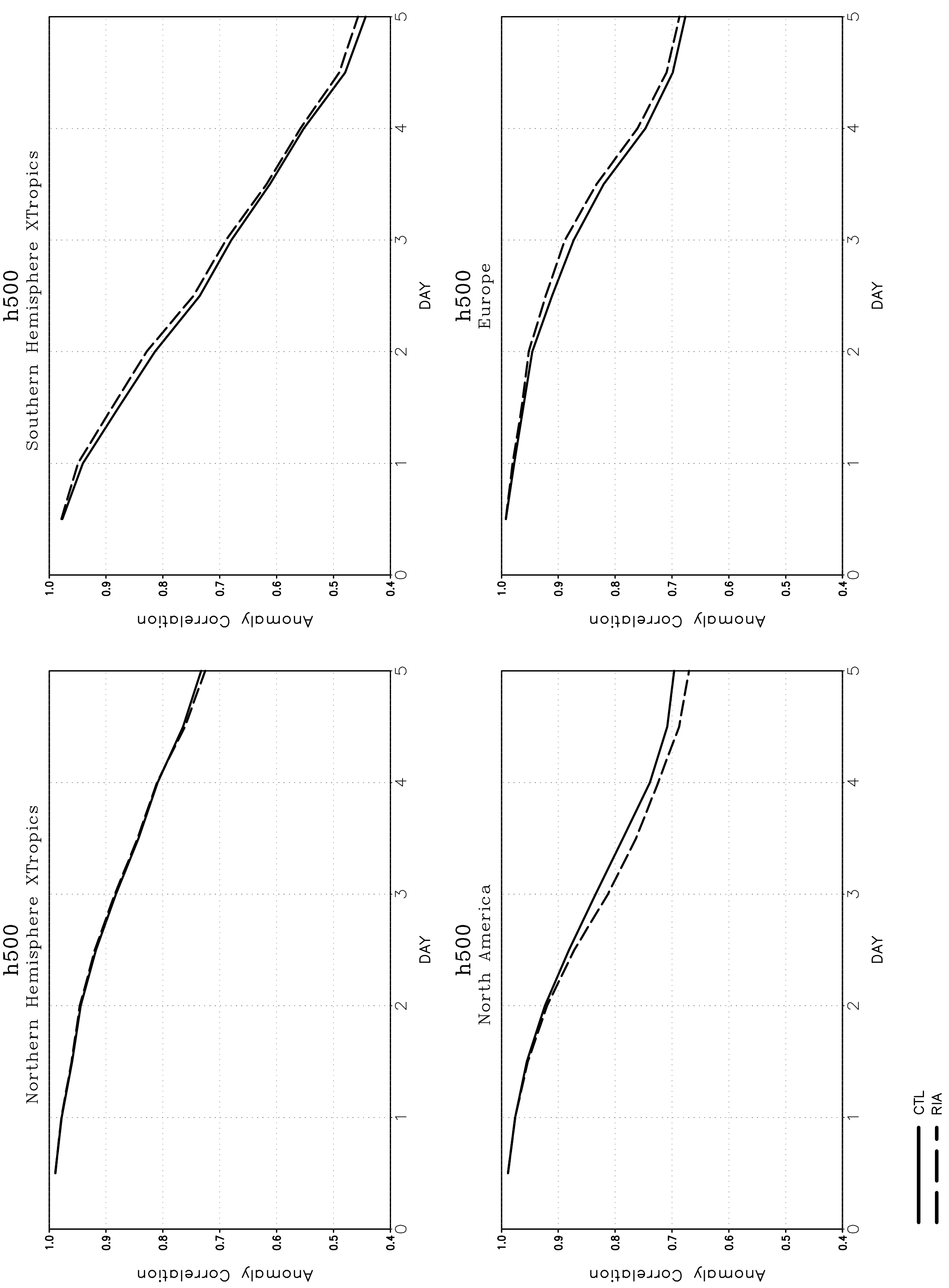


Figure 14: Anomaly correlations for 500 mb geopotential heights for CTL (solid curves) and RIA (dashed curves) experiments.

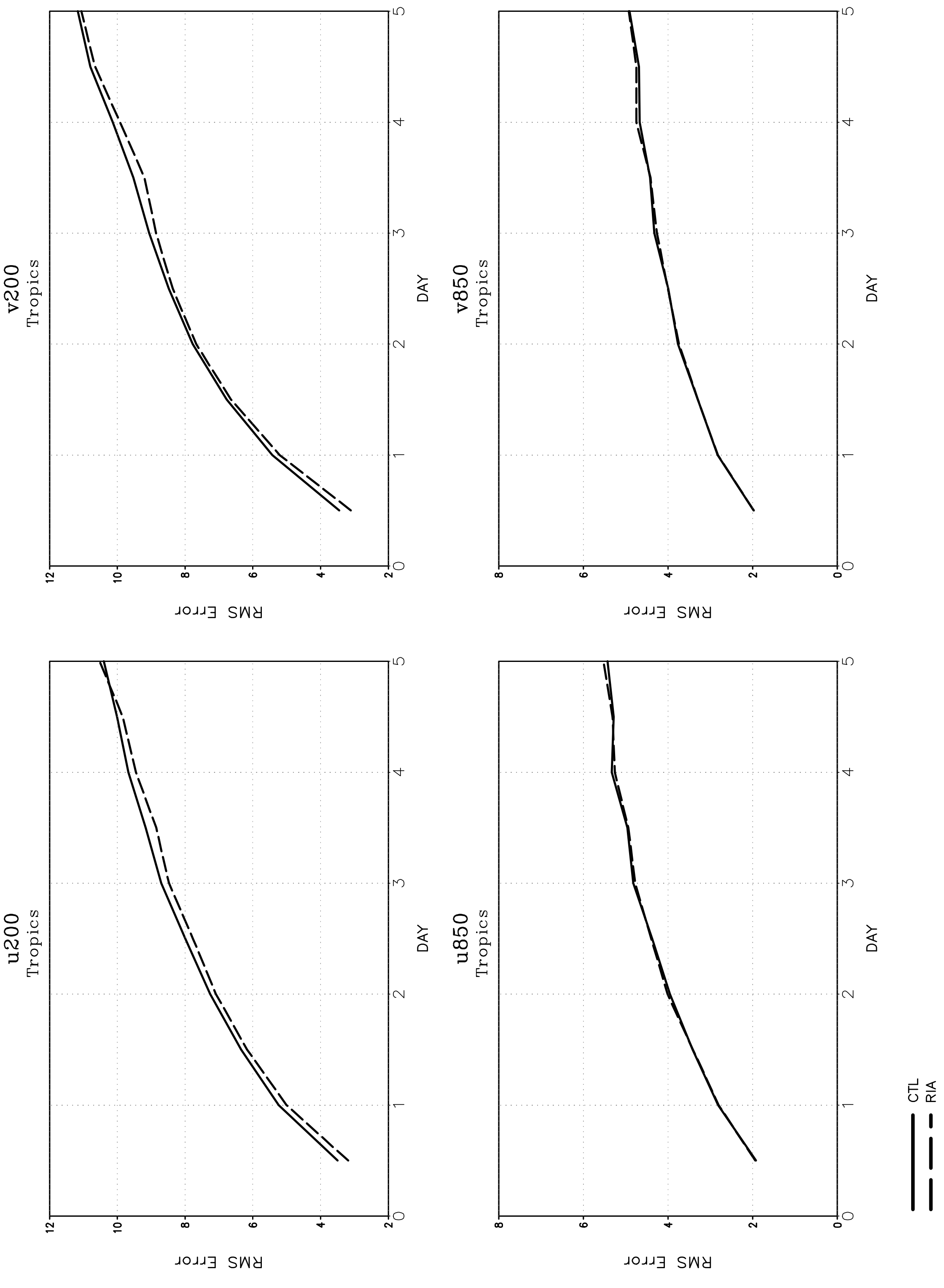


Figure 15: Root-mean-square error from forecast skill study for the zonal (left) and meridional (right) components of the wind at 200 mb (top) and 850 mb (bottom). Solid curves are for CTL experiment; dashed curves are for RIA experiment.