The GEOS-3 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

(Manuscript received 10 May 2002, in final form 10 March 2003)

ABSTRACT

The fixed-lag Kalman smoother (FLKS) has been proposed as a framework to construct data assimilation procedures capable of producing high-quality climate research datasets. FLKS-based systems, referred to as retrospective data assimilation systems, are 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 retrospective assimilation procedures practical.

In this article, an FLKS-based retrospective data assimilation system implementation for the Goddard Earth Observing System Data Assimilation System is presented. The practicality of this implementation comes from the practicality of its underlying (filter) analysis system, that is, the physical-space statistical analysis system (PSAS). The behavior of two schemes is studied here. The first retrospective analysis (RA) scheme is designed simply to update the regular PSAS analyses with observations available at times ahead of the regular analysis times. Results are presented for when observations 6-h ahead of the analysis time are used to update the PSAS analyses and thereby to calculate the so-called lag-1 retrospective analyses. Consistency tests for this RA scheme show that the lag-1 retrospective analyses indeed have better 6-h predictive skill than the predictions from the regular analyses. This motivates the introduction of the second retrospective analysis scheme, which, at each analysis time, uses the 6-h retrospective analysis to create a new forecast to replace the forecast normally used in the PSAS analysis, and therefore allows the calculation of a revised (filter) PSAS analysis. 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 quality of the assimilation.

1. Introduction

The concept of retrospective data assimilation, as invoked in the present article, was introduced by Cohn et al. (1994; CST94 hereafter) to refer to the calculation of refined analyses that use observations *past* each analysis time. Retrospective data assimilation is possible when analyses are not required in real time, for example, during the generation of reanalysis datasets for climate research.

In estimation theory, estimates of the state of a system produced from observations on both sides of the analysis time are known as smoother estimates. In sequential data assimilation a natural smoothing technique to consider is that of fixed-point smoothing. In this case, the usual filter estimate at a fixed time, which is produced from 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 shows their impact to be useful. The idea of estimating the state of a system at a fixed time over and over again as 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. In our fixed-lag smoother formulation, when observations up to one data assimilation update period beyond the time of each regular filter estimate are used to calculate the (refined) smoother estimate, we say that we are calculating the lag-1 (typically 6 h) retrospective analysis; when observations up to two update periods beyond are used to calculate another (even more refined) smoother estimate, we say that we are calculating the lag-2 (typically 12 h) retrospective analysis; and so on. For linear systems under the typical assumption of unbiased Gaussian-distributed errors, the fixed-lag Kalman smoother (FLKS) provides the best unbiased estimate of the state of the system at a sequence of given times using observations in the past, present, and at a time lag-l beyond the time of each estimate.

The FLKS is composed of two major components:

* Additional affiliation: Science Applications International Corporation, Beltsville, Maryland.

+ Additional affiliation: Department of Mathematics and CSIT, Florida State University, Tallahassee, Florida.

Corresponding author address: Dr. Yanqiu Zhu, Data Assimilation Office, NASA GSFC, Code 910.3, Greenbelt, MD 20771.
E-mail: yzhu@dao.gsfc.nasa.gov

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 resulting from 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 portions to study the behavior of a variety of combinations of filter and smoother approximations to the linear FLKS. One particular approximation studied there, namely, the adaptive retrospective data assimilation scheme using a constant forecast error covariance filter (CCF), was seen as having the potential for being implemented in practice. It replaces the filter portion of the FLKS by a constant forecast error covariance filter much like that of operational three-dimensional variational analysis systems. Examples of such systems are those of Parrish and Derber (1992), Courtier et al. (1998), Daley and Barker (2001), and the physical-space statistical analysis system (PSAS) of Cohn et al. (1998), which is central to the work in the present article.

To 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 results 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. Four-dimensional variational data assimilation (4DVAR) 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 based on 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 version 3 (GEOS-3) Data Assimilation System (DAS). In the RA system, there is no feedback from the retrospective analyses into the system. Since the GEOS-3 DAS forecast error covariance matrix of PSAS varies slowly in time, we can identify the suboptimal RA procedure studied here with the CCF scheme of Todling et al. (1998). Our RA implementation in GEOS-3 DAS is general and applicable to any number of time lags, but in the present article we concentrate on results for the 6-h, that is, 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, which introduces a feedback component into the procedure. In the RIA, the lag-1 retrospective analysis at a given time t_{k-1} is used to produce a new forecast at time t_k that is used to revise the filter (PSAS) analysis at the same time t_k. In the RIA the final analysis is the second (iterated) analysis calculated using the forecast 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 improve the overall quality of the analyses. This lag-1 RIA scheme makes the retrospective procedure resemble a 4DVAR cycle (e.g., Courtier et al. 1994; Rabier et al. 2000; Li and Navon 2001).

Indeed, the original FLKS-based retrospective analysis formulation of CST94, and the RIA here, can be viewed as alternative approaches to 4DVAR. 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 3DVAR that take into account observations within a time interval. Ménard and Daley (1996) have shown the equivalence of 4DVAR and fixed-interval smoothing. Similarly, for linear dynamics, the FLKS is algebraically equivalent to 4DVAR and can be derived from the 4DVAR cost function by solving a two-point boundary value problem (Zhu et al. 1999). The main distinction between 4DVAR and the FLKS is in their computational approaches. The former involves an iterative optimization procedure to arrive at 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 account for model error. As pointed out by Todling et al. (1998), FLKS-based assimilation schemes directly inherit any model error covariance parameterization embedded in the filter portion. Various techniques to account for model error in 4DVAR can be formulated by using the dynamical model as a weak constraint on the optimization problem (e.g., Derber 1989; Bennett et al. 1996; Zupanski 1997). Another point to make relates to one of the advantages of 4DVAR over 3DVAR-like procedures, namely, that the former uses the observations nearly at their proper times (e.g., Fisher and Andersson 2001; Rabier et al. 2000), whereas in the latter it is more common to bundle the observations into 6-h batches. This can be resolved by using a rapid update cycle (RUC) strategy. Though this is not explored in the present article, since in GEOS-3 DAS the observations are bundled into 6-h batches, we should point out that there is no intrinsic difficulty in building an FLKS-based retrospective analysis system

under an RUC strategy. Furthermore, since the calculation of retrospective analyses mainly involves applications of PSAS (Zhu et al. 1999), we can expect cost reduction for an FLKS-based retrospective analysis system under RUC (Lyster et al. 2003, manuscript submitted to *Mon. Wea. Rev.*).

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 toward adding a retrospective component to GEOS-3 DAS; 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 on 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:

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

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

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

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

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

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 estimation theory, which is mostly suitable to the development of smoothers. This subscript notation is also particularly helpful in reminding us that for linear systems perturbed by Gaussian-distributed noise the forecast $\mathbf{w}^f_{k\,|\,k-1}$ and filter analysis $\mathbf{w}^a_{k\,|\,k}$ state vectors are actually *conditional means* of the true state n-vector $\mathbf{w}^t_k$, that is,

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

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

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

The EKF, like the KF, depends on the residual p_k-vector $\mathbf{v}_k$ in (1b) formed by 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 observation 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). It is only when the filter is optimal that the time series of residual vectors $\mathbf{v}_k$ can be identified with the innovation sequence (e.g., Anderson and Moore 1979, section 5.3). Although the expression for the weighting matrix $\mathbf{K}_{k\,|\,k}$ in the EKF is similar in form to its linear KF equivalent, contrary to the linear case, $\mathbf{K}_{k\,|\,k}$ in (1c) is now state dependent. The OMF residuals covariance matrix $\mathbf{\Gamma}_k$ is given by

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

for uncorrelated observation and forecast errors. Here, $\mathbf{R}_k$ is the $p_k \times p_k$ observation error covariance matrix, $\mathbf{H}_k$ is the Jacobian matrix of the observation operator $\mathcal{H}_k$, and $\mathbf{P}^f_{k\,|\,k-1}$ is the state-dependent $n \times n$ forecast error covariance matrix. The forecast error covariance matrix depends on the model error covariance matrix $\mathbf{Q}_k$ and evolves from the $n \times n$ analysis error covariance matrix $\mathbf{P}^a_{k-1\,|\,k-1}$ through the Jacobian matrix $\mathbf{A}_{k,k-1}$ of the dynamics operator $\mathcal{A}_{k,k-1}$ according to (1d).

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 filter analyses and retrospective analyses at previous times t_{k-l} using an update equation similar to the state update expression (1b) of the filter portion. The lag-l FLKS retrospective analyses based on observations newly available at time t_k are calculated by

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

for $l = 1, 2, \ldots, \min(k, L)$ for a maximum desired lag $l = L$. Each retrospective analysis for a fixed time t_{k-l} is an ''incremental'' correction to an estimate of the state calculated previously. Each lag of the algorithm introduces corrections to the state estimate by using observations at times further and further beyond the retrospective analysis time, up to the maximum desired lag L. The mechanism for correcting consecutive state estimates at a fixed time with successive smoother calculations makes the FLKS algorithm resemble the fixed-point smoother. Indeed, the FLKS of CST94 and TC96 can be derived from a fixed-point smoother formulation using the approach of state augmentation (e.g., Biswas and Mahalanabis 1973).

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 $\mathbf{K}_{k-l\,|\,k}$ depends on the OMF residual covariance matrix $\mathbf{\Gamma}_k$ in (4). Furthermore, $\mathbf{K}_{k-l\,|\,k}$ also depends on the $n \times p_k$ matrix $\mathbf{H}_k^{\mathrm{T}}$, the transpose of the Jacobian of the observation operator, and on the $n \times n$ forecast–analysis cross-covariance matrix $\mathbf{P}^{fa}_{k,k-l\,|\,k-1}$, through the EKF-based expression

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

as derived in TC96. The forecast-analysis cross-covariance $\mathbf{P}^{fa}_{k,k-l\,|\,k-1}$ 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

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

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

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

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 retrospective analysis at time t_{k-l} is

$$\mathbf{w}^{a}_{k-l|k} = \mathcal{E}\{\mathbf{w}^{t}_{k-l} | \mathbf{w}^{o}_{k}, \mathbf{w}^{o}_{k-1}, \ldots, \mathbf{w}^{o}_{1}\}, \tag{8}$$

where now, in contrast to the filter estimates (2), the expectation is conditioned on all observations before, at, and after time t_{k-l} 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. In general, independently of nonlinearities, 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 arises directly from its being formulated on the basis of an underlying filter is that it incorporates model error covariances naturally and automatically (see also the appendix). In fact, (5)–(7) do not depend explicitly on the model error covariance. A variety of techniques exists to incorporate model error in 4DVAR (e.g., Derber 1989; Bennett et al. 1996; Zupanski 1997). Because 4DVAR is algebraically equivalent to fixed-interval smoothing (see Ménard and Daley 1996; Zhu et al. 1999) and for all practical purposes we can always choose a lag L in fixed-lag smoothing that accomplishes the same benefit as fixed-interval smoothing (Moore 1973), FLKS-based assimilation procedures present a potential alternative to 4DVAR. Since we currently lack the knowledge necessary to parameterize model error covariances, this advantage of the FLKS over 4DVAR is not very significant, but it may prove to be relevant in the future.

3. Practical framework: GEOS-3 DAS considerations

The algorithm described in the previous section serves mainly as a guide to help design feasible data assimilation procedures. It is well known that the computational cost of evolving full covariances is excessive for filtering, let alone for smoothing as in (7), and likely not justifiable because of our relative lack of knowledge of the required input model and observation error statistics. This has 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 our implementation of the FLKS-based retrospective procedure for the GEOS-3 DAS. Before describing the retrospective analysis portion of the implementation we summarize the current GEOS-3 DAS that approximates, in principle, the filter portion of the algorithm.

a. The GEOS-3 analysis and data assimilation system

The Data Assimilation Office (DAO) operational GEOS-3 data assimilation system consists of three major components: an atmospheric general circulation model (GCM), the PSAS, and the incremental analysis update (IAU) procedure. At the so-called analysis times, the GCM provides a forecast field to PSAS so it can process OMF residuals and generate the analysis state. The physical space statistical analysis system is an implementation of the EKF equations (1b)–(1c), obtaining

the analysis state as a correction to the model forecast. The error covariance evolution expressions (1d) and (1e) are neglected and, therefore, PSAS functions as a suboptimal filter, as in the case of other operational 3DVAR systems. 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-h period around the analysis time. The GCM trajectory obtained during the IAU integration is known as the *assimilated* trajectory.

In GEOS-3 DAS the state space of the GCM is different from 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* as the background field transformed to the analysis space. The model and analysis spaces are different because their state variables and grids are different. 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, the zonal and meridional components of the sea level wind, the 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 DAO 1996, for details).

We designate an m-dimensional sigma-coordinate GCM state vector by $\mathbf{y}(\sigma)$ and an n-dimensional pressure-coordinate analysis state vector by $\mathbf{w}(p)$, to emphasize explicitly the vertical coordinate system of these states. For our purposes, we can represent a GCM integration as

$$\frac{d\mathbf{y}(\sigma)}{dt} = \mathcal{M}[\mathbf{y}(\sigma)] + \alpha \mathbf{f}(\sigma). \tag{9}$$

Here, $\mathcal{M}$ is the nonlinear GCM operator and the second term on the right-hand side corresponds to the constant IAU forcing term applied to the GCM during the IAU integration period. The parameter α controls when and how the IAU forcing $\mathbf{f}(\sigma)$ affects the integrations. For the 6-h IAU time interval $[t_{k-1/2}, t_{k+1/2}]$ we set $\tau^{iau} = t_{k+1/2} - t_{k-1/2}$ and $\alpha = 1/\tau^{iau}$, and during the 3-h GCM background integration time interval $[t_{k+1/2}, t_{k+1}]$ 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 forecast $\mathbf{w}^f_{k|k-1}(p)$ through the operation

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

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. This operator can be absorbed into the definition of the state vector and become transparent in the description of the filter and smoother equations. However, to make clear the connection between the mathematical description and the actual implementation of these procedures we opt to refer to Π explicitly.

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) into two steps. The first step is to solve the 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

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

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 (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. Furthermore 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-3 IAU assimilation, the analysis in (12) is converted back to the model space through a conversion operator, Π^+,

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

which is then used finally to construct the filter IAU forcing $\mathbf{f}_{k|k}(\sigma)$ to be used in (9),

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

with $\mathbf{f}_{k|k}(\sigma)$ being, in fact, the model-space filter analysis increment. The actual implementation of Π^+ is such that it renders minimal the difference between a field $\mathbf{w}(p)$ in the analysis space and the field resulting from transforming $\mathbf{w}(p)$ to the model space using Π^+ and, subsequently, transforming the resulting vector back to the analysis space using Π.

The GEOS-3 data assimilation system does not have an explicit initialization scheme. It relies on IAU, a process of gradually introducing the analysis increments over a certain period around the analysis time, to control imbalances resulting from the introduction of the increments. The time filtering properties of IAU have been

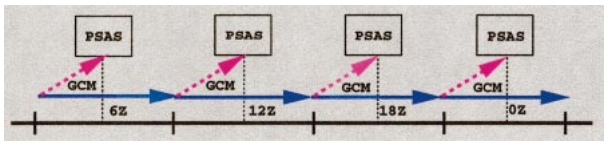

FIG. 1. Schematic representation of the IAU procedure. The dashed arrows represent the 3-h GCM integration that provides the forecast to PSAS. At each analysis time PSAS uses OMF residuals to calculate an updated state estimate (analysis; vertical dotted lines). The analysis-minus-forecast difference is converted to a model-space tendency term used to force the GCM during a 6-h integration around the analysis time; this is the IAU period represented by the solid thick arrows. The cycle is repeated with a 3-h GCM integration, without the IAU tendency term, to provide the forecast for the next analysis time. The line formed by the solid arrows represents a time-continuous IAU trajectory, referred to as the assimilation. (Similar to Fig. 1 of Bloom et al. 1996.)

established for a linear system with analysis given by the sum of a model forecast and an arbitrary analysis increment (Bloom et al. 1996). IAU acts only on the contribution to the assimilation state by the increment. Other strategies employed in 3DVAR-like systems to control imbalances introduced by the analysis procedure include those of Gauthier et al. (1999) and Lynch and Huang (1992). A schematic representation of the IAU assimilation procedure is shown in Fig. 1. In GEOS-3 DAS, observations are processed in 6-h intervals, which in the IAU framework implies that the GCM is integrated for 6 h starting 3 h before the analysis time. Going from left to right in the diagram, at an analysis time, say $t = 6$Z (0600 UTC), observations and a 3-h model forecast (represented by the right-upward-pointing dashed arrow) are combined in PSAS to calculate the filter analysis. This analysis is used to construct the IAU forcing (14) and the model is integrated forward, forced by the IAU tendency starting from $t = 3$Z (0300 UTC) up to $t = 9$Z (0900 UTC). Beyond this time, the IAU forcing is set to 0 and the model runs ''free'' for the next 3 h. At the end of this free 3-h integration the GCM provides the forecast 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 rightward-pointing arrows.

b. The GEOS-3 retrospective analysis

We now have 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 analysis increments

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

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

$$\mathbf{K}_{k-l|k} = \mathbf{P}^f_{k-l|k-l-1}\left[\prod_{j=k-l+1}^{k}(\mathbf{I} - \mathbf{K}_{j-1|j-1}\mathbf{H}_{j-1})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}_{j,j-1}\right]\mathbf{H}^{\mathrm{T}}_k\boldsymbol{\Gamma}^{-1}_k, \quad (16)$$

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

$$\delta\mathbf{w}^a_{k-l|k}(p) = \mathbf{P}^f_{k-l|k-l-1}\left[\prod_{j=k-l+1}^{k}(\mathbf{I} - \mathbf{H}^{\mathrm{T}}_{j-1}\boldsymbol{\Gamma}^{-1}_{j-1}\mathbf{H}_{j-1}\mathbf{P}^f_{j-1|j-2})\mathbf{A}^{\mathrm{T}}_{j,j-1}\right]\mathbf{H}^{\mathrm{T}}_k\mathbf{x}_k, \quad (17)$$

where we used (11) to replace $\boldsymbol{\Gamma}^{-1}_k\mathbf{v}_k$ with $\mathbf{x}_k$. We see from this expression that the lag-l retrospective increment is a linear combination of the columns of the forecast error covariance $\mathbf{P}^f_{k-l\,|\,k-l-1}$, as is the original filter increment. 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 $\mathbf{P}^{fa}_{k,k-l\,|\,k-1}$ and $\mathbf{P}^{aa}_{k,k-l\,|\,k}$, and smoother error covariance $\mathbf{P}^a_{k-l\,|\,k}$ do not 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 $\mathbf{z}_{k\,|\,k}$ as

$$\mathbf{z}_{k\,|\,k} \equiv \mathbf{H}^{\mathrm{T}}_k\mathbf{x}_k, \quad (18)$$

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

$$j = k$$

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

$$\mathbf{z}^{p}_{j-1|k} = \mathbf{A}^{\mathrm{T}}_{j,j-1}\mathbf{z}_{j|k} \tag{19a}$$

$$\mathbf{\Gamma}_{j-1}\mathbf{x}^{r}_{j-1|k} = \mathbf{H}_{j-1}\mathbf{P}^{f}_{j-1|j-2}\mathbf{z}^{p}_{j-1|k} \tag{19b}$$

$$\mathbf{z}_{j-1|k} = \mathbf{z}^{p}_{j-1|k} - \mathbf{H}^{\mathrm{T}}_{j-1}\mathbf{x}^{r}_{j-1|k} \tag{19c}$$

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

$$j = j - 1$$

endwhile

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

Notice that the entire retrospective analysis algorithm (18)–(19) works in the analysis space. In particular, the propagation operator $\mathbf{A}^{\mathrm{T}}_{k,k-1} = \mathbf{A}^{\mathrm{T}}_{k,k-1}(p)$ in (19a) is defined in pressure coordinates and 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, where we recall that m is the dimension of a model state vector and n is the dimension of an analysis state vector.

A few remarks are appropriate 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 such a simple manner that none of the terms on the right-hand side of (1d) are taken into account. However, when the expressions (7) for the smoother error cross-covariances are bypassed and the 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 $\mathbf{P}^{a}_{k-l|k}$. 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 particular region earlier on in the assimilation. That is, when at time t_{k-1}, say, there are no observations over a certain region, the filter analysis at this time will essentially equal the forecast over that region—aside from possible contributions by farther-away regions through the forecast error correlations. If at time t_k, say, observations become available over the region in question, or information from observations at nearby downstream regions get propagated through the adjoint of the tangent linear dynamics $\mathbf{A}^{\mathrm{T}}_{k,k-1}$ into the region in question, this new information will be used to calculate a correction to the filter analysis at time t_{k-1} as the lag-1 retrospective analysis represented in (17). In these cases, it is the first term in the square bracket of (17) that mostly contributes to the correction to the filter analysis.
- The linear system (19b) solved within the retrospective analysis algorithm involves exactly the same operators required to calculate the sensitivity of forecasts to observation changes, as measured by some prespecified cost function, as in the approach of Baker and Daley [2000; cf. 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 4DVAR are closely related to operations required to study forecast sensitivity; the same is true of the operations in FLKS-based retrospective analysis schemes.
- A simple approximation to the retrospective analysis portion just described is to replace the adjoint operator in (19a) by the identity operator. 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 as having a time-independent forecast error covariance, one might expect that replacing the adjoint by the identity operator in (19a) would 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-3 DAS and has found improvement in the mean error due to lag $l = 1$ and even to lag $l = 2$ retrospective analyses when compared with the corresponding PSAS (filter) results.

c. The GEOS-3 lag-1 retrospective-based iterated analysis

When the system is nonlinear, the idea to feed the filter estimate back into the analysis equation is particularly attractive, since we expect the filter analysis to be a better estimate of the state of the system than the forecast 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 for such operators. Iterative 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 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 of these iterative procedures are inspired by Newton-type methods for solving systems of nonlinear equations [see Navon and Legler (1987) and Zou et al. (1993b) for reviews of these methods].

Motivated by the above, we introduce here a procedure to use the retrospective analysis to improve the overall GEOS-3 IAU-based assimilation. At first, the algorithm is based only on the lag-1 retrospective analyses. At any given 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 retrospective IAU forcing as

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

which is similar to (14) but is constructed using observations one lag ahead of time t_k. The lag-1 retrospective IAU forcing $\mathbf{f}_{k\,|\,k+1}(\sigma)$ is the increment constructed by subtracting the background from the model-space retrospective analysis, which is different from the retrospective analysis increment in (15) constructed by subtracting the filter (or previous retrospective) analysis from the retrospective analysis. This lag-1 retrospective IAU forcing 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 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 = 12Z (1200 UTC), we calculate a retrospective analysis by first integrating the transformed PSAS solution vector in (18) back in time using the adjoint operation (19a); this is represented in the diagram by the left-downward-pointing dashed arrow. A new PSAS-like linear system problem can then be solved as in (19b) with the corresponding update (19c), and the lag-1 retrospective analysis constructed using (19d), as represented in the diagram by the box tagged ''Retro ANA.'' In the end, a lag-1 retrospective IAU forcing at t = 6Z (0600 UTC) is constructed as indicated in (23), and the GCM is integrated for 6 h using this forcing term as in (9). From this point on, the procedure follows the regular IAU schematic until it is time to process the observations at t = 18Z (1800 UTC) when the lag-1 retrospective analysis at t = 12Z (1200 UTC) can be calculated and the whole cycle repeated. The thin blue arrows in Fig. 2 correspond to the retrospective trajectory. In the RIA scheme we concentrate on the iterated filter-smoother trajectory represented in the figure 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).

The IAU filtering properties depend only on how the analysis increments are applied given a background field. The standard GEOS-3 assimilation scheme and the RIA scheme use the same background but different analyses to construct an IAU forcing: the PSAS analysis as in (14), and the retrospective analysis as in (23), respectively. Therefore, the GEOS-3 RIA automatically benefits from the filtering properties of IAU. Initialization procedures can also be incorporated in smoother schemes such as 4DVAR (e.g., Thépaut et al. 1993; Polavarapu et al. 2000; Zou et al. 1993a; Gauthier and Thépaut 2001).

From the diagram in Fig. 2 we see that the retrospective-based iterated analysis results in 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-h GCM integration and two extra PSAS analyses. Such an increase in the computational cost can only be justified if the procedure results in considerably improved analyses. One way to reduce the computational burden is by calculating some of the steps in (19) at different resolutions. Similarly to the strategy of incremental 4DVAR 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. This type of approach to reduce computational cost involves the development of additional interpolation operators and their corresponding adjoints.

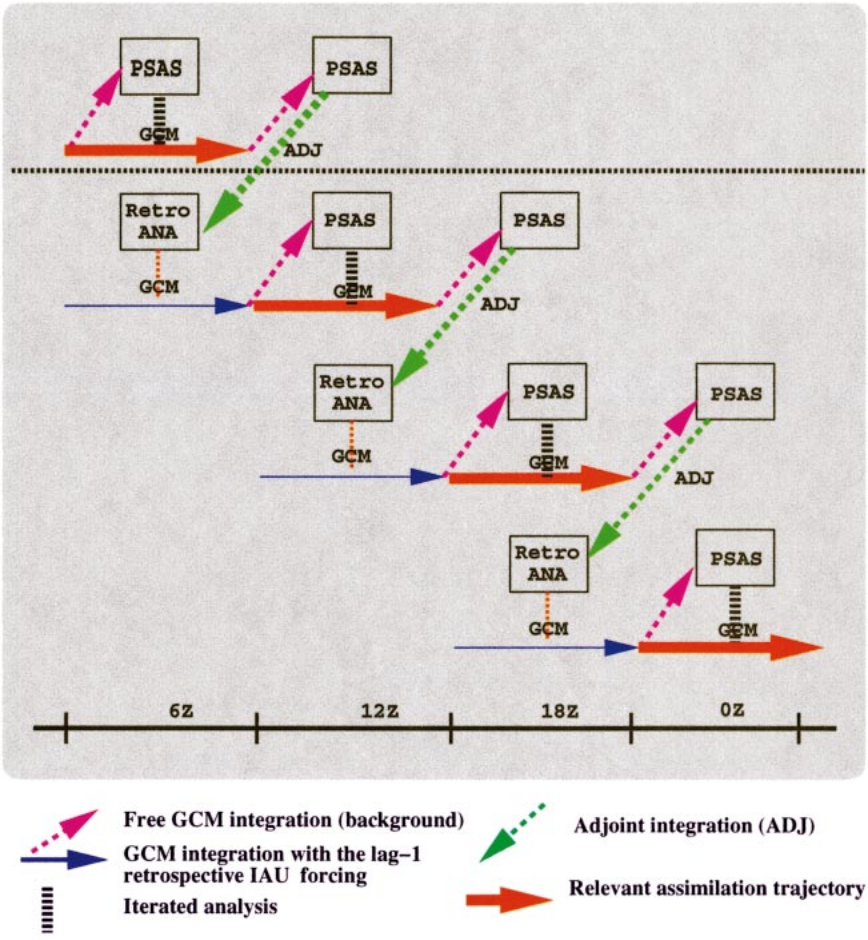

FIG. 2. Schematic representation of the lag-1 iterated retrospective data assimilation procedure. Dashed right-upward-pointing arrows represent GCM forecast integration; solid rightward-pointing arrows represent GCM integration with IAU (thick) and retrospective IAU (thin) forcing. Dashed left-downward-pointing arrows represent 6-h adjoint model integrations. The boxes labeled ''Retro ANA'' stand for the PSAS application in (19b). The retrospective assimilation is used to provide a revised forecast that is further used to revise the filter analysis at each synoptic time.

4. GEOS-3 experimental results

a. Configuration and experimental setup

The retrospective analysis procedures of the previous section were implemented as an extension to GEOS-3 DAS. The retrospective portion of the GEOS-3 software is compatible with the *first* operational version of GEOS-3 DAS, designed to support the National Aeronautics and Space Administration's (NASA's) Earth Observing System mission and its *Terra* satellite. In this section we refer to this earlier operational version as GEOS-3′ to avoid possible confusion with the final version of GEOS-3 operational at the time of this writing. The GEOS-3′ GCM operates at a resolution of 1° latitude by 1° longitude and 48 vertical sigma levels, with a dynamical core essentially that of Suarez and Takacs (1995). At the synoptic hours, PSAS calculates the analysis at a resolution of 2° latitude by 2.5° longitude on 20 pressure levels. Details on the implementation of PSAS can be found in da Silva and Guo (1996), Guo et al. (1998), and Larson et al. (1998). 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. For expediency, the experiments performed for the present article used both the GCM and PSAS at

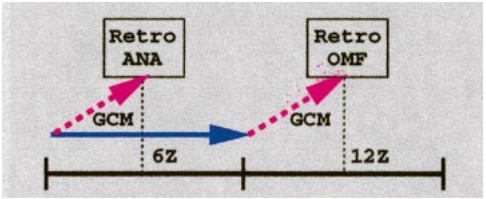

FIG. 3. Schematic of the procedure to issue forecasts from retrospective analyses using the IAU framework. Arrows are similar to those in Fig. 1. The main purpose of the retrospective forecast is the calculation of the OMF residuals indicated by the ''Retro OMF'' box.

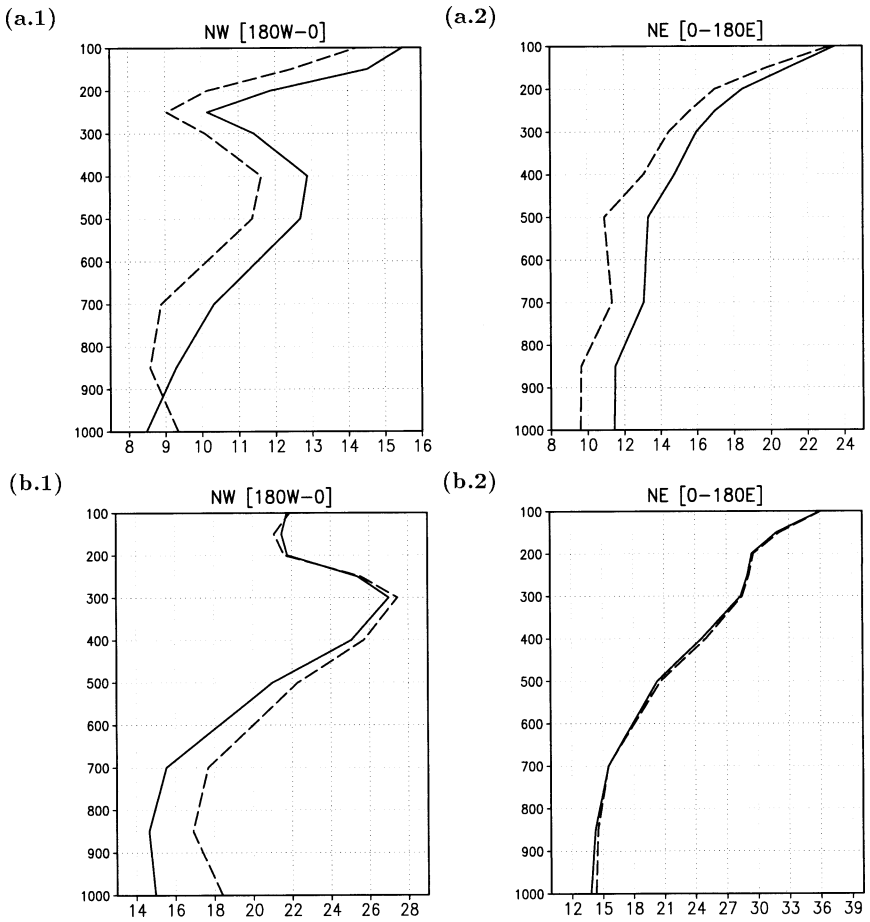

FIG. 4. Spatial rms of the (a) time-averaged bias and (b) std dev for the radiosonde geopotential height OMF residuals for the control experiment (solid curves) and for forecasts from the lag-1 retrospective analyses from the RA experiment (dashed curves): (left) the northwestern quadrant of the globe defined between longitudes 180°W and 0° and between latitudes 20° and 90°N; (right) the northeastern quadrant of the globe between longitudes 0° and 180°E and latitudes 20° and 90°N. Units are in m on the abscissa and hPa on the ordinate.

the coarse horizontal resolution of 4° latitude by 5° longitude; the GCM and PSAS vertical resolutions were kept unchanged. We also simplify the experimental configuration by updating the GCM trajectory needed during the adjoint integrations only every 6 h. Except for sea-wind satellite observations, all observation data types used in GEOS-3′ are included in our experiments. Conventional observations from ships, environmental and drifting buoys, surface stations, winds from pilot balloons, aircraft reports, and radiosonde stations are used. Cloud track wind retrievals and Television Infrared Observational Satellite Operational Vertical Sounding (TOVS) geopotential height retrievals are used as well. Furthermore, the Wentz (1997) special sensor microwave/imager (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 GEOS-3 DAS: the adjoint of the tangent linear GCM; the additional PSAS-like operators involved in (19b); the linear operator (22a) taking model-space variables into analysis-space variables; and the linear operator (22b) taking analysis-space variables into model-space variables. Presently, the adjoint of the GCM includes only the hydrodynamics adjoint and the adjoint of a simple diffusion scheme. Most modifications required to PSAS were quite simple 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), because we strove to make sure that the back and forth operations would render minimal error. Some of this work was done by hand, and some was done using the automatic differentiation tool of Giering and Kaminski (1998).

In the present article, only results for the lag-1 (6 h) retrospective analysis are discussed. We compare the results of three experiments conducted over the month of January 1998. To minimize possible differences due to spinup issues, the experiments are actually started on

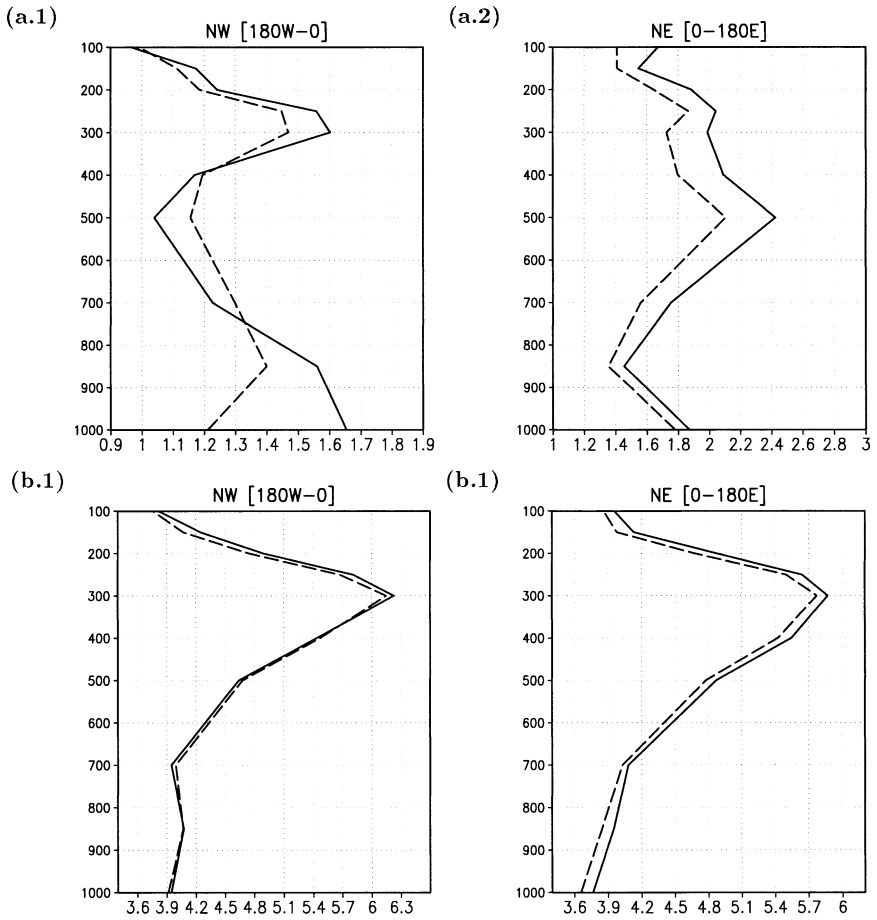

FIG. 5. As in Fig. 4, but for zonal wind radiosondes OMF residuals. Units on the abscissa are here m s^{-1}.

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. In the second experiment, referred to as the RA experiment, we also calculate lag-1 (6 h) retrospective analyses for the entire month of January 1998. Since there is no feedback in this experiment, it uses the same background fields and OMF time series of the control experiment. The third experiment is aimed at evaluating the lag-1 (6 h) retrospective-based iterated analysis procedure introduced in the previous section and is referred to as the RIA experiment.

We evaluate the RA and RIA experiments mainly by examining the time series statistics of their corresponding residuals. That is, depending on the case, we calculate root-mean-square (rms) of 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 4° latitude by 5° longitude cells on the 20 pressure levels of the analysis space before calculating any statistics. We calculate statistics only from grid boxes containing 15 or more reports during the month. In the discussion that follows, we concentrate on results obtained in the troposphere.

b. Evaluation of the 6-h 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 the 6-h 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, one can show that in the linear optimal case, that is, when the filter gain is the Kalman gain,

$$\mathcal{E}\{(\mathbf{w}_k^0 - \mathbf{H}_k\mathbf{w}_{k|k+1}^a)(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k+1}^a)^{\mathrm{T}}\} < \mathcal{E}\{(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k}^a)(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k}^a)^{\mathrm{T}}\}. \quad (24)$$

Although there is no guarantee of this holding in general

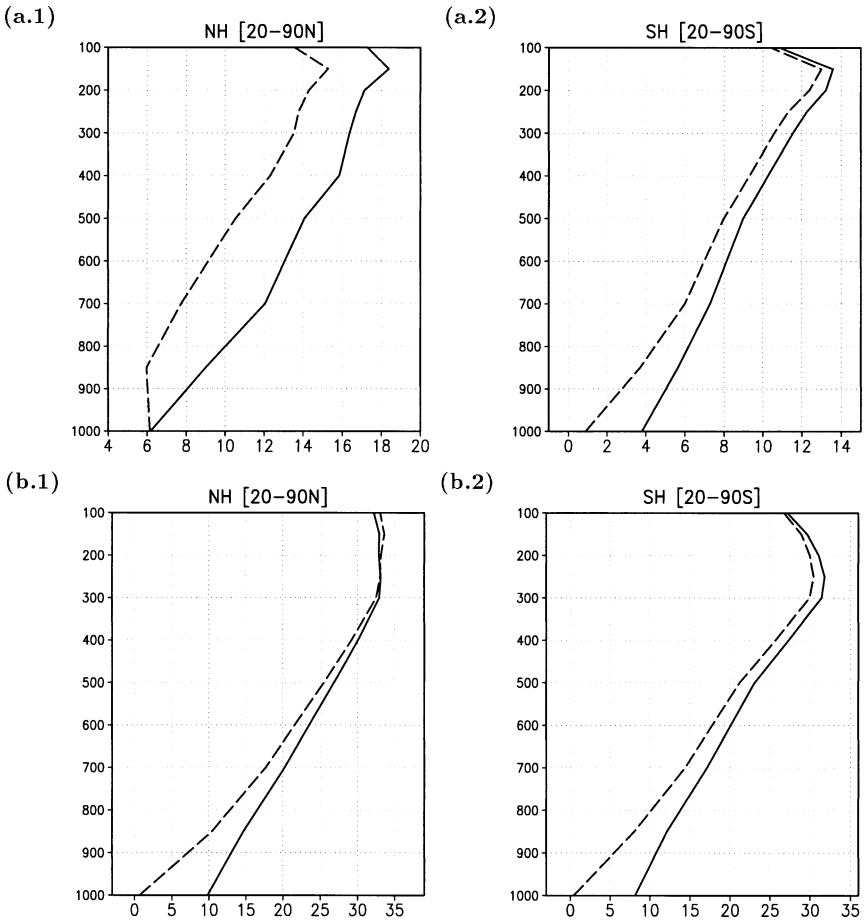

FIG. 6. Similar to Fig. 4, but for TOVS geopotential height OMF residuals and different regions: (left) the Northern Hemisphere, (right) the Southern Hemisphere.

for the suboptimal nonlinear case under study, we would like to examine the extent to which it does. In practice, short of perturbing the observational data, to assess this quantity we must make the usual ergodic assumption and replace the expectation by a time average. Examination of the spatial rms of the time-averaged biases and standard deviations of the OMA residuals when the analyses are either the regular filter analyses of the CTL experiment or the lag-1 retrospective analyses of the RA experiment has shown them to be virtually identical (results not shown). Therefore, from this point of view we might be led to think that there is no payoff in calculating lag-1 retrospective analyses.

Another way of comparing the quality of two sets of analyses is to compare the skill of forecasts issued from them. We expect forecasts issued from retrospective analyses to be superior to regular forecasts for at least their total lag period, 6 h in the lag-1 case here, since their initial conditions have had the benefit of observations that far into each forecast. Hence, we compare the OMF residual statistics of the regular filter forecasts of the CTL experiment and of the so-called retrospective forecasts issued from the lag-1 retrospective analyses. Since the OMF residuals from a regular GEOS-3 DAS run, such as the CTL experiment, involve 6-h forecasts that are produced from partly integrating the GCM with the filter IAU forcing for 3 h and partly integrating the GCM for another 3 h without the influence of the IAU tendencies (see Fig. 1), we must use the retrospective analyses carefully when constructing OMF residuals from them. To make a fair comparison, we calculate OMF residuals from the 6-h 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 schematic shown in Fig. 3. As illustrated in the figure, the retrospective OMF residuals at, say, 12Z (1200 UTC), are calculated by converting the 6Z (0600 UTC) retrospective analysis to the model space and constructing the corresponding retrospective IAU forcing, following (23). This retrospective IAU forcing is used as a tendency term during a 6-h GCM integration, started at 3Z (0300 UTC). At the end of the 6-h integration the retrospective tendency term is turned off, by setting $\alpha = 0$ in (9), and the model is left to run free for another

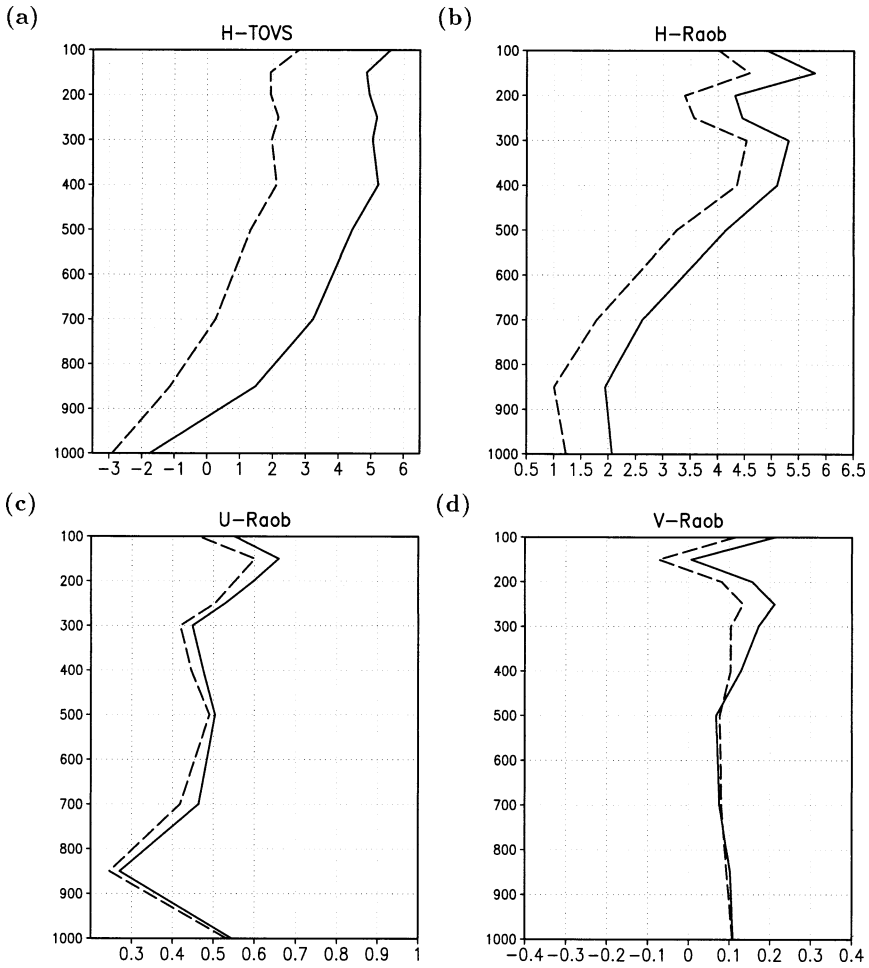

FIG. 7. Globally averaged time-mean OMF residuals for the CTL experiment (solid curves) and for the retrospective forecasts (dashed curves): (a) the geopotential height TOVS retrievals residuals; (b)–(d) the geopotential height, zonal wind, and meridional wind radiosondes residuals, respectively. Units as in Figs. 4 and 5.

3 h, after which the OMF residuals at 12Z (1200 UTC) can be calculated using the observations at that time.

Using these retrospective forecasts, Fig. 4 shows the spatial rms of the (top) time-averaged bias and (bottom) standard deviation for the radiosonde geopotential height OMF residuals for the CTL (solid curves) and RA (dashed curves) experiments averaged over the (left) western and (right) eastern quadrants of the Northern Hemisphere, for latitudes higher than 20°N. These two domains are chosen because they represent the largest concentration of radiosondes over the globe. We see from the top panels that, in the time-averaged bias sense, the forecasts from the lag-1 retrospective analyses correspond to a considerable improvement over the regular GEOS-3 DAS analyses. However, the bottom-left panel shows that for the standard deviations the retrospective forecasts are considerably degraded compared to 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 radiosonde OMF residuals. Except in the northwestern region between pressure levels 700 and 400 hPa, the time-averaged bias for the zonal wind radiosonde OMF residuals is improved when forecasts are issued from the lag-1 retrospective analyses. In this same region, the zonal wind OMF standard deviation (Fig. 5b.1) shows a minor deterioration at levels below 400 hPa, much less than that seen in the OMF heights in Fig. 4b.1; minor improvement in the standard deviations are seen above 400 hPa. Over the northeastern region a minor but consistent improvement is observed in both the bias and standard deviation, as indicated in 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 spatial rms of the (top) time-averaged bias and (bottom) standard deviation for the TOVS geopotential height OMF residuals. Since TOVS provides global coverage in the course of a single day, the spatial

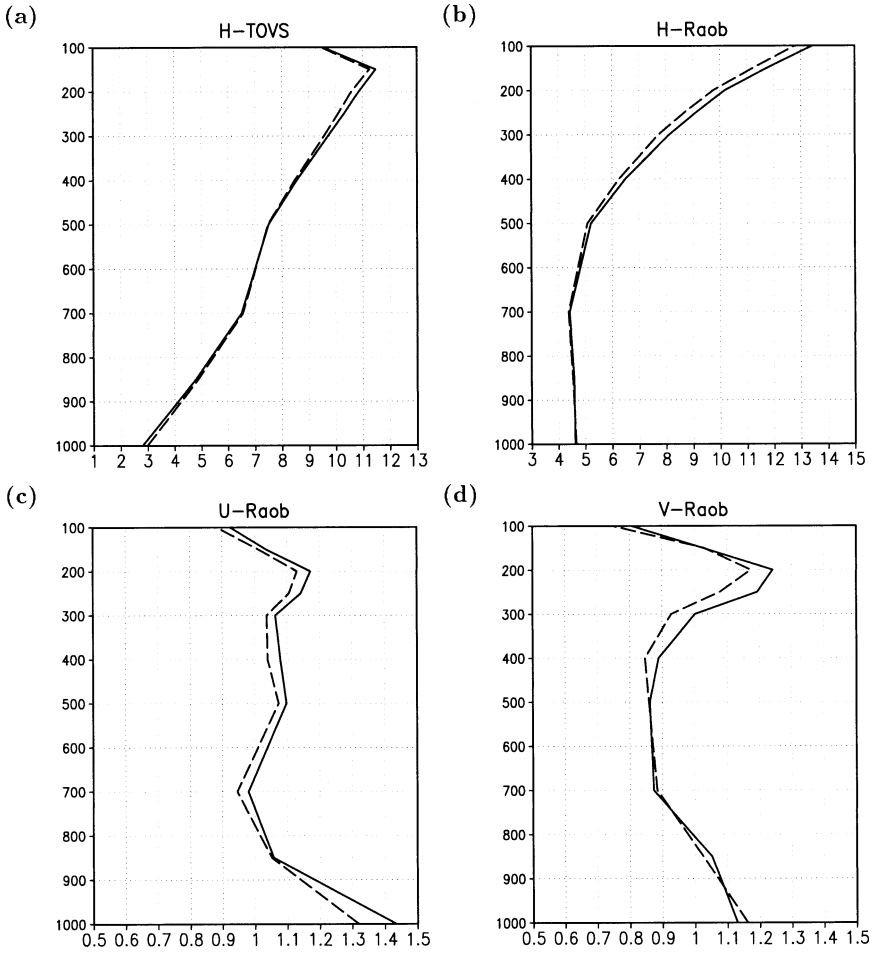

FIG. 8. Spatial rms of the time-averaged bias of OMA residuals for the CTL experiment (solid curves) and retrospective assimilation from the RIA experiment: (a) the TOVS geopotential height residuals; (b)–(d) the radiosonde geopotential height, zonal wind, and meridional wind residuals, respectively. Units as in Figs. 4 and 5.

averages now cover the entire (left) Northern Hemisphere and (right) Southern Hemisphere. 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 (Fig. 6b.1) contradict the deterioration observed in the radiosonde geopotential height OMF residuals (Fig. 4b.1). We attribute this contradiction over North America 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 time-mean residuals. 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-h forecasts and the lag-1 retrospective forecasts. The globally averaged time means for TOVS and radiosonde geopotential height OMF residuals are displayed in Figs. 7a,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 hPa for the TOVS residuals. The zonal and meridional wind components of the radiosonde OMF residuals are displayed in Figs. 7c,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 hPa.

In terms of the metrics presented here for the nonlinear suboptimal case of the GEOS-3 application, we see clear benefit in producing the 6-h forecasts from the lag-1 retrospective analyses over the regular GEOS-3 DAS forecasts. This serves to indicate improved analysis quality with the RA scheme. This also serves as further motivation to consider the iterated retrospective

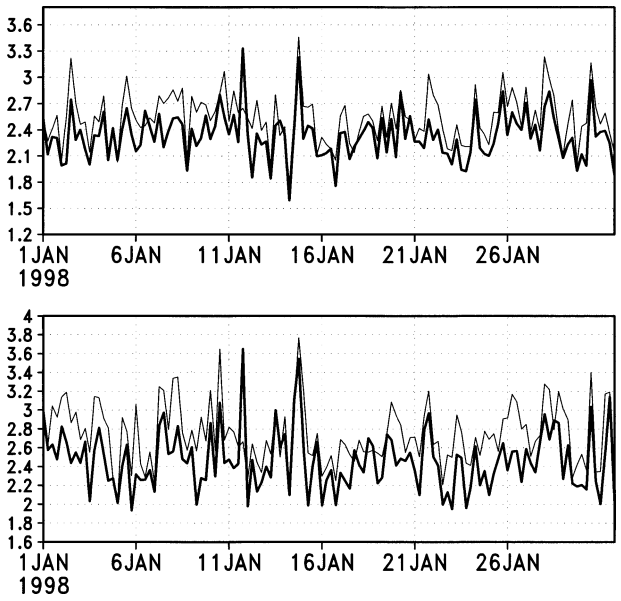

FIG. 9. Time series of the globally averaged bias for the (top) zonal and (bottom) meridional component of the cloud-track wind OMA residuals at 200 hPa. The thin curves are for the CTL experiment, and the thick curves are for the RIA experiment. Units are m s^{-1}.

analysis procedure proposed in the previous section, since it makes direct use of the retrospective forecasts (see Fig. 2).

c. *Evaluation of the 6-h retrospective-based iterated analysis*

We now evaluate the performance of the 6-h (lag 1) retrospective-based iterated analysis scheme. We start by comparing the OMA residuals between the CTL and the RIA experiments. Figure 8 shows the spatial rms of the time-averaged bias for TOVS and radiosonde geopotential height OMA residuals (Figs. 8a,b, respectively), and for the zonal and meridional components of the radiosonde winds (Figs. 8c,d, respectively). Although small, we now actually see improvement in the OMA residuals due to the iterated analysis. To the extent that the expectation can be replaced by the time mean, the inequality (24) holds when $\mathbf{w}^a_{k\,|\,k+1}$ corresponds to the iterated analysis, at least in a globally averaged sense. We have also examined the spatial rms of standard deviations of the OMA residuals of both the CTL and RIA experiments and have found it to change very insignificantly.

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

Frequently, changes made to assimilation systems are evaluated and validated by making comparisons with independent observations, that is, observations that are not assimilated by the system. 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 validate the change in the 200-hPa 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 aircraft wind observations from the Aircraft Communications, Addressing, and Reporting System (ACARS). Neither of these observation types was used in our assimilation experiments and, therefore, they provide independent checks.

The GADS wind observations have been shown by Rukhovets et al. (1998) to be of value to GEOS-3 DAS if used regularly in the PSAS analyses. This suggests that any changes made to GEOS-3 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 of the January 1998 GADS observations to construct OMA residuals for the analyses of both the CTL and RIA experiments. Figure 10 shows maps of standard deviation of the (top) zonal and (bottom) meridional GADS winds OMA gridded residuals for the CTL experiment subtracted from the same quantity for the RIA experiment (RIA minus CTL). 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 the GADS observations are closer to the analyses of the RIA experiment than to those of the CTL experiment.

Similarly, Fig. 11 shows the differences of 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 dwell too much on the

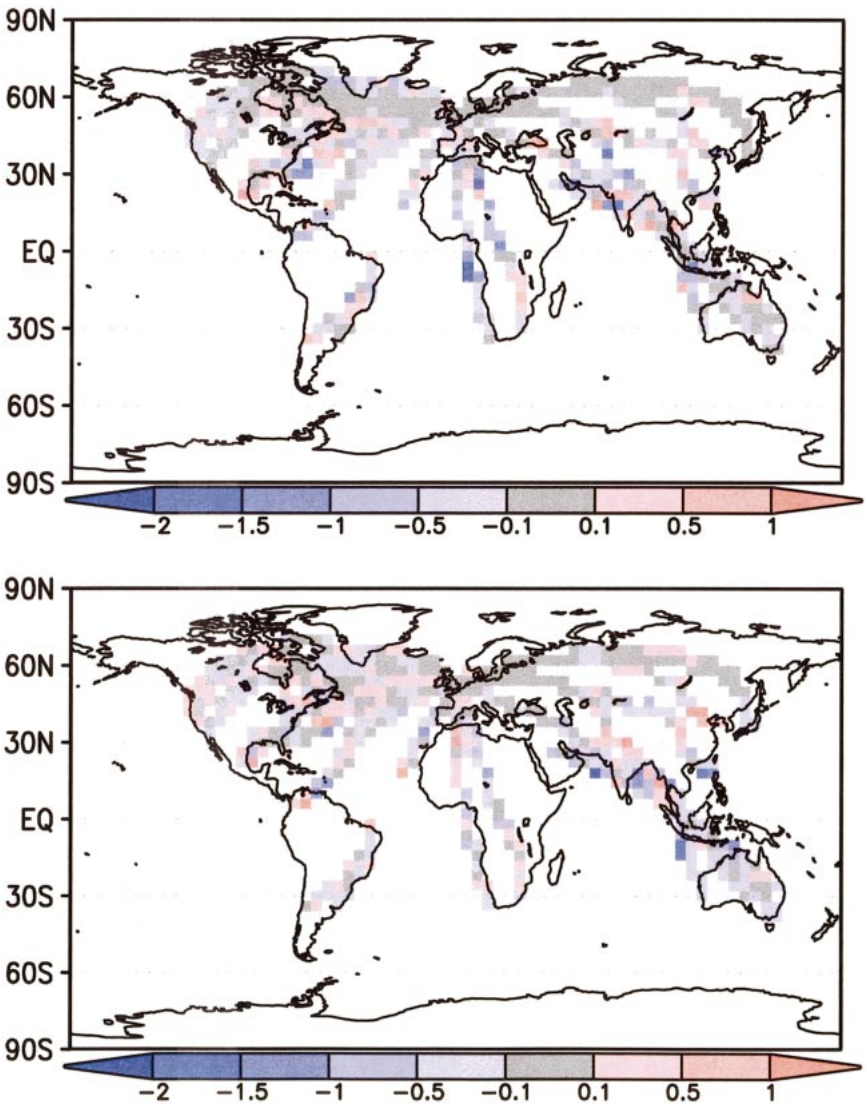

FIG. 10. Maps of the 200-hPa std dev of the GADS winds OMA residuals of the CTL experiment subtracted from the same quantity for the RIA experiment. The top map is for the zonal wind and the bottom map is for the meridional wind. Units are in m s^{-1}.

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 impact on the upper-level winds can be seen more clearly by looking directly at the monthly averaged winds. For instance, Fig. 12 shows (top) the zonally averaged January 1998 monthly mean meridional wind for the RIA experiment and (bottom) its difference from the CTL experiment. 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 affects the Hadley circulation. To see the meridional circulation, we calculate the mass streamfunction ψ by integrating the zonally averaged monthly mean meridional wind using the expression

$$\psi = \frac{2\pi R \cos\phi}{g} \int_{p_{\text{top}}}^{p} [\overline{v}]\, dp', \tag{25}$$

where v stands for the meridional wind, the operator $\overline{\cdot}$ represents the time mean, the operator $[\,\cdot\,]$ represents the zonal average, R is the mean radius of the earth, g is the gravity constant, ϕ is the latitude, and the integral is from $p_{\text{top}} = 10$ hPa to a pressure p. Figure 13 shows the January 1998 mass streamfunction for both the (top) CTL and (bottom) RIA experiments. We see a clear

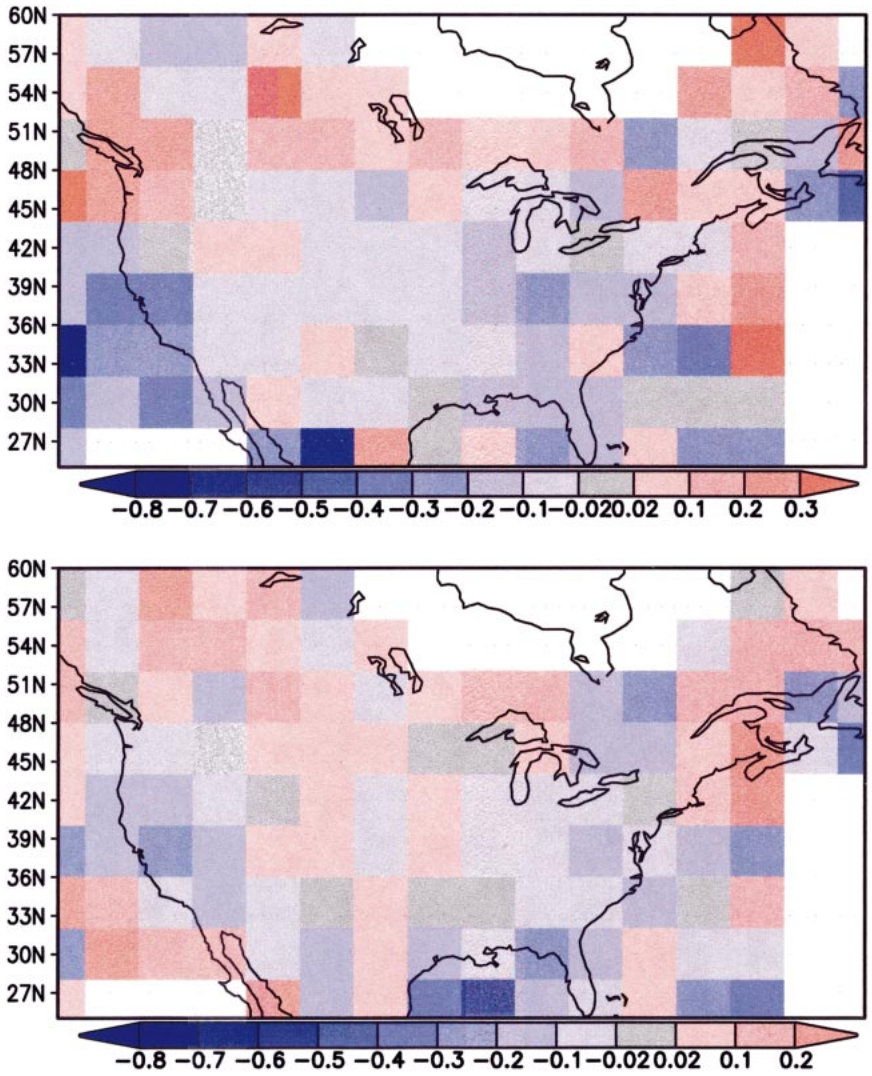

FIG. 11. Same as in Fig. 10, but for the ACARS wind OMA residuals. Only North America is displayed since it corresponds to the area where the bulk of these observations are.

enhancement of the Hadley circulation when the RIA procedure is used, with the mass streamfunction 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 to be well represented at the coarse resolution we use in our experiments here, it is much closer to the circulation pattern of the full-resolution, 1° latitude by 1° longitude GEOS-3 DAS (not shown), with its tropical circulation peaking at 18×10^{10} kg s^{-1}. This suggests that the RIA scheme has the potential for improving climatologically relevant features.

Finally, we compare the skill of 5-day forecasts issued from the CTL and the RIA analyzed fields. These are initialized as in Fig. 3. Since our experiments are confined to the single month of January 1998, we have few independent samples for this comparison. We issued 5-day forecasts starting from 2 January 1998 every 3 days until 26 January 1998, to have a small sample of nine 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 down to five members and performing cross validation. As a measure of forecast skill we calculated anomaly correlations and rms errors [e.g., von Storch and Zwiers (1999), Eqs. (18.17) and (18.18)]. Both the CTL and RIA forecasts were verified against their own analyses. Anomalies were calculated using a 10-yr climatology obtained from European Centre for Medium-Range Weather Forecasts (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 scores of the operational GEOS-3 data assimilation system, which operates at higher resolution than that of the experiments here.

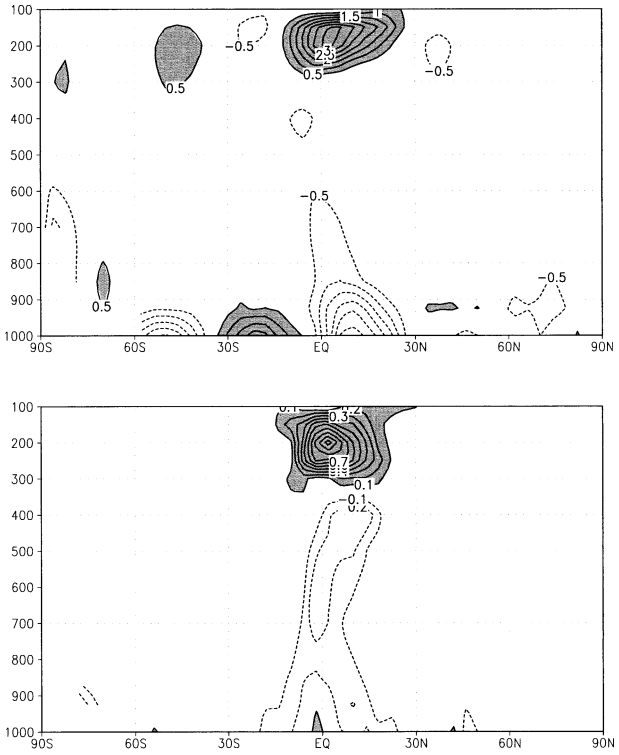

FIG. 12. (top) The Jan 1998 zonally averaged meridional wind for the RIA experiment and (bottom) its difference from the zonally averaged meridional wind of the CTL experiment. Units are in m s^{-1}.

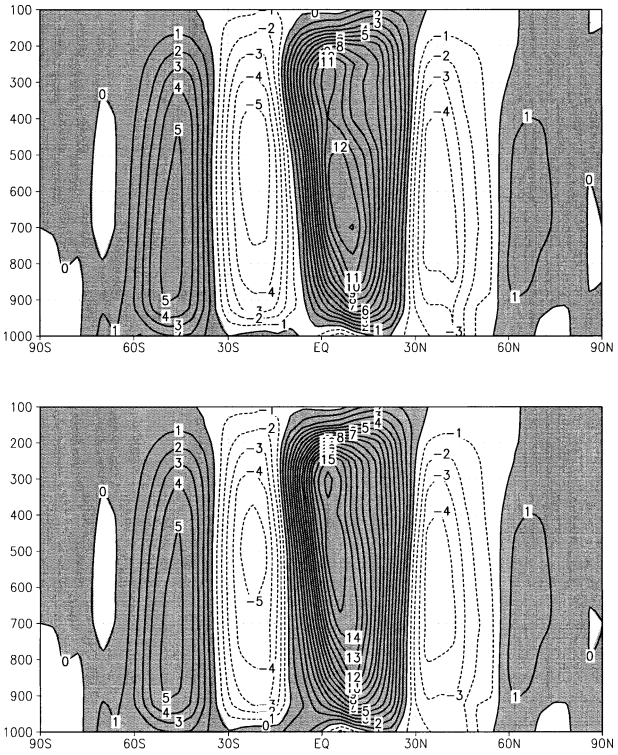

FIG. 13. Mass streamfunction for (top) CTL and (bottom) RIA experiments. Units are in 10^{10} kg s^{-1}.

Figure 14 shows the anomaly correlations for the 500-hPa 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 (top left) Northern Hemisphere extratropics forecasts from RIA analyses are of similar skill as forecasts from the control analyses, at least up to day 4. Over (bottom left) North America the forecast skill from RIA analyses shows some deterioration when compared against the skill of the CTL forecasts. As when studying the OMF residuals obtained from the retrospective forecasts using the RA analyses, this deterioration over North America might be related to contradictions in the observing system over this area (see Figs. 4b.1, 6b.1). In fact, this seems to be an issue confined to this region since, for example, over the (top right) Southern Hemisphere and (bottom right) Europe we see improvement in skill when the 5-day forecasts are issued from the RIA analyses.

As a final illustration comparing the 5-day forecast skill from the CTL and RIA experiments we examine the rms error of the tropical wind fields at 850 and 200 hPa. Figure 15 displays these quantities for both the (left) zonal and (right) meridional components of the wind. The rms errors at 850 hPa are virtually identical, while at 200 hPa 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

A central purpose of atmospheric data assimilation is to produce the best possible estimate of 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. In the context of sequential data assimilation, the fixed-lag Kalman smoother (FLKS) provides a particularly attractive framework. The FLKS formulation is based fully on the underlying filtering strategy. Its formulation requires no error covariance information beyond what is required 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. The only difference between these

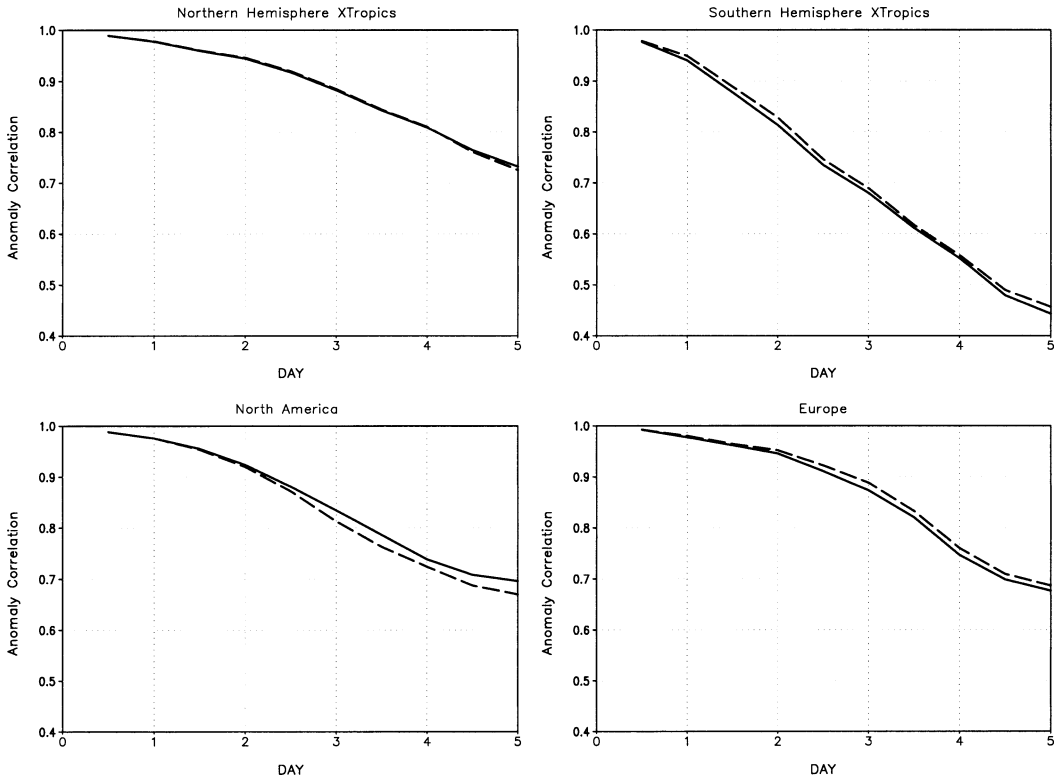

FIG. 14. Anomaly correlations for 500-hPa geopotential heights for CTL (solid curves) and RIA (dashed curves) experiments.

two procedures is that the lag-1 retrospective analysis of the latter is used to produce a revised filter analysis that is supposedly of better quality since it uses observation information from the future. Both of these procedures are implemented as an extension of the Goddard Earth Observing System version 3 (GEOS-3) Data Assimilation System (DAS). The GEOS-3 retrospective procedures implement smoothers that are optimal to the extent that the underlying filter strategy is optimal. In the GEOS-3 DAS the physical-space statistical analysis system (PSAS) is the underlying filter, which is suboptimal given the formulation of its error covariances; as a consequence, the GEOS-3 retrospective procedures are suboptimal as well. The new components required for implementing a retrospective capability in GEOS-3 DAS are the adjoint of the tangent linear model of the GEOS-3 general circulation model (GCM); the rearrangement of a few operators already available in the PSAS of GEOS-3 DAS; and the development of the tangent linear and adjoint operators responsible for transforming between model-space variables and analysis-space variables 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-h (lag 1) retrospective analysis were studied here. Although close examination of the OMA residuals seemed to suggest a rather neutral benefit from the lag-1 retrospective analysis, we saw improvement in the 6-h forecasts issued from these lag-1 retrospective analyses: these so-called retrospective forecasts are a closer match to the observations than the regular GEOS-3 DAS forecasts. This improved 6-h forecast skill motivated the investigation of the RIA scheme since this scheme makes explicit use of the retrospective forecasts. Evaluation of the analyses from the RIA procedure indicated them to be closer to the observations than the usual PSAS analyses. The OMA residuals for independent observations not used during the assimilation further confirmed some of the improvements due to the RIA scheme. More significant improvements were seen when examining climatologically important fields such as the mass streamfunction describing the meridional wind circulation. Finally, anomaly correlations and root-mean-square errors from a small sample of 5-day forecasts indicated a mild improvement in skill scores when analyses from the RIA procedure were used for the 5-day forecasts instead of the regular GEOS-3 analyses. Although the skill scores were not improved everywhere over the globe, they were improved generally.

Much study remains to be done to show that retrospective analysis is a worthwhile extension of PSAS to GEOS-3 DAS. Examples of such studies include the use of a high-resolution retrospective procedure and ex-

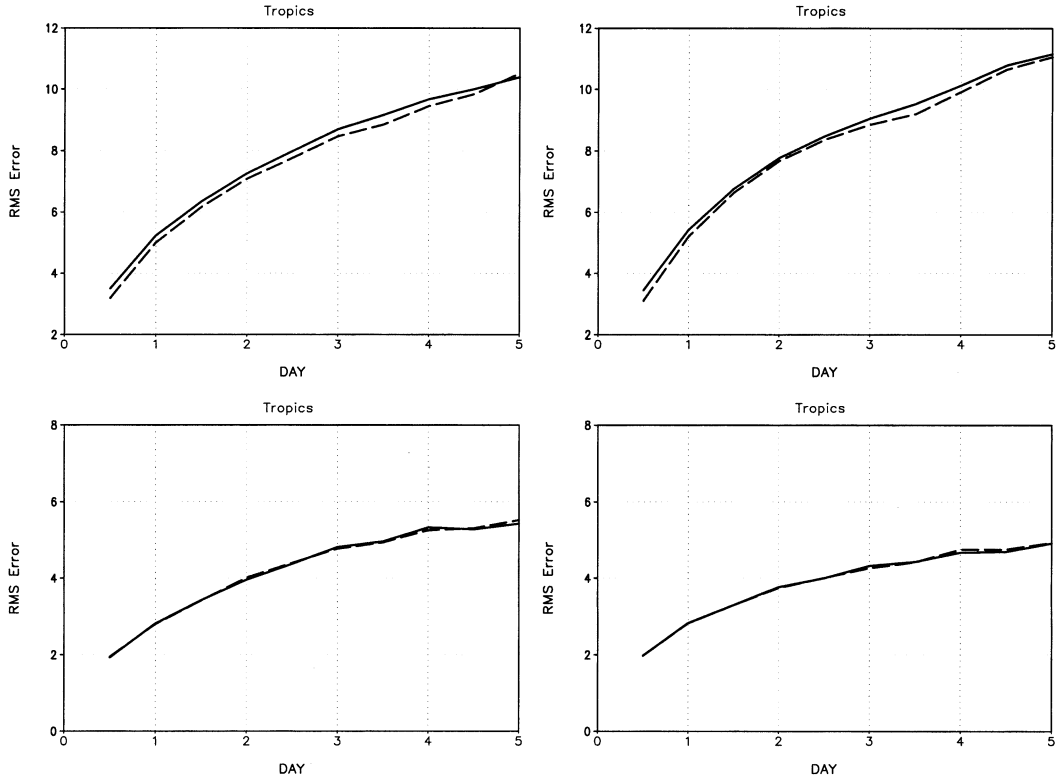

FIG. 15. The rms error from forecast skill study for the (left) zonal and (right) meridional components of the wind at (top) 200 and (bottom) 850 hPa. Solid curves are for CTL experiment; dashed curves are for RIA experiment. Units are in m s^{-1}.

amination of both the overall statistical behavior of the results as well as their quality with respect to representing synoptically relevant situations. However, in light of the newly operational GEOS-4 DAS, which replaces the GEOS-3 GCM, we plan to postpone these studies until the retrospective assimilation capability is brought to this new 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. Comments by the two anonymous reviewers helped to improve the final version of this manuscript. The numerical results in this work were obtained on an SGI Origin 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. Preliminary results of this investigation were presented at the Second International Symposium on Frontiers of Time Series Modeling: Nonparametric Approach to Knowledge Discovery, held in Nara, Japan.

APPENDIX

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 $l = 1, 2, \ldots, j$ we have

$$\mathbf{K}_{k-1|k} = \mathbf{P}^{a}_{k-1|k-1}\mathbf{A}^{\mathrm{T}}_{k,k-1}\mathbf{H}^{\mathrm{T}}_{k}\mathbf{\Gamma}^{-1}_{k} = \mathbf{P}^{f}_{k-1|k-2}[(\mathbf{I} - \mathbf{K}_{k-1|k-1}\mathbf{H}_{k-1})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}_{k,k-1}]\mathbf{H}^{\mathrm{T}}_{k}\mathbf{\Gamma}^{-1}_{k}, \tag{A1a}$$

$$\begin{aligned}\mathbf{K}_{k-2|k} &= (\mathbf{P}^{aa}_{k-1,k-2|k-1})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}_{k,k-1}\mathbf{H}^{\mathrm{T}}_{k}\mathbf{\Gamma}^{-1}_{k} = \mathbf{P}^{a}_{k-2|k-2}\mathbf{A}^{\mathrm{T}}_{k-1,k-2}[(\mathbf{I} - \mathbf{K}_{k-1|k-1}\mathbf{H}_{k-1})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}_{k,k-1}]\mathbf{H}^{\mathrm{T}}_{k}\mathbf{\Gamma}^{-1}_{k} \\ &= \mathbf{P}^{f}_{k-2|k-3}[(\mathbf{I} - \mathbf{K}_{k-2|k-2}\mathbf{H}_{k-2})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}_{k-1,k-2}][(\mathbf{I} - \mathbf{K}_{k-1|k-1}\mathbf{H}_{k-1})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}_{k,k-1}]\mathbf{H}^{\mathrm{T}}_{k}\mathbf{\Gamma}^{-1}_{k},\end{aligned} \tag{A1b}$$

$$\mathbf{K}_{k-j|k} = (\mathbf{P}^{aa}_{k-1,k-j|k-1})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}_{k,k-1}\mathbf{H}^{\mathrm{T}}_{k}\mathbf{\Gamma}^{-1}_{k}$$
$$= \mathbf{P}^{f}_{k-j|k-j-1}[(\mathbf{I} - \mathbf{K}_{k-j|k-j}\mathbf{H}_{k-j})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}_{k-j+1,k-j}] \cdots [(\mathbf{I} - \mathbf{K}_{k-1|k-1}\mathbf{H}_{k-1})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}_{k,k-1}]\mathbf{H}^{\mathrm{T}}_{k}\mathbf{\Gamma}^{-1}_{k}. \quad \text{(A1c)}$$

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

$$\mathbf{K}_{k-l|k} = \mathbf{P}^{f}_{k-l|k-l-1}\left[\prod_{j=k-l+1}^{k}(\mathbf{I} - \mathbf{H}^{\mathrm{T}}_{j-1}\mathbf{\Gamma}^{-1}_{j-1}\mathbf{H}_{j-1}\mathbf{P}^{f}_{j-1|j-2})\mathbf{A}^{\mathrm{T}}_{j,j-1}\right]\mathbf{H}^{\mathrm{T}}_{k}\mathbf{\Gamma}^{-1}_{k}, \quad \text{(A2)}$$

which shows that, as pointed out in the main text, the retrospective gains depend only on filter quantities. The equation above also indicates that the retrospective portion of the FLKS implicitly accounts for model error (Todling et al. 1998).

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. *Quart. J. Roy. Meteor. 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. Meteor. Soc. Japan,* **75** (1B), 257–288.

——, and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Japan,* **74,** 63–75.

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

——, 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. *Quart. J. Roy. Meteor. Soc.,* **120,** 1367–1387.

——, and Coauthors, 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. *Quart. J. Roy. Meteor. Soc.,* **124,** 1783–1808.

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

DAO, 1996: Algorithm theoretical basis document version 1.01. DAO, NASA Goddard Space Flight Center, Greenbelt, MD, 276 pp.

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 Note 96-2, NASA Goddard Space Flight Center, Greenbelt, MD, 66 pp.

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.

Gauthier, P., and J.-N. Thépaut, 2001: Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Wea. Rev.,* **129,** 2089–2102.

——, C. Charette, L. Fillion, P. Koclas, and S. Laroche, 1999: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Center. Part I: The global analysis. *Atmos.–Ocean,* **37,** 103–156.

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 Note 96-04, NASA Goddard Space Flight Center, Greenbelt, MD, 27 pp.

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 Note 98-05, NASA Goddard Space Flight Center, Greenbelt, MD, 189 pp.

Li, Z., and I. M. Navon, 2001: Optimality of 4D-Var and its relationship with Kalman filter and Kalman smoother. *Quart. J. Roy. Meteor. Soc.,* **127,** 661–684.

Lynch, P., and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.,* **120,** 1019–1034.

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.

Polavarapu, S., M. Tanguay, and L. Fillion, 2000: Four-dimensional variational data assimilation with digital filter initialization. *Mon. Wea. Rev.,* **128,** 2491–2510.

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.]

Thépaut, J.-N., D. Vasiljevic, P. Courtier, and J. Pailleux, 1993: Variational assimilation of conventional meteorological observations with a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.,* **119,** 153–186.

Todling, R., 2000: Retrospective data assimilation schemes: Fixed-lag smoothing. *Proc. Second Int. Symp. on Frontiers of Time Series Modeling: Nonparametric Approach to Knowledge Discovery*, Nara, Japan, Institute of Statistical Mathematics, 155–173.

——, and S. E. Cohn, 1996: Some strategies for Kalman filtering and smoothing. *Proc. ECMWF Seminar on Data Assimilation*, Reading, United Kingdom, ECMWF, 91–111.

——, ——, 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, 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.,* **102,** 8703–8718.

Zhu, Y., R. Todling, and S. E. Cohn, 1999: Technical remarks on smoother algorithms. DAO Note 99-02, NASA Goddard Space Flight Center, Greenbelt, MD, 45 pp.

Zou, X., I. M. Navon, and J. Sela, 1993a: Control of gravitational oscillations in variational data assimilation. *Mon. Wea. Rev.,* **121,** 272–289.

——, ——, M. Berger, M. K. Phua, T. Schlick, and F. X. LeDimet, 1993b: Numerical experience with limited-memory, quasi-Newton methods for large-scale unconstrained nonlinear minimization. *SIAM J. Optimization,* **3,** 582–608.

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