DAO Office Note 96-25

Office Note Series on
Global Modeling and Data Assimilation

Richard B. Rood, Head
Data Assimilation Office
Goddard Space Flight Center
Greenbelt, Maryland

A Simple Bias Correction Algorithm for use in Data Assimilation

Lawrence L. Takacs*

Data Assimilation Office, Goddard Laboratory for Atmospheres
** General Sciences Corporation,*
A Subsidiary of Science Applications International Corporation, Laurel, Maryland

Goddard Space Flight Center
Greenbelt, Maryland 20771
December 1996

Abstract

This report examines the effect of model bias within a data assimilation system. For this study we use the Goddard Earth Observing System (GEOS) Data Assimilation System (DAS) which employs the Incremental Analysis Updating (IAU) procedure. In such a procedure, the after-analysis minus first guess increments are used as external forcing within a general circulation model (GCM) to provide corrective adjustments for data assimilation. It is shown that for a GCM which contains a climatological model bias, the climatology of the *assimilation* will be biased from the climatology of the *after-analysis.* A modified IAU procedure is presented which effectively removes the bias from the assimilation. The procedure may also be modified for incorporation into traditional intermittent data assimilation approaches.

Contents

List of Figures

1 Introduction

The Data Assimilation Offfice (DAO) at NASA's Goddard Space Flight Center is currently producing a multi-year global atmospheric re-analysis using historic observations for the period 1980-1994 (see Schubert et al. 1993 and Schubert et al. 1995). This system incorporates Version 1 of the Goddard Earth Observing System (GEOS) general circulation model (GCM) and optimal interpolation (OI) analysis scheme. The GEOS GCM and GEOS Data Assimilation System (DAS) are part of NASA's Earth Observing System program whose primary objective is to produce research quality datasets from the satellite and surface measurements of the earth system that will be available at the turn of the century. In this report we examine the effect of model biases within the GEOS DAS.

Version 1 of the GEOS DAS has been used for shorter assimilations in support of the Coupled Ocean-Atmosphere Response Experiment (COARE), of subprojects of the Global Energy and Water Cycle Experiment (GEWEX), and of the Airborne Southern Hemisphere Ozone Experiment (ASHOE) mission. In addition to these data assimilation applications, the GEOS GCM is being used for climate simulations, such as the DAO's participation in the Atmospheric Model Intercomparison Project (AMIP) (Gates, 1992).

Because of the wide dissemination of the GEOS products, particularly those within data assimilation, it is important that the behavior of the atmospheric GCM and the analysis components of the GEOS DAS be thoroughly analyzed and documented. The behavior of physical processes in the GCM, such as moist convection, radiative heating, surface fluxes, and the hydrological cycle are the focus of Molod *et al.*(1996), while the dynamical aspects of the GCM's climate are discussed in Takacs and Suarez (1996). In both studies, an examination of model biases was emphasized due to their significant impact on the analysis through systematic errors in the model provided first guess. This information is necessary both for a critical evaluation of the assimilation products and to identify model deficiencies that need to be addressed in future versions of the system.

The Incremental Analysis Updating (IAU) procedure, described by Bloom *et al.*(1996), has had significant beneficial impact upon the analysis/assimilation technique. This technique allows for relatively small adjustments (ie. corrections) to be made continuously within a general circulation model to force a data assimilation system, and has been shown to produce improved error statistics when compared to the intermittent analysis approach. In addition the IAU technique acts like a high-frequency time filter, thereby minimizing adverse reactive effects caused by imbalances created from the analysis increments.

The analysis technique used by the DAO's re-analysis effort (see Pfaendtner *et al.*, 1995) assumes that the model provided first guess, as well as the observations, are unbiased with respect to nature. As such, the analysis increments should be composed of adjustments designed to address statistically random errors made by the forecasting system. As pointed out by Daley (1991), however, a misspecification of background error (in this case the assumption of an unbiased first guess) will produce an analysis which is sub-optimal. As shown in Molod *et al.*(1996), the model biases from the GEOS-1 DAS had significant impact on the assimilated products, particulary those quantities or diagnostics not directly assimilated. An examination of the time-averaged analysis increments reveal the systematic nature of the analysis corrections being made, and can be used to assess deficiencies in the various forcing terms within the model (see Schubert and Chang, 1996). Figure 1a shows the climatological January zonal mean specific humidity analysis increment from the GEOS-1 DAS assimilation. The analysis increment shown is the actual forcing used within the GCM during the assimilation (at model grid-points interpolated to pressure surfaces). We can see that there exists a systematic moistening in the lower levels of the atmosphere, maximized at 850 mb, and a systematic drying aloft. These increments are consistent with that obtained from the January mean observation minus first guess (OMF) mixing ratio statistics averaged over the tropics at the observation points (not shown). Figure 1b shows the January mean specific humidity analysis increment at 850 mb. We can see the extensive areas where the model was systematically dry. These systematic errors revealed by the time-mean analysis increments significantly impact the assimlated products themselves (eg. specific humidity) as well as added value products such as precipitation and diabatic heating.

Preliminary efforts have been made to address model bias both during forecasts and data assimilation. Saha (1992) used a constant forcing term (estimated from an ensemble of 1-day forecast errors) to "correct" model similations, and has shown significant improvement in 5-day forecast skill. Da Silva *et al.*(1996) have formulated an on-line estimation and correction of forecast bias based on existing operational statistical analysis methods, and have examined their effect in simple linear modeling tests.

Section 2 contains an overview of the IAU procedure used in Version 1 of the GEOS-DAS, and its impact on the long-term climatology of the assimilation. Section 3 develops a modified IAU procedure which eliminates long-term errors within the assimilation due to model bias. Section 4 presents results using a simple linear model, while Section 5 examines the new IAU procedure within the full GEOS-DAS system. A summary is presented in Section 6.

2 An examination of the GEOS-1 IAU Procedure

We begin by reviewing the IAU procedure which has been used for the multi-year GEOS-1 DAS re-analysis. This examination is performed using a simple linearization of the data assimilation paradigm. The IAU assimilation cycle is a 6-hour period in which the GCM and analysis schemes are employed to produce the assimilated products. Defining t_a as the synoptic time of the analysis (0,6,12,18z), the 6-hour IAU assimilation cycle can be expressed symbolically between $(t_a - 3hr)$ and $(t_a + 3hr)$ as

$$GCM_{fg}(t_a) = GCM_{das}(t_a - 3hr) + \left(\frac{\partial g}{\partial t}\right) 3hr \tag{1}$$

$$\begin{aligned} AMF(t_a) &= ANA(t_a) - GCM_{fg}(t_a) \\ IAU &= AMF(t_a)/\tau_{iau} \end{aligned}$$

$$GCM_{das}(t_a - 3hr + \Delta t) = GCM_{das}(t_a - 3hr) + \left[\left(\frac{\partial g}{\partial t}\right) + IAU\right] \Delta t \tag{2}$$

where $GCM_{das}(t_a - 3hr)$ is the assimilation restart 3 hours prior to the synoptic analysis time, and $\left(\frac{\partial g}{\partial t}\right)$ is the unforced GCM model tendency. In this algorithm, $ANA(t_a)$ is the state computed from the analysis scheme which optimally combines the model first guess with observational data. The timescale associated with the IAU forcing is given as τ_{iau}. From (1) and (2) we see that a first guess is provided to the analysis by simply running the GCM in simulation (unforced) mode for 3 hours. The model first guess is then subtracted from the after-analysis state to provide the analysis minus first guess increment (AMF). Dividing the AMF by the IAU timescale τ_{iau} defines the IAU tendency which is then used as additional forcing to the GCM during the assimilation cycle. The assimilation system is backed up to the original restart 3 hours prior to the analysis time, and subsequently run in forced mode for 6 hours producing the assimilation product.

For this derivation we define the model assimilated state, $GCM_{das}(t_a - 3hr + \Delta t)$, as being composed of a desired "true" assimilated state, GCM_{tru}, plus an error, GCM_{err}:

$$\begin{aligned} GCM_{das}(t_a - 3hr + \Delta t) &= GCM_{tru}(t_a - 3hr + \Delta t) \\ &+ GCM_{err}(t_a - 3hr + \Delta t). \end{aligned} \tag{3}$$

Further, we define the model tendency as being composed of a desired "true" assimilation tendency plus two error terms associated with the model's bias and random error processes:

$$\left(\frac{\partial g}{\partial t}\right) = \left(\frac{\partial g}{\partial t}\right)_{tru} + \left(\frac{\partial g}{\partial t}\right)_{bias} + \left(\frac{\partial g}{\partial t}\right)_{ran}. \tag{4}$$

Of course, the "true" assimilated state and the "true" assimilation tendency are, in general, unkown but are symbolically used here to gain an assessment of assimilation errors. From these definitions, the time evolution of the *true* assimilated state is given by

$$GCM_{tru}(t_a - 3hr + \Delta t) = GCM_{tru}(t_a - 3hr) + \Delta t \left(\frac{\partial g}{\partial t}\right)_{tru}. \tag{5}$$

Finally, we define the *true* assimilated state at the synoptic analysis time t_a as being that produced from the analysis scheme:

$$GCM_{tru}(t_a) \equiv ANA(t_a). \tag{6}$$

For an intermittent analysis approach (6) is trivally satisfied (ignoring initialization subsequent to the analysis) since the assimilated state at the analysis time is *defined* by the after-analysis state. For the IAU approach this is not necessarily true. We see, therefore, that we are simply using the after-analysis state provided by the analysis scheme as a defining goal of the true assimilated state, that is, one which reproduces the after-analysis state at the synoptic analysis time. The after-analysis state may further be defined as:

$$ANA(t_a) = K\,Obs(t_a) + (1-K)\,GCM_{fg}\,(t_a)\,, \tag{7}$$

where K is a *gain matrix* which takes into account the relative accuracies of the model first guess and the observations, Obs. Since K is not generally equal to one, we are not defining the *true* assimilated state as one which *a priori* matches the observations. It is useful to note here that if there are any biases in either $GCM_{fg}\,(t_a)$ or $Obs(t_a)$, the resulting analysis will also be biased to some degree. This is true for both the intermittent and IAU analsyis approaches.

For this derivation the tendency errors associated with model bias are modeled as

$$\left(\frac{\partial g}{\partial t}\right)_{bias} = \frac{g_c - g}{\tau_b}, \tag{8}$$

where g_c is the model climatology, g is the instantanious model state, and τ_b is the timescale associated with model bias. The time varying solution for such a system is given by

$$g(t) = g(t_0)e^{-t/\tau_b} + g_c\left[1 - e^{-t/\tau_b}\right] \tag{9}$$

$$g(t) = g(t_0) \qquad \text{for} \quad t = 0 \tag{10}$$

$$g(t) = g_c \qquad \text{for} \quad t \to \infty \tag{11}$$

As can be seen, as the model solution approaches its climatology the model's bias tendency tends toward zero. Using (3) through (8) in (1) and (2), we may write the errors in the assimilated state as:

$$\begin{aligned} GCM_{err}\,(t_a - 3hr + \Delta t) = GCM_{err}\,(t_a - 3hr)\left(1 - \frac{\Delta t}{\tau_{iau}}\right) \\ + \left(\frac{\partial g}{\partial t}\right)_{ran}\left(1 - \frac{3}{\tau_{iau}}\right)\Delta t \\ + \left[\frac{GCM_c - GCM_{tru}\,(t_a - 3hr) - GCM_{err}\,(t_a - 3hr)}{\tau_b}\right]\left(1 - \frac{3}{\tau_{iau}}\right)\Delta t. \end{aligned} \tag{12}$$

In (12) the "error" refers to the difference between the actual assimilated state and that defined by the desired "true" assimilated state, (5). At the end of the assimilation period ($\Delta t = 6hr$), the time ensemble mean of the error in the assimilated state becomes

$$\overline{GCM_{err}}(t_a + 3hr) = \left(\frac{GCM_c - \overline{ANA}}{\tau_b}\right)\left[\frac{\tau_b(\tau_{iau} - 3)}{\tau_{iau} - 3 + \tau_b}\right]. \tag{13}$$

Here, $\overline{(\ \ )}$ is defined as the long-term ensemble average. Notice in (13) that the time-averaged random error tendency has been eliminated. In obtaining (13) we have used the relations that $\overline{GCM_{err}}(t_a + 3hr) = \overline{GCM_{err}}(t_a - 3hr)$ and $\overline{GCM_{tru}} \equiv \overline{ANA}$. Combining (13) with (3) we obtain an expression for the climatology of the assimilated state:

$$\overline{GCM_{das}} = \frac{\tau_b \overline{ANA} + (\tau_{iau} - 3)GCM_c}{\tau_b + (\tau_{iau} - 3)}. \tag{14}$$

We see that the climatology of the assimilation is equal to a weighted average between the climatology of the after-analysis and the climatology of the GCM. Using (7) we may also express the climatology of the assimilated state as related to the climatology of the observations:

$$\overline{GCM_{das}} = \frac{\tau_b \overline{Obs} + (\tau_{iau}/K - 3)GCM_c}{\tau_b + (\tau_{iau}/K - 3)}. \tag{15}$$

We may summarize the affects of various parameters by the following table:

Parameter	Value	Climatology
τ_b	$\rightarrow \infty$ (no model bias)	$GCM_{das} = ANA$
K	$\rightarrow 0$	$GCM_{das} = GCM_c$
τ_{iau}	$\rightarrow \infty$ (no IAU correction)	$GCM_{das} = GCM_c$
τ_b	$\rightarrow 0$	$GCM_{das} = GCM_c$!

We find that for very large values of timescales associated with model bias, $\tau_b \rightarrow \infty$, there is not enough time between analysis synoptic periods for the model bias to significantly influence the assimilated state. Thus, the climatology of the assimilation is equal to that of the after-analysis. For a gain matrix K that highly favors the model first guess, or as τ_{iau} becomes large (effectively causing no corrective adjustment), we see that the climatology of the assimilation is equal to that of the model. This is not surprising since the model is effectively running in simulation mode. The most disturbing result from the above table, however, is that for model bias timescales which are very short, $\tau_b \rightarrow 0$, the climatology of the assimilation again is equal to that of the model! Therefore, regardless of the corrective action taken by the analysis, the climatology of the assimilated state will be biased.

The after-analysis minus first guess increment may be expressed as:

$$AMF = -\Bigg[GCM_{err}(t_a - 3hr) + 3\Bigg(\left(\frac{\partial g}{\partial t}\right)_{ran}$$

$$+ \quad \frac{GCM_c - GCM_{tru}(t_a - 3hr) - GCM_{err}(t_a - 3hr)}{\tau_b}\Bigg)\Bigg], \tag{16}$$

while the ensemble time-mean AMF and IAU are given by:

$$\overline{AMF} = -\left(\frac{GCM_c - \overline{ANA}}{\tau_b}\right)\left[\frac{\tau_{iau}\tau_b}{\tau_{iau} - 3 + \tau_b}\right] \tag{17}$$

$$= -\left(\frac{GCM_c - \overline{Obs}}{\tau_b}\right)\left[\frac{\tau_{iau}\tau_b}{\tau_{iau}/K - 3 + \tau_b}\right] \tag{18}$$

and

$$\overline{IAU} = -\left(\frac{GCM_c - \overline{ANA}}{\tau_{iau} - 3 + \tau_b}\right) \tag{19}$$

$$= -\left(\frac{GCM_c - \overline{Obs}}{\tau_{iau}/K - 3 + \tau_b}\right). \tag{20}$$

We see that the ensemble time-mean AMF will be zero (indicating statistically unbiased analysis adjustments) only if the model's climatology is equal that of the after-analysis. In general, for GCMs containing model bias, this will not be the case.

The mean square error at the end of the 6-hour assimilation cycle, as well as the mean square after-analysis minus first guess, may also be examined. Assuming, for simplicity, that the desired *true* assimilated state is given by the constant $ANA(t_a)$, it may be shown that:

$$\overline{GCM^2}_{err}(t_a + 3) = \begin{cases} \left[\left(\frac{\partial g}{\partial t}\right)^2_{ran} + \left(\frac{GCM_c - ANA}{\tau_b}\right)^2\right]\left(\frac{\tau_b(\tau_{iau}-3)}{\tau_{iau}-3+\tau_b}\right)^2 & \text{for} \quad \tau_{iau} \neq 3 \\ \overline{GCM^2}_{err}(t_a - 3) = \ldots = \overline{GCM^2}_{err}(t_0) & \text{for} \quad \tau_{iau} = 3 \end{cases} \tag{21}$$

and

$$\overline{AMF^2} = \left[\left(\frac{\partial g}{\partial t}\right)^2_{ran} + \left(\frac{GCM_c - ANA}{\tau_b}\right)^2\right]\left(\frac{\tau_{iau}\tau_b}{\tau_{iau} - 3 + \tau_b}\right)^2. \tag{22}$$

We see from (12) that, for $\tau_{iau} = 3$, the error at the end of the assimilation cycle ($\Delta t = 6hr$) is equal to the error at the beginning of the period but with the opposite sign. This produces the undesirable feature in (21) that the mean square error is proportional to the error in the initial condition. We also see that the mean square AMF contains contributions from both the model's random error processes as well as the model's bias. In summary, it has been shown that the IAU procedure used in the GEOS-1 DAS results in an assimilated product which is biased from the analyzed state produced from the anaysis scheme, and produces non-zero time-mean analysis minus first-guess increments.

3 IAU with Memory

In this section we develop a modified technique for the IAU process which removes the bias from the data assimilation system. Consider the modified IAU algorithm given by

$$
\begin{aligned}
GCM_{fg}\left(t_a\right) &= GCM_{das}\left(t_a - 3hr\right) + \left[\left(\frac{\partial g}{\partial t}\right) + IAU_{fg}\right] 3hr \\
AMF(t_a) &= ANA(t_a) - GCM_{fg}\left(t_a\right) \\
IAU &= AMF(t_a)/\tau_{iau} + IAU_{fg} \\
GCM\left(t_a - 3hr + \Delta t\right) &= GCM_{das}\left(t_a - 3hr\right) + \left[\left(\frac{\partial g}{\partial t}\right) + IAU\right] \Delta t
\end{aligned}
\tag{23}
$$

In this modification of the IAU algorithm, a "first guess" is used for the IAU increment in obtaining the GCM first guess for the analysis. The final, or actual, IAU forcing used in the assimilation is then computed as the original *first guess* increment plus the current after-analysis minus first guess correction. The form of IAU_{fg} will be derived using the constraint that the climatology of the assimilated state be equal to the climatology of the after-analysis. Note that for $IAU_{fg} = 0$, the modified scheme reduces to that of the original method.

Using the same procedure as in section 2, we may express the errors in the assimilated state from (23) as:

$$
\begin{aligned}
GCM_{err}\left(t_a - 3hr + \Delta t\right) = GCM_{err}\left(t_a - 3hr\right)\left(1 - \frac{\Delta t}{\tau_{iau}}\right) \\
+ \left(\frac{\partial g}{\partial t}\right)_{ran}\left(1 - \frac{3}{\tau_{iau}}\right)\Delta t \\
+ \left[\frac{GCM_c - GCM_{tru}\left(t_a - 3hr\right) - GCM_{err}\left(t_a - 3hr\right)}{\tau_b}\right]\left(1 - \frac{3}{\tau_{iau}}\right)\Delta t \\
+ IAU_{fg}\left(1 - \frac{3}{\tau_{iau}}\right)\Delta t.
\end{aligned}
\tag{24}
$$

Taking the ensemble time mean of (24) at the end of the assimilation cycle ($\Delta t = 6hr$), it may be shown that the mean assimilated state using the modified IAU algorithm is given by

$$
\overline{GCM_{err}}\left(t_a + 3hr\right) = \left(\frac{GCM_c - \overline{ANA}}{\tau_b} + IAU_{fg}\right)\left[\frac{\tau_b(\tau_{iau} - 3)}{\tau_{iau} - 3 + \tau_b}\right]. \tag{25}
$$

We see that we can remove the time-mean error in the assimilated state by using a *first guess* IAU increment defined by:

$$IAU_{fg} \equiv -\left(\frac{GCM_c - \overline{ANA}}{\tau_b}\right) \qquad \rightarrow \qquad \overline{GCM_{err}}\,(t_a + 3hr) = 0. \tag{26}$$

Using (23) and (24) we may show that the AMF is given by:

$$\begin{aligned} AMF &= -\Bigg[GCM_{err}\,(t_a - 3hr) + 3\Bigg(\left(\frac{\partial g}{\partial t}\right)_{ran} \\ &+ \frac{GCM_c - GCM_{tru}\,(t_a - 3hr) - GCM_{err}\,(t_a - 3hr)}{\tau_b} + IAU_{fg}\Bigg)\Bigg]. \end{aligned} \tag{27}$$

Using (26) it is easy to show that the ensemble mean IAU and AMF become

$$\overline{IAU} = -\left(\frac{GCM_c - \overline{ANA}}{\tau_b}\right), \tag{28}$$

$$\overline{AMF} = 0. \tag{29}$$

Equation (28) states that for the IAU with memory algorithm the time-mean analysis increment automatically becomes that required for use as IAU_{fg} to eliminate biases in the time-mean assimilated state. Notice that the time-mean analysis increment in this modified IAU algorithm is larger than that derived in section 2 for GEOS-1. This is due to the fact that, as mentioned in section 2, as a model reaches its climatological state the model bias tendency tends toward zero. The mean IAU increment is a measure of the model's bias tendency. For assimilated states which are near the model's climatology, the mean IAU increment will be small. For assimilated states which are far from the model's climatology, the mean IAU increment will be large. It should be stated again that for $GCM_c = \overline{ANA}$, this scheme reduces to the original form presented in Bloom *et al.*(1996).

As in section 2 the mean square error and the mean square after-analysis minus first guess may be examined at the end of the 6-hour assimilation cycle. Assuming, again for simplicity, that the desired *true* assimilated state is given by the constant $ANA(t_a)$, it may be shown that:

$$\overline{GCM^2}_{err}(t_a + 3) = \overline{\left(\frac{\partial g}{\partial t}\right)^2_{ran}} \left(\frac{\tau_b(\tau_{iau} - 3)}{\tau_{iau} - 3 + \tau_b}\right)^2 \qquad \text{for} \qquad \tau_{iau} \neq 3 \tag{30}$$

and

$$\overline{AMF^2} = \overline{\left(\frac{\partial g}{\partial t}\right)^2_{ran}} \left(\frac{\tau_{iau}\tau_b}{\tau_{iau} - 3 + \tau_b}\right)^2. \tag{31}$$

We see that, compared to the mean square error and mean square AMF obtained from the IAU technique used in GEOS-1 (eqs. 21 and 22), the contribution from the model bias has been eliminated. In summary we find that by using the IAU algorithm with memory, the ensemble mean IAU increment is a truer measure (ie., independent of τ_{iau} and the gain matrix K) of the model's climatological bias tendency while at the same time, the ensemble mean assimilated state and AMF are unbiased.

4 Simple Model

To test the impact of using IAU with memory, we consider a simple single-wave system defined by the following parameters:

$$\left(\frac{\partial g}{\partial t}\right)_{tru} = 2\omega\cos(\omega t), \tag{32}$$

$$\left(\frac{\partial g}{\partial t}\right)_{bias} = -\left(\frac{g - g_c}{\tau_b}\right), \tag{33}$$

$$\left(\frac{\partial g}{\partial t}\right)_{ran} = \pm\frac{1}{3} \quad hr^{-1} \quad \text{uniformly distributed}, \tag{34}$$

where

$$\begin{aligned} \omega &= 2\pi/(7 \quad \text{days}) \\ \tau_b &= 24 \quad \text{hours} \\ \tau_{iau} &= 6 \quad \text{hours} \\ g_0 &= 12 \\ g_c &= 4 \\ K &= \frac{1}{2}. \end{aligned} \tag{35}$$

Figure (2) shows a time series of nature as well as a straight model run. For these runs, a new random forcing was introduced every 6 simulated hours centered at the analysis time. We see that the time-mean of the nature run is centered about it's climatology of 12, while the model simulation is centered about 4. In figure (3) we see the impact of data assimilation. Both the traditional intermittent approach and the IAU technique are presented. For the intermittent approach (see fig.3b), the model is run in forecast mode (light solid line) for 6 hours starting from the current after-analysis state (open circles). This produces a first guess (crosses) for the analysis scheme at the end of the 6-hour forecast. The after-analysis state is computed as a weighted average between the first guess and the observations (dark solid line). For these examples, the weights (gain matrix K) are set equal to $\frac{1}{2}$. This after-analysis state then becomes the initial condition for the next 6-hour forecast. We see that through the use of intermittent data assimilation there exits an improved correlation between the assimilation and nature. However, we also see that due to the model's bias (which systematically pulls the 6-hour forecasts away from nature), a significant bias is produced between the assimilation and nature, fig.(3a). In addition, the intermittent assimilation approach also introduces spurious temporal high frequencies into the assimilation product. In figs.(3c,d), we see the impact of using the IAU approach. Starting from the mid-analysis times (3,9,15,21z), the model is run in forecast mode for 3 hours (crosses) producing a first guess for the analysis. The after-analysis state (open circes) is then computed as the weighted average between the first guess and the observations. An IAU tendency is then computed as equal to the after-analysis state minus the first guess, divided by 6 hours. This tendency is then used as additional forcing during the model integration, which is performed by backing up to the mid-analysis time and running for 6 hours creating the assimilation product. From fig.(3c) we see that the IAU approach also produces an assimilation which is highly correlated with the observations. However it, too, produces a significant bias with respect to nature. In fact, the bias is slightly worse than the intermittent approach since the after-analysis states are never fully realized. The significant advantage of the IAU technique is the removal of the spurious high frequencies introduced by intermittent data assimilation.

Figure (4) repeats these experiments using the bias correction algorithm described in section 3. For both the intermittent approach and the IAU technique, a 10-day running mean of the analysis tendency (after-analysis minus first guess divided by 6 hours) is computed and used as forcing during the model's first guess integration. For simplicity, the running mean is approximated as

$$\overline{IAU}^n = \alpha \overline{IAU}^{n-1} + (1-\alpha) IAU^n, \tag{36}$$

where

$$\alpha = \frac{4\mathrm{N}-1}{4\mathrm{N}},$$

N is the number of days used for the running mean, and n refers to a given assimilation cycle. During the first 10 days of the assimilations, the running mean was simply being generated. Beyond 10 days, the current value of the running mean is used to provide an *unbiased* first guess for the analysis. We see that for both techniques, the bias correction algorithm works exceedingly well. Even though the gain matrics has been held constant at $\frac{1}{2}$, the resulting assimilation closely parallels the observations due to the removal of the bias in the first guess.

Finally fig.(5) shows the time-series of the AMF and IAU tendency for both the control IAU and the IAU with Memory assimilation. We see that due to the model's significant bias, the AMF from the control IAU assimilation remains biased while that from the IAU with Memory assimilation tends toward a mean of zero. In addition, as predicted by (19) and (28), the magnitude of the mean IAU tendency increases with use of the bias correction algorithm. The calculated time-mean values of the assimilated state, AMF, and IAU for the Control IAU and IAU with memory are tabulated in Table 2:

Quantity	Control IAU	IAU with Memory
GCM_{das}	9.818	12.000
AMF	1.455	0.000
IAU	0.242	0.333

Although not evident from the examples shown here, it is interesting to note that for model bias timescales which are very short ($\sim$ 1-6 hrs.), a significant length of time is required (several months) to remove the bias from the assimilation product. We will return to this aspect of the algorithm in the next section.

5 GEOS-DAS using IAU with Memory

5.1 Model Description

The GEOS GCM used for this study is a development version based on that described by Takacs *et al.*(1994). This model uses the fourth-order version of the Aries/GEOS dynamical core described in Suarez and Takacs (1995). This core is a modular, Eulerian, finite-difference dynamics package used for many global modeling applications at Goddard. The equations are finite-differenced on an Arakawa C-grid in the horizontal and a Lorenz grid on a standard σ coordinate in the vertical (46-Levels, $p_{top} = 0.1mb$). The vertical distribution of the sigma levels is chosen so as to provide enhanced resolution in the planetary boundary layer and near the tropopause.

The physics package includes a full set of sub-grid parameterizations. Penetrative and shallow cumulus convection are parameterized using the Relaxed Arakawa-Schubert scheme of Moorthi and Suarez (1992), coupled with a Kessler-type scheme for the re-evaporation of falling rain (Sud and Molod, 1988). The thermal and solar radiation parameterizations are from Chou and Suarez (1994) and are accurate to 0.01 mb. Cloudiness is diagnosed using a simple scheme based on liquid water detrained during cumulus convection and large-scale condensation. Turbulent eddy fluxes of momentum, heat, and moisture in the surface layer are calculated using stability-dependent bulk formulas based on Monin-Obukhov similarity functions. Above the surface layer, turbulent fluxes of momentum, heat, and moisture are calculated by the Level 2.5 Mellor-Yamada type closure scheme of Helfand and Labraga (1988), which predicts turbulent kinetic energy and determines the eddy transfer coefficients used for a bulk formulation. The gravity wave drag scheme follows a parameterization described by Zhou *et al.*(1996).

5.2 Analysis Description

The analysis scheme used for this study is based on that used in Version 1 of the GEOS DAS (see Pfaendtner *et al.*(1995), Baker *et al.*(1987)). This method employs an optimal interpolation (OI) analysis scheme at 18 mandatory pressure levels, and the IAU procedure already discussed.

The OI is a three-dimensional (multivariate in z, u, v; univariate in mixing ratio), statistical objective analysis scheme employing damped cosine horizontal autocorrelation functions for model prediction error, and a multivariate oceanic surface analysis incorporating an Ekman balance for the sea-level pressure and winds. Analyzed heights are converted to temperatures increments through the hypsometric equation. Observational data for the surface analysis consists of surface land, ship, and buoy reports. The upper-air analysis incorporates data from rawinsondes, dropwindsondes, rocketsondes, aircraft winds, cloud tracked winds, and thicknesses from the Tiros Operational Vertical Sounder (TOVS).

The OI is performed every 6 hours using observations from a $\pm$ 3-hour data window centered on the analysis times (0000, 0600, 1200, and 1800 UTC). Following Bloom *et al.*(1996), an analysis tendency (OI "After-Analysis" minus GCM "First Guess" over 6-hours) is computed on the GCM sigma surfaces and used as a constant model forcing during the assimilation. With this method the model is not re-initialized at the analysis intervals, and the entire assimilation may be viewed as a continuous integration of the GCM in which the analysis increments act as another of the physical parameterizations.

5.3 Experiment Design and Results

To test the impact of using the IAU with memory algorithm within the GEOS DAS, two assimilations were run for a period of 4 months. For simplicity, the experiments were conducted using 4° x 5° horizontal resolution. The assimilations were started on December 4th, 1992. For the IAU with memory experiment, a 25-day running mean IAU tendency was computed following the technique described by (36). Using the simple linear model shown in section 4 with a prescribed seasonally dependent bias, it was found that 25 days was sufficient to capture the model bias without destroying its seasonal impact. After completion of the spin-up period, the mean IAU tendency was used as forcing during the first-guess phase of the assimilation cycle.

Figure (6a,b) shows the zonally averaged *analysis-first guess* height and specific humdity for the control experiment averaged from January through March, 1993. In this control run we see a significant negative height bias at upper levels in the tropics which is associated with excessive diabatic heating from convection and diagnosed clouds during the first guess forecast. There is also a significant positive bias of heights at mid-latitudes. The moisture bias in the control run is similar to that found in the GEOS-1 DAS where a significant dry bias appears in the first guess forecast at 850 mb, and a significant moist bias near the surface. The upper-level moist bias in the GEOS-1 DAS results has been reduced.

As demonstrated in section 4 the biases in the first guess should impact both the *after-analysis* state as well as the IAU driven *assimilated* state. Figure (6c,d) shows the control zonally averaged *analysis-assimilation* height and specific humdity. We see that the biases in the first guess have significantly impacted the assimation product of heights and specific humidity. While the *analysis-assimilation* differences are smaller than those in the *analysis-first guess* , the very strong signature of the model biases are still present.

Figure (7a,b) shows the zonally averaged *analysis-first guess* height and specific humdity for the IAU with memory experiment averaged from January through March,

1993. In this run we see a significant reduction in the zonally averaged AMF fields. These differences are further reduced for the zonally averaged *analysis-assimilation* (AMA) fields, fig (7c,d). We see that by using the bias correction term in the first-guess component of the assimilation cycle, the systematic difference between the *after-analysis* state and the *first guess* state has been substantially reduced. As a result the "climate", or monthly mean time average, of the *assimilated* state product is also much closer to the "climate" of the analyzed state.

Following the results from sections 2 and 3, the time mean IAU tendency should increase with use of the memory algorithm. Figure (8) depicts the zonally averaged IAU tendencies of temperature and specific humidity for the two assimilations. We see in both fields a factor of ~ 2 increase in the maximum value. This is consistant with the analysis in section 4 for bias timescales on the order of several hours.

Figure (9) shows the impact of using the IAU with memory algorithm at the level of maximum moisture bias, 850 mb, averaged from January through March, 1993. The left column depicts the control run, while the right shows the impact of IAU with memory. The top panels show the specific humidity *analysis-first guess* field derived from the pressurel level analysis. The bias from the control run is very similar to that shown for the GEOS-1 DAS assimilation, fig. (1). The *analysis-first guess* field for the IAU with memory experiment, however, shows little evidence of systematic bias in the first guess forecast. The middle panels show the actual IAU tendency (scaled to 6 hours, interpolated to pressure) used to force the data assimilation system. For the control run, this tendency is practically identical (to within vertical interpolation errors) to the *analysis-first guess* increment produced from the pressure level analysis. However, for the IAU with memory experiment we see a significant increase in the magnitude of the IAU forcing as predicted in sections 2 and 3. The lower panels show the *analysis-assimilation* fields from the DAS products. We see that the control assimilation contains a significant portion of the moisture bias signature evident in the first guess forecast, while that from the IAU with memory experiment shows no bias when compared to the *analysis* state.

Figure (10) repeats this same examination but for the analyzed heights at 150 mb and 250 mb temperature increments. Similar to the moisture analysis, there is a considerable increase in the magnitude of the IAU temperature forcing in conjunction with IAU with memory. This leads to a significant reduction in the systematic bias in the *analysis-first guess* and *analysis-assimilation* height fields.

Figures (11) and (12) depict time-series of the *analysis-first guess* fields and their corresponding IAU tendencies for a location of maximum bias ($120° longitude$, $-2° latitude$). We see that during the spin-up period, the two assimilations are identical. After the spin-up period is completed and the mean IAU tendency is used in the first guess component of the assimilation we see that the *analysis-first guess* difference tends toward zero. In addition, the IAU tendencies begin to grow in magnitude. Even after $3\frac{1}{2}$ months the IAU tendency has not yet reach a level of saturation. As indicated in section 4, this is indicative of biases with very short timescales.

To assess the impact of the IAU with memory algorithm on the actual analyses and final assimilated states, we next compare the final products of the height and moisture fields between the two runs. Figure (13a,b) depicts the differences between the two *after-analysis* states, while fig.(13c,d) show the differences between the two *assimilated* states. We see that the two *after analysis* states for heights are similar in northern mid-latidues between the two runs. This implies that the *after-analysis* state in this region is a realiable assessment of the true atmosphere and is largely determined by observational data. The difference between the two *assimilated* height states, however, is quite significant and has the character of the model bias. Since we have shown from figs.(6) and (7) that that the "climate" from the IAU with memory assimilation is much closer to the "climate" of the *after-analysis*, the IAU with memory algorithm provides a distinct improvement to depicting the actual state of the atmosphere.

The situation is quite different for heights in the southern hemisphere and for moisture in general. We see in fig.(13a,b) that there exists significant differences between the *after-analysis* states between the two runs, especially for moisture. The moisture difference is actually more than the difference in the control *analysis-first guess* (note the change in contour level). This implies that in these data poor regions where the model provided first guess plays a more important role, the model bias more strongly influences the determination of the *after-analysis* state. By greatly reducing the systematic bias from the first guess, the results from the IAU with memory algorithm show that the after-analysis itself has shifted away from the model's dry climatology toward the wetter observations. Since the only data source used for the moisture analysis are rawindsonde measurements, it is informative to examine the *observations-first guess* statistics between the two runs for this field. Figure (14) shows the OMF statistics for mixing ratio averaged in the tropics from January through March, 1993. We see that similar to the AMF shown in figs.(6) and (7) there exists a considerable decrease in the mean OMF error when using the IAU with memory algorithm. A smaller decrease is also evident in the OMF standard deviation. By providing a first guess which is less biased, the analysis scheme automatically draws closer to the observations. As a result, the assimilation produced using this wetter analysis is also significantly wetter than the corresponding assimilation using the control IAU, fig.(13c,d).

6 Summary

In this study we have examined the impact of model bias on the GEOS data assimilation system. Through the use of a linear analysis of the data assimilation paradigm, it was shown that the IAU technique used in Version 1 of the GEOS DAS leaves the assimilated state biased from the climatology of the after-analysis. This bias may be effectively removed by introducing a systematic correction during the first guess phase of the assimilation, in conjunction with a slightly modified form of the IAU forcing. Using simple linear models, the technique worked equally well within the traditional intermittent and IAU assimilation approaches. The systematic correction was derived to be simply equal to the time ensemble mean analysis increment.

When examined within the full GEOS DAS system, it was shown that the control IAU technique clearly left the assimilated states biased from the climatology of the after-analysis. These biases were significantly reduced when the IAU with memory algorithm was applied. In regions where the model bias was strongest, the actual after-analysis state was also significantly affected.

The mean IAU increment used as forcing within the first guess phase of the assimilation is computed from a 25-day running mean from previous analysis periods. The system is adaptive in the sense that as the model climatology improves, and model bias diminishes, the technique adjusts to the orignal form developed by Bloom *et al.*(1996). While it has been shown that the mean IAU forcing is an effective measure to remove model bias, it is also clear that this estimate of model bias will be contaminated by biases in the observing system. For future work, da Silva *et al.*(1996) propose to improve this estimate by using time histories of *observations -first guess* increments from a subset of the observing system for which bias is negligible compared to that of the forecast model.

References

Baker, W.E., S.C. Bloom, J.S. Woollen, M.S. Nester, E. Brin, T.W. Schlatter, and G.W. Branstator, 1987: Experiments with a Three-Dimensional Statistical Objective Analysis Scheme Using FGGE Data. *Mon. Wea. Rev.*, **115**, 272-296.

Bloom, S.C., L.L. Takacs, A.M. DaSilva, and D. Ledvina, 1996: Data Assimilation Using Incremental Analysis Updates. *Mon. Wea. Rev.*, **124**, 1256-1271.

Chou, M.-D. and M.J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Technical Memorandum 104606 Volume 3, Goddard Space Flight Center, Greenbelt, MD 20771.

Daley, R., 1991: Atmospheric Data Analysis. Cambridge University Press, pp.457.

da Silva, A., D. Dee, and L. Takacs, 1996: On-line estimation and correction of forecast bias for data assimilation. DAO Office Note 96-11, Goddard Space Flight Center, Greenbelt, MD 20771.

Gates, W. Lawrence, 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Am Met. Soc.*, **73**, 1962-1970

Helfand, H. M., and J. C. Labraga, 1988: Design of a non-singular level 2.5 second-order closure model for the prediction of atmospheric turbulence. *J. Atmos. Sci.*, **45**, 113-132.

Pfaendtner, J., S. Bloom, D. Lamich, M. Seablom, M. Sienkiewicz, J. Stobie, and A. da Silva, 1995: Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System - Version 1, NASA Technical Memorandum 104606 Volume 4, Goddard Space Flight Center, Greenbelt, MD 20771, 58 pp.

Molod, A., Helfand, H. M., and L. L. Takacs, 1996: The Climatology of Parameterized Physical Processes in the GEOS-1 GCM and their Impact on the GEOS-1 Data Assimilation System. *J. Climate*, **9**, 764-785.

Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa Schubert: A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, **120**, 978-1002.

Pfaendtner, J., S. Bloom, D. Lamich, M. Seablom, M. Sienkiewicz, J. Stobie, and A. da Silva, 1995: Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System - Version 1, NASA Technical Memorandum 104606 Volume 4, Goddard Space Flight Center, Greenbelt, MD 20771, 58 pp.

Saha, S. 1992: Response of the NMC MRF Model to systematic-error correction within integration. *Mon.Wea.Rev.*, **120**, 345-360.

Schubert, S. D., and Y. Chang, 1996: An objective method for inferring sources of model error. *Mon.Wea.Rev.*, **124**, 325-340.

Schubert, S.D., C.K. Park, C.Y. Wu, W. Higgins, Y. Kondratyeva, A. Molod, L. Takacs, M. Seablom, and R. Rood, 1995: A Multiyear Assimilation with the GEOS-1 System: Overview and Results, NASA Technical Memorandum 104606 Volume 6, Goddard Space Flight Center, Greenbelt, MD 20771, 207 pp.

Schubert, S. D., J. Pfaendtner and R. Rood, 1993: An assimilated data set for Earth Science applications. *Bull. Am Met. Soc.*, **74**, 2331-2342.

Suarez, M. J., and L. L. Takacs, 1995: Documentation of the Aries/GEOS Dynamical Core Version 2, NASA Technical Memorandum 104606 Volume 5, Goddard Space Flight Center, Greenbelt, MD 20771, 58 pp.

Sud, Y. C., and A. Molod, 1988: The roles of dry convection, cloud-radiation feedback processes and the influence of recent improvements in the parameterization of convection in the GLA GCM. *Mon. Wea. Rev.*, **116**, 2366-2387.

Takacs, L. L. , A. Molod, and T. Wang, 1994: Documentation of the Goddard Earth Observing System (GEOS) General Circulation Model-Version 1. NASA Technical Memorandum 104606 Volume 1, Goddard Space Flight Center, Greenbelt, MD 20771, 97 pp.

Takacs, L. L. and M.J. Suarez, 1996: Dynamical aspects of climate simulations using the GEOS General Circulation Model. NASA Technical Memorandum 104606 Volume 10, Goddard Space Flight Center, Greenbelt, MD 20771, 70 pp.

Zhou, J., Y.C. Sud, and K.-M. Lau, 1996: Impact of Orographically Induced Gravity Wave Drag in the GLA GCM, *Quart. J. Roy. Meteor. Soc.*, **122**, 903-927.

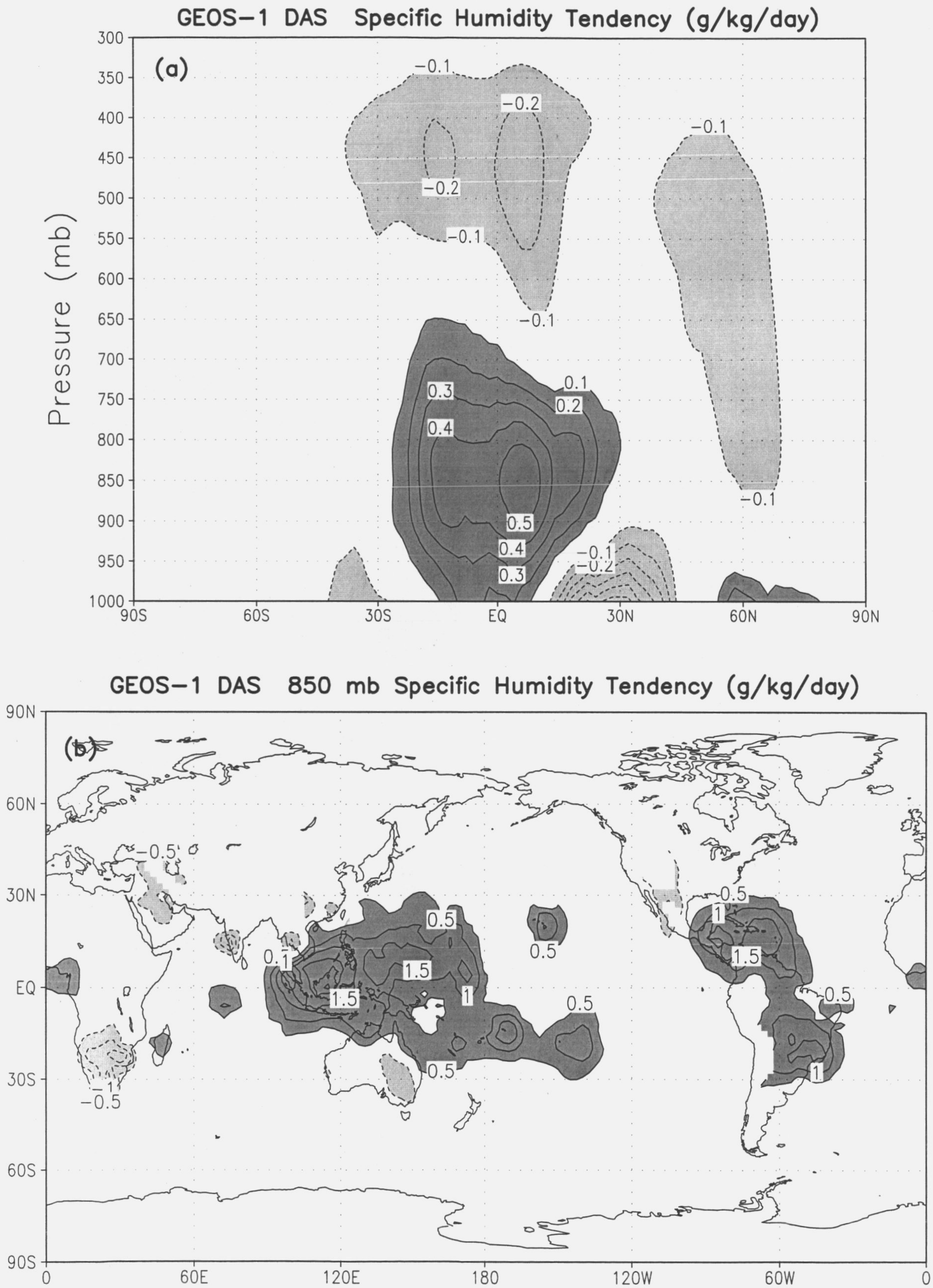
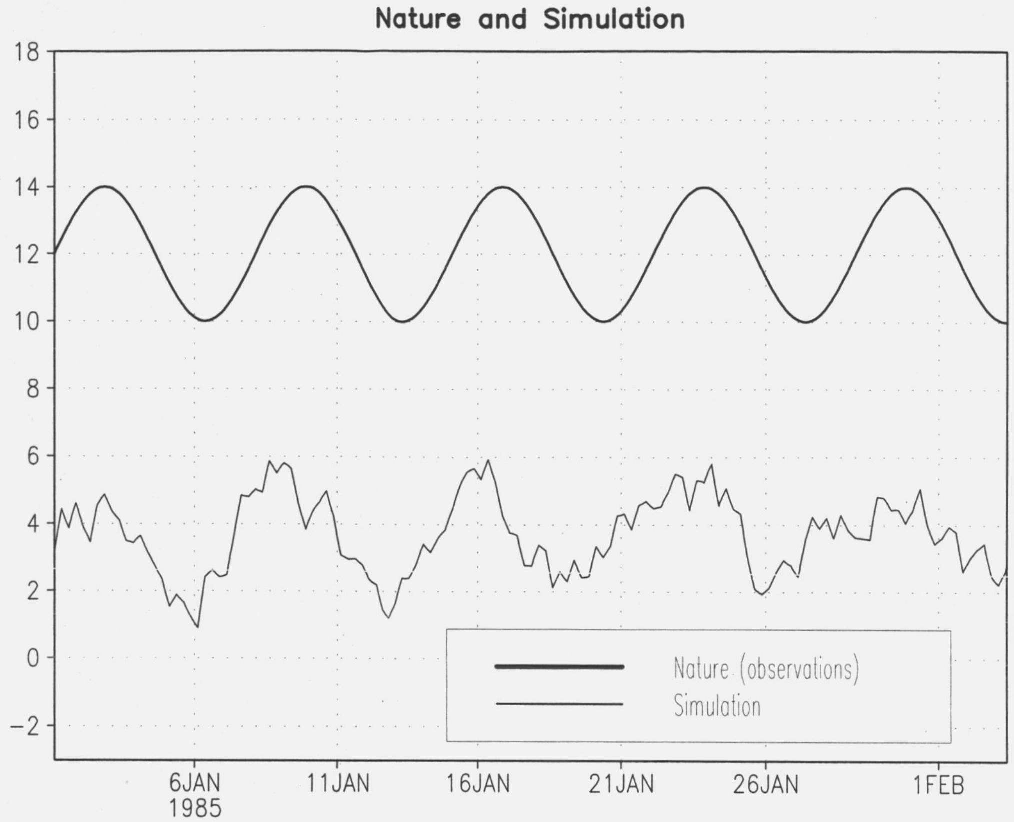

GEOS−1 DAS Specific Humidity Tendency (g/kg/day)
(a)
Pressure (mb)
300
350
400
450
500
550
600
650
700
750
800
850
900
950
1000
90S
60S
30S
EQ
30N
60N
90N
−0.1
−0.2
−0.2
−0.1
−0.1
−0.1
−0.1
0.1
0.2
0.3
0.4
0.5
0.4
0.3
−0.1
−0.2
GEOS−1 DAS 850 mb Specific Humidity Tendency (g/kg/day)
(b)
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0
−0.5
0.5
0.1
1.5
1.5
1
0.5
0.5
0.5
0.5
1
1.5
−0.5
1
−1
−0.5

Figure 1:



Nature and Simulation
18
16
14
12
10
8
6
4
2
0
−2
6JAN
1985
11JAN
16JAN
21JAN
26JAN
1FEB
Nature (observations)
Simulation

Figure 2:

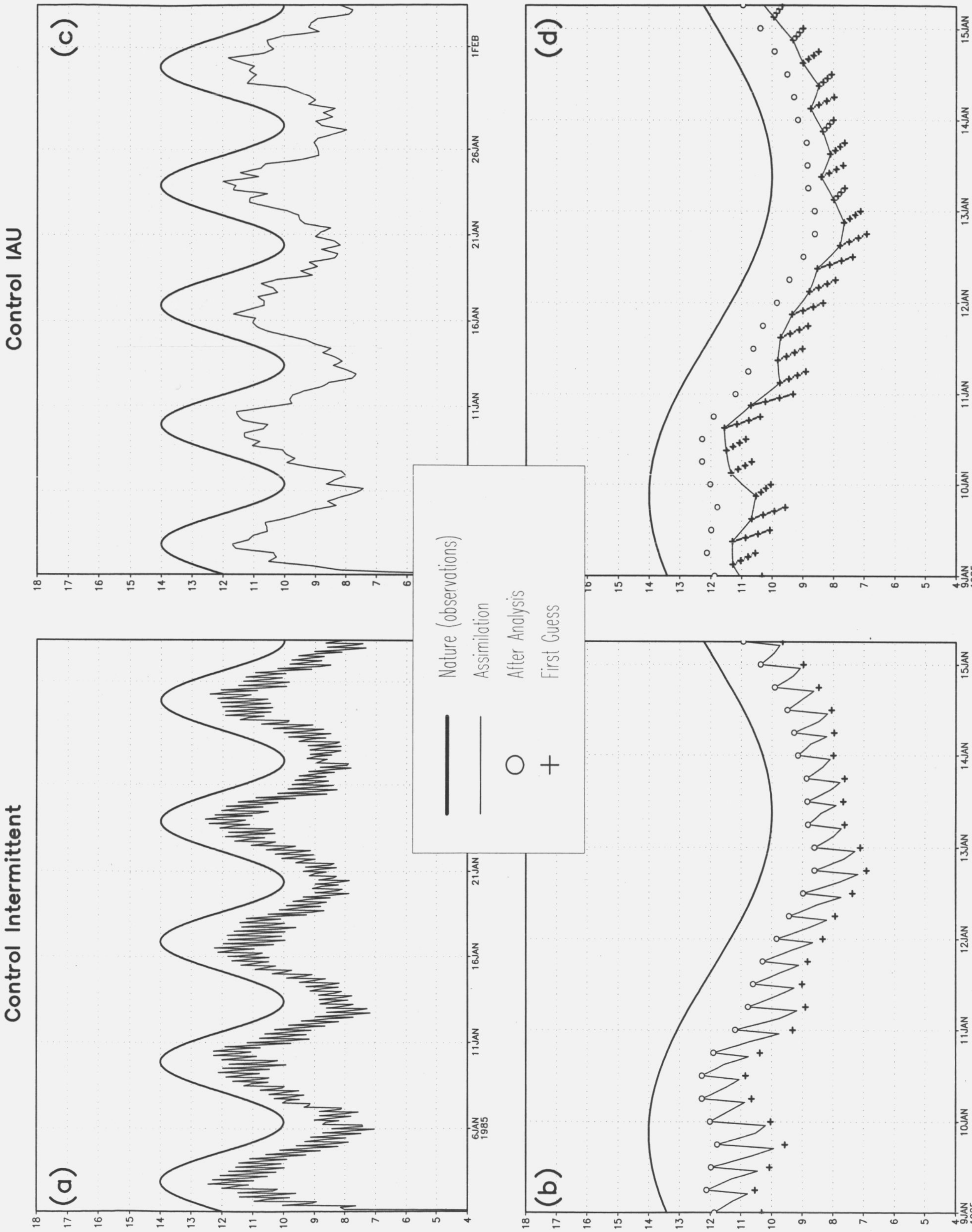

Control Intermittent
Control IAU
(a)
(b)
(c)
(d)
Nature (observations)
Assimilation
After Analysis
First Guess

Figure 3:

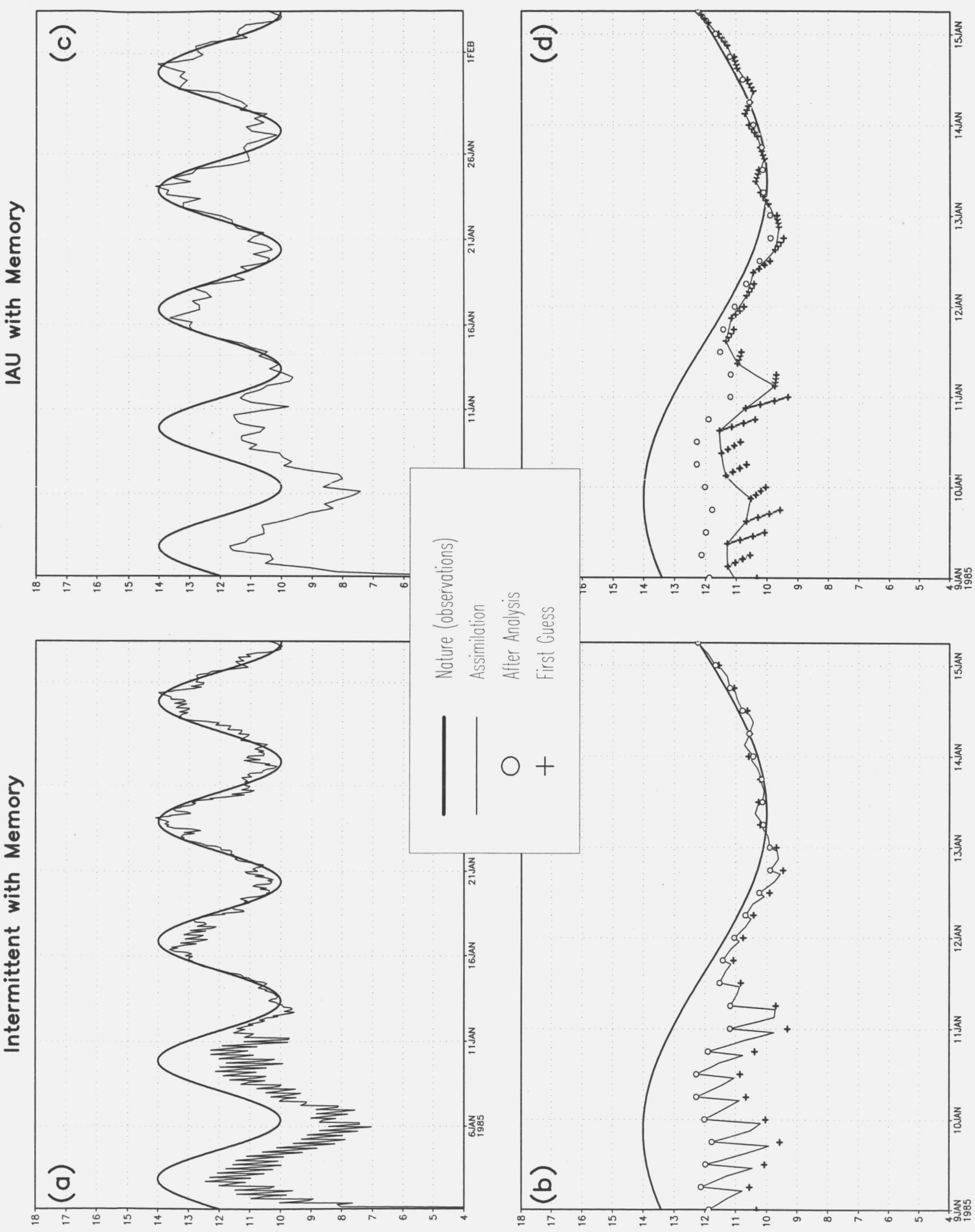
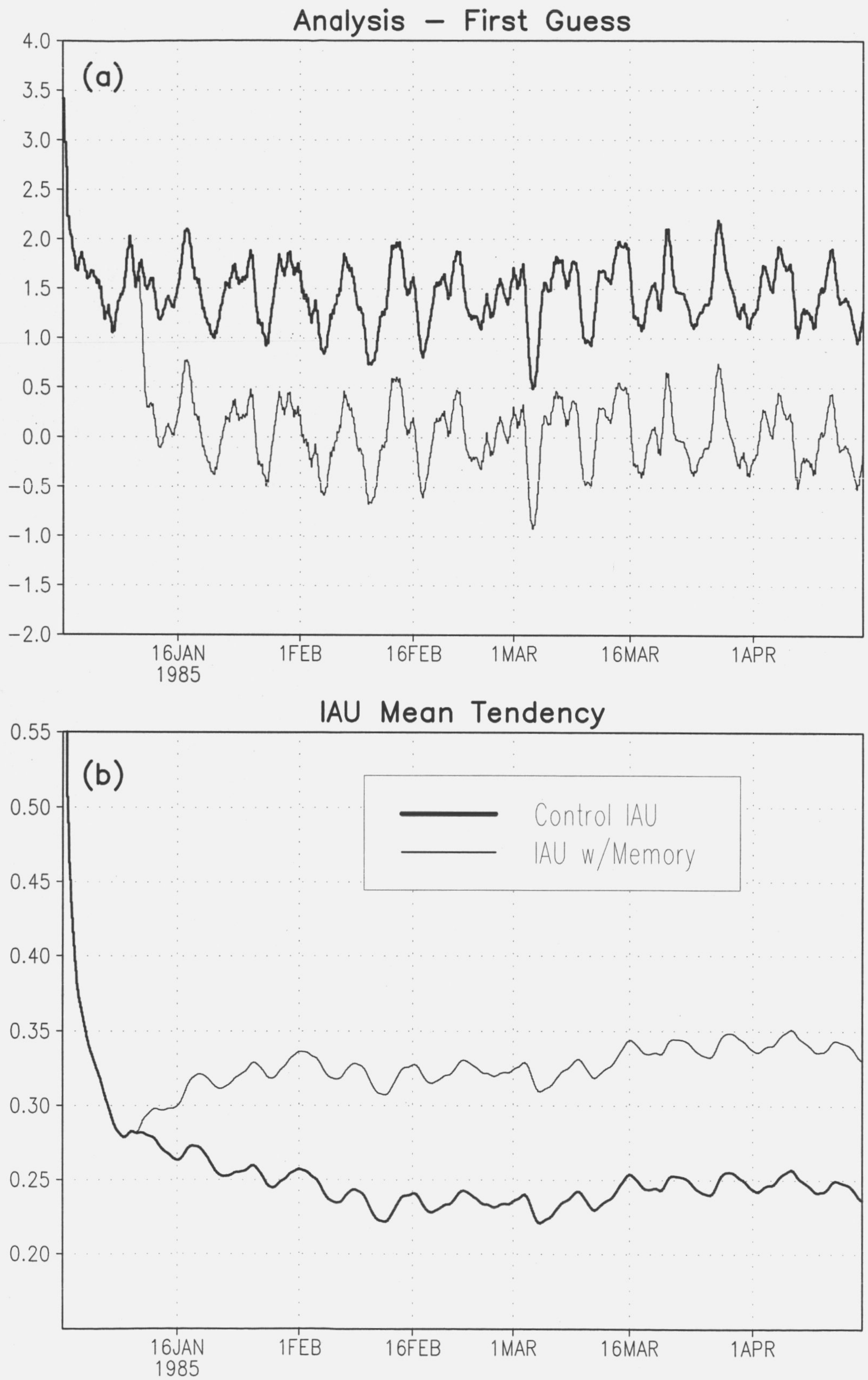

Intermittent with Memory
IAU with Memory
(a)
(b)
(c)
(d)
Nature (observations)
Assimilation
After Analysis
First Guess

Figure 4:



Analysis – First Guess
(a)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
16JAN 1985
1FEB
16FEB
1MAR
16MAR
1APR
IAU Mean Tendency
(b)
Control IAU
IAU w/Memory
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
16JAN 1985
1FEB
16FEB
1MAR
16MAR
1APR

Figure 5:

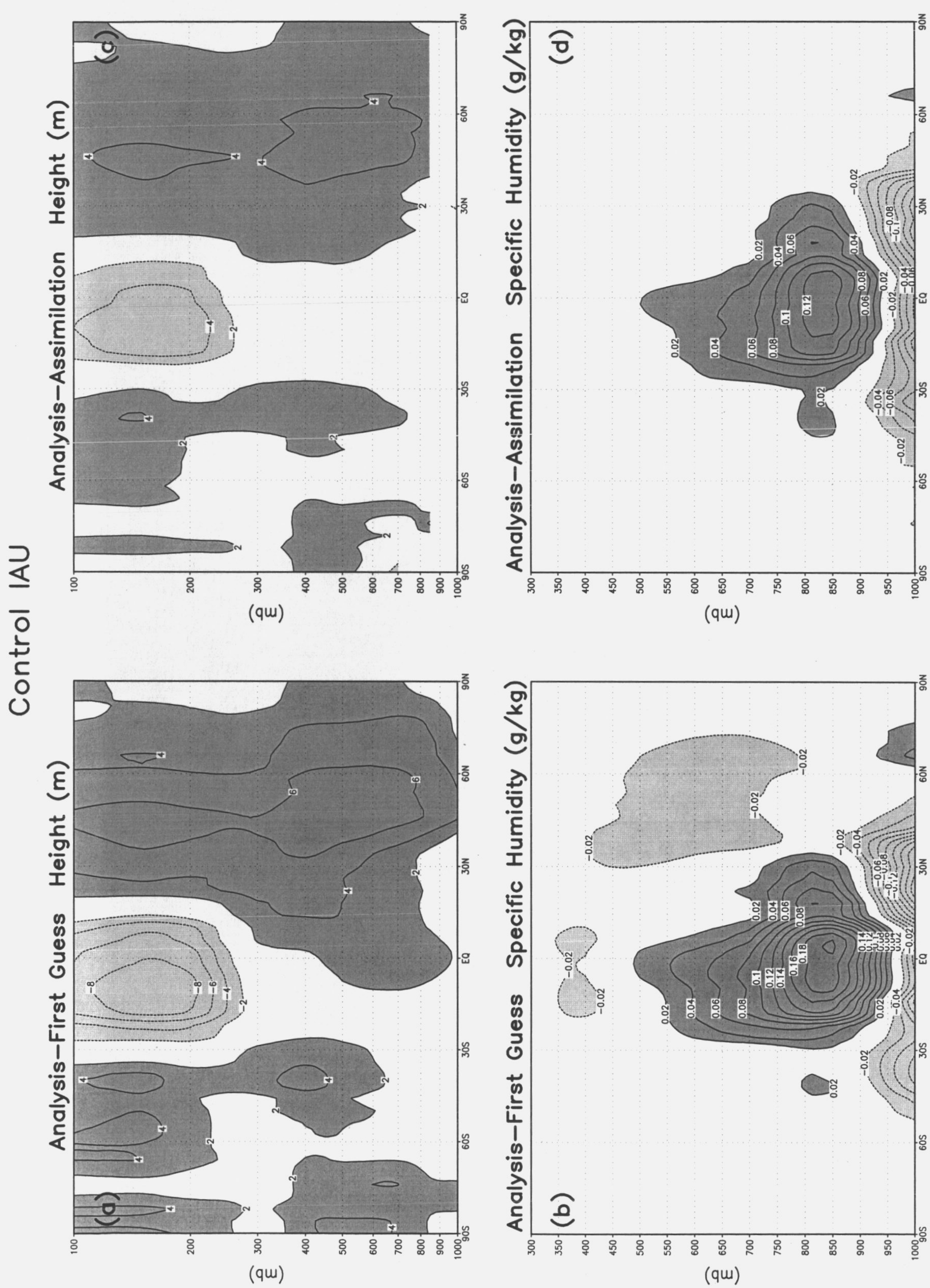
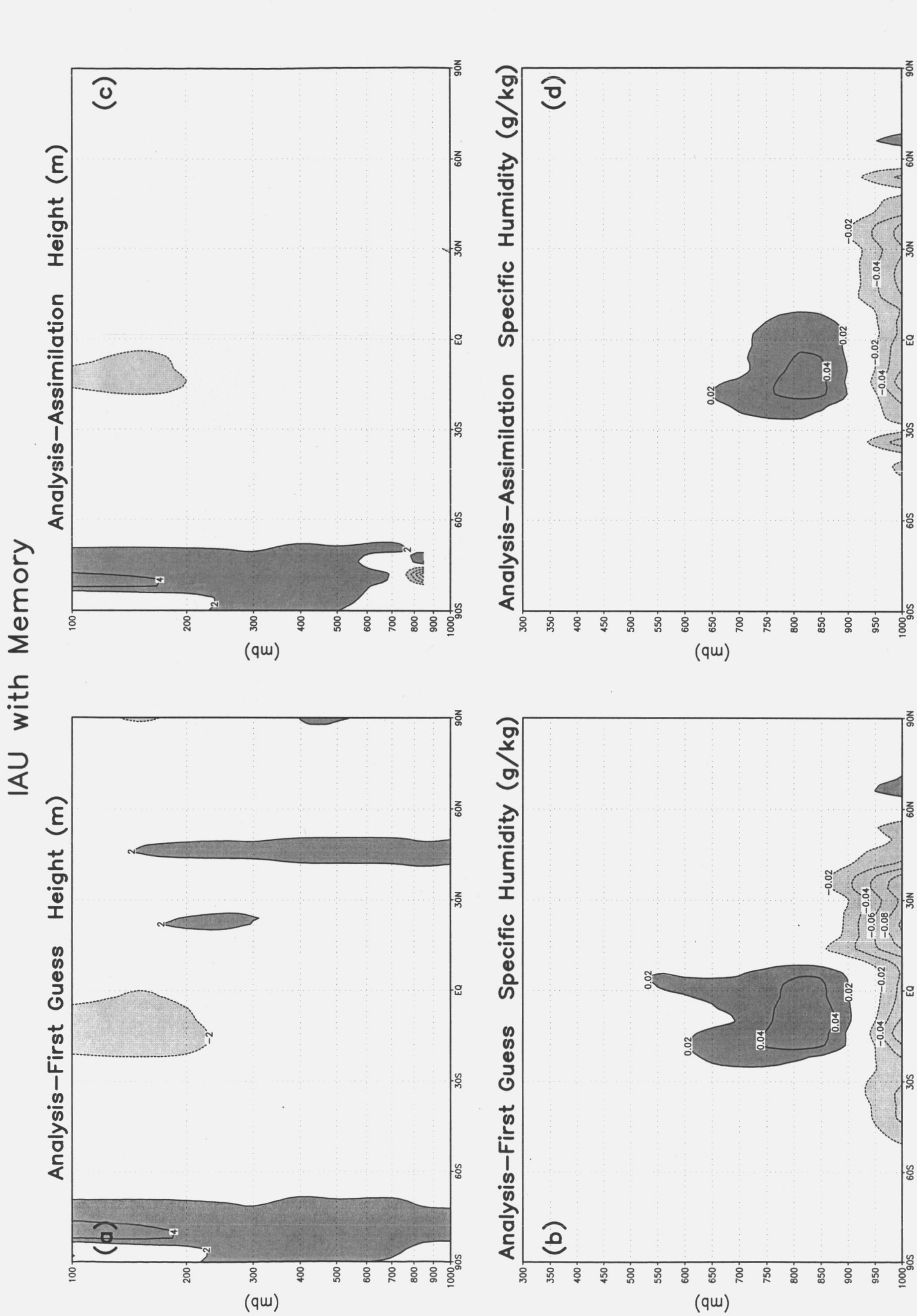
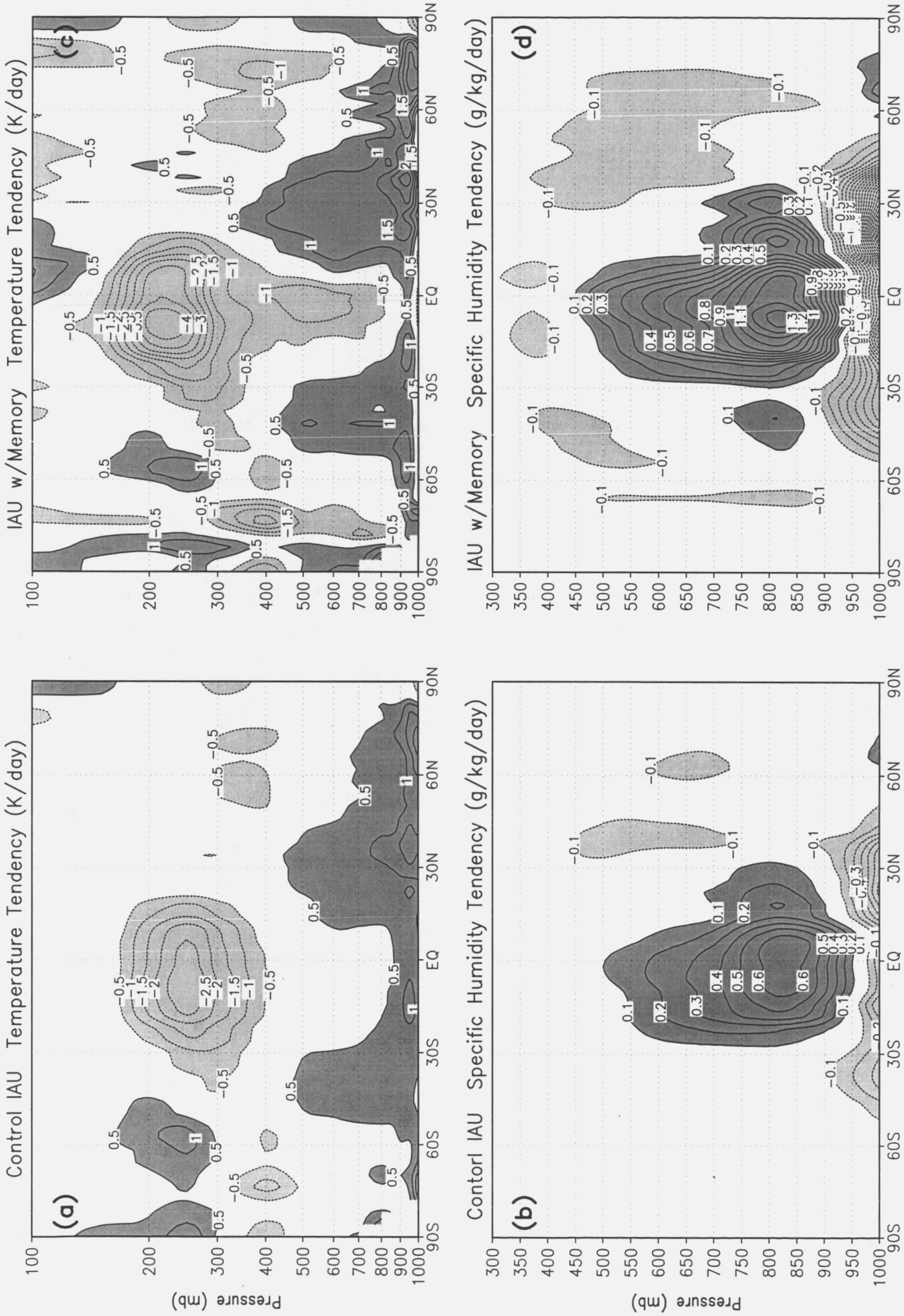

Control IAU
Analysis–First Guess Height (m)
(a)
Analysis–Assimilation Height (m)
(c)
Analysis–First Guess Specific Humidity (g/kg)
(b)
Analysis–Assimilation Specific Humidity (g/kg)
(d)
(mb)

Figure 6:



IAU with Memory
Analysis–First Guess Height (m)
(a)
Analysis–Assimilation Height (m)
(c)
Analysis–First Guess Specific Humidity (g/kg)
(b)
Analysis–Assimilation Specific Humidity (g/kg)
(d)
(mb)
90S
60S
30S
EQ
30N
60N
90N

Figure 7:



(a)
Control IAU Temperature Tendency (K/day)
(b)
Contorl IAU Specific Humidity Tendency (g/kg/day)
(c)
IAU w/Memory Temperature Tendency (K/day)
(d)
IAU w/Memory Specific Humidity Tendency (g/kg/day)
Pressure (mb)

Figure 8:

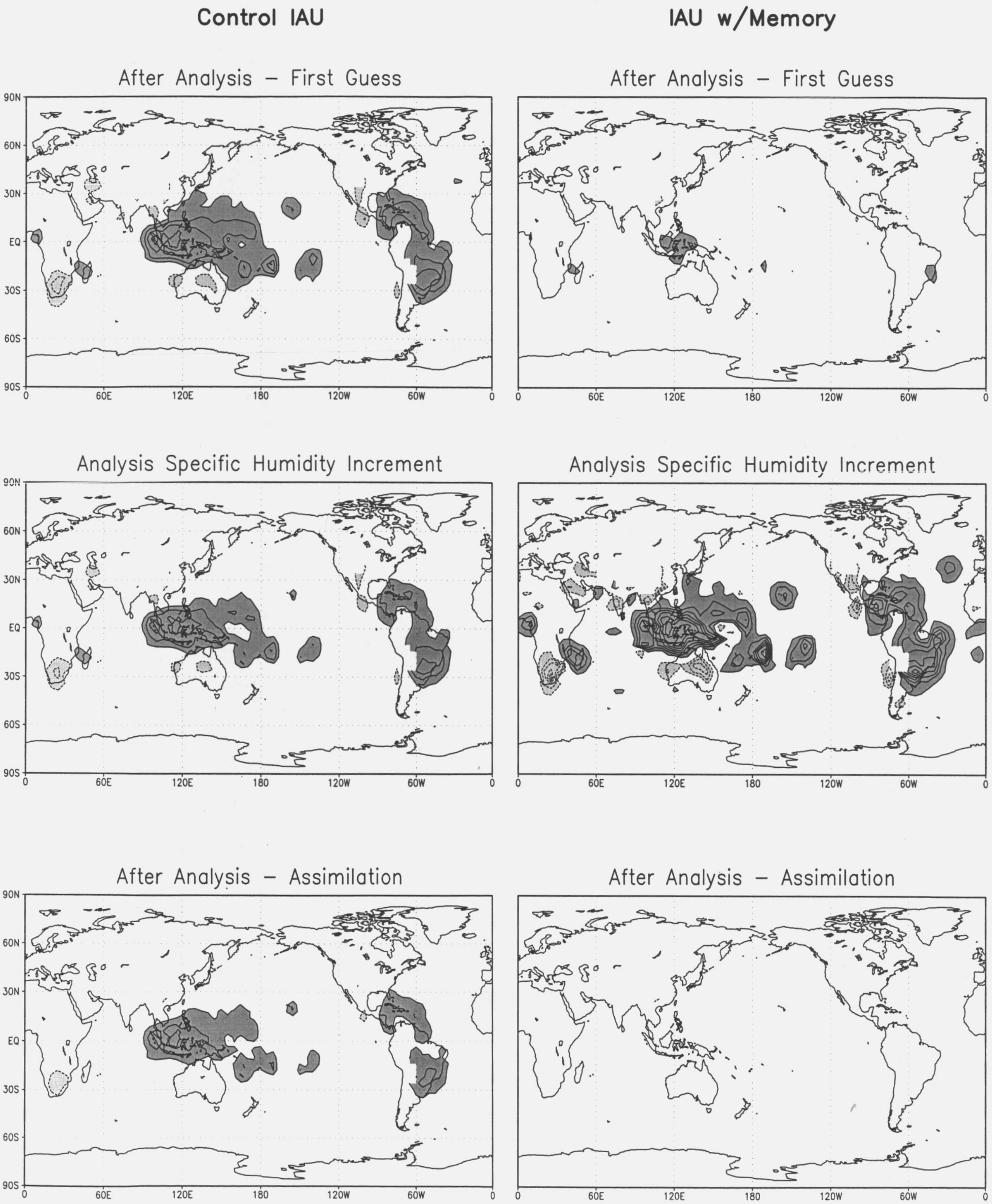

Control IAU
IAU w/Memory
After Analysis – First Guess
After Analysis – First Guess
Analysis Specific Humidity Increment
Analysis Specific Humidity Increment
After Analysis – Assimilation
After Analysis – Assimilation
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0

Figure 9:

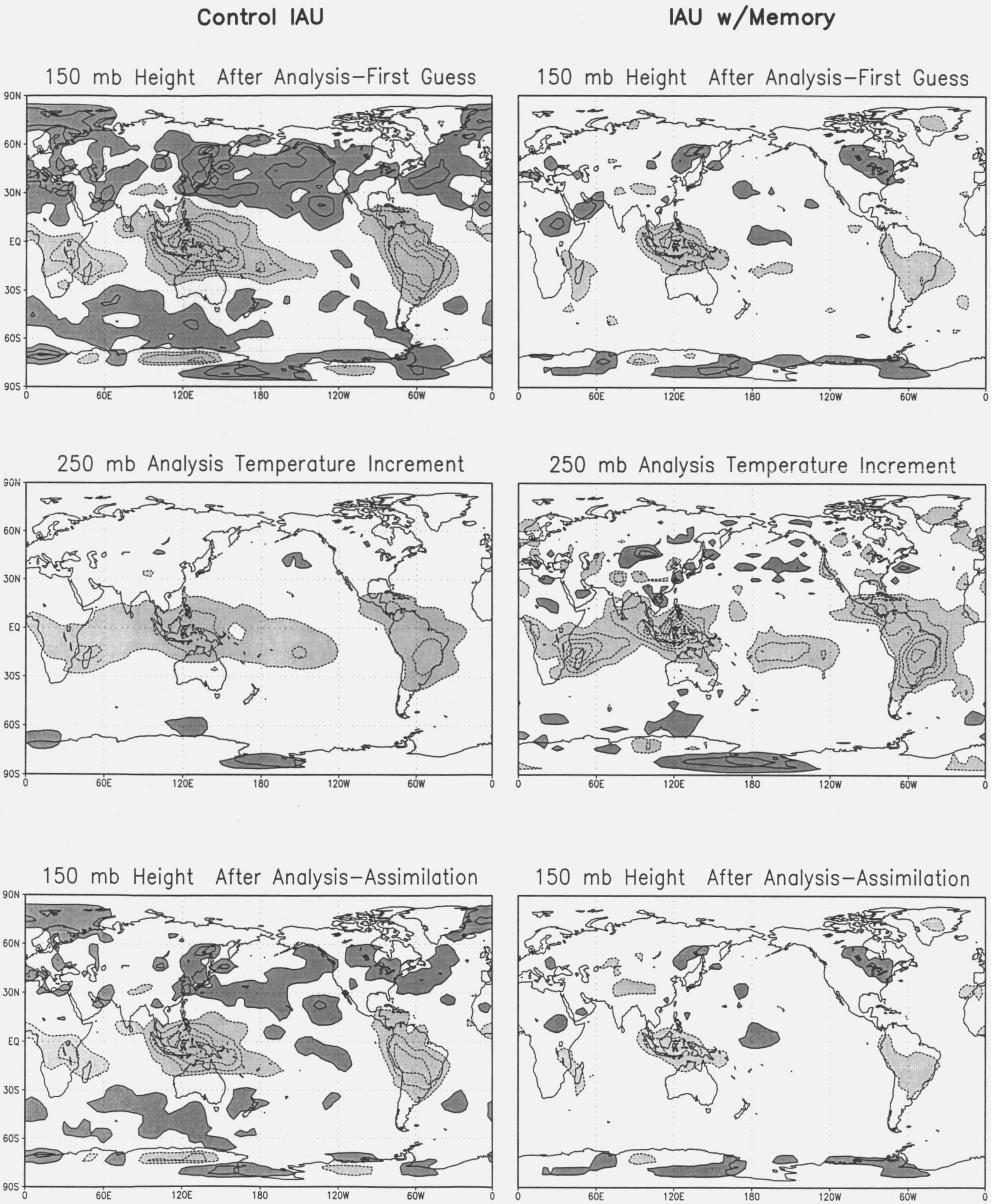

Control IAU
IAU w/Memory
150 mb Height After Analysis−First Guess
150 mb Height After Analysis−First Guess
250 mb Analysis Temperature Increment
250 mb Analysis Temperature Increment
150 mb Height After Analysis−Assimilation
150 mb Height After Analysis−Assimilation
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0

Figure 10:

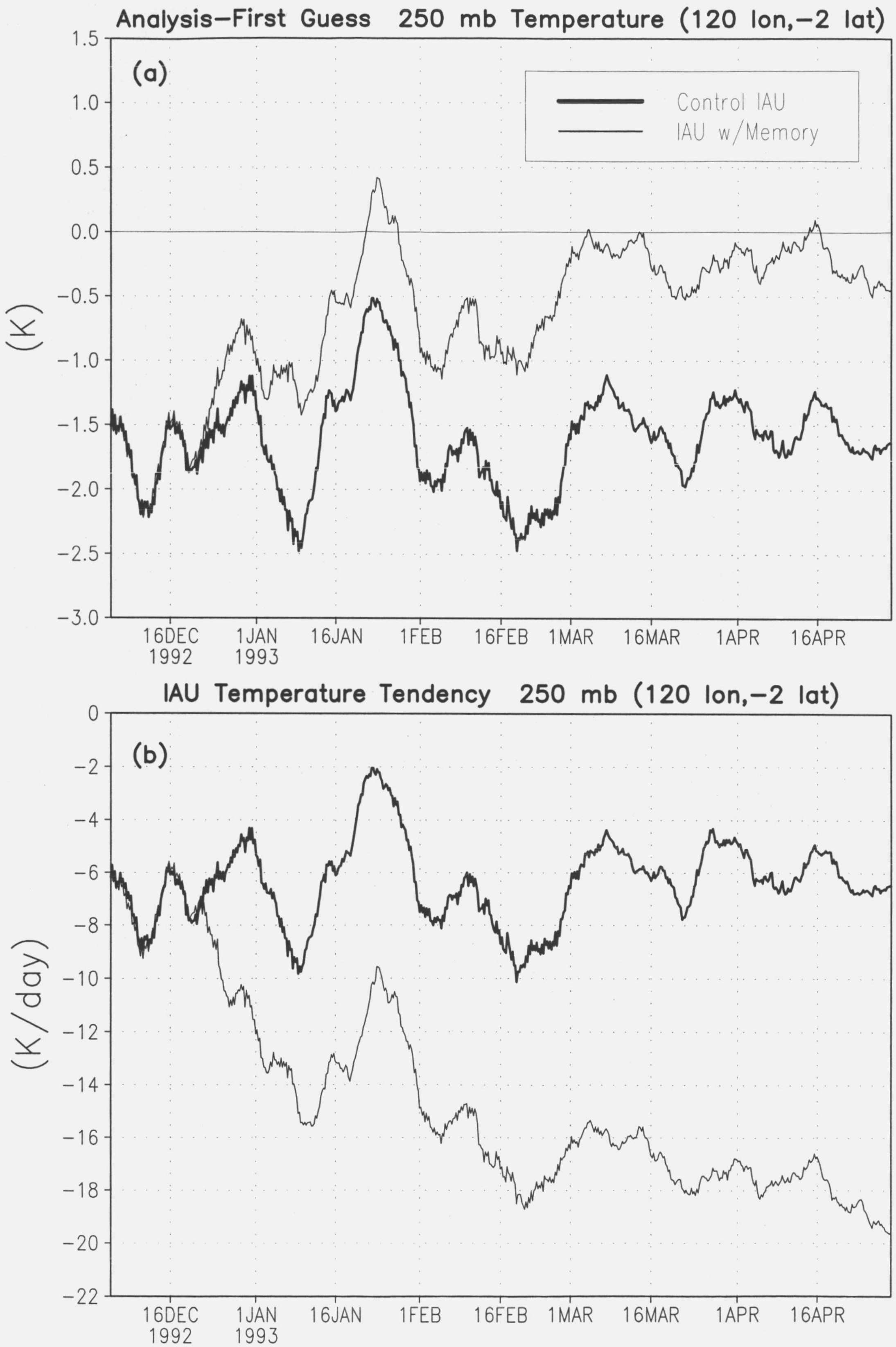

Analysis–First Guess 250 mb Temperature (120 lon,–2 lat)
(a)
Control IAU
IAU w/Memory
(K)
1.5
1.0
0.5
0.0
–0.5
–1.0
–1.5
–2.0
–2.5
–3.0
16DEC 1992
1JAN 1993
16JAN
1FEB
16FEB
1MAR
16MAR
1APR
16APR
IAU Temperature Tendency 250 mb (120 lon,–2 lat)
(b)
(K/day)
0
–2
–4
–6
–8
–10
–12
–14
–16
–18
–20
–22
16DEC 1992
1JAN 1993
16JAN
1FEB
16FEB
1MAR
16MAR
1APR
16APR

Figure 11:

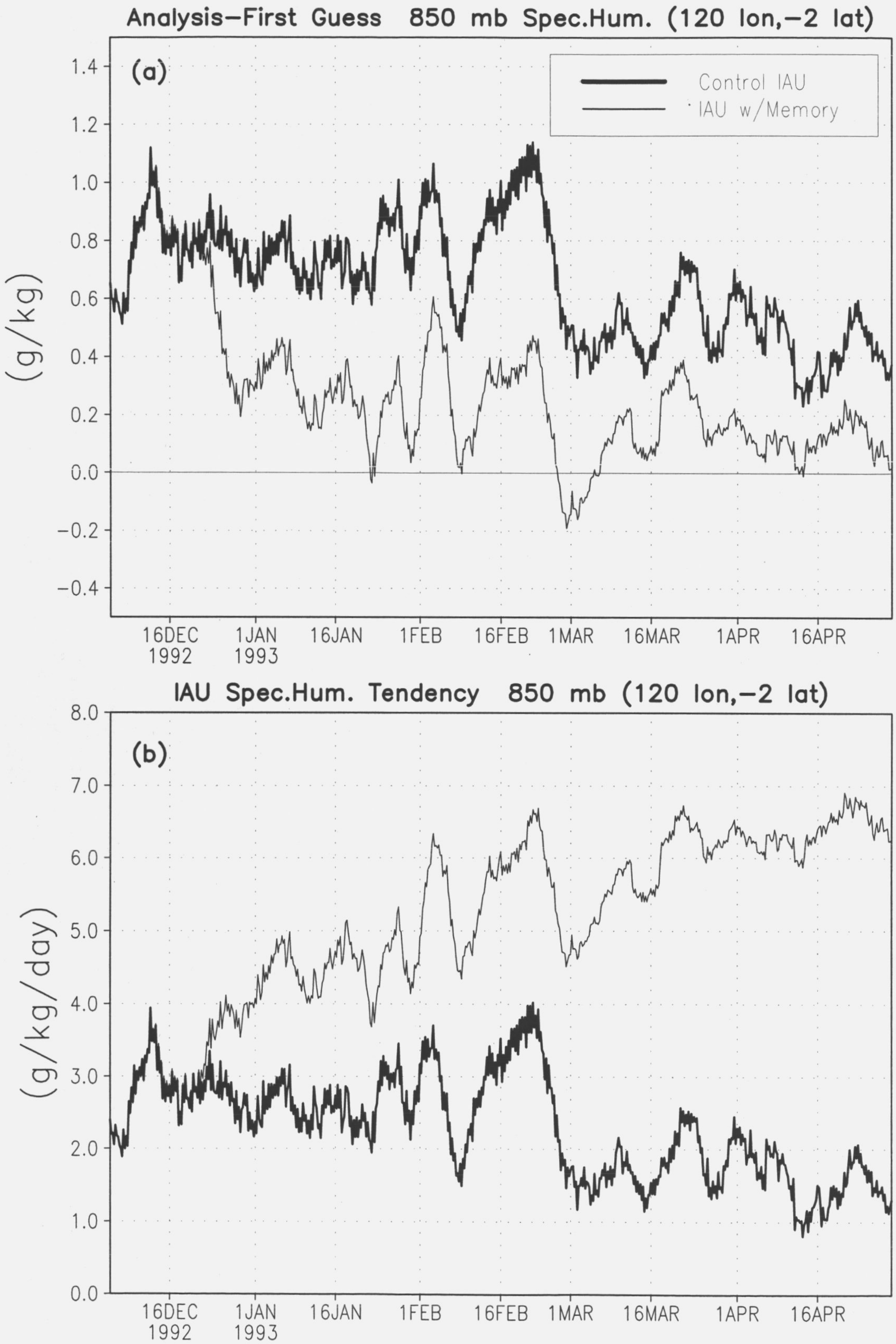
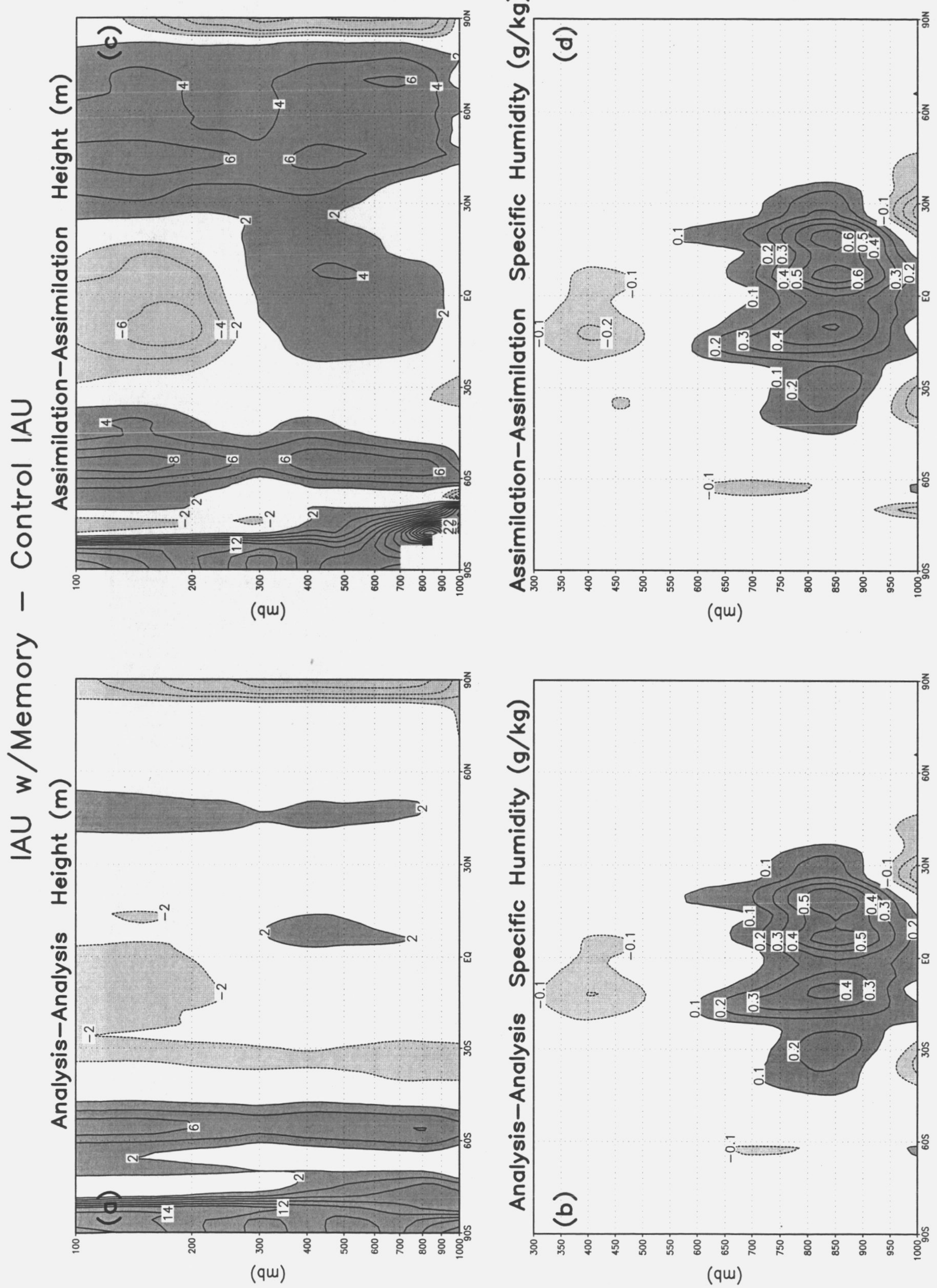

Analysis–First Guess 850 mb Spec.Hum. (120 lon,–2 lat)
(a)
Control IAU
IAU w/Memory
(g/kg)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
–0.2
–0.4
16DEC 1992
1JAN 1993
16JAN
1FEB
16FEB
1MAR
16MAR
1APR
16APR
IAU Spec.Hum. Tendency 850 mb (120 lon,–2 lat)
(b)
(g/kg/day)
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
16DEC 1992
1JAN 1993
16JAN
1FEB
16FEB
1MAR
16MAR
1APR
16APR

Figure 12:



IAU w/Memory – Control IAU
Analysis-Analysis Height (m)
Assimilation-Assimilation Height (m)
Analysis-Analysis Specific Humidity (g/kg)
Assimilation-Assimilation Specific Humidity (g/kg)
(a)
(b)
(c)
(d)
(mb)

Figure 13:

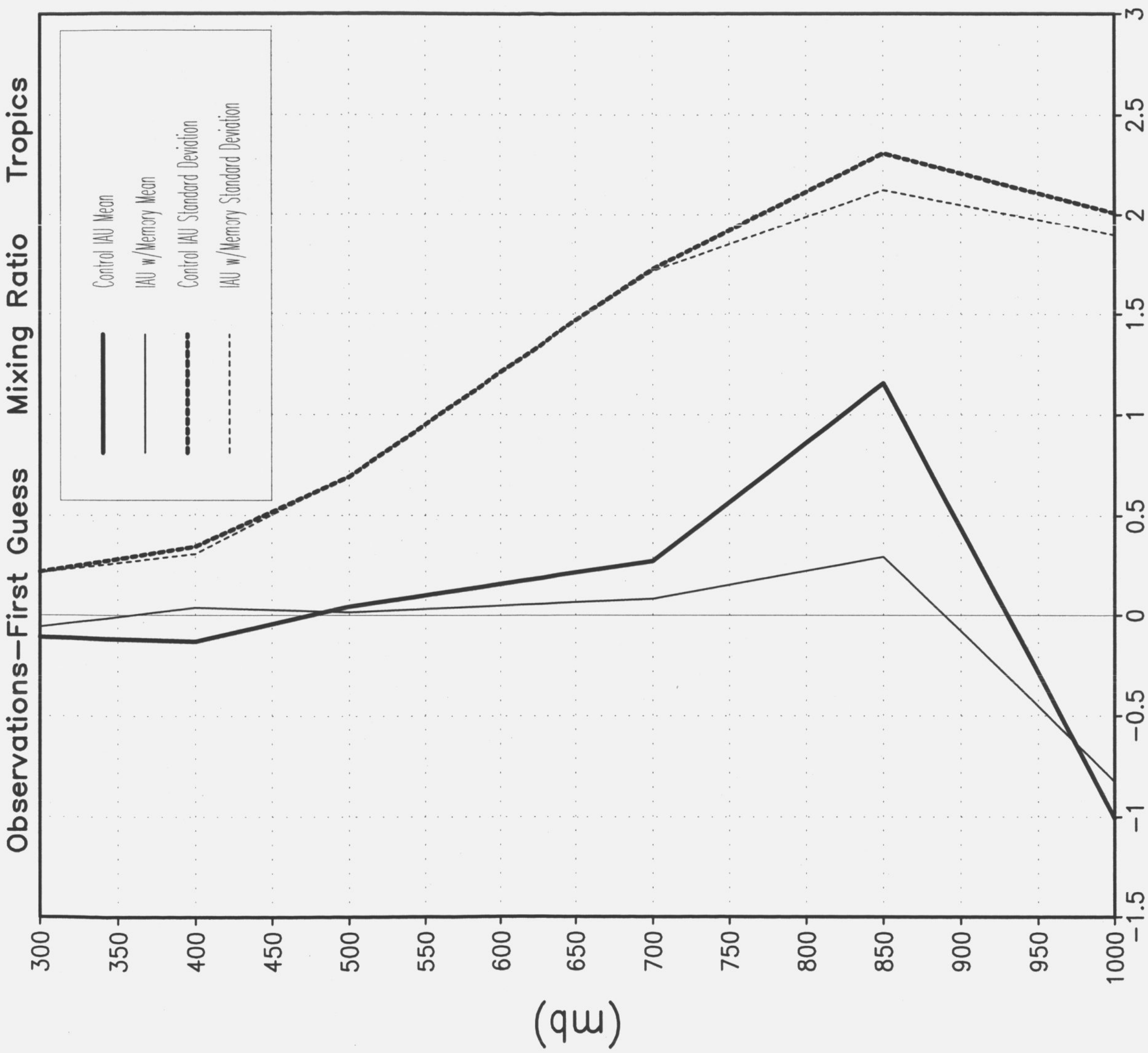

Observations−First Guess Mixing Ratio Tropics
Control IAU Mean
IAU w/Memory Mean
Control IAU Standard Deviation
IAU w/Memory Standard Deviation
(mb)

Figure 14: