Data Assimilation in the Presence of Forecast Bias: The GEOS Moisture Analysis

DICK P. DEE AND RICARDO TODLING

Data Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

ABSTRACT

The authors describe the application of the unbiased sequential analysis algorithm developed by Dee and da Silva to the Goddard Earth Observing System moisture analysis. The algorithm estimates the slowly varying, systematic component of model error from rawinsonde observations and adjusts the first-guess moisture field accordingly. Results of two seasonal data assimilation cycles show that moisture analysis bias is almost completely eliminated in all observed regions. The improved analyses cause a sizable reduction in the 6-h forecast bias and a marginal improvement in the error standard deviations.

1. Introduction

Hidden beneath the computational complexities of atmospheric data assimilation systems lies a multitude of assumptions about the errors associated with observing and predicting atmospheric fields. Most of these assumptions are there for practical reasons, either because there is not enough information to remove them or because they result in critical computational simplifications. Some, however, are known to be false and could be relaxed without too much difficulty, with a potentially large benefit to analysis accuracy.

A case in point is the standard assumption that short-term model forecasts, which are used as first-guess fields for the analyses, are unbiased. There is plenty of evidence to the contrary. For example, Fig. 1 shows the means and standard deviations of the differences between observed and forecast atmospheric water vapor mixing ratios, computed from January 1998 rawinsonde station data in four separate regions. The 6-h forecasts were produced by the Goddard Earth Observing System Data Assimilation System, version 2.8 (GEOS 2.8), which we describe in section 4. The statistics show that the systematic component of the observed-minus-forecast residuals ($O - F$) is not insignificant, especially in the Tropics and throughout the upper troposphere. Assuming that the mean observation error is small, this invalidates the assumption that the forecast is unbiased.

Details of the forecast bias obviously depend on the particulars of the general circulation model (GCM) that produced the forecast. They also depend on the accuracy of the analysis used to initialize the model, which, in turn, is partly determined by the quality and types of observations that entered into the analysis. However, the magnitude of GEOS moisture bias is not atypical. Monthly statistics of specific humidity $O - F$s produced by the operational data assimilation system (DAS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) show comparable biases and standard deviations (F. Lalaurette 1999, personal communication). The ECMWF DAS operates at a higher spatial resolution than GEOS, it uses a different analysis method, and it assimilates Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) as well as rawinsonde moisture data. The main cause of bias in the moisture fields produced by the two systems appears to be related to the model parameterizations of moist processes, which are perhaps inadequate in all current-generation GCMs (Chen et al. 1998).

The analysis produced by GEOS 2.8 is a weighted average of the 6-h forecast and the available observations. Analysis weights are derived from assumptions about the relative accuracies of these two sources of information, which, especially in the case of moisture, are not very well known. Regardless, the analysis inherits a fraction of the forecast bias, simply because of averaging. To illustrate, we also show in Fig. 1 the means and standard deviations of the observed-minus-analysis differences ($O - A$). Although the amplitude of the bias has been reduced by half, its sign is everywhere the same as that of the forecast bias. Applying more weight to the observations will reduce the bias but increase the random component of analysis error. It is generally not possible to produce an unbiased analysis from a biased forecast, unless a reasonable estimate of the forecast bias is available.

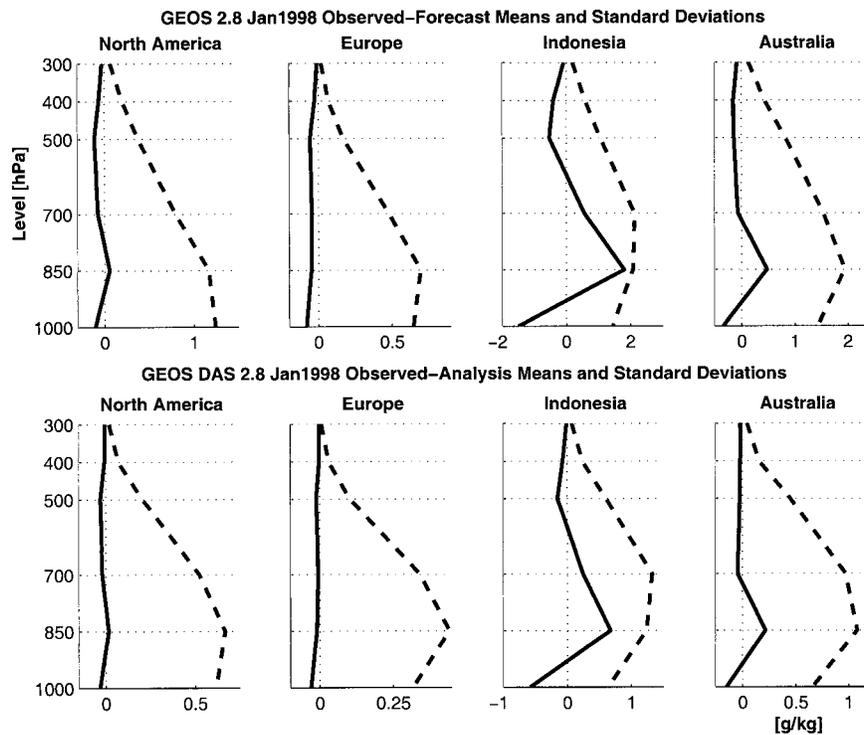The purpose of this paper is to describe the imple-

FIG. 1. Jan 1998 means (solid) and standard deviations (dashed) of atmospheric water vapor mixing ratio observed-minus-forecast (top panels) and observed-minus-analysis (bottom panels) residuals, as a function of pressure, for four different regions. Observations consist of quality-controlled rawinsonde reports; 6-h forecasts and analyses were produced by GEOS 2.8.

mentation in GEOS of the unbiased sequential analysis algorithm developed by Dee and da Silva (1998, hereafter DdS). The algorithm estimates forecast bias from observations and corrects the first guess accordingly. Since the bias estimate is continuously updated, it is more accurately described as an estimate of the slowly varying component of forecast error. The present implementation, which we refer to as GEOS UNB, is limited to the moisture field and based on rawinsonde data only.

As it turns out, the errors in the 6-h moisture forecasts contain relatively large persistent components, which are easily captured by the algorithm. Figure 2 displays the January 1998 $O - F$ and $O - A$ statistics for GEOS UNB, for the same regions as before. The bias has all but disappeared, and even the standard deviations are reduced, albeit by a very small amount. These results were obtained simply by incorporating the forecast bias estimation in the analysis. No other modifications were made to the DAS; in particular, the forecast and observation error covariance models were left unchanged.

Some clarification of terms may be helpful at this point. *Bias* generally refers to a nonzero mean error. In theory the mean and other statistics are defined in terms of a hypothetical ensemble of realizations and its associated probability density. In practical applications, however, bias is usually defined as a time average of a

single realization of the error taken over a finite time interval. This quantity is spatially variable and, if the errors are bounded, it can evolve on a timescale that is comparable with the length of the averaging interval. This length may vary, but our results are usually stated in terms of monthly means. The term *systematic error* is loosely applied in this paper to any type of error caused by an inherent, persistent deficiency in the model or in the observing system. In a nonlinear system systematic errors are necessarily state dependent. Bias is a particular manifestation of systematic errors.

Earlier work addressing systematic forecast errors in the context of data assimilation was done at the former National Meteorological Center (now the National Centers for Environmental Prediction) by Thiébaux and Morone (1990) and Saha (1992), and at the National Aeronautics and Space Administration's (NASA) Data Assimilation Office by Takacs (1996). In each of these studies, forecast bias estimates were derived from analyses rather than from observations. Griffith and Nichols (1996) consider the treatment of model error, and the bias problem in particular, by means of adjoint methods. DelSole and Hou (1999) explore the possibility of constructing state-dependent empirical corrections, also derived from analyses, in order to account for systematic errors in the forecast model.

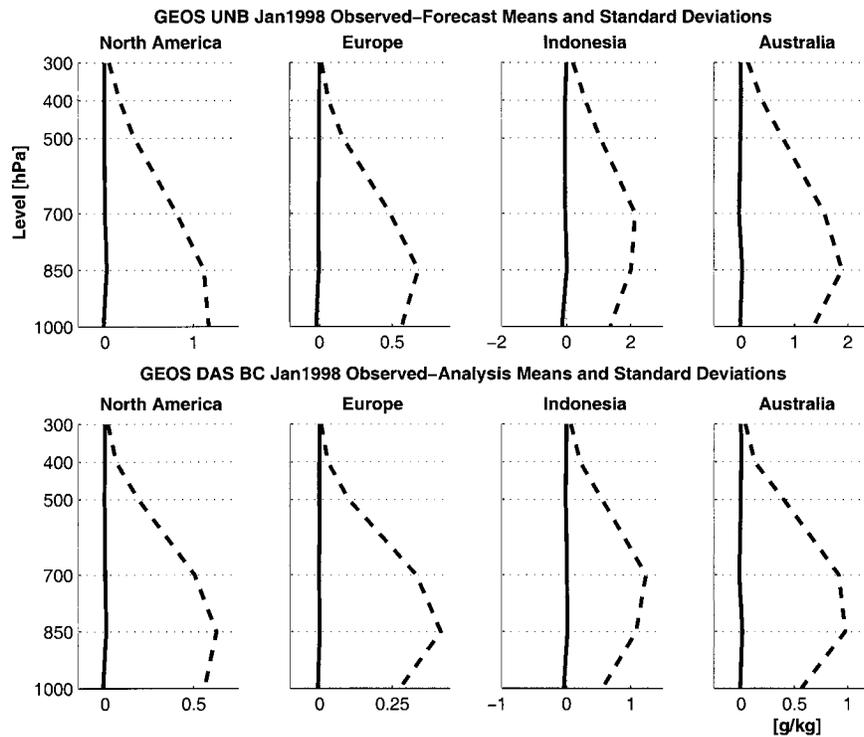The remainder of this paper explains the details of

FIG. 2. As in Fig. 1, but for GEOS UNB. For reference, the GEOS 2.8 observed-minus-forecast standard deviations are reproduced in the top four panels (dotted curves).

our GEOS UNB implementation and experimental results. In section 2 we take a closer look at time series of mixing ratio residuals produced by GEOS 2.8, to better understand the manifestations of systematic errors in the model forecasts. We briefly review the unbiased sequential analysis algorithm in section 3. There we also discuss the specification of a covariance model for the bias estimation errors, as well as other general implementation aspects. In an appendix we derive the optimal weights for the unbiased analysis equations under a reasonable set of assumptions; this issue was not settled in the original presentation of the algorithm in DdS. In section 4 we describe the specifics of the implementation in GEOS and present results of two seasonal data assimilation experiments with the new method. Section 5 contains our conclusions and plans for future work.

2. Evidence of systematic model errors

We can detect systematic model errors by comparing forecasts with observations. Nonzero mean residuals, computed over suitably long time periods and over a reasonably large set of stations, can be attributed to systematic model errors, provided that the mean observation errors are small. This will be the case if systematic errors in the data have been effectively removed. It is standard practice, for example, to correct rawinsonde height data at high altitudes for the effects of solar and infrared radiation (Mitchell et al. 1996). Re-

cent work by Zipser and Johnson (1998) shows that humidity soundings may also be contaminated by instrument-dependent systematic errors, but there is currently no practical way to correct these data in real time. Nevertheless, quality-controlled rawinsonde observations continue to serve as a benchmark for all other estimates of atmospheric moisture content.

The mixing ratio residual statistics shown in Fig. 1 represent averages over a month of station data from four specific regions. The monthly means and standard deviations are similar during other periods and for other station selections, but they are usually largest in the Tropics and smaller (but seasonably dependent) in the extratropics. This reflects a strong dependence of moisture forecast errors, and probably of observation errors as well, on the amount of moisture present in the atmosphere and on its local variability. To see this more clearly, it is helpful to look at individual time series of station data and corresponding 6-h forecasts.

We plotted in Fig. 3 the January 1998 mixing ratio observations, forecasts, and residuals, for three pressure levels at Singapore (1.37°N, 103.98°E). The lower panel shows that the 850-hPa forecasts are consistently drier than the observations. Most of the dots (observations) are above the thin curve (forecasts), and accordingly the residuals (thick curve) tend to be positive. At higher levels the situation is not as obvious. The forecasts at 300 hPa are too wet during several periods lasting a few days or more. This is consistent with the regional mean
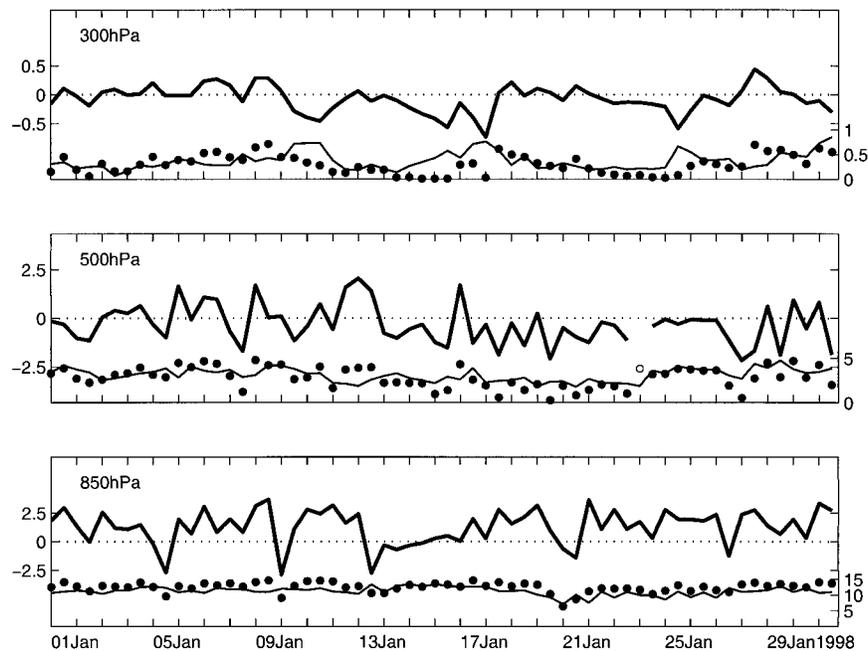
FIG. 3. GEOS 2.8 water vapor mixing ratio (g kg^{-1}) observed-minus-forecast residuals (thick solid curves), rawinsonde observations (dots), and 6-h forecasts (thin solid curves) at 300, 500, and 850 hPa in Singapore (1.37°N, 103.98°E), for the month January 1998. The scales for the residuals are on the left; those for the observed and forecast values on the right. The empty dot in the center panel represents an observation that was rejected by quality control.

wet bias in the forecast model at this level, but there are also periods with several consecutive dry forecasts.

We have looked at many such plots for different time periods and stations in various locations. In contrast with the monthly mean statistics, which are primarily a function of region, altitude, and season, the time-dependent characteristics of systematic forecast errors are not so easily quantified. They are best described as having a tendency to persist for a while: successive 6-h forecasts often remain either too wet or too dry for a few days or more. Although surely there are underlying physical explanations, the onset of such spells seems to occur randomly.

These appear to be manifestations of serially correlated model errors. In a theoretical study, Daley (1992) used a Kalman filter on a simple one-dimensional linear quasigeostrophic model to examine the potential impact of such errors on analysis accuracy. He also pointed out the practical difficulty of distinguishing between model bias and serially correlated model errors, even though the two are conceptually different (the mean of a serially correlated time series may or may not be zero). In our analysis of water vapor mixing ratio observed-minus-forecast residuals we have seen evidence of both types of phenomena.

Serial correlation of the residuals shows up clearly in the time spectra. We plotted normalized power spectra, as a function of wave period, for tropical, Northern Hemispheric, and Southern Hemispheric rawinsonde stations in Fig. 4. There is an excess of power in waves with periods longer than about 5 days, in each of the regions and at all levels; note that serially uncorrelated errors would result in flat spectra. These average spectra were obtained by (i) scaling the residuals at each station and at each level, so that the means are zero and the standard deviations one; (ii) computing the spectrum of each scaled time series; and (iii) averaging the spectra for all time series consisting of at least 50 residuals. Since there are gaps in the data, we used a spectral analysis algorithm for unevenly spaced data (Press et al. 1992, their section 13.8).

We will show in the next section how to use this type of spectral analysis for calibrating the parameters in the bias correction algorithm.

3. Online bias correction

Standard analysis methods are *bias blind,* in the sense that they ignore biases in the first-guess field. If the first guess is actually biased, then so will the analysis be biased. An unbiased analysis method must therefore include a scheme for estimating and removing the first-guess bias.

a. The bias-blind analysis equation

In a sequential statistical data assimilation system such as GEOS, analyses are produced at regular (typically 6 h) intervals. Each analysis combines quality-controlled observations with a forecast issued from an
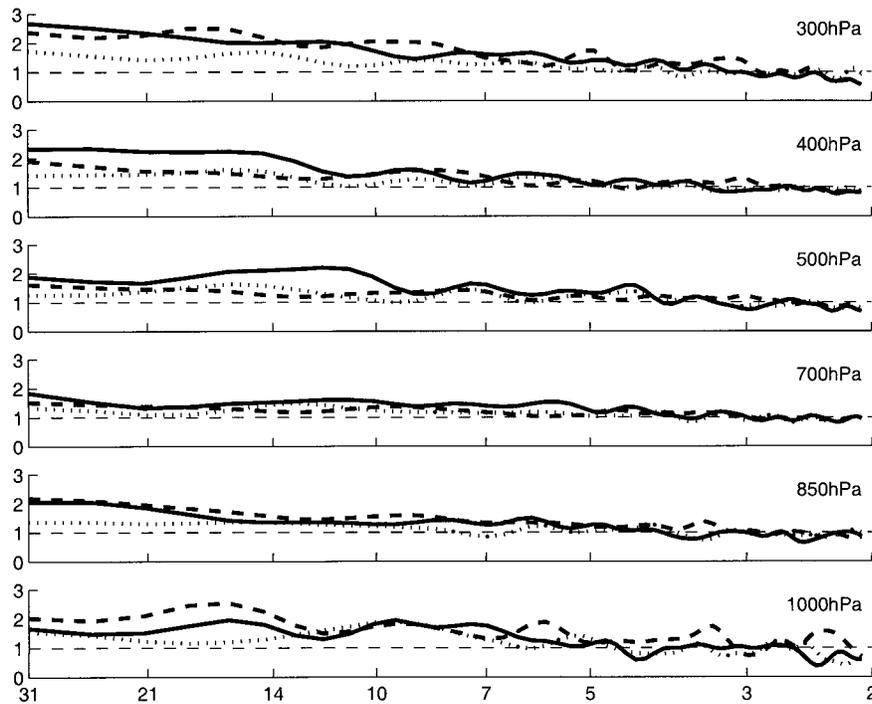
FIG. 4. Average normalized power spectra of GEOS 2.8 water vapor mixing ratio observed-minus-forecast residual time series for Jan 1998 rawinsonde reports in the Tropics (solid), Northern Hemisphere (dashed), and Southern Hemisphere (dotted). Horizontal axis is wave period, in days; vertical axis is normalized power. White noise would have a flat spectrum with power 1 at all wavelengths.

initial state derived from the previous analysis. Symbolically, for $k = 1, 2, \ldots,$

$$\mathbf{w}_k^a = \mathbf{w}_k^f + \mathbf{K}_k[\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f], \qquad (1)$$

where the $n \times 1$ vectors $\mathbf{w}_k^a$, $\mathbf{w}_k^f$ are the analysis and forecast at time t_k, respectively, and the $p \times 1$ vector $\mathbf{w}_k^o$ contains the observations. The $p \times n$ matrix $\mathbf{H}_k$ is the observation operator, which maps model variables to observables, and $\mathbf{K}_k$ is an $n \times p$ matrix of analysis weights.

If both forecast and observations are unbiased, and their errors are mutually independent, then the optimal (least squares) analysis obtains for

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T[\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k]^{-1}, \qquad (2)$$

with $\mathbf{P}_k^f$ and $\mathbf{R}_k$ the forecast and observation error covariances, respectively (Jazwinski 1970).

The development of adequate covariance models is an area of active research. In (linear) theory, $\mathbf{P}_k^f$ and $\mathbf{R}_k$ can be computed explicitly using knowledge of the joint probability distribution of model and observation errors. In reality there is not nearly enough information available to warrant the staggering expense that such a computation would entail. Furthermore, the value of a brute-force approach is questionable, since many of the assumptions that render (2) optimal are incorrect in any case (Dee 1991, 1995). Thus, operational data assimilation systems use highly simplified representations of

the required error covariances. These are usually obtained by a combination of statistical data analysis, careful consideration of multivariate balance requirements, optimal use of computational resources, and artful tuning of covariance parameters.

In this paper we largely ignore these issues, because (1) implies that the analysis $\mathbf{w}_k^a$ will be biased if the forecast $\mathbf{w}_k^f$ is biased, *regardless of the analysis weights* $\mathbf{K}_k$. See DdS (section 2) for more discussion of the transfer of forecast bias to the analysis and of the effect of adjusting the covariance models.

b. The unbiased analysis equations

DdS showed how to produce unbiased analyses in a sequential data assimilation system when the forecast is biased. The idea is to provide a running estimate of the bias and to correct the forecast accordingly. The result is the replacement of (1) by the following two-step algorithm:

$$\hat{\mathbf{b}}_k = \hat{\mathbf{b}}_{k-1} - \mathbf{L}_k[\mathbf{w}_k^o - \mathbf{H}_k(\mathbf{w}_k^f - \hat{\mathbf{b}}_{k-1})], \qquad (3)$$

$$\mathbf{w}_k^a = (\mathbf{w}_k^f - \hat{\mathbf{b}}_k) + \mathbf{K}_k[\mathbf{w}_k^o - \mathbf{H}_k(\mathbf{w}_k^f - \hat{\mathbf{b}}_k)]. \qquad (4)$$

The $n \times 1$ vector $\hat{\mathbf{b}}_k$ is the estimated forecast bias at time t_k. Note that (4) and (1) are identical when $\hat{\mathbf{b}}_k = 0$. The $n \times p$ matrix $\mathbf{L}_k$, which we will specify below,

defines the weighting coefficients for the bias update equation.

The two requirements for the analysis $\mathbf{w}_k^a$ to be unbiased are that (i) the observations $\mathbf{w}_k^o$ are unbiased, and (ii) $\hat{\mathbf{b}}_{k-1}$ is an unbiased prediction of the forecast bias at t_k. This statement follows simply from linearity and holds regardless of the specification of the gain matrices $\mathbf{L}_k$ and $\mathbf{K}_k$. The second condition is reasonable when forecast errors tend to persist at fixed locations. It is not reasonable, for example, when the errors are dominated by large, systematic displacements occurring on synoptic timescales. In that case the bias is not a slowly varying function of time, and its estimation requires a better (e.g., state dependent) bias prediction to replace $\hat{\mathbf{b}}_{k-1}$ in (3). In practical terms, using persistence to predict the bias means that our estimates can only capture the slowly varying component of forecast error, while the remaining components are treated as random.

Whether $\hat{\mathbf{b}}_{k-1}$ provides an unbiased estimate of the forecast bias at t_k also depends on the data coverage, both in space and time. Clearly, if a particular region remains unobserved for a while, then it is not possible to obtain a meaningful bias estimate there, unless additional information (e.g., about the spatial structure of the forecast bias) is used for the bias prediction. As it stands, the bias estimate generated by (3) will remain constant in regions devoid of observations. It is therefore important to control the impact of an occasional observation in a poorly observed region. This can be done by relaxing the bias estimate to its initial state in the absence of observations or by careful data selection. With some abuse of notation, therefore, the observations $\mathbf{w}_k^o$ (and associated $\mathbf{H}_k$) used in (3) and (4) may be different.

Lacking any a priori information about forecast bias, we take

$$\hat{\mathbf{b}}_0 = 0. \tag{5}$$

Consequently the bias estimate will remain zero in unobserved regions, and the analysis produced by (4) will differ from the bias-blind analysis (1) only where data exist. If clearly discernible permanent spatial structures show up in the bias estimates, and there is reason to believe they can be extrapolated, then (5) should be modified accordingly.

We show in the appendix that, within reasonable approximation, the optimal weights for the bias estimator (3) are

$$\mathbf{L}_k = \mathbf{P}_k^b \mathbf{H}_k^{\mathrm{T}}[\mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^{\mathrm{T}} + \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k]^{-1}, \tag{6}$$

if $\mathbf{P}_k^b$ is the error covariance of the bias prediction $\hat{\mathbf{b}}_{k-1}$. The best choice of analysis weights $\mathbf{K}_k$ is not obvious, since the the same observations used for the bias estimation are used again in the analysis equation (4). The bias-corrected forecast $\mathbf{w}_k^f - \hat{\mathbf{b}}_k$ is therefore not independent of the observations $\mathbf{w}_k^o$. However, in the appendix we derive the remarkable result that, if $\mathbf{L}_k$ is

defined by (6), then the optimal analysis weights for (4) are still defined by (2).

The cost of solving the unbiased analysis equations is roughly double that of computing the standard, bias-blind analysis. Any analysis system designed to solve (1) can be used to solve both (3) and (4), simply by changing the background field and the analysis weights. Possible ways to economize without significant loss of accuracy are (i) to use only a subset of the observations for the bias estimation, (ii) to estimate the forecast bias at a reduced spatial resolution, and (iii) to update the forecast bias estimates less frequently.

c. Specification of the error covariances

The bias estimator requires specification of the error covariances $\mathbf{P}_k^b$ of the bias estimates, in addition to the forecast and observation error covariances $\mathbf{P}_k^f$ and $\mathbf{R}_k$ that are needed for the analysis. We are primarily interested in incorporating the unbiased analysis equations into an existing operational data assimilation system. Initially, therefore, we regard the forecast and observation error covariances as given and use a very simple model for the bias estimation error covariance. Our approach is to first concentrate on reducing the mean analysis errors; once that has been achieved we can hope to further improve the analyses by introducing better covariance models.

We are faced with the reality that the actual error characteristics of the bias estimates are not known. DdS proposed the following model for the error covariances of the bias estimates:

$$\mathbf{P}_k^b = \gamma \mathbf{P}_k^f, \tag{7}$$

with γ constant. This model assumes that the spatial correlations of the bias estimation errors are identical to those of the random component of the forecast errors, and (in the multivariate case) that the two types of errors are balanced in the same way. This is an attractive starting point since it takes full advantage of the effort invested in formulating and implementing a forecast error covariance model that produces reasonable results in an operational setting. There are obvious ways to generalize, for example, by allowing γ to depend on space and/or time, or by adjusting the correlation models incorporated in $\mathbf{P}_k^f$.

To arrive at a means for determining an appropriate value for the parameter γ, we study the behavior of the bias estimator at a single observation location that coincides with a model grid point. The bias gain (6) is then

$$L_k = \lambda = \sigma_b^2/(\sigma_b^2 + \sigma_f^2 + \sigma_o^2), \tag{8}$$

with σ_b, σ_f, and σ_o the error standard deviations for the bias estimate, the forecast, and the observation, respectively, which we take to be stationary for the moment. Equation (3) becomes

$$\hat{b}_k = \hat{b}_{k-1} - \lambda[w_k^o - (w_k^f - \hat{b}_{k-1})]$$

$$= -\lambda \sum_{j=0}^{k-1} (1 - \lambda)^j v_{k-j}, \tag{9}$$

with $v_k = w_k^o - w_k^f$, and we used $\hat{b}_0 = 0$.

In case of a constant forecast bias $b_k \equiv b$ it is easy to show from (9) and the assumption that observations are unbiased that $\lim_{k\to\infty} \langle \hat{b}_k \rangle = b$ when $0 < \gamma < 2$. Therefore the mean bias estimate over any sufficiently long time interval will converge to the mean forecast error over that interval. More generally, the asymptotic first-moment properties of an estimator for a linear, stationary system are not sensitive to the covariance specifications, as long as the system is completely observable and controllable (Jazwinski 1970, section 7.6). Correct specification of the error covariances only improves the rate of convergence to the asymptotic estimate. This is consistent with our earlier statement that the analysis equations [(3)–(4)] are unbiased regardless of the error covariance specifications. The covariances, or the value of γ in this scalar case, determine the response of the estimator to errors at shorter timescales.

We showed in section 2 that, in the absence of forecast bias correction, the time series of observed-minus-forecast residuals typically have colored spectra. We would like to determine a value for the parameter λ such that the spectra of the bias-corrected observed-minus-forecast residuals become as flat as possible, in some well-defined sense. In the time-frequency domain, (9) corresponds to

$$\beta_n = R_n(\lambda)v_n, \tag{10}$$

where β_n, v_n are the Fourier coefficients for wavenumber $n > 0$ of the time series $\hat{b}_k$, v_k, respectively. The response function R_n is

$$R_n(\lambda) = -\lambda/[1 - (1 - \lambda)e^{-2\pi i \Delta t/n}], \tag{11}$$

with Δt the time interval between observations. A flat spectrum of $v_k + \hat{b}_k$ corresponds to

$$|v_n + \beta_n| = |[1 + R_n(\lambda)]v_n| = \text{const.} \tag{12}$$

Clearly (12) cannot be satisfied exactly by manipulating the single free parameter λ. Instead we can use λ to reduce the energy in the long-wave portion of the spectrum of observed-minus-forecast residuals. A practical method for estimating λ is to compute the average normalized power spectrum P_n of the residuals for a set of stations (see section 2), and then to find λ that minimizes the functional

$$f(\lambda) = \sum_n n^2\{|[1 + R_n(\lambda)]P_n| - 1\}^2. \tag{13}$$

The factor n^2 serves to emphasize the impact on the long-wave portion of the spectrum, and we prefer to use average spectra in order to increase the sample size. A value of λ can thus be computed, say, separately for data at fixed pressure levels from stations in selected regions. Having determined λ, and with σ_f and σ_o given, (7)–(8) imply

$$\gamma = \frac{\lambda}{1 - \lambda} \frac{\sigma_f^2 + \sigma_o^2}{\sigma_f^2}, \tag{14}$$

which, in conjunction with (7), completes the specification of the bias estimation error covariance model. This is sufficient for our present purposes.

We briefly outline what would be the next step, namely, the reestimation of forecast error standard deviations and other covariance parameters, which, after all, are likely to change as a result of the introduction of forecast bias correction. The bias-corrected observed-minus-forecast residuals are

$$\mathbf{v}_k \equiv \mathbf{w}_k^o - \mathbf{H}_k(\mathbf{w}_k^f - \hat{\mathbf{b}}_{k-1}). \tag{15}$$

Using (A2), (A1), (A4), (A7) of the appendix,

$$\mathbf{v}_k = \boldsymbol{\varepsilon}_k^o - \mathbf{H}_k(\boldsymbol{\varepsilon}_k^f - \boldsymbol{\varepsilon}_k^b), \tag{16}$$

where $\boldsymbol{\varepsilon}_k^o$, $\boldsymbol{\varepsilon}_k^f$, $\boldsymbol{\varepsilon}_k^b$ are the errors in the observations, in the bias-corrected forecast, and in the bias estimate, respectively. The covariances of the residuals are therefore

$$\langle \mathbf{v}_k \mathbf{v}_k^T \rangle = \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^T + \mathbf{R}_k, \tag{17}$$

where we used (A3), (A5), (A6). This relation between the data and the covariance models provides the basis for estimating parameters of $\mathbf{P}_k^f$, $\mathbf{P}_k^b$, and $\mathbf{R}_k$ by, for example, maximum-likelihood techniques (Dee 1995; Dee and da Silva 1999). Parameters of forecast and bias estimation error covariances are probably not separately identifiable, so that a model such as (7) will still be needed to close the problem.

4. Implementation in GEOS

As a first test of online forecast bias estimation and correction in an operational data assimilation system, we modified the GEOS 2.8 moisture analysis and computed data assimilation cycles for two seasons (December 1997–February 1998 and June–August 1998). Here we briefly describe the relevant characteristics of the system and then discuss the moisture bias correction experiments. In order to save space we show results for January 1998 only. Results for other months are qualitatively similar and lead to identical conclusions.

a. Description of GEOS 2.8

GEOS 2.8 produces global atmospheric datasets at 6-hourly intervals (3-hourly for surface fields) on a 2° × 2.5° latitude–longitude grid, at 48 vertical levels in both pressure and sigma coordinates. The core of the system consists of an atmospheric GCM (Suarez and Takacs 1995), the Physical-Space Statistical Analysis System (PSAS) (Cohn et al. 1998), and various interface functions including, for example, quality control of observations. The final, assimilated data products are obtained from the analyzed fields by means of the incre-

mental analysis update (IAU) procedure (Bloom et al. 1996).

Apart from conventional atmospheric observations, the system accepts geopotential heights retrieved from TOVS data, cloud-drift wind retrievals, and surface winds obtained from Special Sensor Microwave/Imager (SSM/I) wind speed data. The only observations of atmospheric moisture entering the system are obtained from rawinsonde soundings, although efforts are currently under way to implement the assimilation of interactive TOVS moisture retrievals (Joiner and Rokke 2000) and SSM/I-derived total precipitable water (Hou et al. 2000).

At 6-hourly intervals during the assimilation, a global analysis is computed in three steps: first for the surface (sea level pressure and 10-m winds), then for the upper-air variables (geopotential heights and winds), and finally for the moisture field (water vapor mixing ratio). In each case PSAS solves the analysis equation (1), combining all available observations taken within 3 h of the analysis time with the first-guess fields produced by the GCM. Analysis weights are defined by (2), based on prescribed forecast and observation error covariance models. The analysis corrections to the first-guess fields are then used in the IAU procedure to force the next 6-h model integration.

The mixing ratio forecast and observation error variances are such that their ratio is a function of pressure only, ranging between a minimum value of 0.76 (at 700 hPa) and a maximum of 2.8 (at 300 hPa). This ratio represents, in a scalar analysis, the weight of a single mixing ratio observation relative to that of the first-guess value at that location. To complete the covariance model specifications, observation errors are assumed uncorrelated in space and time, while forecast error correlations are represented by a separable function of horizontal distance and pressure. The variances and correlation parameters have been estimated from observed-minus-forecast residuals by maximum-likelihood techniques (Dee et al. 1999). Clearly, these exceedingly simple models leave ample room for improvement.

b. Description of GEOS UNB

The experimental system GEOS UNB is identical to GEOS 2.8 in all respects, except that the two-step algorithm (3, 4) replaces the moisture analysis equation (1). Both steps are solved with PSAS, using the same forecast and observation error covariance specifications. For the bias estimation step (3), we only use observations in the vicinity of stations that report at least once a day. The computational cost of the moisture analysis is then roughly twice that in the original system. Since the number of moisture observations is relatively small (about 7000 per day), the additional expense is insignificant in the context of the total DAS computation.

The error covariances for the bias estimates, needed to define the weights (6), are modeled by (7). We tuned the parameter γ using the spectral estimation procedure described in section 3, separately for each pressure level, from time series of GEOS 2.8 observed-minus-forecast residuals restricted to different time periods and regions in space. The estimated values varied somewhat, tending to be largest at the upper levels, where the serial correlation of forecast errors is most pronounced. We ended up using a constant value $\gamma = 0.22$ for our experiments, feeling that further refinement may not be worthwhile until improved covariance models can be implemented.

c. January 1998 results

We first examine the time spectra of the GEOS UNB observed-minus-forecast residuals in order to validate our choice of the parameter γ. As discussed in section 3, the sequential bias estimator acts as a first-order linear filter on the data residuals. Its behavior at a single station location can be characterized by a response function in the time-frequency domain and (with forecast and observation error variances given) is determined by the parameter γ. An optimal analysis scheme would produce white (serially uncorrelated) observed-minus-forecast residuals (Kailath 1968). Figure 5 shows that the regionally averaged spectra of the GEOS UNB residuals are in fact much flatter than those for GEOS 2.8 (shown in Fig. 4).

In an earlier experiment with $\gamma = 0.5$ we found that the spectra had too little power in the low wavenumbers (i.e., the opposite of Fig. 4). The removal of forecast and analysis bias was equally effective in that experiment (in the monthly mean sense), but the random component of forecast error was slightly larger. This is consistent with the scalar theory presented in section 3 and supports our contention that the analyses are unbiased regardless of the covariance specifications. However, increasing the value of γ has the effect of contaminating the bias estimates with noise, and this deteriorates the analyses even if it does not change their time-mean properties. We find our procedure for estimating the parameter γ based on a spectral analysis of observed-minus-forecast residuals to be quite effective.

We now take a look at the impact of the forecast bias correction at a single station. Figure 6 shows time series data at 300, 500, and 850 hPa for the Singapore rawinsonde station; this should be compared with Fig. 3. The thick solid curves are the corrected observed-minus-forecast residuals $\mathbf{w}_k^o - (\mathbf{w}_k^f - \hat{\mathbf{b}}_{k-1})$. The dashed curves show the forecast bias predictions $\hat{\mathbf{b}}_{k-1}$; they indicate that the 6-h forecasts tend to be too wet at the upper levels and too dry below. The bias estimates vary slowly with time; adjustments are made whenever there are persistent differences between the bias-corrected forecasts and the observations. The bias estimates may remain quite small for extended periods, for example at 850 hPa between 14 and 25 January.

Residual time series at a single station fluctuate wildly because of random (or small scale) observation and fore-
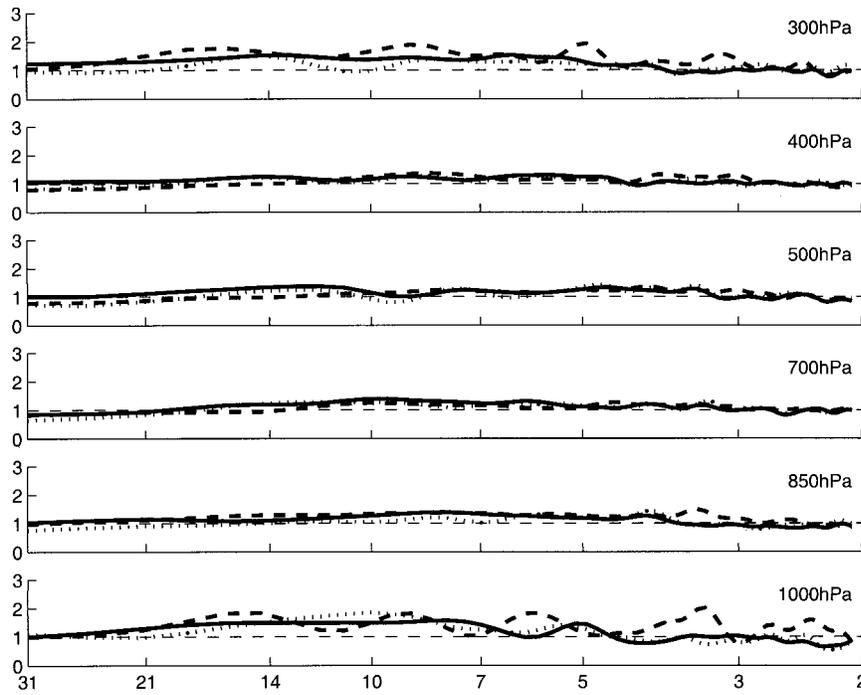
FIG. 5. As in Fig. 4, but for GEOS UNB.

cast errors. It is easier to see the impact of the bias correction by averaging over several stations. For the Indonesian region, for example, we show in Fig. 7 the time evolution of the mean observed-minus-forecast re-

siduals, for GEOS 2.8 (thin curves) and for the modified system (thick curves). A point on each curve is obtained simply by averaging the residuals at all stations in the region. Since the bias-corrected forecasts $(\mathbf{w}_k^f - \hat{\mathbf{b}}_{k-1})$
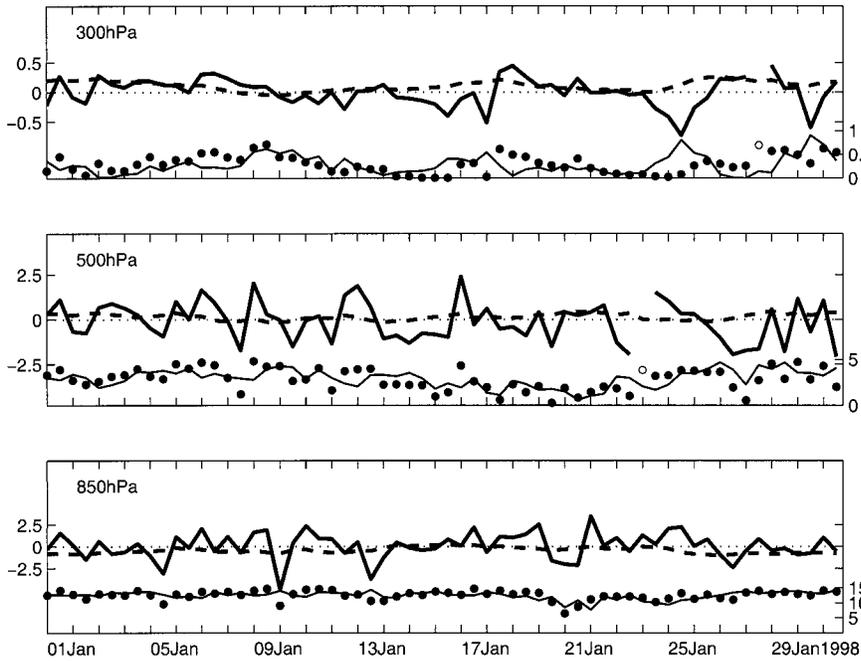
FIG. 6. As in Fig. 3, but for GEOS UNB. The dashed curves in each panel correspond to the forecast bias estimates at the station location; scales for these estimates are indicated on the left vertical axes.
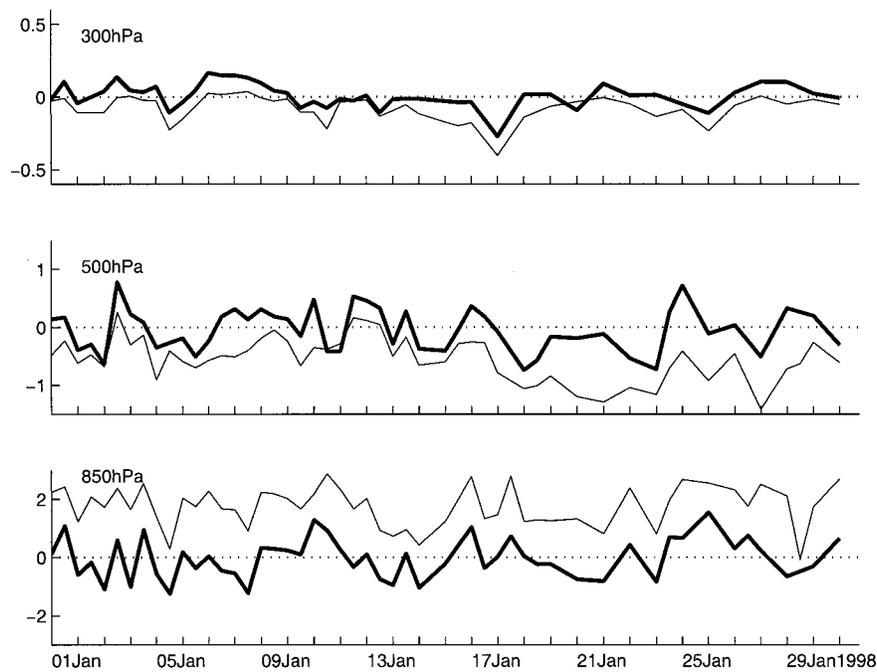
FIG. 7. Time evolution of mean mixing ratio (g kg⁻¹) observed-forecast residuals for Indonesian rawinsonde stations, at 300, 500, and 850 hPa. Forecasts are from GEOS 2.8 (thin curves) and GEOS UNB (thick curves).

are truly forecasts, in the sense that they do not depend on data at t_k, these plots demonstrate that the algorithm is able to predict the forecast bias relative to future observations.

Figure 2, briefly discussed in the introduction, shows the monthly means and standard deviations of both the observed-minus-forecast and observed-minus-analysis residuals, for four different regions. Again, these statistics apply to the bias-corrected 6-h forecasts ($\mathbf{w}_k^f - \hat{\mathbf{b}}_{k-1}$); they show that, by being able to predict the forecast bias, the algorithm succeeds in producing unbiased analyses. An interesting question is whether the improved analyses have a positive impact on the uncorrected forecasts $\mathbf{w}_k^f$ as well. This is by no means self-

evident, since the mechanisms by which model errors are generated are not well understood, although almost certainly nonlinear. Figure 8 shows that, in most cases, the mean errors in the uncorrected forecasts produced from the unbiased analyses in GEOS UNB are in fact much smaller than the errors in the GEOS 2.8 forecasts. This proves that the improvements in the unbiased GEOS UNB analyses do in fact result in significantly better (in the mean sense) short-term forecasts, at least in observed regions.

In Fig. 9 we show an example of a moisture forecast, the forecast bias estimate, and the analysis produced by the algorithm. This is a snapshot for a particular region (Indonesia), level (300 hPa), and synoptic time (0000
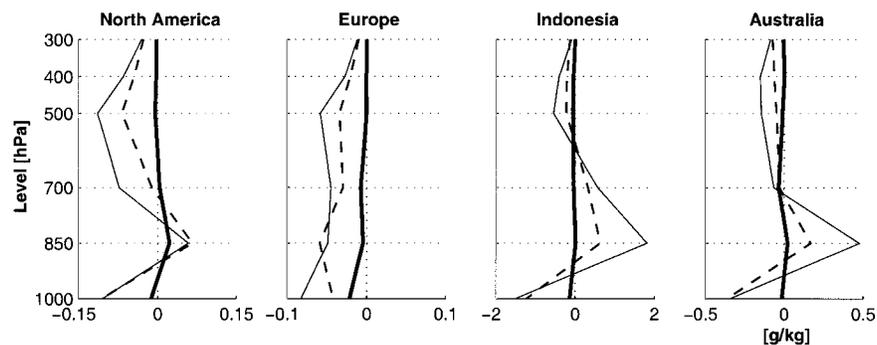

FIG. 8. Monthly mean mixing ratio (g kg⁻¹) observed-minus-forecast residuals for rawinsonde stations in four different regions. Forecasts are from GEOS 2.8 (thin solid curves) and from GEOS UNB before (dashed curves) and after (thick solid curves) bias correction.
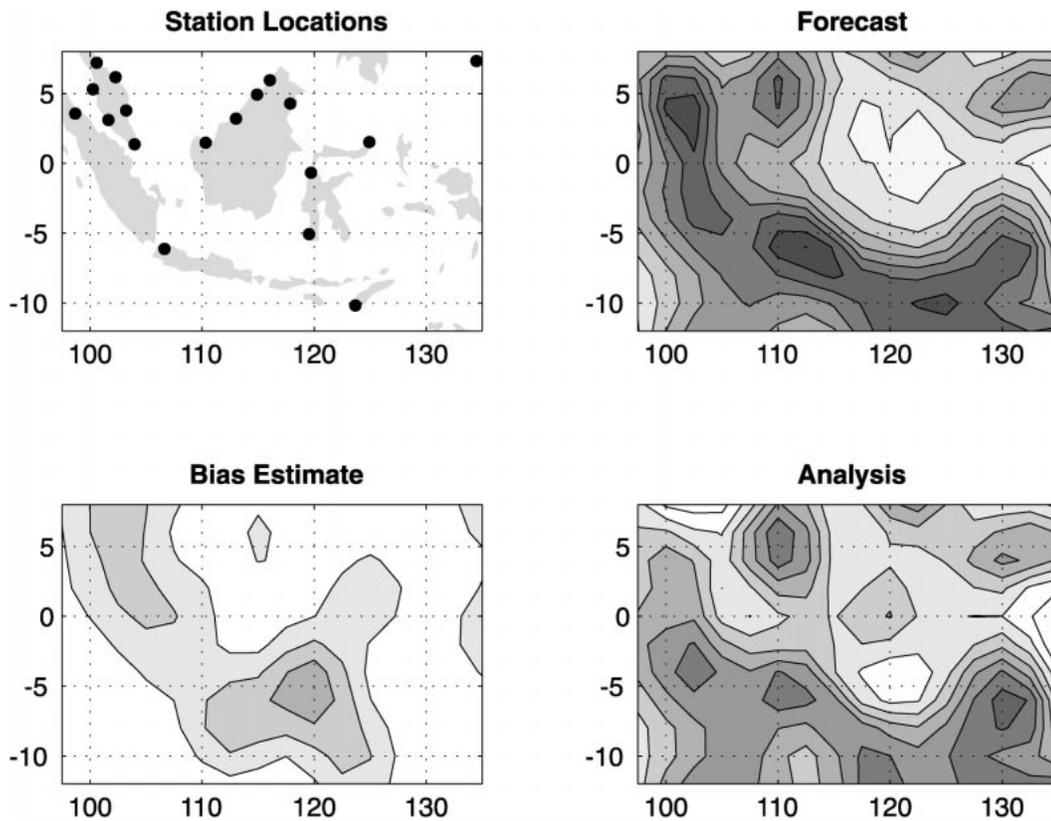
FIG. 9. Station locations, 300-hPa moisture forecast, bias estimate, and final moisture analysis near Indonesia for 0000 UTC 1 Jan 1998. Contour values range from 0 g kg^{-1} (lightest) to 0.8 g kg^{-1} (darkest) by steps of 0.1 g kg^{-1}.

UTC 1 January 1998). The bias estimate produced by our method represents a spatial and time average of recent observed-minus-forecast residuals at the stations shown. The estimate indicates that the forecasts at that level have been consistently wet lately, according to rawinsonde soundings. The analysis further adjusts the bias-corrected forecast based on the current observations.

The spatial structure of the forecast bias estimate shown in Fig. 9 clearly reflects the distribution of station locations. There is simply no information in our scheme about forecast bias other than that contained in the observations. Far enough away from station locations the bias estimates remain at their initial values, which are zero in our experiments. This limits the total impact of the bias correction on the global moisture analysis. Furthermore, the final GEOS data products are based on the output of the IAU procedure, which consists of an integration of the GCM forced by the analysis increments. It is conceivable that this procedure lessens the effect of the bias correction, even in the vicinity of data locations, as long as the GCM remains defective.

To illustrate these points more clearly, the upper left panel of Fig. 10 shows the January 1998 total precipitable water (TPW, in mm) according to the GEOS 2.8 assimilation. The upper right panel shows the relative

error in this assimilated dataset with respect to TPW derived from SSM/I data (Wentz 1997). These estimates are available over the oceans only, and since we use no moisture data other than those obtained from rawinsonde soundings, the errors are quite large. Cyan/blue shades indicate that, relative to the SSM/I data, GEOS is too wet; green/yellow shades show that GEOS is too dry. This panel gives an indication of the degree of uncertainty in the GEOS moisture estimates. The lower left panel shows the relative difference (in percent) between the GEOS 2.8 assimilated TPW and the TPW computed by vertically integrating the GEOS UNB moisture analyses. We regard this as a measure of the total impact of the forecast bias correction on the moisture analyses. The impact is not very large (compared to the uncertainty implied by the Wentz data) and mostly restricted to land. The lower right panel shows the relative impact of the bias correction on TPW in the assimilation, which appears to be slightly damped by the IAU procedure.

5. Conclusions

The object of an analysis is to make the best possible use of the available observations, given a model forecast and whatever is known about the errors. The observa-
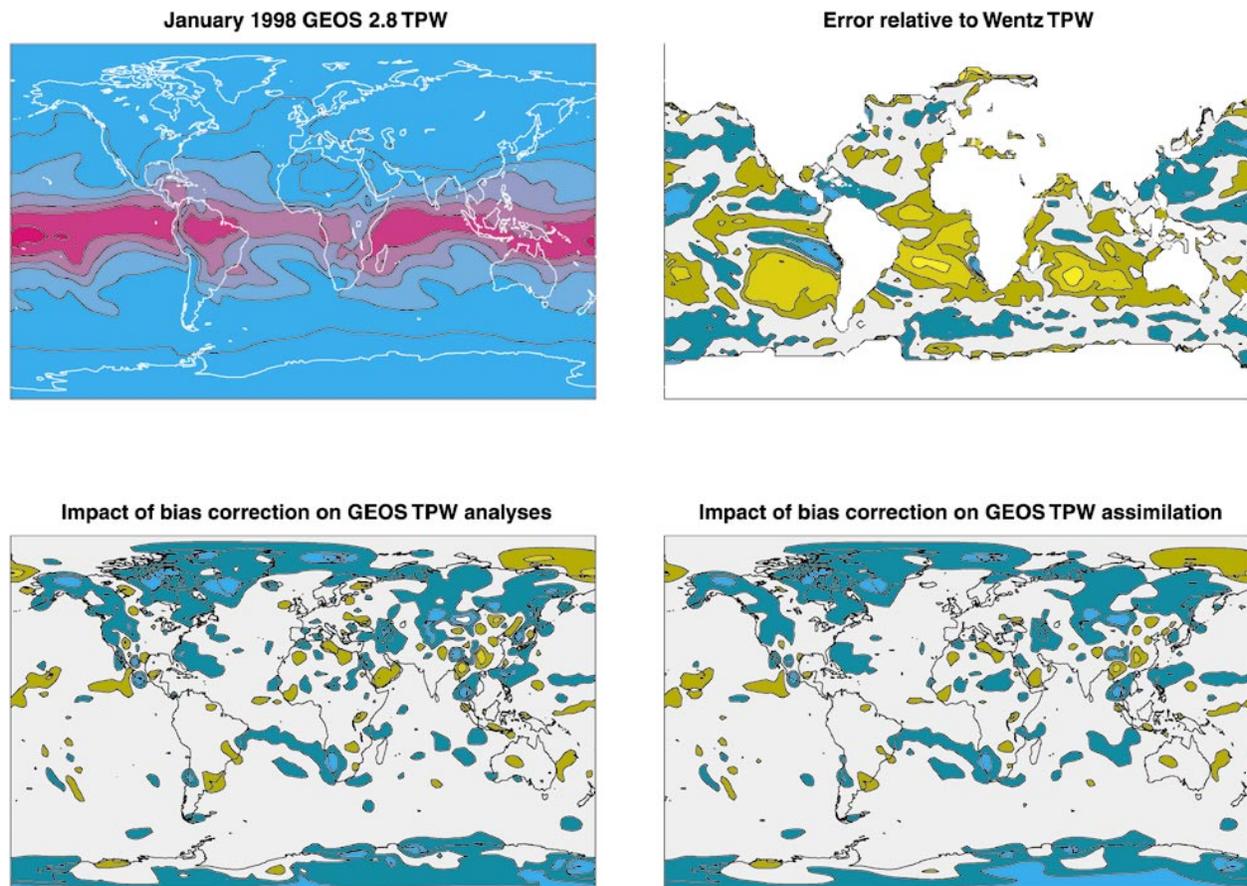
FIG. 10. Upper left panel: Jan 1998 mean total precipitable water in the GEOS 2.8 assimilation. Contour values range from 0 (cyan) to 6 mm (magenta) by steps of 1 mm. Upper right panel: the relative error of the GEOS 2.8 assimilation with respect to the Wentz retrievals. Contour values are −30%, −15%, and −5% (cyan) and 5%, 15%, and 30% (green/yellow). Lower panels, using the same contour values: the relative impact of the forecast bias correction on total precipitable water in the GEOS UNB moisture analyses (left), and the relative impact of the forecast bias correction on total precipitable water in the GEOS UNB moisture assimilation (right).

tions indicate that model errors contain a large persistent component. This means that the data contain useful information about likely errors in subsequent forecasts. A human forecaster would notice and take advantage of this information, being more naturally inclined to think in terms of systematic rather than random model errors. Of course, the representation of model errors by zero-mean white stochastic processes in an automated analysis algorithm is strictly a mathematical device, inspired by computational convenience rather than empirical knowledge.

The unbiased analysis method, introduced in DdS and tested here in the context of the GEOS moisture analysis, includes a simple scheme for estimating forecast bias. Analyses are produced as usual, but only after removing the bias from the forecast. The method relies completely on observations for estimating the bias, although it can be easily generalized to incorporate any additional information about forecast bias that may be available. As explained in DdS, forecast bias estimation is a data assimilation problem in its own right; it requires

unbiased observations (in this case, rawinsonde soundings) supplemented by a model for the bias (in this case, persistence).

This view clearly exposes the main limitations of the method. First, exclusive reliance on observations can be dangerous, since it presumes that effective quality control and bias correction algorithms for the observations are in place. In practice, forecast bias estimates at individual stations must be carefully monitored in order to detect problems with the observations themselves. It is also important to control the impact of an occasional observation in a poorly observed region. Second, in the absence of additional information about forecast bias, the estimates are only meaningful when and where observations are regularly available. This limits the impact of the bias correction in a global analysis system, as long as rawinsonde soundings only are used to estimate the bias. Ultimately, of course, observations that are deemed sufficiently accurate for an analysis should be useful for bias estimation as well. In the moisture case, for example, we intend to investigate the use

of TPW retrievals to obtain nearly global estimates of tropospheric moisture bias.

Our experiments show that the method is extremely effective in predicting moisture forecast bias based on recent observed-minus-forecast residuals. It is therefore able to produce unbiased moisture analyses at all levels in the observed regions. This leads to improved initial conditions for the forecast model and, as a result, a reduction in forecast errors. Specifically, we showed that the 6-h forecast bias is significantly reduced in many cases. Having eliminated analysis bias allows one to identify mean forecast errors with mean model errors, and so the bias estimates produced by the method are in fact estimates of mean model error. These estimates could prove very useful in quantitative studies of model error, which are important both for model development and for advances in data assimilation.

The upshot of our method is that it uses observations more effectively, by removing the assumption that the forecast model is unbiased. It might be argued that this approach offers only a temporary solution tailored to a poorly performing model. Surely, errors will get smaller as models and data improve. However, requirements for analysis and forecast accuracy can be expected to increase accordingly. The need for analysis methods that efficiently extract small-scale information from observations will grow. Such methods must involve adequate descriptions of model errors arising from the treatment of topography, convective parameterizations, etc. It is not at all obvious that the most effective approximation will represent model errors by zero-mean white noise. In fact, we believe that the treatment of systematic discrepancies between model and data will be even more important in future high-performance data assimilation systems.

We plan to extend this work to the complete GEOS analysis system. There is evidence of significant biases and serially correlated model errors in data residuals for all atmospheric variables. The main problems to be addressed in this context are the data selection from nonstationary observers and a reduction of computational expense. We also hope to be able to produce forecast bias estimates for different lead times, and examine whether, say, 24-h forecast skill can be improved by making use of these estimates in real time. Finally, we would like to explore ways to use the forecast bias estimates for identifying specific sources of model error and thereby help improve prediction models.

APPENDIX

Optimal Gains

In order to make claims about consistency and optimality of an estimator we need to be precise about what we wish to estimate and about the errors in our sources of information. The unknowns at time t_k are $\mathbf{w}_k^t$, the true state of the atmosphere, and $\mathbf{b}_k$, the deterministic component of forecast error. Our assumptions on forecast and observations are

$$\mathbf{w}_k^f = \mathbf{w}_k^t + \mathbf{b}_k + \boldsymbol{\varepsilon}_k^f, \quad \langle \boldsymbol{\varepsilon}_k^f \rangle = \mathbf{0}, \quad \langle \boldsymbol{\varepsilon}_k^f (\boldsymbol{\varepsilon}_k^f)^{\mathrm{T}} \rangle = \mathbf{P}_k^f \tag{A1}$$

$$\mathbf{w}_k^o = \mathbf{H}_k \mathbf{w}_k^t + \boldsymbol{\varepsilon}_k^o, \quad \langle \boldsymbol{\varepsilon}_k^o \rangle = \mathbf{0}, \quad \langle \boldsymbol{\varepsilon}_k^o (\boldsymbol{\varepsilon}_k^o)^{\mathrm{T}} \rangle = \mathbf{R}_k \tag{A2}$$

$$\langle \boldsymbol{\varepsilon}_k^o (\boldsymbol{\varepsilon}_k^f)^{\mathrm{T}} \rangle = \mathbf{0}, \tag{A3}$$

and that both noise sequences $\boldsymbol{\varepsilon}_k^o$, $\boldsymbol{\varepsilon}_k^f$ are white. With $\mathbf{b}_k = 0$ these are the standard assumptions for the derivation of the Kalman filter (Jazwinski 1970, section 7.3). Bias parameters were first introduced by Friedland (1969), who derived a sequential bias estimator that is closely related to our unbiased analysis equations. See DdS for a detailed discussion of the relationship between Friedland's algorithm and ours.

We further assume that a prediction $\mathbf{b}_k^p$ of the bias $\mathbf{b}_k$ is available such that

$$\mathbf{b}_k^p = \mathbf{b}_k + \boldsymbol{\varepsilon}_k^b, \quad \langle \boldsymbol{\varepsilon}_k^b \rangle = 0, \quad \langle \boldsymbol{\varepsilon}_k^b (\boldsymbol{\varepsilon}_k^b)^{\mathrm{T}} \rangle = \mathbf{P}_k^b \tag{A4}$$

$$\langle \boldsymbol{\varepsilon}_k^o (\boldsymbol{\varepsilon}_k^b)^{\mathrm{T}} \rangle = \mathbf{0} \tag{A5}$$

$$\langle \boldsymbol{\varepsilon}_k^f (\boldsymbol{\varepsilon}_k^b)^{\mathrm{T}} \rangle = \mathbf{0}. \tag{A6}$$

Assumption (A5) is analogous to (A3); it is clearly true for white observation errors when $\mathbf{b}_k^p$ is a prediction in the true sense of the word, that is, an estimate that does not depend on future data. The cross-covariance assumption (A6) is more troublesome, however, and should be regarded as an approximation. In particular, for our unbiased analysis equations [(3)–(4)] the bias prediction is

$$\mathbf{b}_k^p = \hat{\mathbf{b}}_{k-1}. \tag{A7}$$

The forecast $\mathbf{w}_k^f$ and the bias prediction $\hat{\mathbf{b}}_{k-1}$ both depend on data at t_{k-1}. Clearly, therefore, their respective errors are *not* independent. On the other hand, the forecast errors $\boldsymbol{\varepsilon}_k^f$ involve predictability and model errors as well, so we expect that the neglected cross covariances are relatively small. Substitution of (A1), (A2), (A4), (A7) into (3) gives

$$\hat{\mathbf{b}}_k = \mathbf{b}_k + [\mathbf{I} - \mathbf{L}_k \mathbf{H}_k] \boldsymbol{\varepsilon}_k^b - \mathbf{L}_k [\boldsymbol{\varepsilon}_k^o - \mathbf{H}_k \boldsymbol{\varepsilon}_k^f], \tag{A8}$$

which shows that $\hat{\mathbf{b}}_k$ is an unbiased estimate of $\mathbf{b}_k$ for any $\mathbf{L}_k$. It is also straightforward to show, using the second-moment assumptions in (A1)–(A6), that the best (minimum variance) linear unbiased estimate obtains when $\mathbf{L}_k$ is given by (6). Optimality depends on the correct specifications of all error covariances; in practice, of course, this will not be the case. In particular, as we already pointed out, (A6) is an approximation.

It is not at all obvious, however, that the analysis weights $\mathbf{K}_k$ defined by (2) are optimal. After all, the

observations $\mathbf{w}_k^o$ are used twice: first, for estimating the bias, and second, for computing the analysis. We can write the analysis equation (4) as

$$\tilde{\mathbf{w}}_k^f = \mathbf{w}_k^f - \hat{\mathbf{b}}_k, \tag{A9}$$

$$\mathbf{w}_k^a = \tilde{\mathbf{w}}_k^f + \mathbf{K}_k[\mathbf{w}_k^o - \mathbf{H}_k\tilde{\mathbf{w}}_k^f]. \tag{A10}$$

The unbiased analysis (A10) combines two sources of information—a bias-corrected forecast and a set of observations—whose errors are not independent. In fact, using (A9), (A1), and (A8),

$$\tilde{\boldsymbol{\varepsilon}}_k^f \equiv \tilde{\mathbf{w}}_k^f - \mathbf{w}_k^t = \mathbf{w}_k^f - \hat{\mathbf{b}}_k - \mathbf{w}_k^t = \boldsymbol{\varepsilon}_k^f + \mathbf{b}_k - \hat{\mathbf{b}}_k$$

$$= [\mathbf{I} - \mathbf{L}_k\mathbf{H}_k](\boldsymbol{\varepsilon}_k^f - \boldsymbol{\varepsilon}_k^b) + \mathbf{L}_k\boldsymbol{\varepsilon}_k^o. \tag{A11}$$

The optimal analysis weights, properly accounting for the cross covariances, are (Jazwinski 1970, section 7.3, example 7.5)

$$\mathbf{K}_k = [\tilde{\mathbf{P}}_k^f\mathbf{H}_k^T - \mathbf{X}_k^T][\mathbf{H}_k\tilde{\mathbf{P}}_k^f\mathbf{H}_k^T + \mathbf{R}_k - \mathbf{H}_k\mathbf{X}_k^T - \mathbf{X}_k\mathbf{H}_k^T]^{-1} \tag{A12}$$

where

$$\tilde{\mathbf{P}}_k^f \equiv \langle \tilde{\boldsymbol{\varepsilon}}_k^f(\tilde{\boldsymbol{\varepsilon}}_k^f)^T \rangle \tag{A13}$$

$$\mathbf{X}_k \equiv \langle \boldsymbol{\varepsilon}_k^o(\tilde{\boldsymbol{\varepsilon}}_k^f)^T \rangle. \tag{A14}$$

With $\mathbf{L}_k$ defined by (6), it turns out that (A12) and (2) are identical. The remainder of this appendix proves this statement.

To simplify notation, we omit the subscript k and define

$$\mathbf{P} \equiv \mathbf{P}^b + \mathbf{P}^f, \tag{A15}$$

$$\mathbf{S} \equiv \mathbf{HPH}^T + \mathbf{R}. \tag{A16}$$

Then (A11) and (6) imply that

$$\tilde{\mathbf{P}}^f = [\mathbf{I} - \mathbf{LH}]\mathbf{P}[\mathbf{I} - \mathbf{LH}]^T + \mathbf{LRL}^T$$

$$= \mathbf{P} - \mathbf{P}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^f - \mathbf{PH}^T\mathbf{S}^{-1}\mathbf{HP}^b, \tag{A17}$$

and that

$$\mathbf{X} = \mathbf{RL}^T = \mathbf{RS}^{-1}\mathbf{HP}^b \tag{A18}$$

where we used (A1)–(A6). The first factor of (A12) is

$$\tilde{\mathbf{P}}^f\mathbf{H}^T - \mathbf{X}^T$$

$$= \mathbf{PH}^T - \mathbf{P}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^f\mathbf{H}^T - \mathbf{PH}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T$$

$$\quad - \mathbf{P}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{R}$$

$$= \mathbf{PH}^T - \mathbf{P}^b\mathbf{H}^T\mathbf{S}^{-1}[\mathbf{HP}^f\mathbf{H}^T + \mathbf{HP}^b\mathbf{H}^T + \mathbf{R}]$$

$$\quad - \mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T$$

$$= \mathbf{PH}^T - \mathbf{P}^b\mathbf{H}^T - \mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T$$

$$= \mathbf{P}^f\mathbf{H}^T[\mathbf{I} - \mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T]$$

$$= \mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}[\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}]. \tag{A19}$$

The inverse of the second factor is

$$\mathbf{H}\tilde{\mathbf{P}}^f\mathbf{H}^T + \mathbf{R} - \mathbf{HX}^T - \mathbf{XH}^T$$

$$= \mathbf{HPH}^T - \mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^f\mathbf{H}^T - \mathbf{HPH}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T$$

$$\quad + \mathbf{R} - \mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{R} - \mathbf{RS}^{-1}\mathbf{HP}^b\mathbf{H}^T$$

$$= \mathbf{S} - \mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}[\mathbf{HP}^f\mathbf{H}^T + \mathbf{HP}^b\mathbf{H}^T + \mathbf{R}]$$

$$\quad - \mathbf{HP}^f\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T - \mathbf{RS}^{-1}\mathbf{HP}^b\mathbf{H}^T$$

$$= \mathbf{S} - \mathbf{HP}^b\mathbf{H}^T - [\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}]\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T$$

$$= \mathbf{S} - 2\mathbf{HP}^b\mathbf{H}^T + \mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T$$

$$= [\mathbf{I} - \mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}][\mathbf{S} - \mathbf{HP}^b\mathbf{H}^T]$$

$$= [\mathbf{I} - \mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}][\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}]. \tag{A20}$$

Combining the two factors,

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}[\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}][\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}]^{-1}$$

$$\quad \times [\mathbf{I} - \mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}]^{-1}$$

$$= \mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}[\mathbf{I} - \mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}]^{-1}$$

$$= \mathbf{P}^f\mathbf{H}^T[\mathbf{S} - \mathbf{HP}^b\mathbf{H}^T]^{-1}$$

$$= \mathbf{P}^f\mathbf{H}^T[\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}]^{-1}. \tag{A21}$$

REFERENCES

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.

Chen, M., R. B. Rood, and W. G. Read, 1998: Upper tropospheric water vapor from GEOS reanalysis and UARS MLS observation. *J. Geophys. Res.,* **103,** 19 587–19 594.

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

Daley, R., 1992: The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Wea. Rev.,* **120,** 164–177.

Dee, D. P., 1991: Simplification of the Kalman filter for meteorological data assimilation. *Quart. J. Roy. Meteor. Soc.,* **117,** 365–384.

——, 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123,** 1128–1145.

——, and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.,* **124,** 269–295.

——, and ——, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.,* **127,** 1822–1834.

——, G. Gaspari, C. Redder, L. Rukhovets, and A. M. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: Applications. *Mon. Wea. Rev.,* **127,** 1835–1849.

DelSole, T., and A. Y. Hou, 1999: Empirical correction of a dynamical model. Part I: Fundamental issues. *Mon. Wea. Rev.,* **127,** 2533–2545.

Friedland, B., 1969: Treatment of bias in recursive filtering. *IEEE Trans. Autom. Contr.,* **AC-14,** 359–367.

Griffith, A. K., and N. K. Nichols, 1996: Accounting for model error in data assimilation using adjoint methods. *Computational Differentiation: Techniques, Applications, and Tools, Proc. Second Int. SIAM Workshop on Computational Differentiation,* Santa Fe, NM, Society for Industrial and Applied Mathematics, 195–204.

Hou, A. Y., D. Ledvina, A. da Silva, S. Zhang, J. Joiner, R. Atlas, G. Huffman, and C. 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.

Joiner, J., and L. Rokke, 2000: Variational cloud-clearing with TOVS data. *Quart. J. Roy. Meteor. Soc.,* **126,** 1–24.

Kailath, T., 1968: An innovations approach to least-squares estimation—Part I: Linear filtering in additive white noise. *IEEE Trans. Autom. Contr.,* **AC-13,** 646–655.

Mitchell, H. L., C. Chouinard, C. Charette, R. Hogue, and S. J. Lambert, 1996: Impact of a revised analysis algorithm on an operational data assimilation system. *Mon. Wea. Rev.,* **124,** 1243–1255.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in FORTRAN: The Art of Scientific Computing.* 2d ed. Cambridge University Press, 963 pp.

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

Suarez, M. J., 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 Goddard Space Flight Center, Greenbelt, MD 20771.]

Takacs, L. L., 1996: A simple bias correction algorithm for use in data assimilation. DAO Office Note 96-25, 31 pp. [Available from Data Assimilation Office, Goddard Space Flight Center, Greenbelt, MD 20771.]

Thiébaux, H. J., and L. L. Morone, 1990: Short-term systematic errors in global forecasts: Their estimation and removal. *Tellus,* **42A,** 209–229.

Wentz, F. J., 1997: A well-calibrated ocean algorithm for SSM/I. *J. Geophys. Res.,* **102,** 8703–8718.

Zipser, E. J., and R. H. Johnson, 1998: Systematic errors in radiosonde humidities: A global problem? Preprints, *10th Symp. Meteorological Observations and Instrumentation,* Phoenix, AZ, Amer. Meteor. Soc., 72–76.