Bias and data assimilation*

Dick P. Dee

ECMWF, Shinfield Park, Reading
RG2 9AX, United Kingdom
D.Dee@ecmwf.int

ABSTRACT

All data assimilation systems are affected by biases, caused by problems with the data, by approximations in the observation operators used to simulate the data, by limitations of the assimilating model, or by the assimilation methodology itself. A clear symptom of bias in the assimilation is the presence of systematic features in the analysis increments, such as large persistent mean values or regularly recurring spatial structures. Bias can also be detected by monitoring statistics of observed-minus-background residuals for different instruments. Bias-aware assimilation methods are designed to estimate and correct systematic errors jointly with the model state variables. Such methods require attribution of a bias to a particular source, and its characterization in terms of some well-defined set of parameters. They can be formulated either in a variational or sequential estimation framework by augmenting the system state with the bias parameters.

1 Introduction

Textbook data assimilation theory is primarily concerned with the problem of optimally combining model predictions with observations in the presence of random, zero-mean errors. In reality, errors in models and data are often systematic rather than random. Model errors caused by inaccurate surface forcing, poor resolution of the boundary layer, simplified representations of moist physics and clouds, and various other imperfections, are not well represented by random noise. Satellite observations contain instrument-dependent biases that are often larger than the amplitude of the useful signal, and approximations in radiative transfer calculations can cause complex, state-dependent systematic errors in the assimilation. Many conventional observations are biased as well, e.g., daytime high-altitude radiosonde temperatures due to solar radiation effects; measurements taken close to the ground due to inaccurate station elevation information and errors in the model's surface representation; cloud-drift derived wind observations due to errors in cloud-top height assignment.

Considerable efforts are made to remove biases from models and observations, particularly at operational centers, yet their effect on the quality of assimilated data products remains significant. In the context of numerical weather prediction, the presence of residual biases means that the available data are not used optimally, and in some cases cannot be used at all. In the realm of climate research based on re-analyzed data sets, it can be extremely difficult to separate real signals and trends from spurious ones caused by biases in models and data. Figure 1 provides a schematic illustration of this problem. If unbiased observations are assimilated using a biased model, then the model drift causes a positive bias in the assimilation. The size of the bias depends on the accuracy as well as the frequency of the observations. As a result, a change in characteristics of the observing system, even if all observations are unbiased, leads to what might be perceived as an apparent change in climate. See Santer et al. 2004 for an interesting account of dealing with these types of complications in attempting to isolate real climate signals using state-of-the-art assimilated data sets and statistical analysis techniques.

We consider the term *bias* to broadly include any type of error that is systematic rather than random. In statistics, bias is a property of an estimator which, on average, under- or over-estimates some quantity. For example, a model which is consistently cold at some location is biased. The bias may be spatially variable, seasonal,

*To appear in *Q. J. R. Meteorol. Soc.*, 2006.

diurnal, or even situation-dependent. If we allow some flexibility with respect to the notion of a model (or an observing system) as an *estimator*, and with the operative definition of *averaging*, then any component of error that is systematic in some well-defined sense can be considered a bias. This is consistent with the usage of human forecasters, who describe, for example, the tendency of a particular model to generate excessive surface lows in certain recurring situations as a bias.

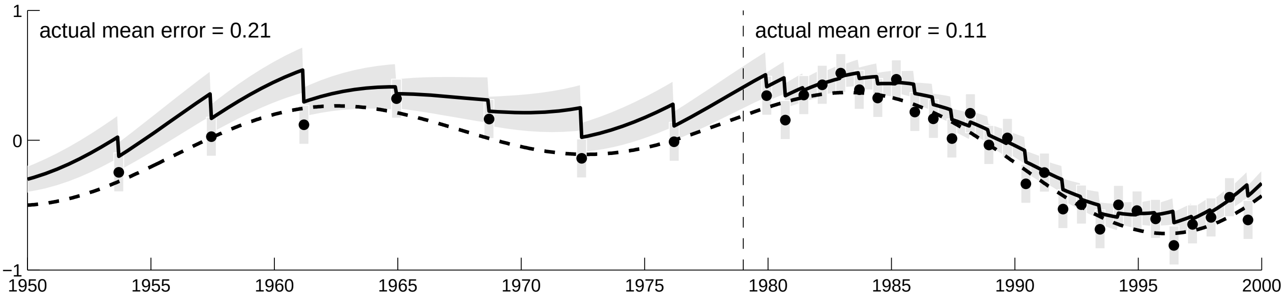

Figure 1: Assimilation of unbiased observations in a biased model, and the effect of observing frequency on the apparent climate. The dashed curve represents the true state evolution, observations are indicated by the dots, and the solid curve is the assimilation.

Data assimilation systems that are designed to correct random, zero-mean errors in a model-generated background estimate based on unbiased observations might be called *bias-blind*. Routine monitoring of observed-minus-background residuals (also known as innovations, background residuals, or background departures) in bias-blind systems invariably shows evidence of biases in either the model, the observations, or both. Similarly, the presence of persistent or repetitive patterns in the analysis increments produced during the assimilation indicates that there are systematic discrepancies between model and observations, and possibly among different components of the observing system as well. To effectively remove those discrepancies during the data assimilation process requires *bias-aware* assimilation methods, which incorporate specific assumptions about the source and nature of (some of) the biases in the system, and are specifically designed to estimate and correct those biases.

2 Bias-blind data assimilation

Data assimilation in practice is essentially a sequential procedure, in which a model integration is periodically adjusted on the basis of actual observations confined to a finite time window. While the length of the window and many other specifics may vary, most assimilation methods are similar in that observations $\mathbf{y}$ are combined with a model-generated state estimate $\mathbf{x}^b$ (the *background*) by minimizing a functional

$$J(\mathbf{x}) = (\mathbf{x}^b - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x}^b - \mathbf{x}) + [\mathbf{y} - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{h}(\mathbf{x})] \tag{1}$$

with respect to the model state $\mathbf{x}$. The function $\mathbf{h}(\cdot)$ denotes a set of observation operators used to express the relationship between model state and observations; this may involve integration of the model in a 4D-Var system. The matrices $\mathbf{B}$ and $\mathbf{R}$ represent covariance operators usually associated with background and observation errors, respectively; the latter includes the effects of approximations in the observation operators. The minimizing solution $\mathbf{x} = \mathbf{x}^a$ (the *analysis*) satisfies the nonlinear equation

$$\mathbf{x}^a - \mathbf{x}^b = \mathbf{B} \left(\left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}^a} \right)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{h}(\mathbf{x}^a)) \tag{2}$$

obtained by setting the gradient of $J(\mathbf{x})$ to zero. An important implication of (2) is that all possible adjustments to the background are confined to the range of $\mathbf{B}$. This explains why the specification of background error covariances is so important to the performance of a data assimilation system and its ability to absorb and retain observational information. Interestingly, this statement does not necessarily depend on whether $\mathbf{B}$ actually provides an accurate representation of background error covariances.

Methods based on (1) are *bias-blind*, since they are designed to correct random errors only. We can see this most clearly by linearizing (2) to obtain the familiar analysis equation

$$\mathbf{dx} = \mathbf{Kdy} \tag{3}$$

where $\mathbf{dx} = \mathbf{x}^a - \mathbf{x}^b$ denotes the analysis increment and $\mathbf{dy} = \mathbf{y} - \mathbf{h}(\mathbf{x}^b)$ is the vector of observed-minus-background residuals, and the gain operator $\mathbf{K}$ (or analysis weight matrix) is given by

$$\mathbf{K} = \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}, \qquad \mathbf{H} = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^b} \tag{4}$$

In terms of the analysis, background, and observation errors defined by

$$\mathbf{e}^a = \mathbf{x}^a - \mathbf{x}, \qquad \mathbf{e}^b = \mathbf{x}^b - \mathbf{x}, \qquad \mathbf{e}^o = \mathbf{y} - \mathbf{h}(\mathbf{x}) \tag{5}$$

with $\mathbf{x}$ the unknown true state (defined in model space), (3) implies

$$\mathbf{e}^a \approx \mathbf{K}\mathbf{e}^o + [\mathbf{I} - \mathbf{KH}]\mathbf{e}^b \tag{6}$$

To first order, therefore, any biases in the model background or in the observations are linearly transferred to the analysis:

$$< \mathbf{e}^a > \approx < \mathbf{K}\mathbf{e}^o > + < [\mathbf{I} - \mathbf{KH}]\mathbf{e}^b > \tag{7}$$

where $< \cdot >$ represents linear averaging over a sufficiently large ensemble. If either the background or the observations are biased, then the analysis is biased regardless of the specification of the gain operator $\mathbf{K}$. In practice one can adjust $\mathbf{K}$ to reduce the bias in the analysis, but this will introduce additional noise as a result. If estimates of the biases are available, Dee and da Silva (1998, Section 2) show how to modify $\mathbf{K}$ in order to minimize the total (root-mean-square) analysis error. However, the resulting rms-optimal analyses are still biased, and noisier than the optimal analysis in a system which is free of bias; see Fig. 2. It is not possible to produce an unbiased analysis from a biased background and/or biased observations with a bias-blind analysis method.

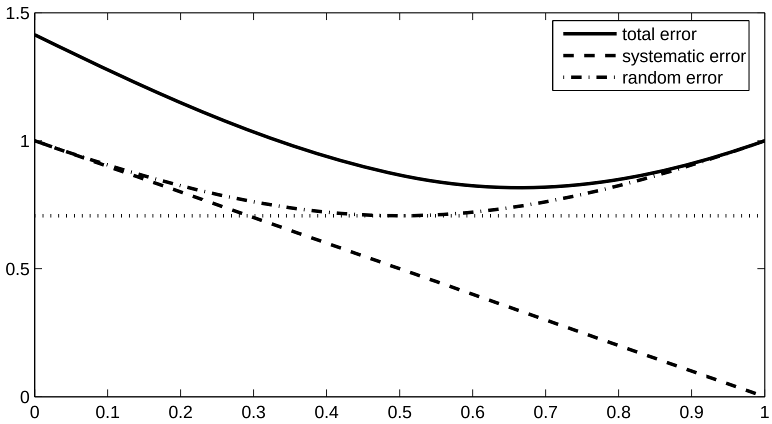

Figure 2: Analysis error as a function of the gain $\mathbf{K}$*, given a single unbiased observation with error standard deviation* σ *and a background estimate with bias* $b = \sigma$ *and error standard deviation* σ*. The dotted horizontal indicates the minimum analysis error obtainable when* $b = 0$*.*

2.1 Bias detection using analysis increments

How prevalent are biases in state-of-the-art assimilation systems? The easiest way to detect the presence of bias is to see whether behavior of the type illustrated in Fig. 1 occurs, i.e., whether the analysis has a tendency to make systematic corrections to the model background. In the ideal (bias-free) situation we should expect mean analysis increments close to zero:

$$< \mathbf{dx} > \approx < \mathbf{Ke}^o > - < \mathbf{KHe}^b > \approx 0 \tag{8}$$

Non-zero mean increments are significant when they are large compared to the size of a typical increment at any given time. A typical increment depends primarily on the configuration of the observing system, but also on the quality of the background. In fact, (3–5) imply that

$$< \mathbf{dx}\, \mathbf{dx}^T > \approx \mathbf{KHB} \tag{9}$$

provided background and observation error covariance specifications are reasonably accurate.

Figure 3 shows the zonal monthly mean for August 2002 of temperature increments produced in the ERA-40 reanalysis. There is clear evidence of biases in the system. The most conspicuous features are the mean stratospheric increments exceeding 1K in an alternating positive-negative pattern, and similar oscillating increments over the Southern high latitudes descending into the troposphere. Maps at stratospheric levels of monthly mean increments at different times of day (not shown here) indicate persistent large-scale biases of opposite sign in different geographic areas, that roughly coincide with the locations of the available satellites. Overall, the mean increments represent a large fraction of typical increments, even in the middle troposphere where the model is relatively skillful and observations are abundant. Similar plots for humidity, ozone, and other variables throughout the ERA-40 period are publicly accessible at the European Centre for Medium-Range Weather Forecasts (ECMWF) web site (http://www.ecmwf.int).

While the mean analysis increments in ERA-40 clearly indicate the presence of substantial biases, additional information is needed to identify their sources. The main stratospheric biases in the assimilation are likely caused by model errors, which are known to be large and systematic in the stratosphere. The problem is complicated by the fact that the available observations are biased there as well. The main source of data in the middle and upper stratosphere used in ERA-40 consists of radiances obtained from TOVS/ATOVS instruments carried on successive generations of NOAA polar orbiting satellites (Hernandez et al. 2004). These have been corrected for biases related to scan angle and air mass using off-line tuning procedures described in Harris and Kelly (2001). The air-mass dependent bias correction is primarily designed to account for inaccuracies in the fast radiative transfer calculations that are used in the assimilation. For lack of a true reference, this correction relies on a small set of predictors that are computed from the model background. It is possible, therefore, that model biases are supported or even reinforced by the radiance assimilation in areas where few other observations exist.

The bias problems discussed here are by no means unique to ERA-40 but appear to exist in many global atmospheric data assimilation systems. Langland (2005, pers. comm.) has noted striking similarities in mean temperature analysis increments produced by the Naval Research Laboratory's Atmospheric Variational Data Assimilation System (NAVDAS). Polavarapu et al. (2006) discuss identical problems with the assimilation of stratospheric data in the Canadian Middle Atmosphere Model (CMAM), and suggest that non-physical features of the increments are closely related to the specification of background error covariances in their system. This is a consequence of the fact that the background covariance operator controls the structures of analysis increments, cf. (2). In particular, the vertical structure of the increments in the Southern Hemisphere, so evident in Fig. 3, may simply reflect extrapolation by the analysis of large corrections made near the stratopause (McNally 2004). Most data assimilation systems are not equipped to handle large, systematic corrections; they were designed to make small adjustments to the background fields that are consistent with the presumed multivariate and spatial structures of random errors.

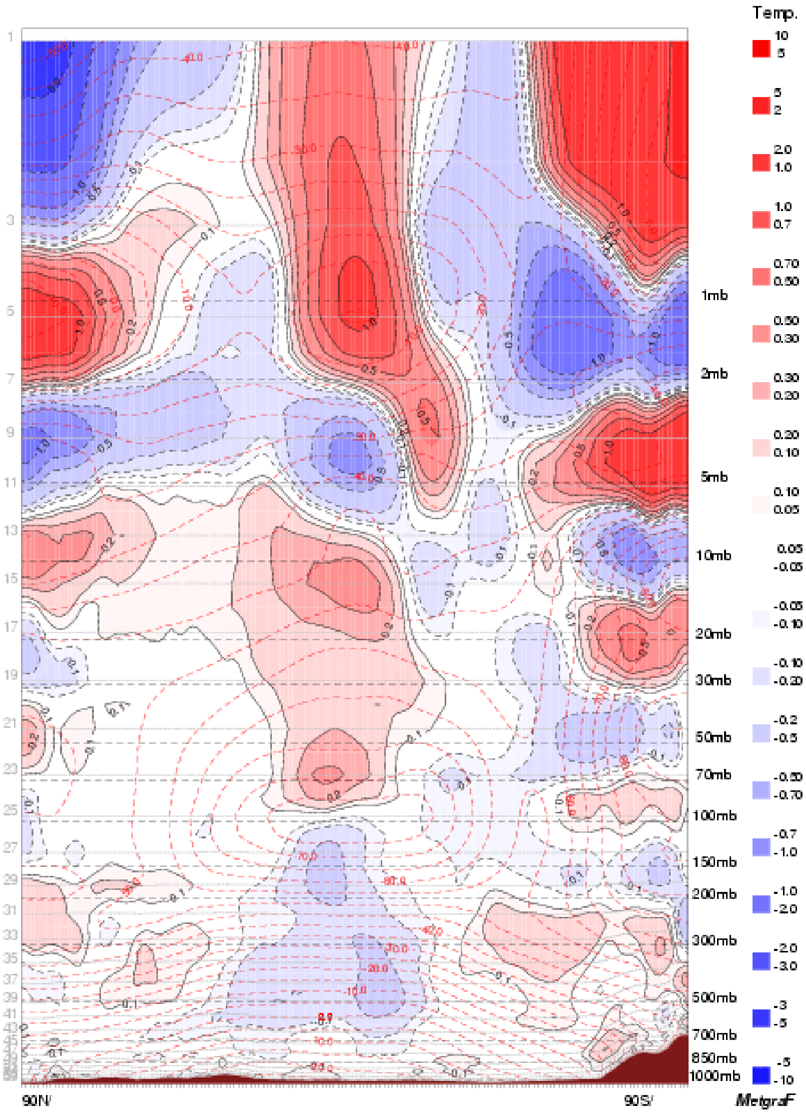

Figure 3: Zonal mean of time-averaged temperature increments [K] produced in the ERA-40 reanalysis for August 2002. Legend for color shading is on the right. Red dashed contours indicate mean analyzed temperatures. Graphic provided by courtesy of the ECMWF.

2.2 Bias detection using background residuals

Statistics of observed-minus-background residuals provide a different, sometimes more informative, view on systematic errors in model or observations. Operational NWP centers routinely monitor time- and space-averaged background residuals associated with different components of the observing system, providing a wealth of information on the quality of the input data as well as on the performance of the assimilation system. In general, small root-mean-square residuals imply that the system is able to accurately predict future observations. Non-zero mean residuals, however, indicate the presence of biases in the observations and/or their model-predicted equivalents, since

$$< \mathbf{dy} > \approx < \mathbf{e}^o > - < \mathbf{He}^b > \tag{10}$$

There is no general method for identifying bias sources based on (10) alone. However, an observed change in the residual mean for a particular component of the observing system may indicate, for example, a developing bias in that component, or even the impending failure of an instrument. Early detection of such problems is, in fact, one of the main functions of an operational monitoring system. More generally, combined information about residual statistics for different (perhaps overlapping) components of the observing system can lead to useful insights into sources of bias, possibly in the model, which can then be further explored.

2.2.1 Weather time scales.

While basic statistics such as the time-mean and standard deviation are useful for detecting persistent errors, additional information can be gleaned from time series of observed-minus-background residuals by considering their spectral properties. The well-known *innovation property* (Anderson and Moore 1979, Theorem 3.1; Daley 1992) states that the background residuals are white in time (not serially correlated) if the analysis is optimal. Dee and Todling (2000) computed normalized power spectra of radiosonde humidity residuals obtained from the Goddard Earth Observing System (GEOS) data assimilation system to show clear evidence of suboptimality and the presence of systematic errors on time scales on the order of 5-10 days in the assimilation. The normalized spectrum for an individual station can be computed using an algorithm designed for unevenly spaced data due to Lomb (1976) as described in Section 13.8 of Press et al. (1992). To illustrate, Fig. 4 shows spectra of radiosonde temperature observed-minus-background residuals obtained from the National Centers for Environmental Prediction (NCEP) global assimilation system, averaged over all Northern-hemisphere stations, plotted at various pressure levels as a function of the wave period in days. Due to the normalization the curves should be flat for white residuals, even when the time-mean and standard deviations vary by station. There is excessive power in periods longer than 10 days, as well as a strong peak in the diurnal cycle. Near the surface this peak may reflect systematic under-estimation by the model of the mean diurnal temperature variation; at higher levels it is probably caused by remaining solar radiation bias in the radiosonde temperature observations.

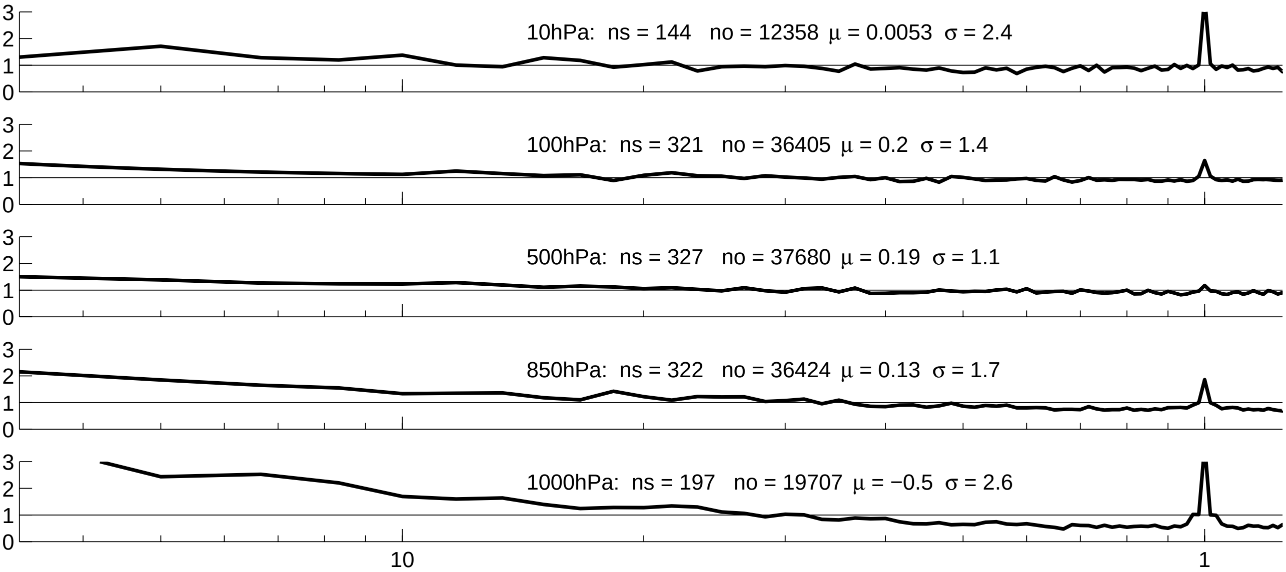

Figure 4: Normalized power spectra of NCEP temperature observed-minus-background residuals for Jan-Feb 2005, averaged over all Northern Hemisphere radiosonde stations with at least 50 reports during the period. The horizontal axis indicates the wave period in days. Printed in each panel are: pressure level; number of stations used; number of observations used; mean of the residuals; standard deviation of the residuals.

2.2.2 Seasonal time scales.

McNally (2004) provides an interesting and convincing example of stratospheric model bias detection, primarily based on a study of residual statistics obtained from Advanced Microwave Sounding Unit A (AMSUA)

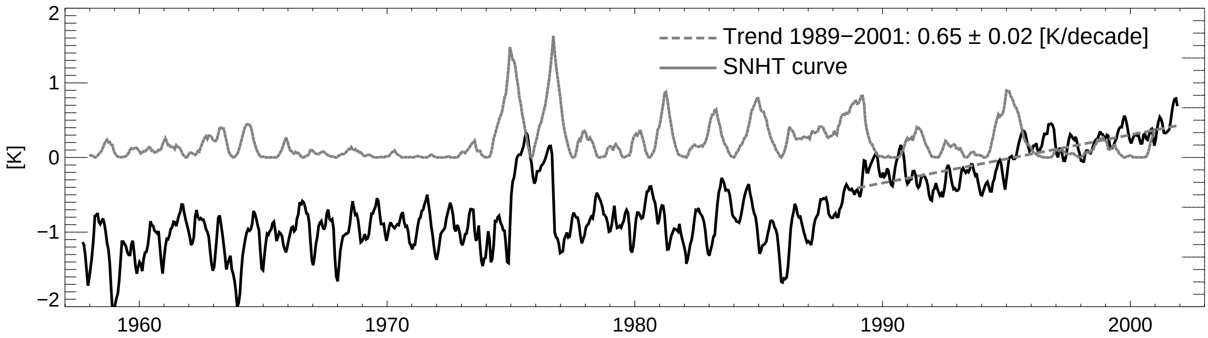

Figure 5: Solid black: Uncorrected background-minus-observed temperatures at 50hPa, averaged over all radiosonde stations south of 25N. Solid gray: Probability of a break in the time series, based on a variant of the Standard Normal Homogeneity Test. Dashed gray: Linear trend in the mean background temperatures during 1989-2001. Graphic provided by courtesy of L. Haimberger.

radiance data. The time evolution of the mean background residuals for AMSUA channel 14 brightness temperatures, which are mainly sensitive to upper stratospheric temperatures, was shown to exhibit a large seasonal variation with an average amplitude of about 3K, and with opposite phases in the two hemispheres. These characteristics strongly point to the model as the dominant source of bias, and this was confirmed by careful cross-comparison with independent research data. McNally (2004) also discusses other, more subtle, aspects of the bias problem related to radiative transfer calculations, as well as the spurious vertical structures in the temperature increments that are imposed by the background error covariance formulation.

2.2.3 Climate time scales.

Haimberger (2005) describes an automated scheme for a posteriori elimination of artificial breaks or jumps in historical radiosonde station data. These breaks in the time series are often caused by equipment changes at individual stations, which have not always been properly documented. The main reference for break detection used in this study is the time sequence of globally averaged background temperature fields produced in the ERA-40 reanalysis. A major challenge in this approach is that the background estimates themselves contain spurious trends, which must be accounted for before corrections can be made to the radiosonde data. Haimberger's Fig. 19, reproduced here as Fig. 5, illustrates this problem quite well. The black curve represents the uncorrected background-minus-observed temperatures at 50hPa, averaged over all stations south of 25N. The solid gray curve is a test statistic used to measure the probability of a break in the time series. Three bias-related problems are clearly noticeable in the uncorrected residuals. First, the jump of about 1K during 1975 and most of 1976 was caused by an erroneous bias correction of the NOAA-4 radiances. Second, the increase in the mean residuals between 1985-1990 is due to the gradual replacement of radiosonde equipment in Australia and the Pacific. Finally, the trend in the 1990s (indicated by the gray dashed line segment) has been identified with warming due to excessive tropical precipitation, associated with the assimilation of increasing amounts of humidity data during this period (Andersson et al. 2005). Haimberger's correction scheme incorporates a method for removing biases from the background reference, which essentially relies on the assumption that these biases are global, as opposed to station biases, which are local. Haimberger (2005) provides highly recommended reading for anyone interested in the difficulties, subtleties, and practical aspects of bias correction in data assimilation.

3 Bias-aware data assimilation

Some data assimilation methods are designed to estimate parameters that represent systematic errors in the system, simultaneously with the model state variables themselves. In the following sections we will describe several examples of such bias-aware methods, but we first raise some general issues that pertain to all of them.

By design, bias-aware assimilation requires assumptions about the nature of the biases: first, the attribution of a bias to a particular source, and second, a characterization of the bias in terms of some well-defined set of parameters. The three diagrams in Fig. 6 roughly indicate what happens to the assimilation when a bias is attributed to the model, to the observations, or to neither, as in a bias-blind assimilation. The need to attribute errors to their proper sources is obvious in any data assimilation system, but becomes especially critical when it involves bias correction. This is because a wrong attribution will force the assimilation to be consistent with a biased source. If the source of a known bias is uncertain, bias-blind assimilation may be the safest option.

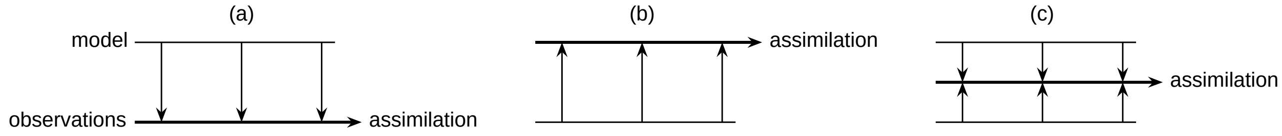

Figure 6: Assimilation with bias attribution to the model (a), to the observations (b), to neither the model nor the observations (c).

In general, bias estimation requires the formulation of a model for the bias, as well as a reference data set from which to estimate the parameters of this bias model. Both requirements involve difficult choices. For example, biases associated with radiative transfer errors are often modeled with flow-dependent predictors (Eyre 1992; Derber and Wu 1998; Harris and Kelly 2001); the bias parameters to be estimated in this case are the predictor coefficients. The choice of predictors for a particular sensor, while clearly important, is far from obvious. Bias modeling for satellite radiances is still an active area of research; alternative models involving physical parameters of the radiative transfer have been proposed by Joiner et al. (1998) and by Watts and McNally (2004).

For biases associated with systematic model errors, such as the stratospheric biases discussed earlier, it is even more difficult to develop useful representations of the biases themselves or of their generation mechanisms. One possibility is to directly model the bias in the background fields by assuming persistence or some other type of prescribed time behavior (Dee and da Silva 1998; Dee and Todling 2000; Radakovich 2001; Lamarque et al. 2004; Chepurin et al. 2005). The advantage of this approach is that background errors are observable (cf. (10)), which makes it relatively straightforward to formulate a consistent bias estimation scheme. Nevertheless, it would be preferable to estimate tendency errors that lead to the bias in the background fields, if this could be used to suppress bias generation during the integration of the model (Derber 1989; Radakovich 2001; Bell et al. 2004; Balmaseda et al. 2006). Much more research is needed in this area, including work in the directions set out by Tsyroulnikov (2006), who has begun to address the problem of developing advanced stochastic representations of model errors that are consistent with the spatial and temporal structures of the forecast errors they generate.

Finally, a true (unbiased) reference is needed to estimate the parameters of a given bias model. In practice it is necessary to resort to surrogates such as independent observations, model background fields, or analyses. For example, radiance bias parameters have been estimated from observed-minus-background residuals collected over time (Harris and Kelly 2001), but also from collocated radiosonde data in a sequential updating procedure (Joiner and Rokke 2000). Apart from sampling issues that must be considered in any statistical estimation scheme, there is a risk that bias in the reference data ultimately gets attributed to the wrong source. If bias parameters are estimated jointly with the model state in a bias-aware assimilation scheme, then the final state estimate (i.e., the analysis) serves as the implicit reference for the bias estimation. This means that all available information is assimilated in a consistent manner, but it does not guarantee that the biases have been attributed correctly.

3.1 Variational analysis methods

It is conceptually straightforward to estimate bias parameters along with the model state in a variational analysis, as long as the relationships among parameters and state components are well-defined. The general idea is to introduce an augmented control vector

$$\mathbf{z}^T = \left[\mathbf{x}^T\, \beta^T\right] \tag{11}$$

that includes the parameters β as well as the model state $\mathbf{x}$. The analysis is then obtained by minimizing

$$J(\mathbf{z}) = (\mathbf{z}^b - \mathbf{z})^T \mathbf{Z}^{-1} (\mathbf{z}^b - \mathbf{z}) + [\mathbf{y} - \tilde{\mathbf{h}}(\mathbf{z})]^T \mathbf{R}^{-1} [\mathbf{y} - \tilde{\mathbf{h}}(\mathbf{z})] \tag{12}$$

with respect to the new control vector $\mathbf{z}$, whose background estimate $\mathbf{z}^b$ must now include a prior estimate β^b of the bias parameters. We use the notation $\tilde{\mathbf{h}}$ to indicate that the observation operator may depend on (some of) the newly introduced parameters. The matrix $\mathbf{Z}$ represents an augmented background error covariance operator, which, in principle, includes cross-covariances among parameters and state vector components. Implementation of this approach in practice requires a workable approximation for these covariances, as well as an efficient minimization algorithm.

3.1.1 Variational bias correction of radiance data.

Variational bias correction of satellite radiances was first implemented at NCEP in their spectral statistical interpolation (SSI) analysis system (Derber and Wu 1998), and more recently at ECMWF (Dee 2004). Both implementations rely on linear predictor models for the air-mass dependent component of the bias, although the choice of predictors differs in the two systems. In (12) we therefore have

$$\tilde{\mathbf{h}}(\mathbf{z}) = \mathbf{h}(\mathbf{x}) + \mathbf{b}(\beta, \mathbf{x}), \qquad \mathbf{b}(\beta, \mathbf{x}) = \sum_{i=0}^{N_p} \beta_i \mathbf{p}_i(\mathbf{x}) \tag{13}$$

where $\mathbf{b}$ is the bias model and the $\mathbf{p}_i$ are the predictors. Typically $\mathbf{p}_0$ is constant while the remaining predictors are functionals of the state at the observation locations, such as tropospheric thickness, integrated lapse rate, etc. Only a few predictors are used in order not to over-fit the biases, but the predictor coefficients for each channel and each sensor are allowed to be different. The total number of radiance bias parameters included in the system is therefore roughly $N = N_p \times N_s \times N_c$, where N_p is the number of predictors used, N_s is the number of sensors being assimilated, and N_c is the number of channels per sensor. The dimension N of the parameter vector is very small compared to the dimension of the state vector $\mathbf{x}$, so it should not be costly to perform the bias correction during the minimization.

The background estimate β^b for the predictor coefficients is usually just the latest estimate obtained from the previous analysis. The errors in this estimate are generally correlated with the state estimation errors, because they depend on the same data. For lack of quantifiable information about these correlations, however, the background error covariances in (12) are specified as

$$\mathbf{Z} = \begin{bmatrix} \mathbf{B}_{\mathbf{x}} & 0 \\ 0 & \mathbf{B}_\beta \end{bmatrix} \tag{14}$$

with $\mathbf{B}_{\mathbf{x}}$ the (state) background error covariances, and $\mathbf{B}_\beta$ the parameter background error covariances. Written in terms of $\mathbf{x}$ and β, (12) then becomes

$$\begin{aligned} J(\mathbf{x}, \beta) = {} & (\mathbf{x}^b - \mathbf{x})^T \mathbf{B}_{\mathbf{x}}^{-1} (\mathbf{x}^b - \mathbf{x}) \\ & + (\beta^b - \beta)^T \mathbf{B}_\beta^{-1} (\beta^b - \beta) \\ & + [\mathbf{y} - \mathbf{h}(\mathbf{x}) - \mathbf{b}(\mathbf{x}, \beta)]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{h}(\mathbf{x}) - \mathbf{b}(\mathbf{x}, \beta)] \end{aligned} \tag{15}$$

The first term is the usual background term for the state vector, cf. (1). The second term represents the background constraint on the bias parameters. It controls the adaptivity of the estimates: a strong constraint means that the parameter updates in each analysis cycle are small, while a weak constraint (or no constraint at all) implies that the parameter estimates respond quickly to the latest observations. The third term is the bias-adjusted observation term.

Efficient minimization of the functional (15) requires the ability to evaluate its gradient with respect to all control variables, including the bias parameters. This means that the adjoint of the bias model must be available, which is a simple matter for the linear additive bias model in (13) but may be more complicated when bias parameters are deeply embedded in the radiative transfer calculations. In addition, the inclusion of bias parameters in the minimization severely affects the conditioning of the problem; see Dee (2004) for further discussion of this issue.

An important practical advantage of an adaptive bias correction system for satellite radiances is that it reduces the need for manual tuning procedures, which are tedious and prone to error, especially in view of the large number and variety of sensors being assimilated (Thépaut 2003). The system will automatically adjust the bias for a given channel in order to maintain consistency with all available information. Adaptive bias correction will compensate for slow drift that may occur in some channels, but can also handle sudden changes due to unexpected events. This is illustrated in Fig. 7, which shows the evolution of the bias corrections, residual statistics, and data counts for channel 3 brightness temperatures of the Microwave Sounding Unit (MSU) on NOAA-9 over a four-month period. The bias in this channel changed abruptly on 1 November 1986 and again on 4 December 1986, possibly due to solar flares. The initial reponse of the assimilation system is to reject most of the observations in the quality control step. However, with the remaining data the analysis immediately begins to adjust the bias estimates, and then more data are gradually returned to the system over the next few days. It can be seen from the plot that the noise in the residuals does not change during this period, which suggests that there may still be useful information in this channel.

3.1.2 Variational correction of systematic model errors.

There is a large body of work concerning variational formulations of the data assimilation problem that can account for model errors, starting with Sasaki (1970). The standard formulation, which assumes that model errors are random, additive, and white (e.g. Ménard and Daley 1996), is not specifically designed to correct systematic model errors. However it is conceivable that the additional degrees of freedom introduced into the system could effectively force the model to an unbiased state. In the Variational Continuous Assimilation technique proposed by Derber (1989), the control variables used for the minimization represent model tendencies rather than the model state itself. Zupanski (1997) developed a regional weak-constraint 4D-Var system in which the control variable includes both the model state at initial time and serially correlated model error represented by a first-order Markov process. Griffith and Nichols (2000) similarly proposed schemes for correcting model errors in a variational framework, including persistent tendency errors. Trémolet (2003) has recently implemented a weak-constraint formulation of the ECMWF operational assimilation system, and preliminary experiments have shown that this system can be effective in reducing the impact of stratospheric temperature model biases. At the Naval Research Laboratory, an observation-space variational data assimilation system that incorporates model error correction terms is in an advanced stage of development (Xu et al. 2005; Rosmond and Xu 2005).

Weak-constraint variational methods offer a great deal of flexibility in configuring a data assimilation system to account for the presence of model errors. All such methods introduce additional controls that can be used to move the assimilation away from a perfect-model trajectory. Due to advances in computing and minimization techniques, the technical issues associated with greatly increasing the size of the control vector do not appear to present an insurmountable obstacle. Outstanding scientific issues are much more formidable: How to design the constraints for the model error terms in the variational formulation? In principle this requires specification of model error covariances, which will determine the types of spatial and multivariate structures of the correc-

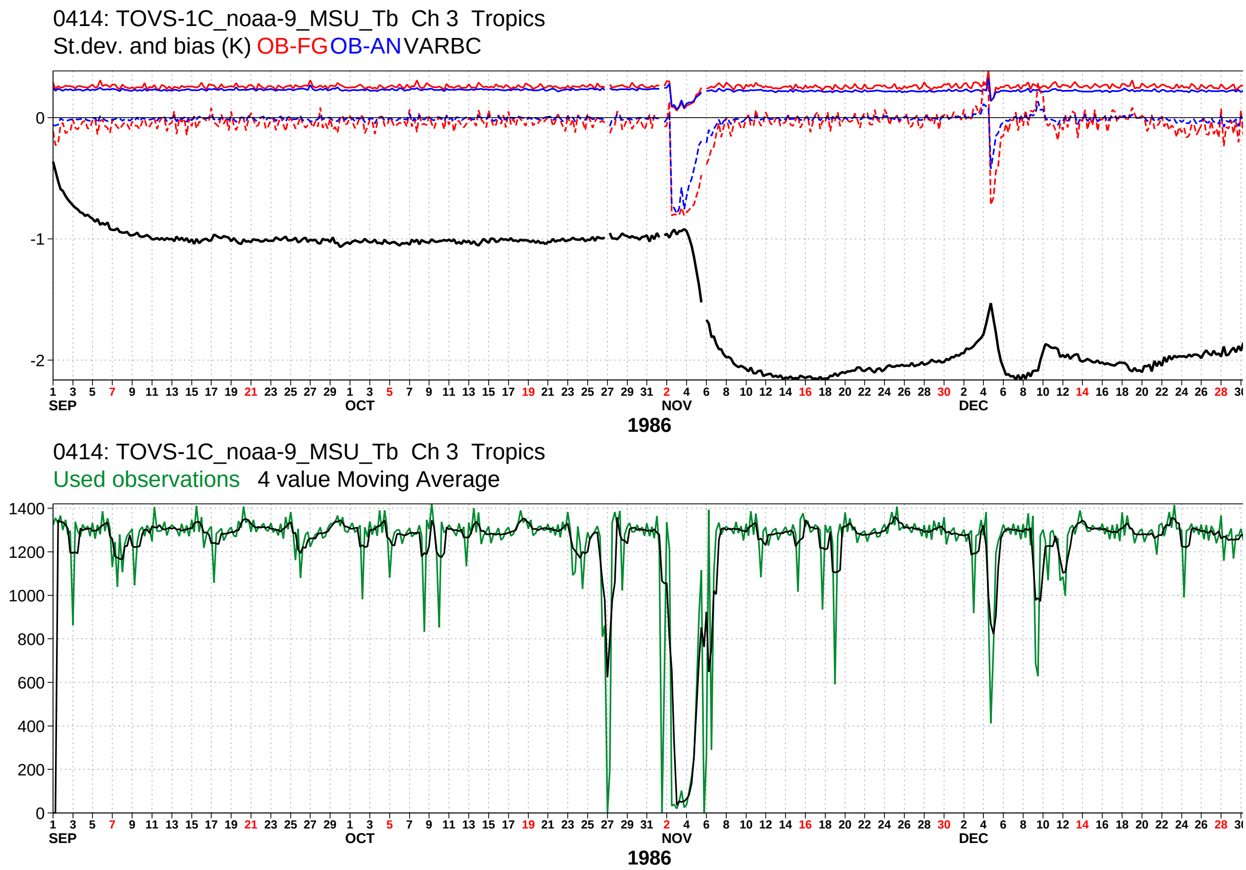

Figure 7: Adaptive bias correction for MSU channel 3, in response to a sudden change in the instrument calibration. The top panel indicates the evolution of the mean and standard deviation of observed-minus-background (red) and observed-minus-analysis (blue) residuals, as well as the rms of the applied bias correction (black). Statistics are computed for data points with latitudes between 20S and 20N, from 1 September 1986 – 31 December 1986. The bottom panel indicates the data count (green) and its 4-value moving average (black).

tions that can be applied to the model during data assimilation. The introduction of many additional degrees of freedom for correcting the model brings with it the undesirable potential for falsely attributing errors in the observations (intermittent and/or systematic) to the model, unless the constraints are somehow designed to prevent this. Perhaps the greatest challenge lies in choosing the right degrees of freedom, i.e., in developing physically meaningful representations of model error that can be clearly distinguished from possible observation errors.

3.2 Sequential estimation methods

The distinction traditionally made in our field between sequential and variational data assimilation is somewhat artificial. All assimilation methods are fundamentally sequential, in the sense that analyses are produced sequentially in time. Each analysis in the sequence is an approximate solution of a variational problem, in which observations are optimally combined with a model-generated background estimate. Clearly there are important differences among systems, in the way ‘optimal’ is defined, in the nature of the approximations made, and in the solution algorithms used, but the underlying statistical concepts are generally similar. A key unresolved issue in data assimilation is the cycling problem, i.e., how to efficiently and accurately propagate error information forward in time. The primary purpose of sequential estimation methods is to address this problem.

Sequential state estimation is often framed in the context of the Kalman filter (Kalman 1960), which provides the optimal solution of the cycling problem for a linear stochastic-dynamic system, albeit under a very restrictive set of assumptions and with unrealistic information requirements (Dee 1991). As in the variational formulation, the standard technique for estimating parameters along with the model state is to augment the state and then to reformulate the estimation problem in terms of this augmented state. Friedland (1969) followed this approach for a class of linear systems in which both model and observations are subject to additive systematic errors, which depend on a set of constant bias parameters. He showed that the Kalman filter for the augmented system in this case is algebraically equivalent to two sets of filter equations: one for the model state estimation, and another for estimating the bias parameters. This constitutes the so-called *separate-bias estimation* scheme for estimating biases in a bias-blind data assimilation system, and for producing bias-corrected state estimates in a separate post-analysis step.

Dee and da Silva (1998) considered the problem of estimating and removing biases in the background fields during data assimilation. They derived several algorithms, including an off-line bias estimator similar to Friedland's, as well as a coupled version in which updated bias estimates are used to produce unbiased analyses during the assimilation. This adaptive bias correction scheme was implemented in the humidity analysis component of the GEOS global data assimilation system (Dee and Todling 2000). We briefly review the key ideas here, and then proceed to discuss some extensions and generalizations.

3.2.1 Correcting persistent bias in the background.

Consider the simple bias model for background errors defined by

$$\mathbf{e}^b = \mathbf{b} + \tilde{\mathbf{e}}^b \quad \text{with} \quad < \tilde{\mathbf{e}}^b >= 0 \tag{16}$$

where the bias vector $\mathbf{b}$ is constant in time but otherwise arbitrary, i.e., unconstrained by any spatial or multivariate structures. Assuming that all available observations are unbiased, a sequential two-step algorithm for estimating the background bias together with the state during data assimilation is

$$\hat{\mathbf{b}}_k = \hat{\mathbf{b}}_{k-1} - \mathbf{K}_k^{\mathbf{b}} \left[\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k^b - \hat{\mathbf{b}}_{k-1}) \right] \tag{17}$$

$$\mathbf{x}_k^a = (\mathbf{x}_k^b - \hat{\mathbf{b}}_k) + \mathbf{K}_k^{\mathbf{x}} \left[\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k^b - \hat{\mathbf{b}}_k) \right] \tag{18}$$

where the subscript k denotes time, and $\mathbf{K}_k^{\mathbf{b}}, \mathbf{K}_k^{\mathbf{x}}$ are the gain matrices for bias and state estimation, respectively. The first step updates a previous bias estimate $\hat{\mathbf{b}}_{k-1}$ based on the latest observations $\mathbf{y}_k$, while the second step produces an unbiased state analysis using the latest bias estimate $\hat{\mathbf{b}}_k$. Setting $\hat{\mathbf{b}} \equiv 0$ gives the standard *bias-blind* linear analysis equation (3). The bias correction in (18) is optional; one can also estimate the bias in the background fields off-line using (17) as a separate diagnostic step. See Dee and da Silva (1998) for a complete derivation and further discussion, as well as some extensions to non-persistent bias models.

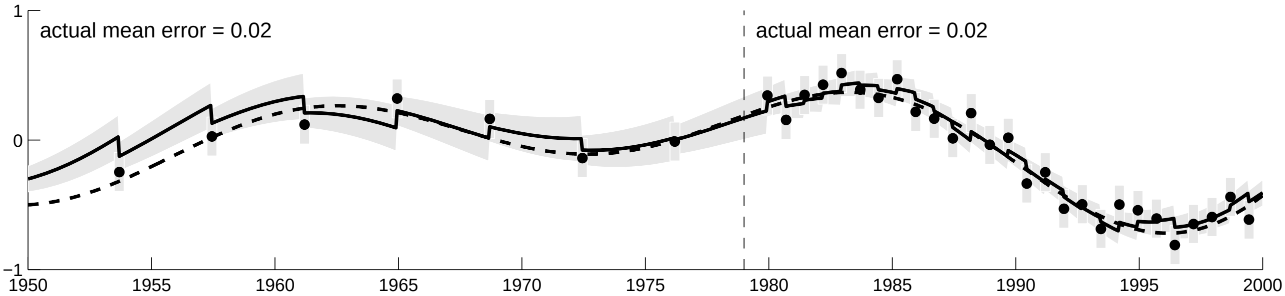

Figure 8: As Fig. 1, but with adaptive model bias correction.

As a trivial illustration we show in Fig. 8 the equivalent of Fig. 1, but now including bias correction on the background. The algorithm learns, after the first few analyses, that the model forecast consistently over-estimates the observation. It then uses this information to adjust subsequent model predictions. As a result, the mean errors rapidly approach zero and become independent of the observing frequency. The bias correction is adaptive: if the bias were to change, then the algorithm would adjust the estimates accordingly.

Dee and Todling (2000) proved optimality of the two-step algorithm, in the sense that it provides unbiased minimum-variance state estimates, when

$$\mathbf{K}^{\mathbf{x}} = \mathbf{B}_{\mathbf{x}}\mathbf{H}^T \left[\mathbf{H}\mathbf{B}_{\mathbf{x}}\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{19}$$

$$\mathbf{K}^{\mathbf{b}} = \mathbf{B}_{\mathbf{b}}\mathbf{H}^T \left[\mathbf{H}\mathbf{B}_{\mathbf{b}}\mathbf{H}^T + \mathbf{H}\mathbf{B}_{\mathbf{x}}\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{20}$$

where $\mathbf{B}_{\mathbf{x}}$ is the background error covariance matrix for the state estimates and $\mathbf{B}_{\mathbf{b}}$ is the error covariance for the bias estimates, i.e.,

$$\mathbf{B}_{\mathbf{x}} = < \tilde{\mathbf{e}}_k^b (\tilde{\mathbf{e}}_k^b)^T > \tag{21}$$

$$\mathbf{B}_{\mathbf{b}} = < (\hat{\mathbf{b}}_{k-1} - \mathbf{b})(\hat{\mathbf{b}}_{k-1} - \mathbf{b})^T > \tag{22}$$

In practice these covariances are unknown, but one can try

$$\mathbf{B}_{\mathbf{b}} = \gamma \mathbf{B}_{\mathbf{x}} \tag{23}$$

if a reasonable specification for $\mathbf{B}_{\mathbf{x}}$ is available. To adopt this model implies that the multivariate and spatial structures of the bias corrections will be similar to those of the analysis increments, which, depending on the application, may or may not be desirable. The scalar parameter γ controls the adaptivity of the bias estimates $\hat{\mathbf{b}}$. With γ small the estimates evolve slowly, and will represent long-term time-averaged background errors. Dee and Todling (2000) discuss the time-behavior of the algorithm in more detail, and present a method for tuning γ based on spectral properties of the observed-minus-background residuals (cf. Fig. 4).

3.2.2 A simplified version of the algorithm.

The cost of the bias update in (17) can be prohibitive, since it requires an extra solution of the analysis equation. The cost can be reduced by using only a subset of the observations for estimating the bias, or by expressing the bias in terms of a relatively small number of parameters, as we show in the next section. Alternatively, with some approximations the algorithm can be considerably simplified when (a) the covariance model (23) is used; (b) the parameter γ is sufficiently small; (c) the same observations used in the analysis are used for the bias estimation.

When $\gamma \ll 1$ the bias estimates will evolve slowly, and we can approximate $\hat{\mathbf{b}}_k$ in (18) by $\hat{\mathbf{b}}_{k-1}$. The terms in brackets on the right-hand sides of (17) and (18) are then identical. Furthermore, using (23) in (19, 20) gives

$$\mathbf{K}^{\mathbf{b}} = \gamma \mathbf{B}_{\mathbf{x}}\mathbf{H}^T \left[(1+\gamma)\mathbf{H}\mathbf{B}_{\mathbf{x}}\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{24}$$

$$\approx \gamma \mathbf{K}^{\mathbf{x}} \tag{25}$$

This approximation seems reasonable in view of the usual uncertainties in $\mathbf{B}_{\mathbf{x}}$ and $\mathbf{R}$. Reversing the order of (17–18) we obtain:

$$\mathbf{x}_k^a = (\mathbf{x}_k^b - \hat{\mathbf{b}}_{k-1}) + \mathbf{K}_k^{\mathbf{x}}[\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k^b - \hat{\mathbf{b}}_{k-1})] \tag{26}$$

$$\hat{\mathbf{b}}_k = \hat{\mathbf{b}}_{k-1} - \gamma \mathbf{K}_k^{\mathbf{x}}[\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k^b - \hat{\mathbf{b}}_{k-1})] \tag{27}$$

It is now trivial to compute the bias update (27), since it involves a previous calculation made in (26). This simplification of the algorithm was first suggested by A. da Silva, and is briefly described in Radakovich et al. (2001).

Any sequential data assimilation system can be easily modified to incorporate this algorithm. This can be seen clearly by arranging (26–27) as

$$\tilde{\mathbf{x}} = \mathbf{x}_k^b - \hat{\mathbf{b}}_{k-1} \tag{28}$$

$$\begin{cases} \mathbf{dy} = \mathbf{y}_k - \mathbf{h}_k(\tilde{\mathbf{x}}) \\ \mathbf{dx} = \mathbf{K}_k^{\mathbf{x}}\mathbf{dy} \\ \mathbf{x}_k^a = \tilde{\mathbf{x}} + \mathbf{dx} \end{cases} \tag{29}$$

$$\hat{\mathbf{b}}_k = \hat{\mathbf{b}}_{k-1} - \gamma\mathbf{dx} \tag{30}$$

The bracketed module (29) represents a standard (bias-blind) analysis scheme; the bias correction step (28) and update step (30) do not depend on the details of this scheme.

3.2.3 Parameterized error models.

A more flexible model for background errors with deterministic components is

$$\mathbf{e}^b = \mathbf{b}(\beta) + \tilde{\mathbf{e}}^b \quad \text{with} \quad < \tilde{\mathbf{e}}^b >= 0 \tag{31}$$

where $\mathbf{b}$ is a known function of a vector β of unknown bias parameters. The bias model $\mathbf{b}$ can comprise spatial, temporal, or multivariate constraints, either explicitly or implicity by means of state-dependent predictors. Estimators for the bias parameters β can be derived by expressing the relationship between the parameters and the observations:

$$\mathbf{y} - \mathbf{h}(\mathbf{x}^b) \approx \mathbf{e}^o - \mathbf{H}\mathbf{e}^b \tag{32}$$

$$= \mathbf{e}^o - \mathbf{H}\mathbf{b}(\beta) - \mathbf{H}\tilde{\mathbf{e}}^b \tag{33}$$

If we define

$$\mathbf{g}(\beta) = \mathbf{H}\mathbf{b}(\beta) \quad \text{and} \quad \tilde{\mathbf{e}} = \mathbf{e}^o - \mathbf{H}\tilde{\mathbf{e}}^b \tag{34}$$

then

$$\mathbf{dy} = \mathbf{g}(\beta) + \tilde{\mathbf{e}} \quad \text{with} \quad \begin{cases} < \tilde{\mathbf{e}} > \approx 0 \\ < \tilde{\mathbf{e}}\tilde{\mathbf{e}}^T > \approx \mathbf{H}\mathbf{B}_{\mathbf{x}}\mathbf{H}^T + \mathbf{R} \end{cases} \tag{35}$$

This defines a complete measurement model for β, describing the information about the parameters β that is implicit in the observations.

There are many ways to derive estimation algorithms for β based on (35), for example by state augmentation in a variational analysis as described in Section 3(a). In the special case when $\mathbf{b}$ is linear in β, i.e.,

$$\mathbf{b}(\beta) = \mathbf{L}_k\beta \tag{36}$$

for some linear operator $\mathbf{L}_k$, the analogue of (17–18) is

$$\hat{\beta}_k = \hat{\beta}_{k-1} - \mathbf{K}_k^{\beta}[\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k^b - \mathbf{L}_k\hat{\beta}_{k-1})] \tag{37}$$

$$\mathbf{x}_k^a = (\mathbf{x}_k^b - \mathbf{L}_k\hat{\beta}_k) + \mathbf{K}_k^{\mathbf{x}}[\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k^b - \mathbf{L}_k\hat{\beta}_k)] \tag{38}$$

This estimator is optimal when $\mathbf{K}^{\mathbf{x}}$ is as given by (19) and

$$\mathbf{K}^{\beta} = \mathbf{B}_{\beta}\mathbf{L}^T\mathbf{H}^T\left[\mathbf{H}\mathbf{L}\mathbf{B}_{\beta}\mathbf{L}^T\mathbf{H}^T + \mathbf{H}\mathbf{B}_{\mathbf{x}}\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{39}$$

where $\mathbf{B}_{\beta}$ is the error covariance matrix for the parameter estimates

$$\mathbf{B}_{\beta} =< (\hat{\beta}_{k-1} - \beta)(\hat{\beta}_{k-1} - \beta)^T > \tag{40}$$

The gain matrix $\mathbf{K}^{\beta}$ for the bias update has one row for each bias parameter and as many columns as there are observations. The algorithm requires $\mathbf{B}_{\beta}$, but the estimates will not be sensitive to its specification if the number of parameters is small compared to the number of available observations. See Chepurin *et al.* (2005) for an application of this algorithm for bias correction in a tropical ocean model.

3.2.4 Correction of model tendency errors.

The sequential estimation schemes discussed so far are designed to correct biases in the background fields whenever an analysis is produced. Such intermittent correction schemes can account for the accumulated effect of model errors, but they cannot prevent the generation of the biases during the integration of the model. To do so requires adjustments to the model tendencies, or, even better, to the model itself.

Ideally, bias estimates obtained in a sequential bias estimation scheme should lead to information about model errors that can be used during model integration. Thus, instead of cycling with

$$\mathbf{x}_k^b = \mathbf{m}_{k,k-1}(\mathbf{x}_{k-1}^a) \tag{41}$$

where $\mathbf{m}_{k,k-1}$ represents an integration of a biased forecast model from t_{k-1} to t_k, one would use a modified version $\tilde{\mathbf{m}}$ of the model:

$$\tilde{\mathbf{x}}_k^b = \tilde{\mathbf{m}}_{k,k-1}(\mathbf{x}_{k-1}^a, \hat{\beta}_{k-1}) \tag{42}$$

such that the resulting background $\tilde{\mathbf{x}}_k^b$ is unbiased. Note the analogy with the modified observation operator $\tilde{\mathbf{h}}$ in (12). If successful, this would obviate the need for a separate bias correction step; for example, one could remove (28) in the simplified algorithm discussed earlier.

The model modification (42) can be implemented using a linear updating scheme similar to the Incremental Analysis Update (IAU) algorithm described in Bloom et al. (1996), by applying a fraction of the correction $\hat{\mathbf{b}}$ at each time step during the model integration. Such an incremental bias correction technique was shown to be very effective for correcting land-surface model bias in a skin temperature assimilation study by Radakovich et al. (2004). More sophisticated techniques for correcting model tendencies based on statistical estimates of biases in the background require an understanding of the physical mechanisms underlying bias generation in the model. For example, Bell et al. (2004) used on-line estimates of subsurface temperature bias in an ocean assimilation system to make adjustments to the model's pressure gradient during the integration of the model; see also Balmaseda et al. (2006).

3.2.5 Prediction of analysis increments.

Finally we sketch an alternative approach to sequential model bias correction during data assimilation that, to our knowledge, has not been previously explored. It was noted in Section 2(a) that a clear symptom of bias in data assimilation is the appearance of systematic patterns in the analysis increments, such as persistent mean values, but also spatial features that correlate with the configuration of the observing system. For example, Figure 9 shows mean total column ozone increments produced in ERA-40, computed separately for all August 2002 analyses at 00 UTC, 06 UTC, 12 UTC, and 18 UTC. The mean increments range from -3 to +3 Dobson units in the Northern hemisphere, and from -11 to +14 Dobson units in the Southern hemisphere. These plots clearly reflect the location of the satellite carrying the sensor, and the patterns suggest that there are persistent discrepancies between the model-predicted ozone and the measurement data.

In the presence of bias, therefore, certain components of the increments are systematic and therefore predictable. Suppose we can concoct a function $\mathbf{f}_k$ that predicts the next analysis increment based on, say, the most recent L increments. Provided the predictable part of the increment can be attributed to model errors, the following algorithm

$$\mathbf{dx}_k^p = \mathbf{f}_k(\mathbf{dx}_{k-L}, \cdots, \mathbf{dx}_{k-1}) \tag{43}$$
$$\mathbf{dx}_k = \mathbf{K}_k(\mathbf{y}_k - \mathbf{h}(\mathbf{x}_k^b - \mathbf{dx}_k^p)) \tag{44}$$

will correct the model background and produce unbiased analyses. Figure 10 provides a simple demonstration of this algorithm, for the case where x is a scalar, and the model error comprises a slowly varying bias, an approximately diurnal cycle, and serially correlated noise. For the increment prediction we used a lag-6

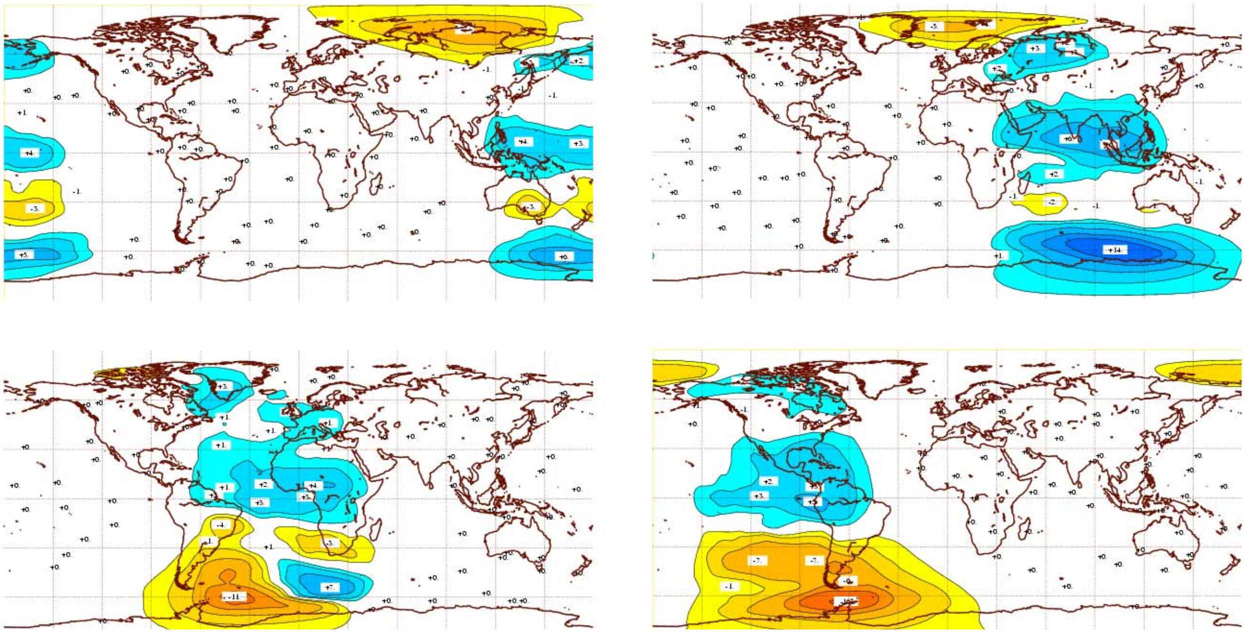

Figure 9: Mean total column ozone increments from ERA-40, computed for August 2002 analyses at (clockwise from top-left) 00 UTC, 06 UTC, 12 UTC, and 18 UTC. Graphics provided by courtesy of the ECMWF.

autoregressive moving average (ARMA) model whose coefficients are continuously updated using a recursive least-squares identification algorithm (Ljung 1999, Eq. 11.12). The thin line in the top panel shows the simulated true state, the dots indicate the 6-hourly observations, and the solid line shows the assimilation obtained with (43, 44). Initially no increment is predicted, i.e., $\mathbf{dx}^p = 0$, and the assimilation is bias-blind. Then, at the time indicated by the vertical dashed line, increment prediction is turned on. The center panel shows the actual analysis increment (thin line) and its prediction (thick line). The adaptive ARMA scheme is clearly effective in predicting the deterministic component of the increment. The lower panel shows the background error (thin line) and the analysis error (thick line), as well as the rms of the analysis error prior to and after the start of increment prediction.

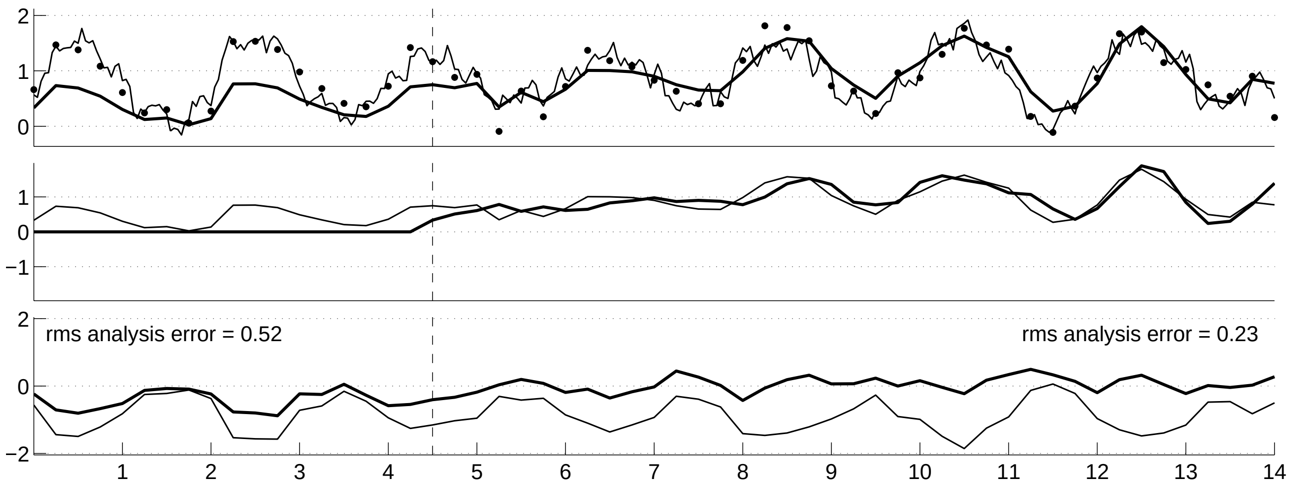

Figure 10: Model bias correction using prediction of analysis increments. See text for explanation.

4 Conclusion

Biases (or, more broadly, systematic errors) are prevalent in data assimilation. All ingredients of a data assimilation system—the forecast model, boundary conditions, observations, observation operators, covariance models—can generate, extrapolate, or enhance biases. The presence of bias can be detected on the input side by monitoring differences between observations and their model-predicted equivalents, and on the output side by examining systematic features of the analysis increments. Separation of different bias sources requires additional information, such as independent observations, knowledge of the underlying causes, or hypotheses about the error characteristics of possible sources.

Most data assimilation systems are not designed to correct bias during the analysis step. In concept it is not very difficult to develop bias-aware assimilation methods. The general approach is to introduce additional parameters in the estimation problem that represent the biases in the system. The main scientific challenge is to correctly attribute a detected bias to its source, and then to develop a useful model for the bias. When different sources produce similar biases, the assimilation may correct the wrong source. This risk increases as more degrees of freedom are added to the system, for example, in a weak-constraint variational analysis supporting model error correction that also contains parameters for radiance bias correction. It is not clear that constraints on the correction terms can be designed in such a way that model bias and observation bias can always be correctly and simultaneously identified in the analysis.

A bias-aware analysis scheme designed to correct bias in either the background or the observations will, by construction, reduce the mean analysis increments, but not necessarily for the right reason. In order to test whether the attribution of the bias is correct one needs to verify that the analysis has actually improved. Fig. 11 shows schematically how a successful bias correction of the background during the assimilation should lead to a better analysis and hence to reduced forecast errors. Unfortunately, in practice, reducing the bias in the initial conditions may not improve the forecast, unless the model itself is changed.

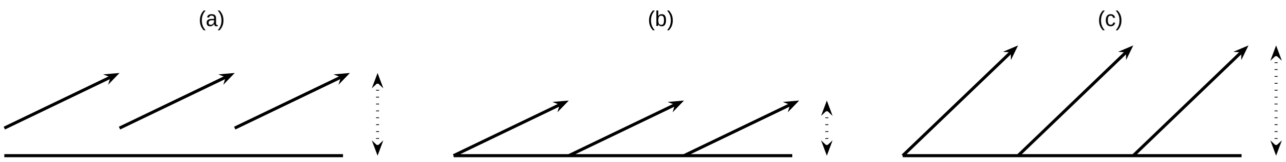

Figure 11: Bias-blind assimilation in the presence of model bias (a), successful bias-aware assimilation leading to reduced forecast bias (b), unsuccessful bias-aware assimilation leading to increased forecast bias (c)

Model bias correction is particularly challenging because it is difficult to develop useful representations for the biases or for the mechanisms that cause them to develop. Intermittent bias correction of background estimates in a sequential estimation scheme does not prevent the generation of the bias during the integration of the model. Incremental bias correction schemes, which use bias estimates to correct model tendencies, may be more effective in guiding the model to an unbiased forecast, provided the corrections are physically meaningful.

References

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal filtering.* Prentice-Hall, Englewood Cliffs.

Andersson, E., Bauer, P., Beljaars, A., Chevallier, F., Hólm, E., Janisková, M., Kållberg, P., Kelly, G., Lopez, P., McNally, A., Moreau, E., Simmons, A. J., Thépaut, J.-N. and Tompkins, A. M. (2005). Assimilation and Modeling of the Atmospheric Hydrological Cycle in the ECMWF Forecasting System. *Bull. Amer. Meteorol. Soc.,* **86**, 387–402.

Bell, M. J., Martin, M. J. and Nichols, N. K. (2004). Assimilation of data into an ocean model with systematic errors near the equator. *Q. J. R. Meteorol. Soc.*, **130**, 873–893.

Balmaseda, M. A., Dee, D. P., Vidard, A. and Anderson, D. L. T. (2006). A multivariate treatment of bias for sequential data assimilation: Application to the tropical oceans. Submitted to *Q. J. R. Meteorol. Soc*.

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

Chepurin, G. A., Carton, J. and Dee, D. P. (2005). Forecast model bias correction in ocean data assimilation. *Mon. Wea. Rev.*, **133**, 1328–1342.

Daley, R. (1992). The lagged innovation covariance: A performance diagnostic for atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 178–196.

Dee, D. P (1991). Simplification of the Kalman filter for meteorological data assimilation. *Q. J. R. Meteorol. Soc.*,**117**, 365–384.

Dee, D. P. (2004). 'Variational bias correction of radiance data in the ECMWF system'. Pp. 97–112 in Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP, 28 June–1 July 2004, Reading, UK.

Dee, D. P. and da Silva, A. M. (1998). Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc.*,**124**, 269–295.

Dee, D. P. and Todling, R. (2000). Data assimilation in the presence of forecast bias: The GEOS moisture analysis. *Mon. Wea. Rev.*,**128**, 3268–3282.

Derber, J. C. (1989). A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437–2446.

Derber, J. C. and Wu, W.-S. (1998). The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287–2299.

Eyre, J. R. (1992). 'A bias correction scheme for simulated TOVS brightness temperatures'. Technical Memorandum 186, ECMWF, Reading, UK.

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

Griffith, A. K. and Nichols, N. K. (2000). Adjoint methods in data assimilation for estimating model error. *Flow Turbulence and Combustion*, **65**, 469–488.

Haimberger, L. (2005). 'Homogenization of radiosonde temperature time series using ERA-40 analysis feedback information'. ERA-40 Project Report Series No. 23, ECMWF, Reading, UK.

Harris, B. A. and Kelly, G. (2001). A satellite radiance-bias correction scheme for data assimilation. *Q. J. R. Meteorol. Soc.*, **127**, 1453–1468.

Hernandez, A., Kelly, G. and Uppala, S. (2004). 'The TOVS/ATOVS observing system in ERA-40'. ERA-40 Project Report Series No. 18, ECMWF, Reading, UK.

Joiner, J., Lee, H. T., Strow, L. L., Bhartia, P. K., Hannon, S., Miller, A. J. and Rokke, L. (1998). Radiative transfer in the 9.6 micron HIRS ozone channel using collocated SBUV-determined ozone abundances. *J. Geophys. Res.*, **103**, 19213–19229..

Joiner, J. and Rokke, L. (2000). Variational cloud-clearing with TOVS data. *Q. J. R. Meteorol. Soc.*, **126**, 725–748.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.*, **82**, 35–45.

Lamarque, J.-F., Khattatov, B., Yudin, Y., Edwards, D. P., Gille, J. C., Emmons, L. K., Deeter, M. N., Warner, J., Ziskin, .D. C., Francis, G. L., Ho, S., Mao, D., Chen, J. and Drummond, J. R. (2004). Application of a bias estimator for the improved assimilation of Measurements of Pollution in the Troposphere (MOPITT) carbon monoxide retrievals. *J. Geophys. Res.*, **109**, D16304, doi:10.1029/2003JD004466.

Ljung, L. (1999). *System identification: Theory for the user.* (2nd ed.) Prentice-Hall, Upper Saddle River.

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, **39**, 447-462.

McNally, A. P (2004). 'The assimilation of stratospheric satellite data at ECMWF'. Pp. 103–106 in Proceedings of the ECMWF/SPARC Workshop on Modelling and Assimilation for the Stratosphere and Tropopause, 23–26 June 2003, Reading, UK.

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

Polavarapu, S., Shepherd, T. G., Rochon, Y. and Ren, S. (2006). Some challenges of middle atmosphere data assimilation. *Q. J. R. Meteorol. Soc.*, in press.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in FORTRAN: The art of scientific computing.* (2nd ed.) Cambridge University Press, Cambridge.

Radakovich, J. D., Bosilovich, M. G., Chern, J.-D., da Silva, A. M., Todling, R., Joiner, J., Wu, M.-L., and Norris, P. (2004). 'Implementation of coupled skin temperature analysis and bias correction in the NASA/GMAO finite-volume data assimilation system (FvDAS)'. P1.3 in Proceedings of the Eighth Symposium on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, 12–15 January 2004, Seattle, WA.

Radakovich, J. D., Houser, P. R., da Silva, A. M. and Bosilovich, M. G. (2001). 'Results from global land-surface data assimilation methods'. Pp. 132–134 in Proceedings of the Fifth Symposium on Integrated Observing Systems, 14–19 January 2001, Albuquerque, NM.

Rosmond, T. and Xu, L. (2005). Development of NAVDAS-AR: Non-linear formulation and outer loop tests. *Tellus A*, in press.

Santer, B. D., Wigley, T. M. L., Simmons, A. J., Kållberg P. W., Kelly, G. A., Uppala, S. M., Ammann, C, Boyle, J. S., Bruggemann, W., Doutriaux, C., Fiorino, M., Mears, C., Meehl, G. A., Sausen ,R., Taylor, K. E., Washington, W. M., Wehner, M. F. and Wentz, F. J. (2004). Identification of anthropogenic climate change using a second-generation reanalysis. *J. Geophys. Res.*, **109**, D21104, doi:10.1029/2004JD005075.

Sasaki, Y. (1970). Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, **98**, 875–883.

Thépaut, J.-N. (2003). 'Satellite Data Assimilation in Numerical Weather Prediction: an overview'. Pp. 75–95 in Proceedings of the ECMWF annual seminar on recent developments in data assimilation for atmosphere and ocean, 8–12 September 2003, Reading, UK.

Trémolet, Y. (2003). 'Model error in variational data assimilation'. Pp. 361–367 in Proceedings of the ECMWF annual seminar on recent developments in data assimilation for atmosphere and ocean, 8–12 September 2003, Reading, UK.

Tsyroulnikov, M. D. (2006). Stochastic modelling of model errors: A simulation study. *Q. J. R. Meteorol. Soc.*, in press.

Watts, P. D. and McNally, A. P. (2004). 'Identification and correction of radiative transfer modelling errors for atmospheric sounders: AIRS and AMSU-A'. Pp. 23-38 in Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP, 28 June–1 July 2004, Reading, UK.

Xu, L., Rosmond, T. and Daley, R. (2005). Development of NAVDAS-AR: Formulation and initial tests of the linear problem. *Tellus A*, in press.

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