

# An adaptive scheme for cycling background error variances during data assimilation

Dick P. Dee

*Data Assimilation Office, NASA/Goddard Space Flight Center  
Greenbelt, Maryland, USA*

## 1 Introduction

Data assimilation systems attempt to combine and extrapolate observational information in order to generate the best possible four-dimensional representation of the state of the atmosphere. Extrapolation in space and time is achieved by blending the observations with model-generated state estimates. The quality of the assimilated product very much depends on the assumptions that are made about the errors in both observations and model predictions. For example, a single observation of a meteorological variable such as temperature or surface pressure typically affects the estimates of those (and other) quantities in a neighborhood of the observation location. The amplitude and spatial extent of this effect is largely determined by assumptions about the errors in the background and observations.

To phrase this point more precisely, consider the variational analysis of a set of observations  $\mathbf{y}^o$ , given a model-generated background estimate  $\mathbf{x}^b$  valid at the same time. The goal is to compute the analysis  $\mathbf{x}^a$  that minimizes

$$J(\mathbf{x}) = (\mathbf{x}^b - \mathbf{x})^T \mathbf{P}^{-1} (\mathbf{x}^b - \mathbf{x}) + (\mathbf{y}^o - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathcal{H}(\mathbf{x})), \quad (1)$$

where  $\mathbf{P}$  and  $\mathbf{R}$  are the background and observation error covariances, respectively. (Additional notation is listed in the table in the next section). In the special case when the observation operator  $\mathcal{H}$  is linear, i.e., when  $\mathcal{H}(\mathbf{x}) = \mathbf{H}\mathbf{x}$ , the minimizing solution is

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{P}\mathbf{H}^T [\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}]^{-1} [\mathbf{y}^o - \mathbf{H}\mathbf{x}^b]. \quad (2)$$

This expression shows that the change to the background due to the observations is in the column space of  $\mathbf{P}$ . The structure of the analysis increment  $\mathbf{x}^a - \mathbf{x}^b$  is therefore strongly determined by the specification of the background error covariances. It also clearly (but less directly) depends on properties of  $\mathbf{R}$ . This is still true when the observation operator is nonlinear, and/or if it involves integration forward in time as in four-dimensional variational (4DVAR) assimilation. In that case the influence of any given observation on the analyzed state is also affected by the model dynamics, but the analysis increment still resides in the column space of the background error covariance operator.

Current practice is to estimate the covariances based on statistics of residuals between different state estimates (model forecasts or analyses) and/or between observations (observed-minus-forecast residuals). These estimates tend to be quasi-stationary (varying on a seasonal time scale) and they always involve some kind of spatial averaging (in the physical and/or spectral domain).

There have been various attempts to develop methods for 'cycling' the background error covariances in a time-dependent fashion. Actual model errors must depend on the flow; this can be rigorously demonstrated (Cohn and Dee 1988) but is also intuitively obvious. We would like to move beyond quasi-stationary covariance representations in order to account for the dynamic evolution and intermittent growth and decay of the errors.

The Kalman filter has often been championed as providing a basis for solving this problem in our field of application. This algorithm was originally derived by Kalman (1960) as the optimal state estimator for a linear

discrete stochastic-dynamic system (prediction model) with additive white Gaussian system noise (model error) and a linear measurement model (observation operator) with Gaussian measurement noise (observation error). Kalman showed that the background and analysis error covariances associated with the optimal estimator evolve according to

$$\mathbf{P}_k^b = \mathbf{M}_k \mathbf{P}_{k-1}^a \mathbf{M}_k^T + \mathbf{Q}_k \quad (3)$$

$$\mathbf{P}_k^a = \mathbf{P}_k^b - \mathbf{P}_k^b \mathbf{H}_k^T \left[ \mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^T + \mathbf{R}_k \right]^{-1} \mathbf{H}_k \mathbf{P}_k^b \quad (4)$$

where the subscript  $k$  denotes time,  $\mathbf{M}$  is the linear model operator, and  $\mathbf{Q}$  is the model error covariance. Other authors derived extensions and approximations to handle weakly nonlinear dynamics and observation operators, serially correlated model errors, unknown bias parameters, and other complications.

Notation	(subscript $k$ denotes time)	
$\mathbf{x}_k^t$	truth in model state representation	$\mathbb{R}^n$
$\mathbf{x}_k^b$	background state estimate	$\mathbb{R}^n$
$\mathbf{x}_k^a$	analyzed state	$\mathbb{R}^n$
$\mathbf{y}_k^o$	generic observation vector	$\mathbb{R}^p$
$\mathbf{x}_k^o$	in-situ observation vector	$\mathbb{R}^p$
$\mathcal{H}_k$	generic observation operator	$\mathbb{R}^n \rightarrow \mathbb{R}^p$
$\mathbf{H}_k$	linear observation operator	$\mathbb{R}^n \times \mathbb{R}^p$
$\mathbf{I}_k$	linear interpolation operator	$\mathbb{R}^n \times \mathbb{R}^p$
$\mathcal{M}_{k,k-1}$	generic model operator	$\mathbb{R}^n \rightarrow \mathbb{R}^n$
$\mathbf{M}_{k,k-1}$	tangent-linear model operator	$\mathbb{R}^n \times \mathbb{R}^n$
$\mathbf{A}_{k,k-1}$	advection operator	$\mathbb{R}^n \times \mathbb{R}^n$
$\mathbf{e}_k^b \equiv \mathbf{x}_k^b - \mathbf{x}_k^t$	background error	$\mathbb{R}^n$
$\mathbf{e}_k^a \equiv \mathbf{x}_k^a - \mathbf{x}_k^t$	analysis error	$\mathbb{R}^n$
$\mathbf{e}_k^o \equiv \mathbf{y}_k^o - \mathcal{H}_k(\mathbf{x}_k^t)$	observation error	$\mathbb{R}^p$
$\mathbf{e}_k^m \equiv \mathbf{e}_k^b - \mathbf{A}_{k,k-1} \mathbf{e}_{k-1}^a$	model error	$\mathbb{R}^n$
$\mathbf{P}_k^b \equiv \langle \mathbf{e}_k^b (\mathbf{e}_k^b)^T \rangle$	background error covariance	$\mathbb{R}^n \times \mathbb{R}^n$
$\mathbf{P}_k^a \equiv \langle \mathbf{e}_k^a (\mathbf{e}_k^a)^T \rangle$	analysis error covariance	$\mathbb{R}^n \times \mathbb{R}^n$
$\mathbf{R}_k \equiv \langle \mathbf{e}_k^o (\mathbf{e}_k^o)^T \rangle$	observation error covariance	$\mathbb{R}^p \times \mathbb{R}^p$
$\mathbf{X}_k \equiv \langle \mathbf{e}_k^o (\mathbf{e}_k^b)^T \rangle$	observation-background error covariance	$\mathbb{R}^p \times \mathbb{R}^n$
$\mathbf{Q}_k \equiv \langle \mathbf{e}_k^m (\mathbf{e}_k^m)^T \rangle$	model error covariance	$\mathbb{R}^n \times \mathbb{R}^n$
$\sigma_k^b \equiv (\text{diag } \mathbf{P}_k^b)^{1/2}$	background error standard deviation	$\mathbb{R}^n$
$\sigma_k^a \equiv (\text{diag } \mathbf{P}_k^a)^{1/2}$	analysis error standard deviation	$\mathbb{R}^n$
$\sigma_k^o \equiv (\text{diag } \mathbf{R}_k)^{1/2}$	observation error standard deviation	$\mathbb{R}^p$
$\sigma_k^m \equiv (\text{diag } \mathbf{Q}_k)^{1/2}$	model error standard deviation	$\mathbb{R}^n$
$\mathbf{d}\mathbf{x}_k^o \equiv \mathbf{x}_k^o - \mathbf{I}_k \mathbf{x}_k^b$	in-situ observation residuals	$\mathbb{R}^p$
$\rho_k$	in-situ observation-background error correlations	$\mathbb{R}^p$
$\mathcal{S}_k$	variance observation operator	$\mathbb{R}^p \rightarrow \mathbb{R}^p$
$\mathbf{e}_k^s$	sampling error for variance observations	$\mathbb{R}^p$
$\mathcal{R}_k$	variance reduction factor	$\mathbb{R}^n \rightarrow \mathbb{R}^n$

The fundamental obstacle in applying the Kalman filter to complex geophysical systems is that it requires a great deal of information about the joint probability distribution of model errors and observation errors that is simply not available. Regrettably, the majority of literature on the application of Kalman filter theory to atmospheric/oceanic data assimilation continues to point to the computational cost of the algorithm as its main drawback. This is unfortunate because it tends to draw attention away from the more fundamental issues, which have to do with information requirements.

In a qualitative sense the Kalman filter reflects three important aspects of error evolution in the assimilation cycle:

- (1) Propagation of initial errors, represented by the first term in (3)
- (2) Error growth due to model defects, represented by the second term in (3)
- (3) Error reduction due to the incorporation of observational information, represented by (4).

However, the derivation of the Kalman filter covariance evolution equations rests on many assumptions that are inappropriate for our application (Dee 1991). Rather than starting with (3,4) and attempting to design a viable computational scheme for these equations, we will formulate separate computational models for each of the three qualitative aspects of error evolution. We then combine these models in a sequential algorithm for estimating and updating background error variances in a cycling data assimilation system.

The main restrictions of the algorithm as presented here are that it is univariate, it is primarily designed for advective quantities, and it estimates error variances only. Correlations must be separately modeled, e.g. by statistical means, or with ensemble methods. The variance estimation relies on observations, and its performance depends primarily on (1) the data coverage and (2) the ability to accurately specify observation errors.

## 2 Error propagation

The effect on short-term forecast accuracy of uncertainties in the initial conditions can, in principle, be calculated very precisely (e.g., by ensemble methods, higher-order extensions of the Kalman filter, or by numerical solution of the Fokker-Planck equation), but this requires access to the probability distribution of the initial errors. What we actually know about initial errors, however, is mostly qualitative; i.e., they tend to be similar to background errors in data-sparse regions, and they tend to be smaller and spatially less correlated than background errors in well-observed regions. The latter depends on the quality of the observations and on the ability of the analysis system to make good use of them. For example, an analysis system that produces geostrophically balanced increments may in fact increase errors locally where the flow is highly ageostrophic. While it may be possible to obtain meaningful statistics for initial errors based on time- and space averaging, these will tell us very little about the initial errors at a given point in time and space.

It is therefore not clear how much can be gained in practice by using the full model equations to compute the short-term effect on the background error variances of the initial errors. The most obvious approximation (used in the extended Kalman filter) is to use the tangent linear model. For the case of atmospheric water vapor it may be just as effective to ignore model physics altogether and approximate the dynamic propagation of initial errors by advection only. The additional error incurred by this simplification can then be folded into the model error term. The same approach could be used for any other scalar quantity which is approximately conserved by the flow.

In a cycling data assimilation system the initial conditions for the model are obtained from the latest analysis, possibly after some type of initialization procedure, or a forward model integration in case of 4DVAR. In what follows we will simply equate initial errors with analysis errors.

Suppose  $\mathbf{A}_{k,k-1}$  denotes the operator which advects a scalar field from  $t_{k-1}$  to  $t_k$ , based on a model reference trajectory (typically the same trajectory used to produce the background estimate). If  $\mathbf{e}^{b,a,m}$  are the background,

analysis, and model errors, respectively, then

$$\mathbf{e}_k^b = \mathbf{A}_{k,k-1} \mathbf{e}_{k-1}^a + \mathbf{e}_k^m \quad (5)$$

Note that the model error  $\mathbf{e}_k^m$  is defined here simply as the component of background error which is not explained by the advection of initial error.

It is not difficult to show (e.g. Cohn 1993) that the error variances satisfy the same equation:

$$(\boldsymbol{\sigma}_k^b)^2 = \mathbf{A}_{k,k-1} (\boldsymbol{\sigma}_{k-1}^a)^2 + (\boldsymbol{\sigma}_k^m)^2 \quad (6)$$

The parentheses here indicate an element-by-element vector operation:  $(\boldsymbol{\sigma})^2$  denotes the vector whose elements are the squares of the elements of the vector  $\boldsymbol{\sigma}$ .

### 3 Error growth

Getting a handle on the contribution of model error to the background error variance is key to being able to predict the background error variances. It may be possible to formulate a flow-dependent model for  $(\boldsymbol{\sigma}_k^m)^2$ , e.g. based on gradients in the background field, surface properties, or other factors. However, for the moment it is not known how to do this, and we will therefore strictly rely on observations to estimate the additive effect of model error on the background error variances.

The intuitive idea behind model error variance estimation based on observations is simply this: if, for some subset of observations, the discrepancies with the background are larger than expected, then this may be an indication of large model errors. One obviously has to be very careful with this idea, because it depends on an adequate notion of 'larger than expected,' and because large residuals can also be caused by bad observations. In practice this requires good estimates of observation error variances, and also puts a heavy burden on observational quality control (see Dee *et al.* 2001).

Here we develop this idea for in-situ observations  $y^o = x^o$  only, for which the observation operator is simply spatial interpolation:

$$\mathcal{H}(\mathbf{x}_k) = \mathbf{I}_k \mathbf{x}_k \quad (7)$$

In practice this represents a (relatively small) subset of the available observations; for the experiments described in the next section we used rawinsonde reports and TOVS retrievals for model error variance estimation. For the in-situ observed-minus-background residuals  $\mathbf{d}\mathbf{x}_k^o$  we have

$$\begin{aligned} \mathbf{d}\mathbf{x}_k^o &\equiv \mathbf{x}_k^o - \mathbf{I}_k \mathbf{x}_k^b \\ &= (\mathbf{I}_k \mathbf{x}_k^t + \mathbf{e}_k^o) - \mathbf{I}_k (\mathbf{x}_k^t + \mathbf{e}_k^t) \\ &= \mathbf{e}_k^o - \mathbf{I}_k \mathbf{e}_k^t \end{aligned} \quad (8)$$

where  $\mathbf{e}^o$  is the observation error. Clearly

$$\langle \mathbf{d}\mathbf{x}_k^o (\mathbf{d}\mathbf{x}_k^o)^T \rangle = \mathbf{V}_k \quad (9)$$

where

$$\mathbf{V}_k \equiv \mathbf{R}_k + \mathbf{I}_k \mathbf{P}_k \mathbf{I}_k^T - \mathbf{I}_k \mathbf{X}_k^T - \mathbf{X}_k \mathbf{I}_k^T \quad (10)$$

and  $\mathbf{P}$ ,  $\mathbf{R}$ ,  $\mathbf{X}$  are the background, observation, and observation-background error covariances, respectively. The diagonal of (9) can be approximated by

$$\langle (\mathbf{d}\mathbf{x}_k^o)^2 \rangle \approx (\boldsymbol{\sigma}_k^o)^2 + \mathbf{I}_k (\boldsymbol{\sigma}_k^b)^2 - 2(\boldsymbol{\rho}_k) (\boldsymbol{\sigma}_k^o) (\mathbf{I}_k \boldsymbol{\sigma}_k^b) \quad (11)$$

where  $\boldsymbol{\rho}$  is the vector of in-situ correlations between observation and interpolated background errors. (Parentheses surrounding vectors again denote element-by-element operations.) This approximation is accurate if it can

be assumed that background errors are strongly correlated over the interpolation distances. This will generally be the case when the spatial resolution of the model is high compared to the correlation length scales.

The variance equation (11) is equivalent to the statement

$$(\mathbf{dx}_k^o)^2 = (\sigma_k^o)^2 + \mathbf{I}_k(\sigma_k^b)^2 - 2(\rho_k)(\sigma_k^o)(\mathbf{I}_k\sigma_k^b) + \mathbf{e}_k^s \quad (12)$$

where  $\mathbf{e}_k^s$  represents the diagonal approximation error incurred in replacing (9), compounded with (and dominated by) sampling error. If the observation error variances  $(\sigma_k^o)^2$  and the cross-correlations  $\rho_k$  are known, then (12) represents an observation model for the vector of background error variances  $(\sigma_k^b)^2$ . The *variance observations* consist of the squared residuals:

$$(\mathbf{dx}_k^o)^2 = \mathcal{S}_k((\sigma_k^b)^2) + \mathbf{e}_k^s \quad (13)$$

and the *variance observation operator*  $\mathcal{S}$  is defined by

$$\mathcal{S}_k((\sigma_k^b)^2) \equiv (\sigma_k^o)^2 + \mathbf{I}_k(\sigma_k^b)^2 - 2(\rho_k)(\sigma_k^o)(\mathbf{I}_k\sigma_k^b) \quad (14)$$

We will refer to  $\mathbf{e}^s$  as the *sampling error* for the variance observations. Under the assumption that the in-situ residuals  $\mathbf{dx}_k^o$  are zero-mean and approximately Gaussian with covariances given by (10), it can be shown that

$$\langle \mathbf{e}_k^s \rangle \approx 0 \quad (15)$$

$$\langle \mathbf{e}_k^s (\mathbf{e}_k^s)^T \rangle \approx 2(\mathbf{V}_k)^2 \quad (16)$$

where  $(\mathbf{V}_k)^2$  is the matrix obtained by squaring each element of  $\mathbf{V}_k$ . Note that  $\mathbf{V}_k$  depends on the background error covariances, which are unknown.

A prediction  $(\sigma_k^b)_{pred}^2$  of the background error variance can be obtained by integrating the variance evolution equation (6). The squared observation residuals can then be used to adjust this prediction by means of a linear estimation scheme of the form

$$(\sigma_k^b)_{est}^2 = (\sigma_k^b)_{pred}^2 + \mathbf{L}_k \left[ (\mathbf{dx}_k^o)^2 - \mathcal{S}_k((\sigma_k^b)_{pred}^2) \right] \quad (17)$$

Such a scheme is inherently suboptimal, because the linear gain  $\mathbf{L}_k$  for an optimal scheme would have to depend on the error covariance (16), which, in turn, depends on the unknown background error variances. The optimal gain would also require specification of the error covariances associated with the variance prediction, which, of course, are also unknown. However, even if suboptimal, this scheme simply averages the variance predictions with the variances as implied by the observations. This will improve the predictions as long as the variance observation operator  $\mathcal{S}_k$  is reasonably accurate.

We can construct a gain matrix  $\mathbf{L}_k$  based on rough estimates of the required error covariances. For the variance prediction errors we assume

$$\left\langle ((\sigma_k^b)_{pred}^2 - (\sigma_k^b)^2)((\sigma_k^b)_{pred}^2 - (\sigma_k^b)^2)^T \right\rangle \approx \mathbf{D}_k \mathbf{C}_k \mathbf{D}_k \quad (18)$$

where

$$\mathbf{D}_k = \alpha \text{diag}(\sigma_k^b)_{pred}^2 \quad (19)$$

$$\mathbf{C}_k = \mathbf{C}(l_h, l_v) \quad (20)$$

with  $\alpha$  a positive scalar and  $\mathbf{C}(l_h, l_v)$  a fixed isotropic correlation matrix with horizontal and vertical length scales  $l_h$  and  $l_v$ , respectively. In place of (16) we use

$$\langle \mathbf{e}_k^s (\mathbf{e}_k^s)^T \rangle \approx 2 \text{diag}((\sigma_k^o)^2 + \mathbf{I}_k(\sigma_k^b)_{pred}^2)^2 \quad (21)$$

Based on these estimates we take

$$\begin{aligned} \mathbf{L}_k &= \mathbf{L}_k(\alpha, l_h, l_v) \\ &= \mathbf{D}_k \mathbf{C}_k \mathbf{D}_k \mathbf{I}_k^T \left[ \mathbf{I}_k \mathbf{D}_k \mathbf{C}_k \mathbf{D}_k \mathbf{I}_k^T + 2 \text{diag}((\sigma_k^o)^2 + \mathbf{I}_k(\sigma_k^b)_{pred}^2)^2 \right]^{-1} \end{aligned} \quad (22)$$

The scalars  $\alpha$ ,  $l_h$ , and  $l_v$  are calibration parameters that control the temporal and spatial smoothing associated with the scheme (17).

## 4 Error reduction

From a mathematical point of view it is not difficult, for any given analysis method, to derive analysis error covariances from the error covariances of the background and of the observations. The required computations are expensive, especially for global analysis systems. Fischer and Courtier (1995) show how covariance estimates of the large-scale components of the analysis errors could be obtained in the context of a variational data assimilation system. Daley and Barker (2001) have formulated an efficient observation-space based scheme for computing analysis error covariances that can produce small-scale features in the variance estimates. Their scheme takes advantage of the preconditioning steps in the iterative conjugent-gradient solver in their analysis system. There are some unresolved questions about the accuracy of their estimates near the boundaries of the observation prisms which they use to localize the computations.

For the purpose of cycling the background error variances we are primarily interested in obtaining a reasonable estimate of the reduction factor  $\mathcal{R}$  in

$$(\sigma^a)^2 = \mathcal{R}(\sigma^b)^2 \quad (23)$$

where we omit the time index for the moment. Our approach will be to process the observations sequentially, and produce iterative variance estimates as follows:

$$(\sigma_{(0)}^a)^2 = (\sigma^b)^2 \quad (24)$$

$$(\sigma_{(i)}^a)^2 = \mathcal{R}_i(\sigma_{(i)}^a)^2, \quad i = 1, \dots, p \quad (25)$$

$$(\sigma^a)^2 = (\sigma_{(p)}^a)^2 \quad (26)$$

where  $\mathcal{R}_i$  is the reduction factor associated with observation  $i$ . If each of the  $\mathcal{R}_i$  are accurately computed, then this algorithm is close to optimal in case of uncorrelated observation errors.

Consider the case of a single in-situ observation  $x_o$  located at grid location  $i$ . We first assume that the error standard deviation for this observation is  $\sigma_i^o$ , and that the observation error is not correlated with the background error. Then we can show that the analysis error variance at location  $j$  is

$$(\sigma_j^a)^2 = [1 - c_{ij}^2 \kappa_i] (\sigma_j^b)^2 \quad (27)$$

where  $c_{ij}$  is the correlation between background errors at locations  $i$  and  $j$ , and

$$\kappa_i = \frac{(\sigma_i^b)^2}{(\sigma_i^b)^2 + (\sigma_i^o)^2} \quad (28)$$

Note that subscripts now indicate spatial locations. We see that the variance reduction is concentrated near the observation location. For example, in case of locally isotropic background error correlations with decorrelation length scale  $l$  we would have

$$c_{ij}^2 = 1 - \left(\frac{r_{ij}}{l}\right)^2 + O\left(\frac{r_{ij}}{l}\right)^4 \quad (29)$$

where  $r_{ij}$  denotes the distance between locations  $i$  and  $j$ . This shows that the effect of the observation on the analysis error variance diminishes with the square of the distance from the observation location. It also demonstrates that the background error correlations must be known to be able to estimate this effect.

This simplest possible case illustrates clearly that the analysis error variances are very sensitive to details of the observation and background error covariances. No computational approach, no matter how sophisticated, can change this fact.

We can generalize the model (27) for the local variance reduction due to a single observation in various ways. For example, if the error in the observation is correlated with the error in the background, then that observation contains less independent information. The analysis error variance will therefore be larger, i.e., less reduced. Instead of (27) we can then take

$$(\sigma_j^a)^2 = [1 - (1 - \rho_i)^2 c_{ij}^2 \kappa_i] (\sigma_j^b)^2 \quad (30)$$

with  $\rho_i$  the observation-background error correlation. The variance reduction factor is then

$$\mathcal{R}_i = [1 - (1 - \rho_i)^2 c_{ij}^2 \kappa_i] \quad (31)$$

Similarly we can formulate semi-empirical models for the local variance reduction due to any single observation associated with a particular instrument. With some additional effort one can model the variance reduction in physical space due to a single satellite sounding. It is not difficult, for example, to model the error reduction of the moisture field due to the analysis of a single observation of total precipitable water in a vertical column. Such a model must depend on the analysis method used, and will have to involve assumptions about the vertical structure of the background errors.

## 5 Sequential estimation

The following sequential algorithm puts together the three main elements of variance estimation that have been discussed in the previous sections:

$$\begin{aligned} (\sigma_k^m)^2_{pred} &= (\sigma_{k-1}^m)^2_{est} && \text{model error prediction} \\ (\sigma_k^b)^2_{pred} &= \mathbf{A}_{k,k-1} (\sigma_{k-1}^a)^2_{est} + (\sigma_k^m)^2_{pred} && \text{variance evolution} \\ (\sigma_k^m)^2_{est} &= (\sigma_k^m)^2_{pred} + \mathbf{L}_k \left[ (\mathbf{d}\mathbf{x}_k^o)^2 - \mathcal{S}_k((\sigma_k^b)^2_{pred}) \right] && \text{adaptive adjustment} \\ (\sigma_k^b)^2_{est} &= (\sigma_k^b)^2_{pred} + \left[ (\sigma_k^m)^2_{est} - (\sigma_k^m)^2_{pred} \right] && \text{variance update} \\ (\sigma_k^a)^2_{est} &= \mathcal{R}_k (\sigma_k^b)^2_{est} && \text{variance reduction} \end{aligned}$$

The algorithm must be initialized with *a priori* estimates  $(\sigma_0^a)^2$  and  $(\sigma_0^m)^2$  for the analysis and model error variances at time  $t = t_0$ .

This particular formulation of the algorithm uses the model error variance estimated at time  $t = t_{k-1}$  as a first guess for the model error variance at time  $t = t_k$ . This presumes that there is some degree of local persistence to the model error variances. One can set  $(\sigma_k^m)^2_{pred} = 0$  instead, which means that model error variances are continuously re-estimated from current observations. If a more useful statistical and/or flow-dependent prediction model for model error variances is available, it can be inserted in the first step of the algorithm.

The adaptive adjustment step is absolutely crucial to the algorithm. First of all, it is needed to remove the dependence of the variance estimates on the initial conditions  $(\sigma_0^a)^2, (\sigma_0^m)^2$ . At least in well-observed regions this dependence should quickly diminish with time. Secondly, the use of observations is supposed to compensate for the approximation errors incurred in each steps of the algorithm. As we have tried to argue, such approximations are inevitable because of lack of detailed information about model and observation errors. Ultimately we have only the observations to rely on for attempting to extract such information.

## 6 Experiments with moisture assimilation

We have implemented the methods described in this article for the moisture analysis component of the Physical-space/Finite-volume Data Assimilation System (fvDAS). This system was recently developed at the Data Assimilation Office at NASA<sup>1</sup>'s Goddard Space Flight Center and is scheduled to become operational in August 2002. It is based on a finite-volume general circulation model (fvGCM) (Lin and Rood 1996, 1998), a statistical quality control system (Dee *et al.* 2001), and uses the Physical-Space Statistical Analysis System (PSAS) to combine six-hour forecasts with observations (Cohn *et al.* 1998). A comprehensive description and evaluation of the fvDAS system will be published elsewhere.

<sup>1</sup>National Aeronautics & Space Administration

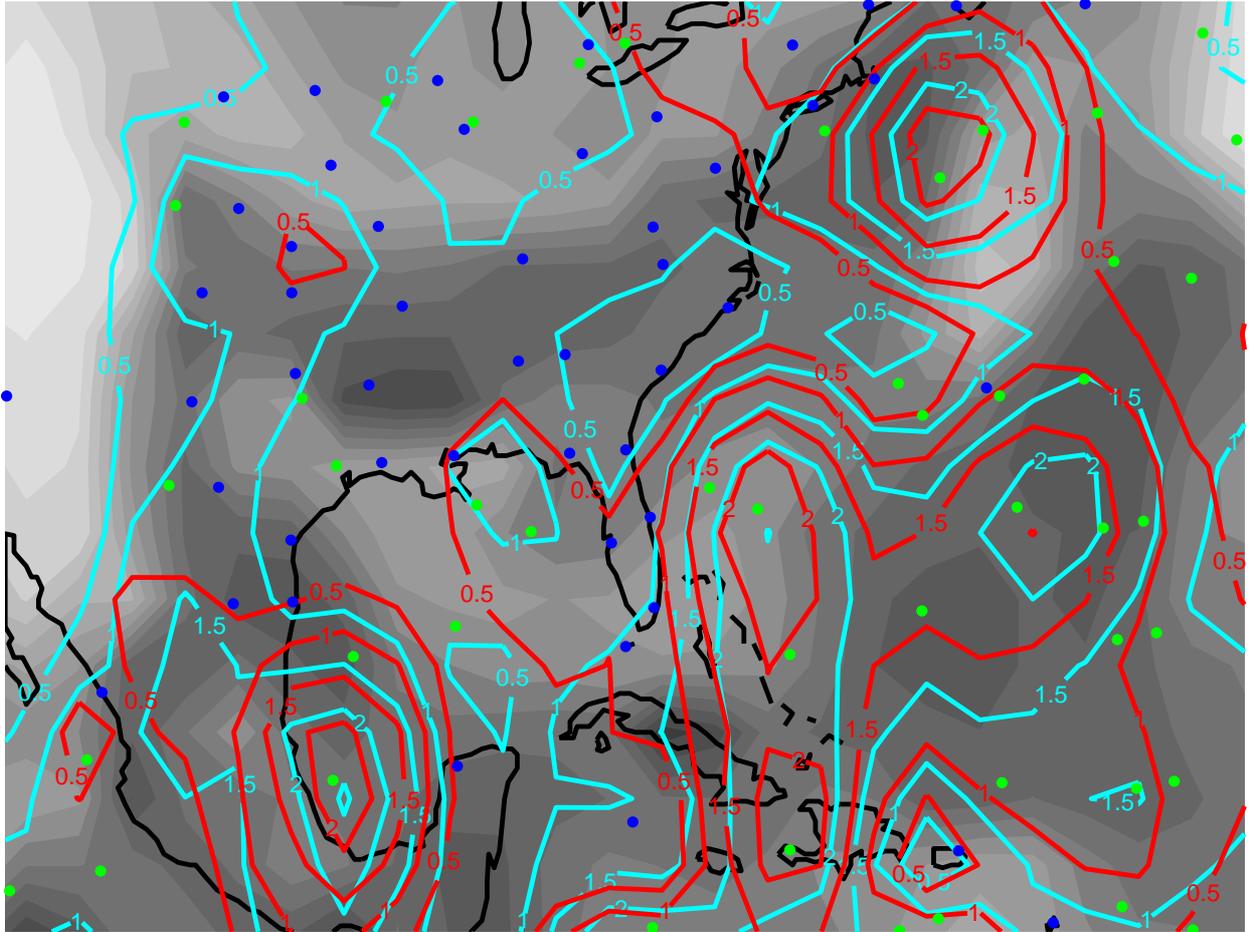


Figure 1: Estimated background error standard deviations (cyan contours) and model error standard deviations (red contours), superimposed on the specific humidity background estimates (shaded) for model layer 4, valid at 1 July 1998 (0 UTC). The region shown includes the East Coast of the United States, Central America, and the Caribbean. Darker shades correspond to higher values of specific humidity. All units are in g/kg. Blue dots indicate rawinsonde station locations; green dots mark the locations of TOVS retrievals.

The moisture analysis is based on the pseudo-relative humidity variable (Dee and da Silva 2002). In-situ moisture observations consist of radiosonde station data and interactive TOVS<sup>2</sup> retrievals (Joiner and Rokke 2000). We prescribe observation error standard deviations (in pseudo-relative humidity) that are a function of pressure only. Translated in terms of specific humidity, this corresponds to the assumption that the observation errors are a simple function of the saturation mixing ratio according to the model, i.e., they are a function of the background temperature field. See Dee and da Silva (2002) for details.

We also analyze estimates of total precipitable water (TPW) derived from SSM/I<sup>3</sup> data (Wentz 1997) in a two-step procedure which is essentially equivalent to a separability assumption on the background errors. We first generate a two-dimensional TPW increment to the background based on the data. This is done globally with PSAS and involves some horizontal smoothing based on the presumed moisture background error correlations. We then compute a three-dimensional correction to the background moisture field which is consistent with the TPW increment. The vertical structure of the correction is based on assumptions about the vertical correlations of the moisture background errors. Based on this analysis procedure we can derive a model for the analysis error reduction factor  $\mathcal{R}_i$  due to a single TPW observation; details of this model will be described elsewhere.

<sup>2</sup>Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder

<sup>3</sup>Special Sensor Microwave/Imager

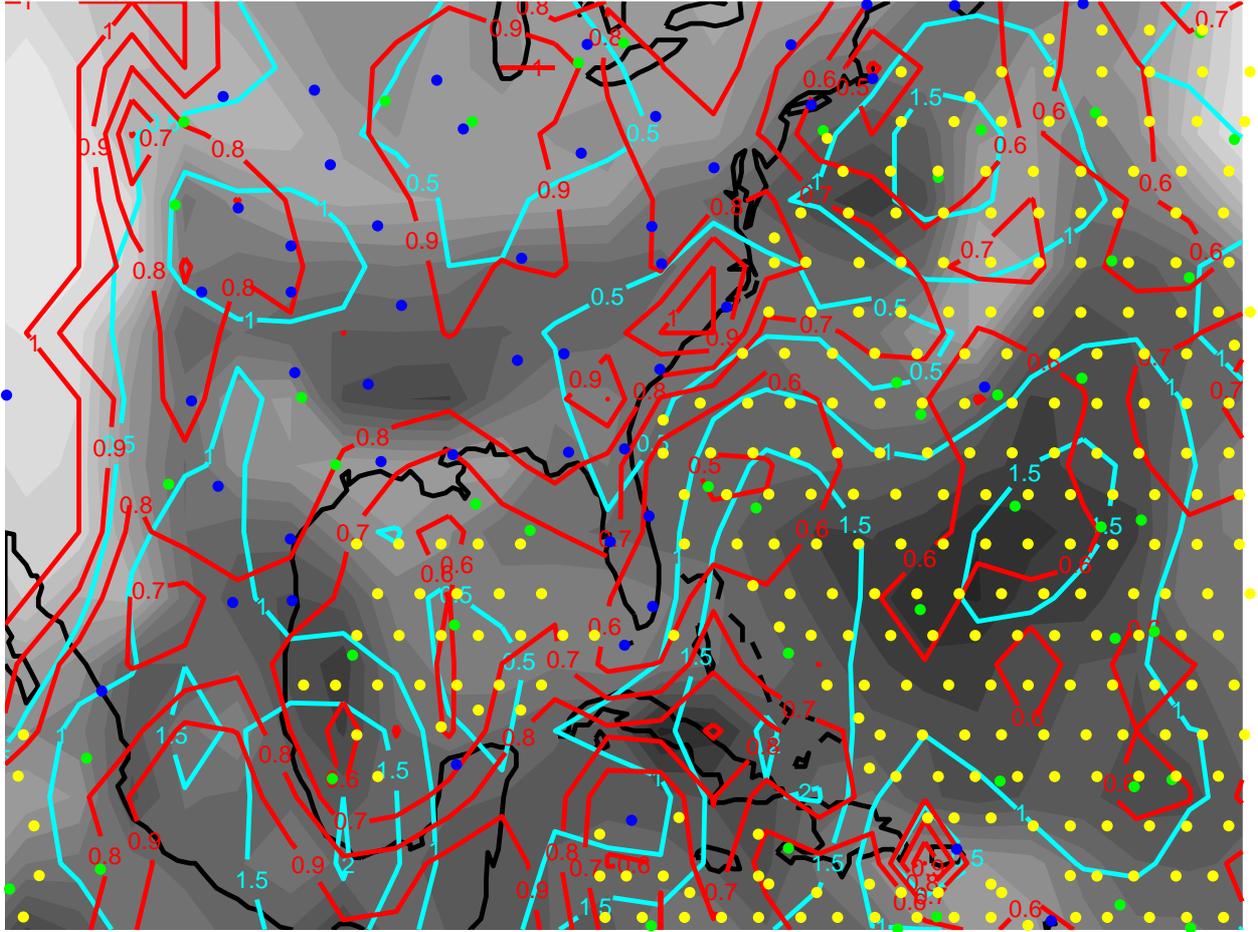


Figure 2: Estimated analysis error standard deviations (cyan contours) and the variance reduction factor (red contours), superimposed on the specific humidity analysis (shaded, using the same scale as in Fig. 1), for model layer 4, valid at 1 July 1998 (0 UTC). Locations of all observations used for the analysis are indicated, with blue dots for rawinsondes; green dots for TOVS retrievals, and yellow dots for TPW observations.

We used a low-resolution ( $2 \times 2.5$  degrees, 55 layers) version of the system for our initial experiments. Propagation of the analysis error variance field, as in (6), is computed with the tracer transport facility built into the fvGCM. We estimate the effect of model errors on the background error variances based on rawinsonde observations and interactive TOVS retrievals, using PSAS to solve the update equation (17). Analysis error computations involve the complete observing system, including TPW.

The computational overhead for the entire scheme amounts to less than 10% of the cost of the moisture analysis. We have not made a serious attempt to optimize software performance. We do not expect the relative cost to increase at higher model resolutions, since the fvGCM tracer transport computation is highly efficient. If necessary, the model error estimation step can be made cheaper by using a localized analysis scheme to solve (17), rather than a global solver such as PSAS. This is justifiable because the statistical estimation is far from optimal in any case. The cost of the analysis error estimation is proportional to the number of observations, and the computations are all local and therefore easily parallelized.

In Fig. 1 we show a snapshot of the estimated background error standard deviations (cyan contours) for specific humidity, superimposed on the background specific humidity field itself (gray shading). These estimates are valid at 0 UTC, 1 July 1998, for model layer 4, which is located on average at about 120 hPa above terrain. The maximum value of specific humidity in the area shown, corresponding to the darkest shade of gray, is almost 15 g/kg. The maximum estimated error standard deviation is about 3.5 g/kg. Also marked are the locations of

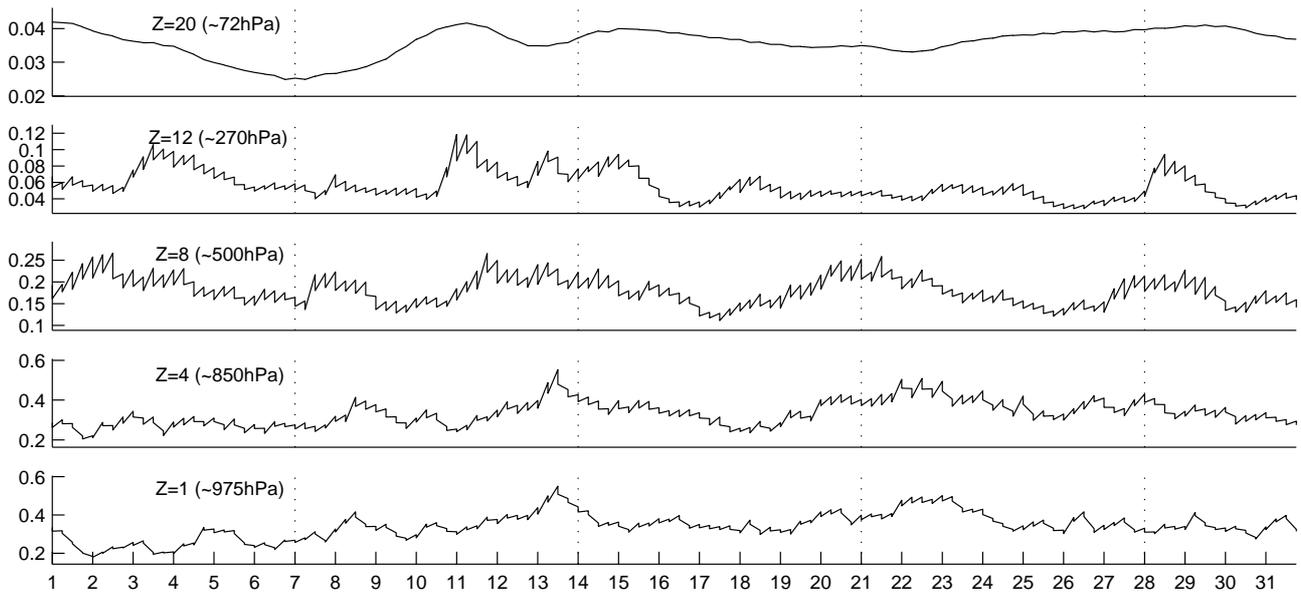


Figure 3: Regional rms of specific humidity error standard deviations for Europe, for several model layers, July 1998

the rawinsonde stations (blue dots) and of the TOVS retrievals (green dots) that were used to estimate the model error variance at that particular time. The red contours indicate the model error standard deviation estimated from those observations.

The analysis for the same time and location is shown in Figure 2, together with the locations of all observations used for the analysis (blue dots: rawinsondes; green dots: TOVS retrievals; yellow dots: TPW observations). Also shown are the estimated analysis error standard deviations (cyan contours) and the variance reduction factor (red contours).

Figure 3 shows the time evolution (for July 1998) of regionally averaged (rms) estimates of error standard deviations of specific humidity at various model levels in Europe. The curves show the 6-hourly cycle of error growth due to model error and error reduction due to the observational information. At the top level shown (at approximately 70 hPa) there are no observations, so that the only visible signal there is due to the advection of information. The estimated values of the error standard deviations at that level are rather sensitive to the initial specifications. At lower levels the available observations are used to anchor the estimates, based on observed-minus-background residuals. There the estimated values are sensitive to the prescribed observation error standard deviations. At the lowest level shown, the influence of the observations is felt only by extrapolation based on the assumed vertical background error correlations.

Figure 4 shows the zonal rms of the estimated background error standard deviations (contours) superimposed on the zonal mean of the background specific humidity field (shaded), at 1 July 1998 (0 UTC). This picture is fairly typical for that month, although the changes in time of the estimates are clearly visible in the zonal means. The largest errors occur in the lowest layers in tropics. The size of the errors is not simply proportional to the magnitude of the field, although there is a clear correlation. A relative maximum appears just below the tropical tropopause, which could be related to active convection and relatively large model errors there. The size of the errors relative to the field values at those altitudes suggest that the background estimates are probably not meaningful there. The estimated errors in the southern hemisphere appear to be too optimistic, possibly due to overestimation of the error reduction due to TPW observations.

Figure 5 shows, for the same time, the difference between the zonal rms of the estimated background error standard deviations and the zonal rms of the propagated analysis error standard deviations. To the extent that the latter were estimated accurately, this difference can be attributed to model errors. As it is, we see small

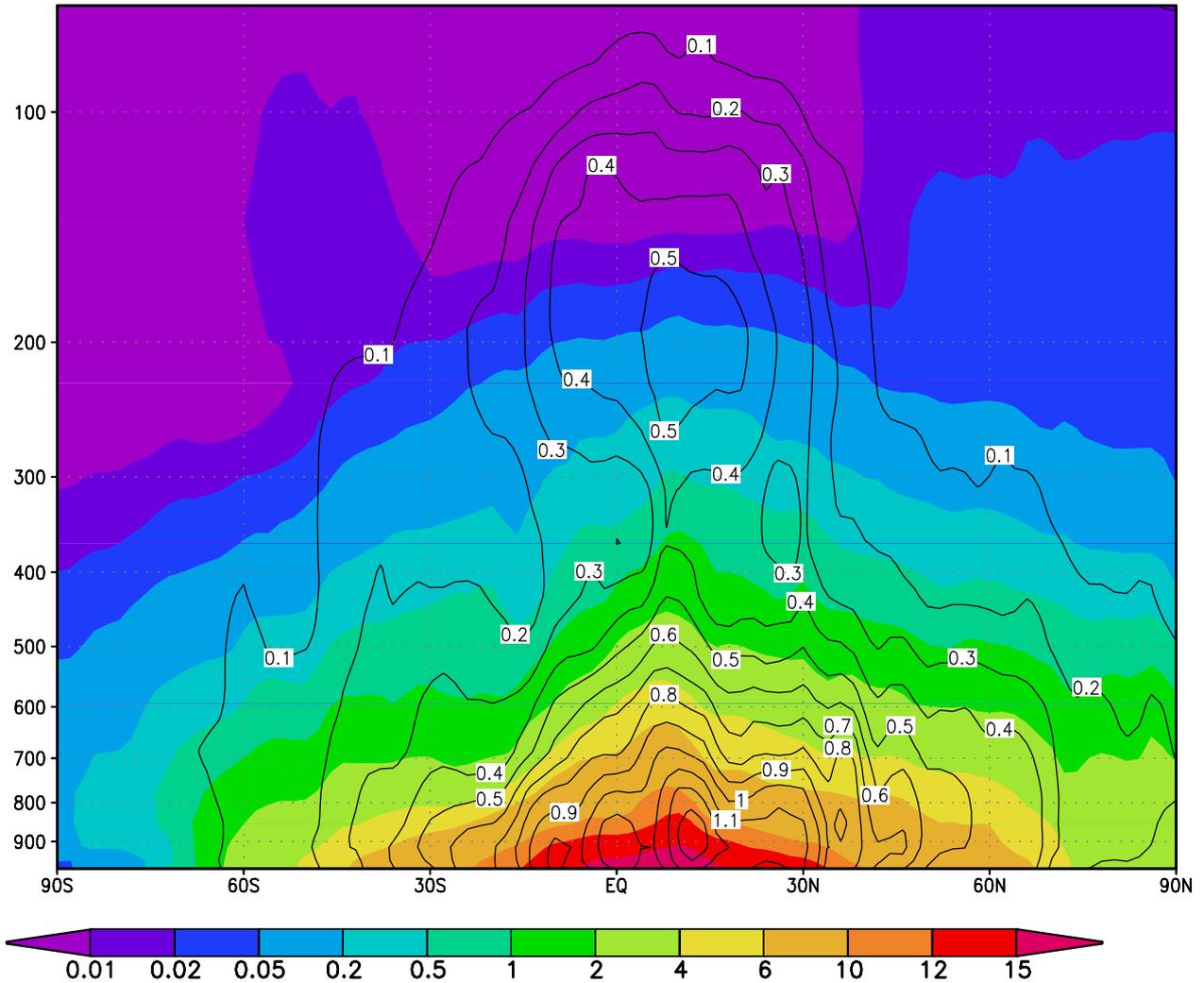


Figure 4: Zonal mean specific humidity background (shaded) and error standard deviations (contours), at 1 July 1998 (0 UTC)

negative adjustments of the estimates at higher altitudes in the tropics, which indicates that effect of propagated initial errors on the background errors was overestimated. The model errors appear to be largest in the tropics, especially in the lowest model layers. The accuracy of these estimates depends primarily on (1) the data coverage, and (2) the validity of the assumptions on observation errors that enter in to the estimation procedure.

## 7 Conclusion

We have presented a sequential algorithm for the estimation and adaptive updating of univariate background error variances, which accounts for the effects of (1) propagation of initial errors, (2) error growth due to model errors, and (3) error reduction due to the use of observations. Essential assumptions and/or restrictions for each of these aspects are that (1) can be modeled by advection, (2) can be estimated from observed-minus-background residuals, and (3) can be estimated from knowledge of the observing system and its error characteristics.

We have implemented the algorithm in a global atmospheric data assimilation system and applied it to the on-line estimation of specific-humidity background error standard deviations. We obtained estimates with realistic features at a reasonable cost. The algorithm is stable, in the sense that different initial specifications of the error

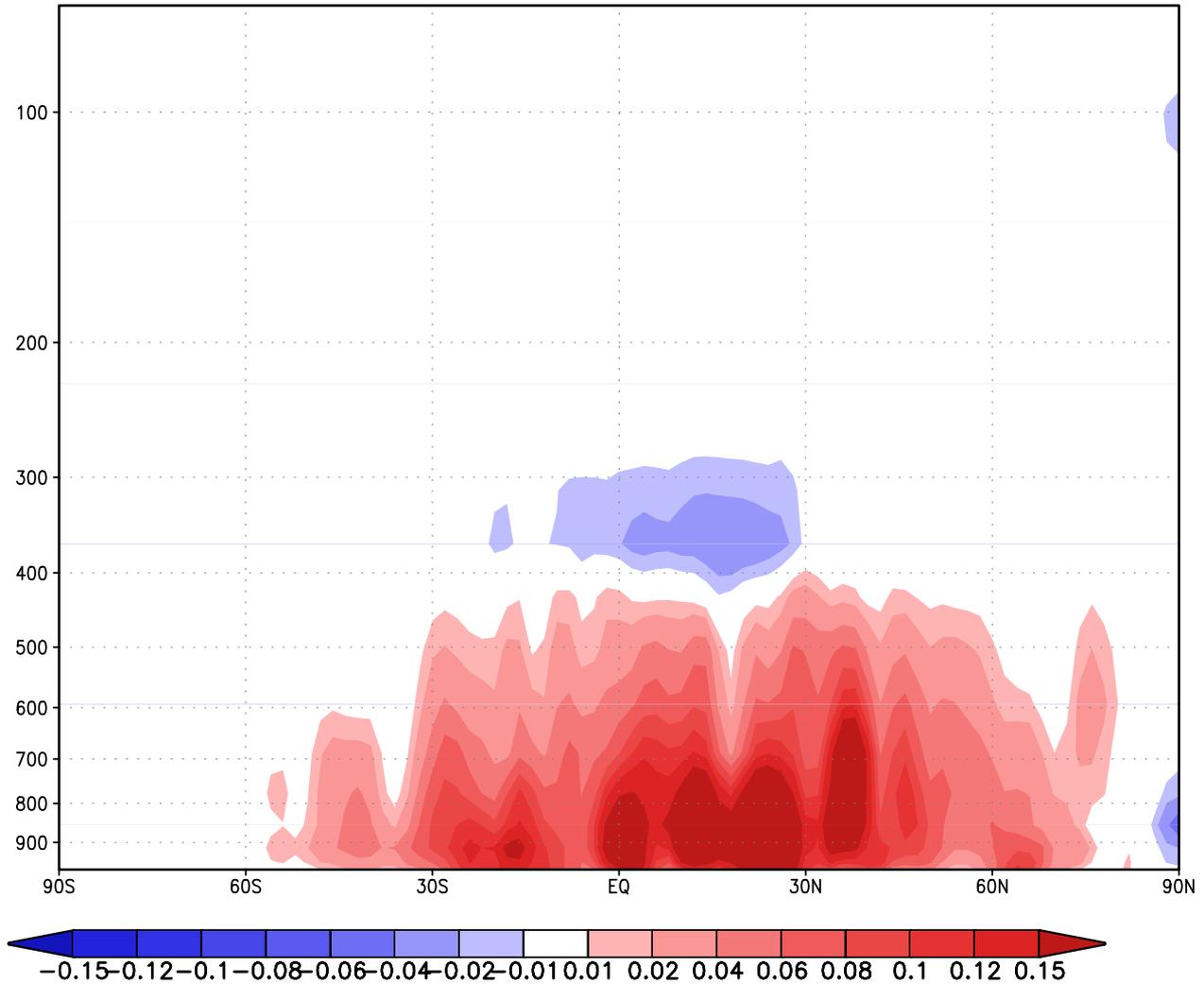


Figure 5: Estimated error growth due to model error: Zonal mean difference between the adjusted estimate of background error standard deviations and the predicted background error standard deviations based on advection of initial errors only, at 1 July 1998 (0 UTC)

variances quickly converge to the same estimates in regions where observations are available. This stability is due to the adaptive feature of the algorithm, which ensures that the estimates are consistent with observational residuals.

It is possible to use this method for cycling the background error variances for the mass variable (virtual potential temperature) as well. Wind error variances can then be estimated by (1) deriving the variance of the balanced component of wind error from the mass error covariance estimates, and (2) estimating the variance of the remaining decoupled component of wind error adaptively from wind observations.

We do not know whether the use of these variance estimates in an operational data assimilation system will lead to measurably better analyses and/or forecasts. The accuracy of the variance estimates produced by our algorithm depends on many uncertain parameters, most notably the parameters that describe the information content of the observations. This brings us back to the heart of the problem in data assimilation: there is no way to do it reliably in the absence of models and observations whose error characteristics are very well understood. The fundamental limitations are informational, rather than computational.

## References

- Cohn, S. E., 1993: Dynamics of short-term univariate forecast error covariances. *Mon. Wea. Rev.*, **121**, 3123–3149.
- 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., and E. Barker, 2001: *NAVDAS: Formulation and diagnostics*. *Mon. Wea. Rev.*, **129**, 869–883.
- Dee, D. P., 1990: Simplified adaptive Kalman filtering for large-scale geophysical models. Pp. 567–574 in M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran (eds.), *Realization and Modelling in System Theory*, Proceedings of the International Symposium MTNS-89, Volume I. Birkhäuser, Boston, 593pp.
- Dee, D. P., 1991: Simplification of the Kalman filter for meteorological data assimilation. *Q. J. R. Meteorol. Soc.*, **117**, 365–384.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.
- Dee, D. P., L. Rukhovets, R. Todling, A. M. da Silva, and J. W. Larson, 2001: An adaptive buddy check for observational quality control. *Q. J. R. Meteorol. Soc.*, **127**, 2451–2471.
- Dee, D. P., and A. M. da Silva, 2002: On the choice of variable for atmospheric moisture analysis. *Mon. Wea. Rev.*, *in press*.
- Fisher, M., and P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF Technical Memorandum*, No. 220, 35pp.
- Joiner, J., and L. Rokke, 2000: Variational cloud-clearing with TOVS data. *Q. J. R. Meteorol. Soc.*, **126**, 1–24.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D., J. Basic Eng.*, **82**, 35–45.
- Lin, S.-J., and R. B. Rood, 1996: Multidimensional flux-form semi-Lagrangian transport schemes. *Mon. Wea. Rev.*, **124**, 2046–2070.
- Lin, S.-J., and R. B. Rood, 1998: A flux-form semi-Lagrangian general circulation model with a Lagrangian control-volume vertical coordinate. *The Rossby-100 symposium*, Stockholm, Sweden.
- Wentz, F. J., 1997: A well-calibrated ocean algorithm for SSM/I. *J. Geophys. Res.*, **102**, 8703–8718.