On the choice of variable for atmospheric moisture analysis

Dick P. Dee[1,2]
Arlindo M. da Silva[2]

[1] *Science Applications International Corporation, Beltsville, Maryland*
[2] *Data Assimilation Office, NASA/Goddard Space Flight Center, Greenbelt, Maryland*

Abstract

The implications of using different control variables for the analysis of moisture observations in a global atmospheric data assimilation system are investigated. A moisture analysis based on either mixing ratio or specific humidity is prone to large extrapolation errors, due to the high variability in space and time of these parameters and to the difficulties in modeling their error covariances. Using the logarithm of specific humidity does not alleviate these problems, and has the further disadvantage that very dry background estimates cannot be effectively corrected by observations. Relative humidity is a better choice from a statistical point of view, because this field is spatially and temporally more coherent and error statistics are therefore easier to obtain. If, however, the analysis is designed to preserve relative humidity in the absence of moisture observations, then the analyzed specific humidity field depends entirely on analyzed temperature changes. If the model has a cool bias in the stratosphere this will lead to an unstable accumulation of excess moisture there.

A *pseudo-relative humidity* can be defined by scaling the mixing ratio by the background saturation mixing ratio. A univariate pseudo-relative humidity analysis will preserve the specific humidity field in the absence of moisture observations. A pseudo-relative humidity analysis is shown to be equivalent to a mixing ratio analysis with flow-dependent covariances. In the presence of multivariate (temperature-moisture) observations it produces analyzed relative humidity values that are nearly identical to those produced by a relative humidity analysis. Based on a timeseries analysis of radiosonde observed-minus-background differences it appears to be more justifiable to neglect specific humidity-temperature correlations (in a univariate pseudo-relative humidity analysis) than to neglect relative humidity-temperature correlations (in a univariate relative humidity analysis). A pseudo-relative humidity analysis is easily implemented in an existing moisture analysis system, by simply scaling observed-minus-background moisture residuals prior to solving the analysis equation, and rescaling the analyzed increments afterward.

1 Introduction

This paper concerns the choice of control variable for the analysis of moisture observations in a global atmospheric data assimilation system. Most analysis methods can be formulated in a variational framework as a procedure for minimizing the distance between model and observations (Courtier 1997). The choice of variable is important because the notion of 'distance' depends on it. This means that the error covariance models that determine the relative weighting of information in the analysis must be provided for the control variables. In practical terms, the analysis is essentially a procedure for extrapolating observational information to the model domain. Some representations of the moisture field are better suited to this procedure than others.

The question of which variable to use to for atmospheric moisture analysis has been around for a long time (Atkins 1974, van Maanen 1981). The major numerical weather prediction centers have arrived at different choices during the development of their assimilation systems. At the time of this writing, relative humidity is analyzed at the Met Office (Lorenc *et al.* 2000) and at the Australian Bureau of Meteorology Research Centre (P. Steinle, *pers. comm.*). The National Centers for Environmental Prediction (Parrish and Derber 1992), the European Centre for Medium-Range Weather Forecasts (Rabier *et al.* 1998), and the Japan Meteorological Agency (K. Onogi, *pers. comm.*) analyze specific humidity. The operational system at the Canadian Meteorological Centre (L. Fillion, *pers. comm.*) and the newly developed Naval Research Laboratory Atmospheric Variational Data Assimilation System (Daley and Barker 2001) use the logarithm of specific humidity. Some of the centers are currently rethinking various aspects of their humidity analysis, including the choice of control variable.

Here we document some of the considerations that went into the development of 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[1]'s Goddard Space Flight Center and is expected to become operational in early 2002. It is based on a finite-volume general circulation model (Lin and Rood 1996) and uses the Physical-Space Statistical Analysis System (PSAS) (Cohn *et al.* 1998). The system obtains information about atmospheric water vapor content from radiosonde soundings (Garand *et al.* 1992), from interactive TOVS[2] retrievals (Joiner and Rokke 2000), and from total precipitable water (TPW) estimates derived from SSM/I[3] data (Wentz 1997). Work is in progress to extract additional humidity information from radio occultation data transmitted by the Global Positioning System (Poli *et al.* 2002). A comprehensive description and evaluation of the fvDAS system will

---

[1] National Aeronautics & Space Administration
[2] Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder
[3] Special Sensor Microwave/Imager

be published elsewhere.

The need to carefully consider the choice of moisture variable became particularly clear to us during the development of the observational quality control component of the fvDAS. The final phase of the quality control relies on an adaptive buddy check algorithm (Dee *et al.* 2001), which involves the statistical estimation of the expected local variability of the flow. We initially found that the algorithm did not perform well for moisture observations, because the statistical estimation applied to mixing ratio observed-minus-background residuals was not meaningful due to the high variability of the field itself. We then developed a simple background-dependent transformation for the residuals that effectively removed this problem. This ultimately led us to use the same transformation to define a new humidity variable for the fvDAS global analysis system.

The outline of this paper is as follows. After briefly reviewing some background material, we discuss each of the currently prevalent choices of control variable: mixing ratio/specific humidity, logarithm of specific humidity, and relative humidity. Our main concern is with the statistical properties of the background and observation errors for each choice, as manifested in observed-minus-background residuals obtained from radiosonde humidity soundings. We then introduce the variable used in the fvDAS, which we refer to as pseudo-relative humidity, and discuss the implementation of a change to this variable in an existing analysis system. We discuss the differences and similarities between relative humidity and pseudo-relative humidity analyses, and finally present some preliminary results obtained with the fvDAS to illustrate the impact that such a change can bring about.

2   Background

The analysis in an atmospheric data assimilation system is constructed by combining a model-generated background estimate with bias-corrected and quality-controlled observations. If the background estimate is represented by the n-vector $\mathbf{x}^b$, the observations by the p-vector $\mathbf{y}^o$, and the mapping from the control variable to the observations by the vector function $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^p$, then the goal is to compute the analysis $\mathbf{x}^a$ that minimizes

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

where the $n \times n$ matrix $\mathbf{P}$ and $p \times p$ matrix $\mathbf{R}$ are the background and observation error covariances, respectively. In the special case when the observation operator $\mathbf{h}$ is linear, i.e., when $\mathbf{h}(\mathbf{x}) = \mathbf{H}\mathbf{x}$ for some $p \times n$ matrix $\mathbf{H}$, the mini-

mizing solution is

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{P}\mathbf{H}^T \left[\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\right]^{-1} \left[\mathbf{y}^o - \mathbf{H}\mathbf{x}^b\right]. \tag{2}$$

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 $\mathbf{x}^a - \mathbf{x}^b$ therefore strongly depends on the specification of the background error covariances. This is true in case of a nonlinear observation operator as well, and/or if the operator involves integration forward in time as in four-dimensional variational (4DVAR) assimilation.

To illustrate this general point with a simple example, consider the one-dimensional analysis of a single observation y^o of total precipitable water in a vertical column. Let $\mathbf{x} = (q_1, q_2, \ldots, q_n)^T$, where q_k is the mean specific humidity for a model layer whose pressure thickness is δp_k. The $1 \times n$ observation operator $\mathbf{H}$ is then defined by

$$\mathbf{H}\mathbf{x} = \frac{1}{g} \sum_{j=1}^n q_j \delta p_j, \tag{3}$$

with g the gravity constant. For simplicity we take a diagonal background error covariance $\mathbf{P} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$. The analyzed specific humidity for layer k is then

$$q_k^a = q_k^b + \frac{\sigma_k^2 \delta p_k}{g s^2} \cdot (y^o - y^b), \tag{4}$$

where $y^b = \mathbf{H}\mathbf{x}^b$ is the total precipitable water in the background column, and $s^2 = (\sigma^o)^2 + \sum_{j=1}^n \sigma_j^2 \delta p_j^2 / g^2$ with σ^o the observation error standard deviation. The vertical structure of the analysis increment $q_k^a - q_k^b$ in this case depends *only* on the assumed background error variances. If we include vertical correlations in $\mathbf{P}$ then the redistribution of moisture in the column will obviously depend on them as well.

In practice, background error covariances are not known and must be modeled. This can be done based on statistical properties of forecast differences (Parrish and Derber 1992), by means of ensemble methods (Evensen 1994) or by explicit parameterization (e.g., Riishøjgaard 1998). The usefulness of a covariance model is not necessarily related to its degree of mathematical sophistication. For example, the covariance evolution associated with the Kalman filter involves many modeling assumptions which are, at best, questionable in the context of atmospheric data assimilation (Dee 1991). A fundamental fact which is still widely misunderstood is that the ability to estimate error covariances is limited not by computing power but by a lack of data. For this reason, none of the existing approaches deals with the model errors (i.e. errors in the atmospheric prediction model) in a satisfactory manner.

These issues are especially pertinent to the analysis of atmospheric moisture, which is strongly affected by mesoscale dynamics and nonlinear physical processes that are poorly resolved by general circulation models. Errors in model predictions of the humidity field can be quite large, involving the deformation and displacement of synoptic features, and misrepresentation of small-scale sources and sinks. Most models exhibit systematic errors, such as a tendency to be consistently dry in certain areas, but these biases vary with season and location. The usual stochastic assumptions about model errors invoked in data assimilation (zero-mean, stationary, white noise) evidently do not apply here.

The question we try to address in this paper is therefore primarily a practical one, concerning the robustness of an analysis scheme in view of inevitable approximations in the representations of the errors. Specifically, we set out to choose a moisture analysis variable such that spatially and/or temporally averaged statistics can provide meaningful estimates of the error covariances. We will also consider whether it is even reasonable to attempt to characterize the error distributions by the covariances—i.e., whether the distribution is approximately Gaussian. These are basic prerequisites for any attempt to formulate more advanced flow-dependent covariance models. It will turn out, of course, that each choice of variable has its drawbacks, and that the most practical solution may depend on the priorities of the center that is operating the data assimilation system.

3   Mixing ratio and specific humidity

The water vapor content of a given volume of air can be expressed in terms of the *mixing ratio*

$$w = \frac{m_v}{m_d},$$ (5)

where m_v is the mass of water vapor and m_d the mass of dry air contained in the volume. The water vapor concentration, or *specific humidity* is

$$q = \frac{m_v}{m_v + m_d}.$$ (6)

Since $q = w/(1 + w)$, mixing ratio and specific humidity are equivalent independent parameters of the moist atmosphere. Their numerical values are nearly identical and rarely exceed $20\,\mathrm{g\,kg^{-1}}$. Peixoto and Oort (1992, Section 12.3) provide a comprehensive overview of the observed climatology of atmospheric water vapor. Specific humidity is a prognostic variable in most general circulation models, being conserved in a parcel of air in the absence of condensation and evaporation.

Figure 1 shows a typical snapshot of the spatial distribution of specific humidity, taken from a 55-layer $1° \times 1\frac{1}{4}°$ fvDAS analysis valid at 0 UTC on 1 January 1998. The central panel shows the layer-mean specific humidity for the fourth model layer, which corresponds to approximately 850 hPa over the oceans. The top panel shows the vertical distribution along the Equator in the lowest eight model layers, from the surface up to about 500 hPa. The right panel shows the vertical distribution along the Greenwich Meridian. The field shows a rapid decrease of magnitude in the vertical direction, and intricate patterns and sharp gradients in horizontal cross-sections. Maximum values occur at low altitudes and high temperatures.

[Figure 1 about here.]

We can obtain some information about the mixing ratio background errors by comparing the background estimates with radiosonde observations. If $\mathbf{w}^o$ is a vector of mixing ratio observations and $\mathbf{Hw}^b$ the corresponding vector of interpolated background values, then

$$\mathbf{w}^o - \mathbf{Hw}^b = \mathbf{e}^o - \mathbf{He}^b, \tag{7}$$

where $\mathbf{e}^o = \mathbf{w}^o - \mathbf{Hw}^t$ is the observation error and $\mathbf{e}^b = \mathbf{w}^b - \mathbf{w}^t$ the background error. Here $\mathbf{w}^t$ denotes a representation of the true mixing ratio field in the model state space, so that this definition of the observation error $\mathbf{e}^o$ includes what is commonly known as representativeness error (Lorenc 1986).

Equation 7 shows that statistical properties of the residuals $\mathbf{w}^o - \mathbf{Hw}^b$ are directly related to those of the background and observation errors. Dee (1995) and Dee and da Silva (1999) exploit this relationship in order to estimate unknown parameters of the covariance models for both types of errors. In the present article we are primarily interested in the degree of statistical homogeneity of the errors, as evidenced by spatially and temporally averaged statistics of the residuals. We will also examine whether the residuals tend to support a representation of the errors by Gaussian distributions with sample statistics based on time and space averaging.

Figure 2 displays root-mean-square (rms) statistics for mixing ratio residuals over a one-year period, taken from an fvDAS assimilation. The three main panels show the rms of the residuals as a function of time and pressure for stations in the Northern Hemisphere, Tropics, and Southern Hemisphere. The resolution in time is one day: each data point represents the rms of all available residuals on that day and at that pressure level. Only residuals associated with quality-controlled radiosonde humidity reports at mandatory levels up to 300 hPa were processed for these plots.

[Figure 2 about here.]

6

Not surprisingly, the variability of the residuals reflects that of the humidity field itself. The solid curves in the leftmost panels of Fig. 2 show, for each region, the rms of all residuals for the month January (blue) and July (red). They clearly indicate the large changes of magnitude with altitude, latitude, and season. The dashed curves in these panels provide an indication of the degree of normality of the residual distribution. They are obtained by computing, at each pressure level, the 68th percentile of the residual magnitudes for the ensemble associated with the corresponding solid curves. The dashed and solid curves should nearly coincide in case of a Gaussian distribution whose mean and standard deviation are functions of pressure only, since roughly 68% of an ensemble would then lie within one standard deviation of the mean.

The extreme variability and changes in scale of the errors and of the field itself causes some difficulties when using mixing ratio (or, equivalently, specific humidity) as the analysis variable for atmospheric moisture. This can be illustrated by considering the mixing ratio analysis at w_j^a at model grid location j obtained from a single observation w_i^o at grid location i, given by

$$w_j^a = w_j^b + \kappa[w_i^o - w_i^b], \tag{8}$$

where w_i^b, w_j^b are the background estimates at locations i, j, respectively. The coefficient κ depends on the specification of mixing ratio background and observation error covariances; compare with (2). In the absence of bias, the best linear unbiased estimator is obtained when

$$\kappa = \frac{\rho_{ij}\sigma_i^b\sigma_j^b}{(\sigma_i^b)^2 + (\sigma_i^o)^2}, \tag{9}$$

with ρ_{ij} the correlation between the background errors at locations i and j, σ_i^b, σ_j^b their standard deviations, and σ_i^o the error standard deviation associated with the observation. With $\kappa \neq 0$, the analysis (8) extrapolates moisture information from the observation location to nearby grid locations.

The ability to perform meaningful extrapolation of observational information in the analysis requires that expected changes in magnitude of the errors between any two locations be accurately represented in the error covariances. Clearly this is problematic for mixing ratio, in view of the large changes in scale over short distances in space and time. Large errors can result from locally inaccurate covariance specifications by extrapolating mixing ratio increments across sharp gradients, especially in the vertical direction. The problem can be made worse if the analysis scheme contains a mechanism for ensuring non-negative humidity estimates, since the distribution of the extrapolation errors will then be asymmetrical, resulting in a tendency to produce excess moisture in the analysis.

One can attempt to formulate a flow-dependent covariance model in order to account for the spatial and temporal variability of the humidity field. Rabier *et*

7

al. (1998) developed an empirical model for the specific humidity background error variances, which depends on the background temperature and relative humidity estimates. They also noted a small but significant accumulation of excess water vapor in the lower stratosphere, found to be due to the inaccurate extrapolation of information from upper tropospheric observations. Their solution to this problem was simply to avoid extrapolation altogether to levels higher than 100 hPa, by setting the error correlations between these and all lower levels to zero.

4   Logarithm of specific humidity

The discrepancy between the dashed and solid curves in the left panels of Fig. 2 indicates that mixing ratio errors are not well represented by a Gaussian distribution with monthly statistics. Phrased in more practical terms, this means that spatially and/or temporally averaged variance estimates may not be very meaningful as statistics of the mixing ratio errors. It has been argued that it is more reasonable to assume that the errors follow a lognormal distribution (e.g., Daley and Barker 2000), based on the idea that errors in tracer transport are perhaps more naturally modeled as an accumulation of independent multiplicative, rather than additive, effects (Ott 1995). That would suggest taking the *logarithm of specific humidity s* as an analysis variable:

$$s = \log q \tag{10}$$

An important feature of a moisture analysis based on s is that the specific humidity $q = e^s$ in the analyzed state is always positive.

Note that additive errors in s correspond to multiplicative errors in q. It might seem reasonable to assume that the magnitudes of both background and observation errors tend to be proportional to the field itself. This would imply that the rms statistics for observation residuals in s are more homogeneous in space and time than those in q (or w). Figure 3, which is like Fig. 2 but for observation residuals $s^o - Hs^b$, shows that the spatial variability of the errors in $\log q$ is still considerable, although not quite as large as that of w. However, the dashed curves in the left panels do not closely match the solid curves, which suggests that the distribution of the errors is *not* lognormal.

[Figure 3 about here.]

The use of an analysis variable that has a singularity at $q = 0$ has a serious drawback, which can be explained by considering the analysis q_i^a of a single observation q_i^o at model grid point i. A linear analysis of $\log q$ gives

$$\log q_i^a = \log q_i^b + \kappa[\log q_i^o - \log q_i^b] \tag{11}$$

8

so that

$$q_i^a = (q_i^b)^{1-\kappa}(q_i^o)^{\kappa}.$$ (12)

Assuming $\epsilon < \kappa < 1-\epsilon$ for some $\epsilon > 0$, this means that the analyzed humidity at a given location will approach zero whenever the background estimate *or* the observation at that location is close to zero. Unfortunately it is not unusual for a model-generated background estimate to be very dry in locations where the observations indicate that, in fact, the atmosphere is wet. The use of $\log q$ as an analysis variable would result in de facto rejection of all observational information in such cases. The singularity at $q = 0$ can be removed by analyzing $\log(c + q)$ for some constant $c > 0$, which would have to be comparable in magnitude to the humidity field itself. However, this would eliminate the advantage of a guaranteed non-negative humidity analysis, and it would defeat any theoretical argument that suggests a lognormal error distribution in the first place.

5   Relative humidity

The mixing ratio w^s of a volume of air which is saturated with water vapor is called the *saturation mixing ratio*. This quantity is a known function of pressure and temperature; see Wallace and Hobbs 1977, Section 2.6, for a precise definition and further details. *Relative humidity* is defined by

$$rh = \frac{w}{w^s(T,p)}.$$ (13)

Relative humidity is a useful dynamic parameter for the description of moist physics in the atmosphere. Its values are affected by all dynamic processes that result in a change of temperature, such as convection and subsidence, absorption of radiation, and release of latent heat. Model parameterizations of cloud formation and the prediction of precipitation are particularly sensitive to changes in the relative humidity. Peixoto and Oort (1996) survey the climatology of relative humidity in the global atmosphere, as derived from radiosonde observations.

The rms statistics of radiosonde-observed relative humidity residuals $(rh)^o - H(rh)^b$ are shown in Fig. 4. Changes in magnitude of the statistics with pressure, latitude, and season are still noticeable but relatively modest. The apparently larger errors at the upper levels in the Tropics are consistent with the increased uncertainties associated with active convection and moisture divergence below the tropopause. The increased noisiness in the lower panel is primarily due to sampling; there are relatively few radiosonde stations in the Southern Hemisphere. The left panels show that the 68th percentiles of the residual magnitudes match the rms statistics reasonably well in each of the

9

cases shown. This indicates that a Gaussian distribution is more plausible for relative humidity errors than for errors in specific humidity or its logarithm.

[Figure 4 about here.]

If the moisture analysis control variable is relative humidity, then either temperature or humidity observations can affect both the analyzed temperature and specific humidity fields. For example, in the absence of humidity observations, a single temperature observation T_i^o at model grid point i gives

$$T_i^a = T_i^b + \kappa[T_i^o - T_i^b] \tag{14}$$

$$(rh)_i^a = (rh)_i^b \tag{15}$$

which implies a change in the mixing ratio at that location:

$$w_i^a \neq w_i^b \tag{16}$$

by virtue of (13) and the fact that the saturation mixing ratio w^s depends on temperature. Therefore, any change in temperature implied by observations will cause the water vapor mixing ratio to be adjusted in such a way that the relative humidity background estimates remain unchanged.

Lorenc et $al.$ (1996) have shown in the context of the Met. Office global data assimilation system that the preservation of relative humidity in the absence of humidity observations can be advantageous in certain meteorologically important situations, and their work has led to a decision to use relative humidity for the moisture analysis in their system (Lorenc et $al.$ 2000). The argument is based on the fact that the cloud parameterizations incorporated in the model respond primarily to relative humidity. If the model has a tendency to be too cool, and this tendency is corrected by observations, then it is preferable to maintain the model's relative humidity by increasing the specific humidity estimates. This was shown to help improve the predicted precipitation in moist, cyclonic situations in mid-latitudes.

By the same reasoning, however, if the model has a cool bias in the stratosphere then the warming effect of temperature data will induce a spurious accumulation of moisture there. In fact, from (13) we have

$$w_i^a \approx w_i^b + \Gamma\left(T_i^a - T_i^b\right) \tag{17}$$

where

$$\Gamma = (rh)_i^b \cdot \left.\frac{\partial w^s}{\partial T}\right|_{T=T_i^b}. \tag{18}$$

If the mean temperature increment $\langle T^a - T^b \rangle$ is positive, then $\langle \Gamma \rangle$ will also be positive and (17) leads to unlimited growth in the assimilation. Stratospheric moisture is long-lived, so the only way to counteract this growth is by

10

introducing an artificial moisture sink in the stratosphere, e.g. by Newtonian relaxation to climatology. Even in the absence of model temperature biases, a slow accumulation of moisture can result from humidity perturbations induced by zero-mean random temperature forcing, if, as discussed earlier, the analysis contains a mechanism for ensuring non-negative humidity estimates.

6   Pseudo-relative humidity

We can define a *pseudo-relative humidity* by

$$\tilde{w} = \frac{w}{w^{sb}} = \frac{w}{w^s(T^b, p)} \tag{19}$$

where T^b is the background temperature. The background pseudo-relative humidity and relative humidity fields are, of course, identical:

$$\tilde{w}^b = \frac{w^b}{w^{sb}} = (rh)^b. \tag{20}$$

However, the observed pseudo-relative humidity is *not* equal to the observed relative humidity:

$$\tilde{w}^o = \frac{w^o}{w^{sb}} \neq (rh)^o, \tag{21}$$

but rather amounts to a flow-dependent transformation of the observed mixing ratio.

Pseudo-relative humidity predicts relative humidity fairly well, depending on the accuracy of the background temperature estimates. The rms error statistics of pseudo-relative humidity and of relative humidity should therefore be similar. This is supported by Fig. 5, which shows the daily rms statistics for radiosonde-observed pseudo-relative humidity residuals $\tilde{w}^o - H\tilde{w}^b$; compare with Fig. 4. The pseudo-relative humidity statistics show a slightly larger vertical gradient, particularly at the highest levels in the Tropics. In all other respects the similarities are striking.

[Figure 5 about here.]

Given a single observation at location i, the pseudo-relative humidity analysis at location j is

$$\tilde{w}_j^a = \tilde{w}_j^b + \tilde{\kappa} \, [\tilde{w}_i^o - \tilde{w}_i^b], \tag{22}$$

11

where $\tilde{\kappa}$ now depends on the covariances for pseudo-relative humidity background and observation errors:

$$\tilde{\kappa} = \frac{\tilde{\rho}_{ij}\tilde{\sigma}_i^b \tilde{\sigma}_j^b}{(\tilde{\sigma}_i^b)^2 + (\tilde{\sigma}_i^o)^2}. \tag{23}$$

Based on the residual statistics shown in Fig. 5, it is not unreasonable to represent pseudo-relative humidity errors using sample statistics based on time and space averaging. Homogeneous pseudo-relative humidity errors correspond to mixing ratio errors that depend primarily on the local saturation mixing ratio as predicted by the model. In physical terms, the uncertainty is expected to be large where the atmosphere has a high capacity for water vapor, i.e., at low levels and high temperatures.

We can show that the analysis of pseudo-relative humidity is equivalent to an analysis of mixing ratio with flow-dependent covariance models for both background and observation errors. It is easy to see this for the special case of a single observation at a model gridpoint, since (22, 23) corresponds to (8, 9) with

$$\sigma_i^o = \tilde{\sigma}_i^o w_i^{sb}, \tag{24}$$

$$\sigma_j^b = \tilde{\sigma}_j^b w_j^{sb}, \tag{25}$$

$$\rho_{ij} = \tilde{\rho}_{ij}, \tag{26}$$

confirming that constant $\tilde{\sigma}_i^o, \tilde{\sigma}_j^b$ correspond to mixing ratio error standard deviations σ_i^o, σ_j^b that are proportional to the background saturation mixing ratios w_i^{sb}, w_j^{sb}, respectively.

More generally, consider the variational analysis for mixing ratio $\mathbf{w}$, obtained by minimizing

$$\begin{aligned} J(\mathbf{w}) = (\mathbf{w}^b - \mathbf{w})^T \mathbf{P}^{-1}(\mathbf{w}^b - \mathbf{w}) \\ + (\mathbf{y}^o - \mathbf{h}(\mathbf{w}))^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{h}(\mathbf{w})). \end{aligned} \tag{27}$$

The pseudo-relative humidity analysis corresponds to a change of variable in model state space

$$\widetilde{\mathbf{w}} = \mathbf{D}^{-1}\mathbf{w}, \qquad \mathbf{D} = \mathrm{diag}(\mathbf{w}^{sb}), \tag{28}$$

and in observation space

$$\widetilde{\mathbf{y}}^o = \mathbf{E}^{-1}\mathbf{y}^o, \qquad \mathbf{E} = \mathrm{diag}(\mathbf{h}(\mathbf{w}^{sb})). \tag{29}$$

Accordingly, let

$$\widetilde{\mathbf{P}} = \mathbf{D}^{-1}\mathbf{P}\mathbf{D}^{-1}, \tag{30}$$

$$\widetilde{\mathbf{R}} = \mathbf{E}^{-1}\mathbf{R}\mathbf{E}^{-1}. \tag{31}$$

Linear algebra then shows that (27) can be written

$$J(\widetilde{\mathbf{w}}) = (\widetilde{\mathbf{w}}^b - \widetilde{\mathbf{w}})^T \widetilde{\mathbf{P}}^{-1}(\widetilde{\mathbf{w}}^b - \widetilde{\mathbf{w}})$$
$$+ (\widetilde{\mathbf{y}}^o - \widetilde{\mathbf{h}}(\widetilde{\mathbf{w}}))^T \widetilde{\mathbf{R}}^{-1}(\widetilde{\mathbf{y}}^o - \widetilde{\mathbf{h}}(\widetilde{\mathbf{w}})), \tag{32}$$

where

$$\widetilde{\mathbf{h}}(\widetilde{\mathbf{w}}) = \mathbf{E}^{-1}\mathbf{h}(\mathbf{D}\widetilde{\mathbf{w}}). \tag{33}$$

This proves the equivalence between a mixing ratio analysis with covariance models $\mathbf{P}, \mathbf{R}$ and a pseudo-relative humidity analysis with covariance models $\widetilde{\mathbf{P}}, \widetilde{\mathbf{R}}$. The diagonal transformations $\mathbf{D}, \mathbf{E}$, which depend on the background temperature field, generate flow-dependent mixing ratio error covariance models even when $\widetilde{\mathbf{P}}, \widetilde{\mathbf{R}}$ are stationary.

Fig. 6 shows a simple example of the flow-dependent mixing ratio increments that can occur in a pseudo-relative humidity analysis. We computed the vertical impact of a moisture observation in case of a background temperature profile with a lapse rate of $8\,\mathrm{K\,km^{-1}}$ below $220\,\mathrm{hPa}$, and constant temperature above. Using a scale height $\mathrm{H} = 7\,\mathrm{km}$, we calculated the background saturation mixing ratio $\mathbf{w}^{sb}$ and 'observed' the mixing ratio

$$w_i^o = w_i^b + \frac{1}{2}w_i^{sb} \tag{34}$$

at a single pressure level. The value and location of the observation residual $w_i^o - w_i^b$ is marked by the dotted lines in each panel. We then computed pseudo-relative humidity increments for this exercise using

$$\tilde{\kappa} = \exp(-20(\log(p_i/p_j))^2) \tag{35}$$

in (22), which corresponds to a vertically homogeneous and isotropic pseudo-relative humidity error covariance model. The curves in each of the panels show the vertical structure of the mixing ratio increments $\mathbf{w}^a - \mathbf{w}^b$ associated with observations at 850, 500, 250, and $200\,\mathrm{hPa}$, respectively. Note the different scales indicated along the horizontal axes: the increments are proportional to the saturation mixing ratios, which decrease rapidly with altitude. Background temperatures are indicated along the vertical axis of the rightmost panel. The increment in this panel shows the effect of an abrupt change in lapse rate at the tropopause, whose location is indicated by the gray horizontal bar. For tropospheric observations the maximum mixing ratio increment generally occurs slightly below the observation location.

[Figure 6 about here.]

13

7 Discussion

A crucial question for the analysis of atmospheric moisture is whether errors in relative humidity can be considered statistically independent of errors in temperature. That would justify a univariate relative humidity analysis, which preserves the relative humidity field in the absence of moisture observations. A univariate pseudo-relative humidity analysis, on the other hand, would preserve the specific humidity field in that case. We earlier mentioned the work by Lorenc *et al.* (1996), who argue in favor of preserving relative humidities, although they were specifically concerned with mid-latitude weather prediction and the performance of their cloud parameterization scheme. We pointed out, however, that a univariate relative humidity analysis can cause an unstable accumulation of excess moisture in the stratosphere. This argument has carried considerable weight at the Data Assimilation Office, where there is a particular emphasis on stratospheric analysis.

In order to shed some light on the underlying question of multivariate moisture-temperature error correlations we examined timeseries of simultaneous relative humidity and temperature observed-minus-background residuals. Using December 1999 radiosonde reports, we computed the correlation coefficient between relative humidity and temperature residuals at each station and at each mandatory level up to 300 hPa. The dashed curves in Fig. 7 show the average timeseries correlations for all stations with at least 30 complete reports in the Northern Hemisphere, Tropics, and Southern Hemisphere. We see significant anti-correlations between the relative humidity and temperature residuals, increasing in magnitude toward the surface. At the lowest levels the correlations are typically about -0.5.

[Figure 7 about here.]

A recent paper by Franke and Barker (2000) shows very similar results. They computed the cross-covariances between relative humidity and temperature errors using residuals between radiosonde observations and 6-hour forecasts from the Naval Operational Global Atmospheric Prediction System (NOGAPS) for the period March–June 1998. The diagonal elements of the cross-correlation matrix they obtained (shown in their Fig. 13) correspond to the dashed curves in our Fig. 7, and are nearly identical if not slightly more negative.

We now turn to the correlations between mixing ratio and temperature residuals, shown in Fig. 7 (solid curves), again computed from December 1999 radiosonde data. These are significantly and consistently smaller than the relative humidity-temperature correlations. As a further check we also calculated spatially (as opposed to temporally) averaged cross-correlations on a daily basis, similar to the computation of the daily rms statistics displayed in Figs. 2–5. The results (not

14

shown) are entirely consistent with both sets of curves in Fig. 7. This suggests that it is more accurate to neglect mixing ratio-temperature error covariances in analyzing tropospheric moisture than to neglect relative humidity-temperature error covariances.

The moisture analysis is obviously less sensitive to the choice of control variable when both moisture and temperature are adequately observed. To illustrate, we show in Fig. 8 the results of a univariate pseudo-relative humidity analysis of mixing ratio observations, combined with a univariate temperature analysis. These analyses are valid for the Eastern United States on 1 January 2002 at 0 UTC, and are based on actual 500 hPa radiosonde observations available at that time. We used homogeneous and isotropic univariate covariance models for this exercise, with $\sigma^b = 2\sigma^o$ and a horizontal decorrelation length scale of 300 km for all variables at all locations. The top panel shows the observed-minus-background mixing ratio residuals with the mixing ratio analysis increments superimposed, and the center panel shows the observed-minus-background temperature residuals and increments. From these and the temperature background field we computed the implied relative humidity increments, shown in the bottom panel.

[Figure 8 about here.]

Figure 9 then shows the result of combining a univariate relative humidity analysis with a univariate temperature analysis, using the same set of radiosonde reports and the same covariance specifications. The top panel shows the relative humidity increments, directly computed from the relative humidity observed-minus-background residuals. The temperature increments were already shown in the previous figure, and the bottom panel now shows the implied mixing ratio increments. Figure 10 shows the difference between the relative humidity increments associated with the two analyses (left panel; this is the difference between the bottom panel of Fig. 8 and the top panel of Fig. 9) and the difference between the mixing ratio increments associated with the two analyses (right panel; this is the difference between the top panel of Fig. 8 and the bottom panel of Fig. 9). We see that the relative humidity increments are equal to within 3% in most places. The differences between the mixing ratio increments increase with the temperature gradient South of Florida.

[Figure 9 about here.]

[Figure 10 about here.]

15

8 Implementation

It is straightforward to implement a change of variable to pseudo-relative humidity in an existing variational analysis system. For convenience we assume that the original control variable is mixing ratio, although the same procedure can be applied if the existing system is based on some other variable, simply by redefining the diagonal transformations $\mathbf{D}$ and $\mathbf{E}$. Writing the analysis equation (32) for pseudo-relative humidity as

$$
\begin{aligned}
J(\widetilde{\mathbf{w}}) = (\widetilde{\mathbf{w}}^b - \widetilde{\mathbf{w}})^T \widetilde{\mathbf{P}}^{-1} (\widetilde{\mathbf{w}}^b - \widetilde{\mathbf{w}}) \\
+ \left[\mathbf{E}^{-1}(\mathbf{y}^o - \mathbf{h}(\mathbf{w})) \right]^T \widetilde{\mathbf{R}}^{-1} \left[\mathbf{E}^{-1}(\mathbf{y}^o - \mathbf{h}(\mathbf{w})) \right],
\end{aligned}
\tag{36}
$$

suggests the following implementation:

1. Compute $\mathbf{w}^{sb}$ and $\mathbf{h}(\mathbf{w}^{sb})$ from the background state;

2. Compute the scaled residuals $\mathbf{E}^{-1}(\mathbf{y}^o - \mathbf{h}(\mathbf{w}))$;

3. Solve the variational problem with covariance models $\widetilde{\mathbf{P}}, \widetilde{\mathbf{R}}$ and observation operator $\widetilde{\mathbf{h}}$;

4. Compute the mixing ratio increment $\mathbf{w}^a - \mathbf{w}^b = \mathbf{D}(\widetilde{\mathbf{w}}^a - \widetilde{\mathbf{w}}^b)$.

In the linear case the solution procedure is expressed by

$$
\mathbf{w}^a = \mathbf{w}^b + \mathbf{D}\widetilde{\mathbf{K}}\mathbf{E}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{w}^b),
\tag{37}
$$

with

$$
\widetilde{\mathbf{K}} = \widetilde{\mathbf{P}}\widetilde{\mathbf{H}}^T \left[\widetilde{\mathbf{H}}\widetilde{\mathbf{P}}\widetilde{\mathbf{H}}^T + \widetilde{\mathbf{R}} \right]^{-1}.
\tag{38}
$$

The change of variable can be regarded as a preconditioner for the variational analysis equation (Lorenc 1988), and it may improve the convergence properties of an iterative solver since the error covariance models for pseudo-relative humidity should be better conditioned than those for mixing ratio.

The modified observation operator $\widetilde{\mathbf{h}}$ as defined by (33) involves multiplying and dividing by background saturation mixing ratios. For some data types it may be more practical, and possibly more accurate, to ignore (33) and reformulate the observation operator directly in terms of pseudo-relative humidity. For example, in case of radiosonde observations

$$
\mathbf{h}(\mathbf{w}) = \mathbf{H}\mathbf{w},
\tag{39}
$$

where $\mathbf{H}$ now represents spatial interpolation from the model state to the observation locations. In this case

$$
\widetilde{\mathbf{h}}(\widetilde{\mathbf{w}}) = \mathbf{H}\widetilde{\mathbf{w}}
\tag{40}
$$

16

is preferable over (33), since the pseudo-relative humidity field is smoother than the mixing ratio field and therefore lends itself better to spatial interpolation.

9 fvDAS results

We briefly summarize some initial results obtained with the pseudo-relative humidity analysis incorporated in the fvDAS. Figure 11 shows bar plots of rms mixing ratio background errors for three separate experiments. Errors are computed with respect to January 1998 radiosonde mixing ratio observations at 1000, 850, 700, 500, and 300 hPa, separately for the Northern Hemisphere, Tropics, and Southern Hemisphere. The top bar (red) in each panel corresponds to a control experiment in which mixing ratio was the analysis variable, and the only source of moisture data consisted of radiosonde observations. The blue bar shows the positive impact on the mixing ratio errors of changing to a pseudo-relative humidity analysis, still analyzing radiosonde observations only. The green bar adds the impact of interactive TOVS humidity retrievals (Joiner and Rokke 2000) to the pseudo-relative humidity analysis.

These results show that the change of variable from mixing ratio to pseudo-relative humidity produces a small but not insignificant decrease of rms background errors at all levels. It is encouraging that the introduction of TOVS moisture information further improves the fit to radiosonde observations.

We did not perform model bias correction (Dee and Todling 2000) for this study, nor did we make an effort to optimize the error covariance models for any of the experiments. In each case we specified background and observation error standard deviations and isotropic background error correlations, as discussed in Dee and Todling (2000), such that the relative weight of an observation at its location (κ in (9) with $i = j$) is a function of pressure only. The mixing ratio and pseudo-relative humidity analyses therefore produce essentially identical results at the observation locations, but extrapolate the observational information differently, as illustrated by the example in Fig. 6.

[Figure 11 about here.]

10 Conclusion

We reviewed several commonly used choices of control variable for the moisture analysis in atmospheric data assimilation. For water vapor mixing ratio and

17

specific humidity, error covariance modeling is complicated by the high variability in space and time of the errors and of the field itself. Inaccurate covariance specifications for these variables can cause extrapolation errors in the analysis which are large compared to the local field values. Furthermore, characterization of the estimation errors in terms of their second moments is not very meaningful, since the error distributions are far from Gaussian. These remarks apply to the logarithm of specific humidity as well, which suffers from the additional drawback that dry background estimates are not corrected well by wet observations.

Relative humidity is a better choice from a statistical point of view, in the sense described above. The relative humidity field is more coherent in space and time, and therefore lends itself better to extrapolation. Since relative humidity depends on temperature, its use as an analysis variable implies that temperature observations directly affect the specific humidity field. We found that this can result in an unrealistic and unstable accumulation of moisture in the stratosphere in the presence of temperature model biases. Fundamentally, a univariate relative humidity analysis involves the assumption that relative humidity errors and temperature errors are statistically independent. However, radiosonde station timeseries of observed-minus-background residuals show a significant negative correlation between relative humidity residuals and temperature residuals. Correlations between specific humidity residuals and temperature residuals are generally much smaller in magnitude.

We defined a pseudo-relative humidity by scaling the mixing ratio by the saturation mixing ratio of the background field. Pseudo-relative humidity is a good predictor of relative humidity, depending on the accuracy of the background temperature estimates. Its statistical properties are therefore similar to those of relative humidity. However, in the absence of moisture observations a univariate pseudo-relative humidity analysis preserves the background specific humidity field, and this prevents stability problems with the stratospheric moisture analysis. A pseudo-relative humidity analysis is equivalent to a mixing ratio analysis with flow-dependent error covariance specifications. It is easily implemented in an existing analysis system by scaling the observed-minus-background residuals prior to solving the analysis equation, and then converting the pseudo-relative humidity analysis increments back to the original humidity variable. Early results with the fvDAS, obtained with extremely simple error covariance models, indicate that this change of variable can lead to a better fit of the background humidity estimates to radiosonde observations.

The nature of the work reported here is primarily practical, in recognition of the fact that actual errors in operational data assimilation systems are not well represented by the covariance specifications. It is important to consider the robustness of algorithmic design decisions to this practical reality. In particular, the control variable should be such that time- and space averaging of the errors gives meaningful results, so that approximate covariance models based on statis-

tical averages do present some useful information about local errors. An optimal moisture analysis scheme would require multivariate moisture-temperature error covariance specifications, but we have not pursued this beyond the blunt statement that it appears to be more justifiable to neglect specific humidity-temperature correlations (in a univariate pseudo-relative humidity analysis) than to neglect relative humidity-temperature correlations (in a univariate relative humidity analysis).

With this as a starting point, however, we feel it may be feasible to improve the description of humidity errors by modeling the three main dynamic effects on the background error covariances in the assimilation cycle: (1) advection of initial errors, (2) error growth due to model defects, and (3) error reduction due to the incorporation of observations. This would represent a simplification of the Kalman filter equations following ideas outlined in earlier work (Dee 1990, 1991). We have begun to implement simple representations of each of these effects in the fvDAS, in order to cycle the moisture background error covariance specifications. We plan to present the results of initial assimilation experiments with this scheme in a separate article.

References

Atkins, M. J., 1974: The objective analysis of relative humidity. *Tellus*, **26**, 663–671.

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.

Courtier, P., 1997: Variational methods. *J. Meteor. Soc. Japan*, **75**, 211–218.

Daley, R., and E. Barker, 2000: NAVDAS Source Book 2000. *NRL Publication NRL/PU/7530--00-418*, Naval Research Laboratory, Monterey, CA.

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., and A. M. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.*, **124**, 1822–1834.

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

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.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162.

Franke, R., and E. Barker, 2000: Vertical correlation functions for temperature and relative humidity errors. *Mon. Wea. Rev.*, **128**, 3962–3981.

Garand, L., C. Grassotti, J. Hallé, and G. L. Klein, 1992: On differences in radiosonde humidity reporting practices and their implications for numerical weather prediction and remote sensing. *Bull. Amer. Meteor. Soc.*, **73**, 1417–1423.

Joiner, J., and L. Rokke, 2000: Variational cloud-clearing with TOVS data. *Q. J. R. Meteorol. Soc.*, **126**, 1–24.

Lin, S.-J., and R. B. Rood, 1996: Multidimensional flux-form semi-Lagrangian transport schemes. *Mon. Wea. Rev.*, **124**, 2046–2070.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194.

Lorenc, A. C., 1988: Optimal nonlinear objective analysis. *Q. J. R. Meteorol. Soc.*, **114**, 205–240.

Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders, 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **126**, 2991–3012.

Lorenc, A. C., D. Barker, R. S. Bell, B. Macpherson, and A. J. Maycock, 1996: On the use of radiosonde humidity observations in mid-latitude NWP. *Meteorol. Atmos. Phys.*, **60**, 3–17.

van Maanen, J., 1981: Objective analysis of humidity by the optimum interpolation method. *Tellus*, **33**, 113–122.

Ott, W. R.,1995: *Environmental Statistics and Data Analysis.* Lewis Publlishers, Boca Raton, Florida, 313pp.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1747–1763.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate.* American Institute of Physics, New York, 520pp.

Peixoto, J. P., and A. H. Oort, 1996: The climatology of relative humidity in the atmosphere. *J. Climate*, **9**, 3443–3463.

Poli, P., J. Joiner, and E. R. Kursinski, 2002: 1DVAR analysis of temperature and humidity using GPS radio occultation data. Submitted to *J. Geophys. Res.*.

Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Undén, J. Eyre, A. Hollingsworth, and F. Bouttier, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Q. J. R. Meteorol. Soc.*, **124**, 1809–1829.

Riishøjgaard, L.-P., 1998: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus*, **50A**, 42-57.

Wallace, J. M., and P. V. Hobbs, 1977: *Atmospheric Science.* Academic Press, New York, 467pp.

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

List of Figures
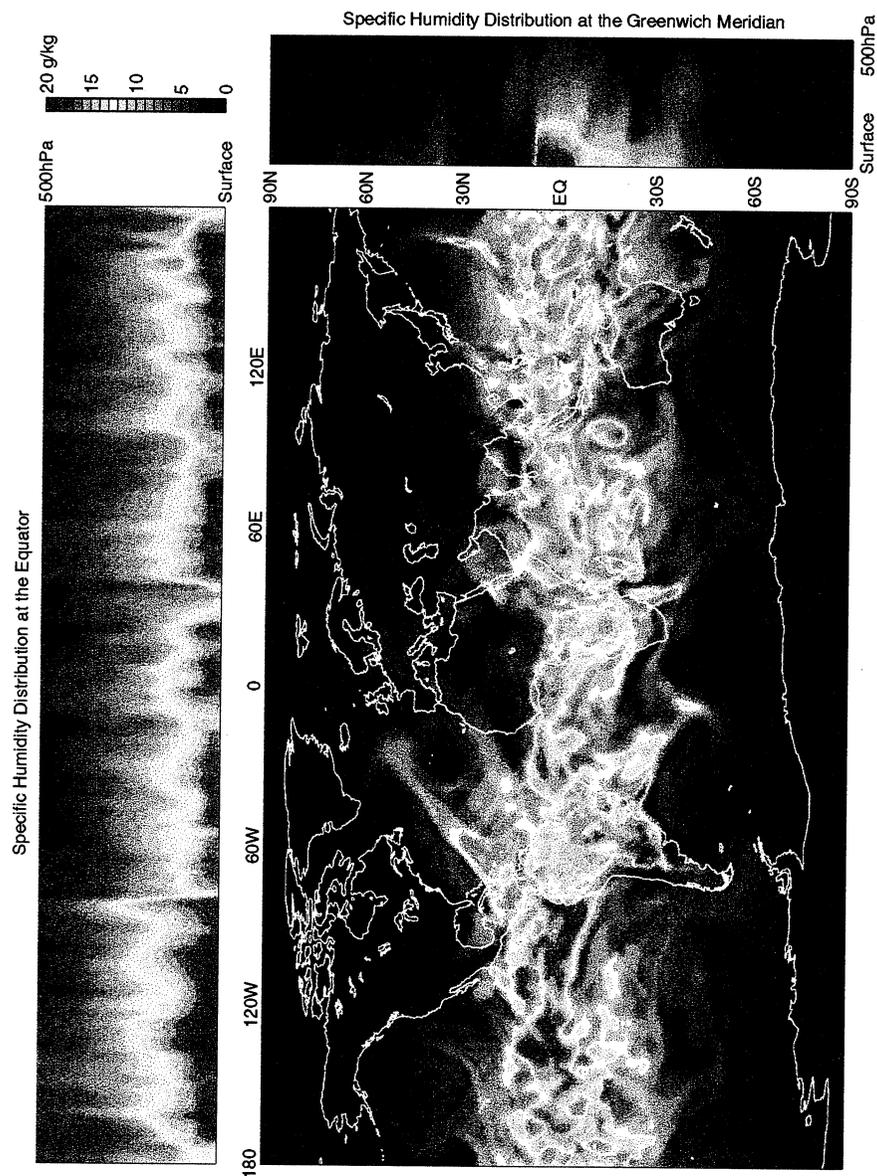
Figure 1: Specific humidity distribution at 0 UTC, January 1, 1998, produced by the fvDAS on a 55-layer $1° \times 1\frac{1}{4}°$ latitude-longitude grid. The larger of the three panels shows the layer-mean specific humidity for the fourth model layer, where most of the water vapor tends to be concentrated (at approximately 850 hPa over the oceans). The top panel shows the vertical distribution along the Equator in the lowest eight model layers, from the surface up to about 500 hPa. The right panel shows the vertical distribution along the Greenwich Meridian.

24
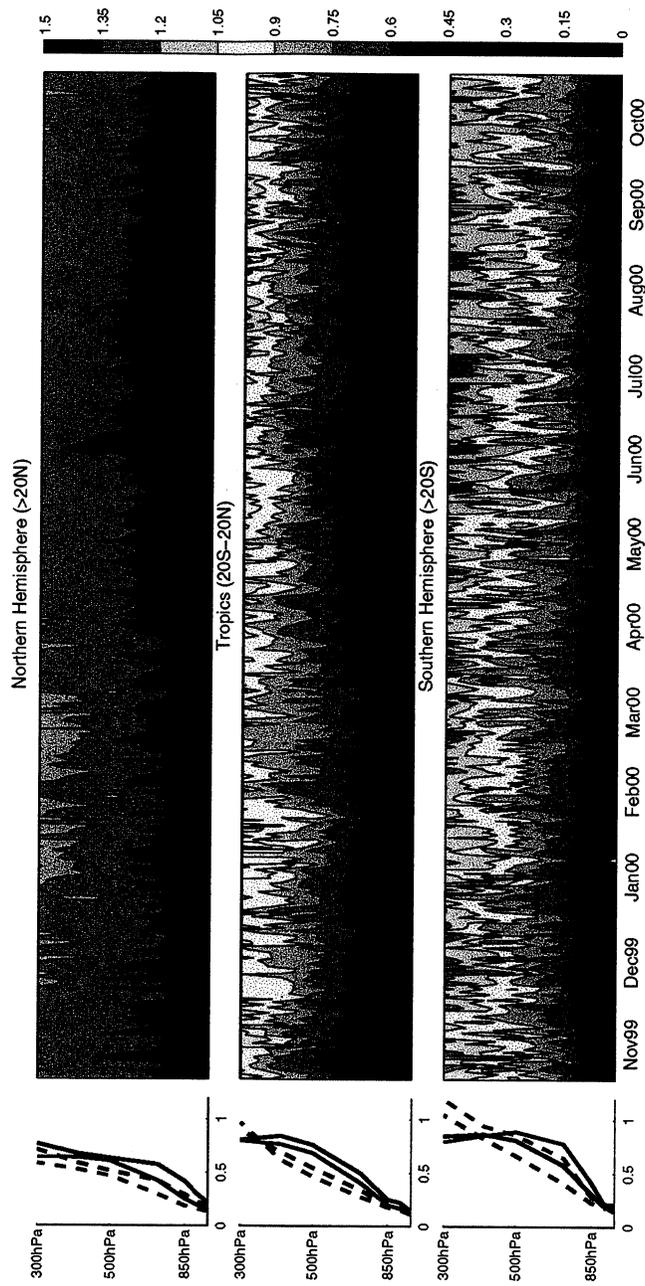
Figure 3: As Fig. 2, but for log q residuals. Units are non-dimensional.
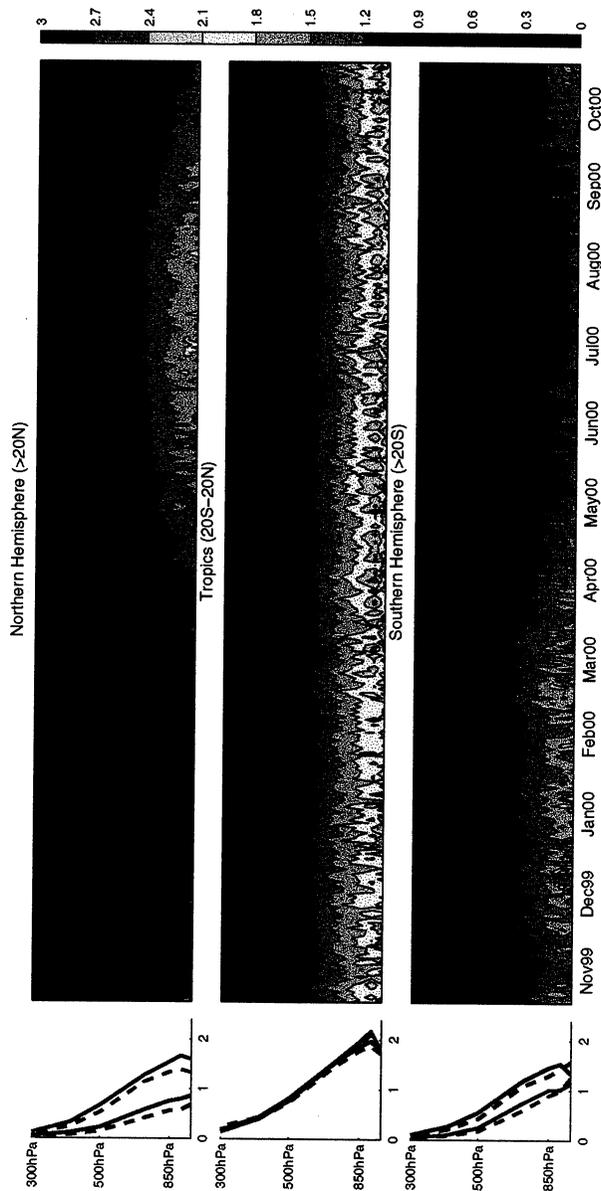
Figure 2: Daily rms statistics of radiosonde observed-minus-background mixing ratio residuals during the period 1 November 1999 – 31 October 2000, produced by the fvDAS. The three shaded contour plots show the rms of the residuals as a function of time and pressure for the Northern Hemisphere, Tropics, and Southern Hemisphere. The color scale is indicated on the right; units are $g\,kg^{-1}$. The panels to the left of each contour plot show the rms of all residuals for the month January 2000 (blue solid curve) and July 2000 (red solid curve), and the corresponding 68th percentiles (dashed curves) of the residual magnitudes.
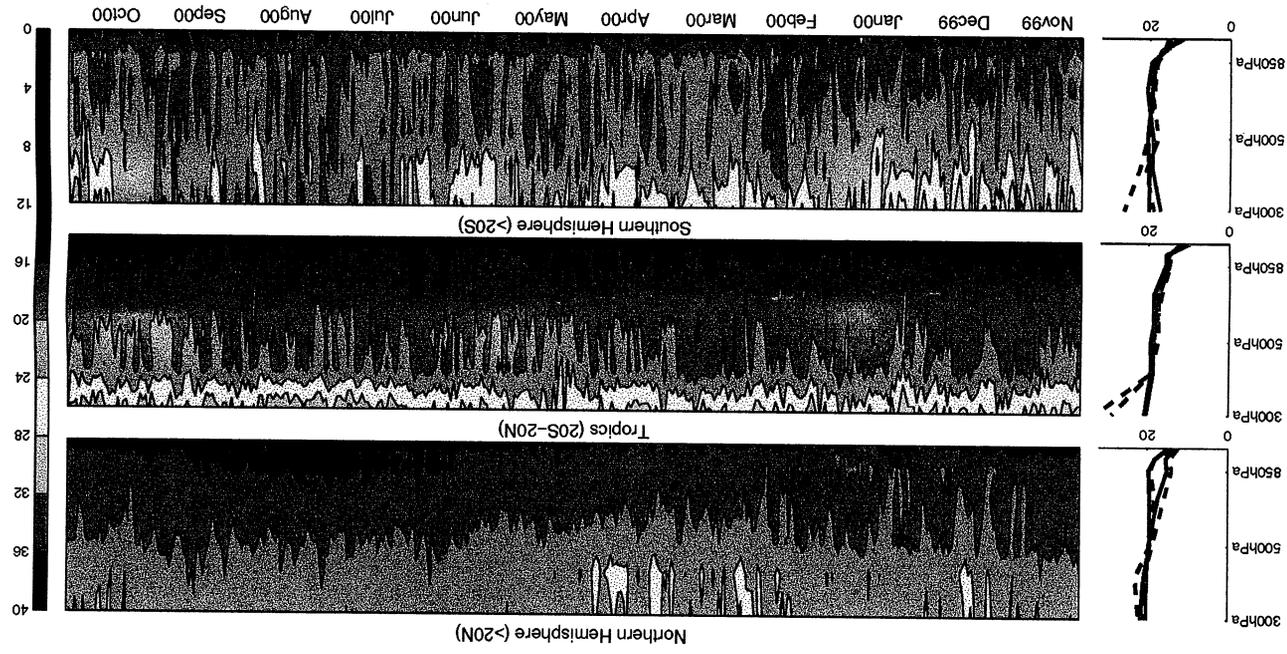
Figure 4: As Fig. 2, but for *rh* residuals. Units are in percents.

27

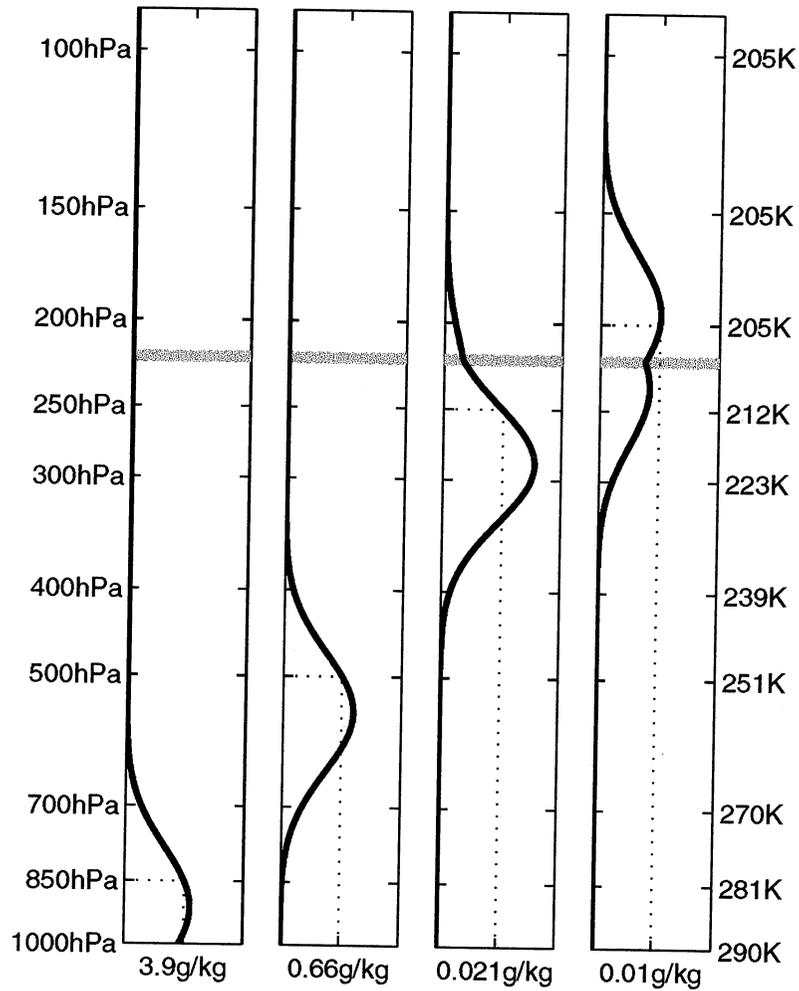Figure 5: As Fig. 2, but for $\tilde{w}$ residuals. Units are non-dimensional.
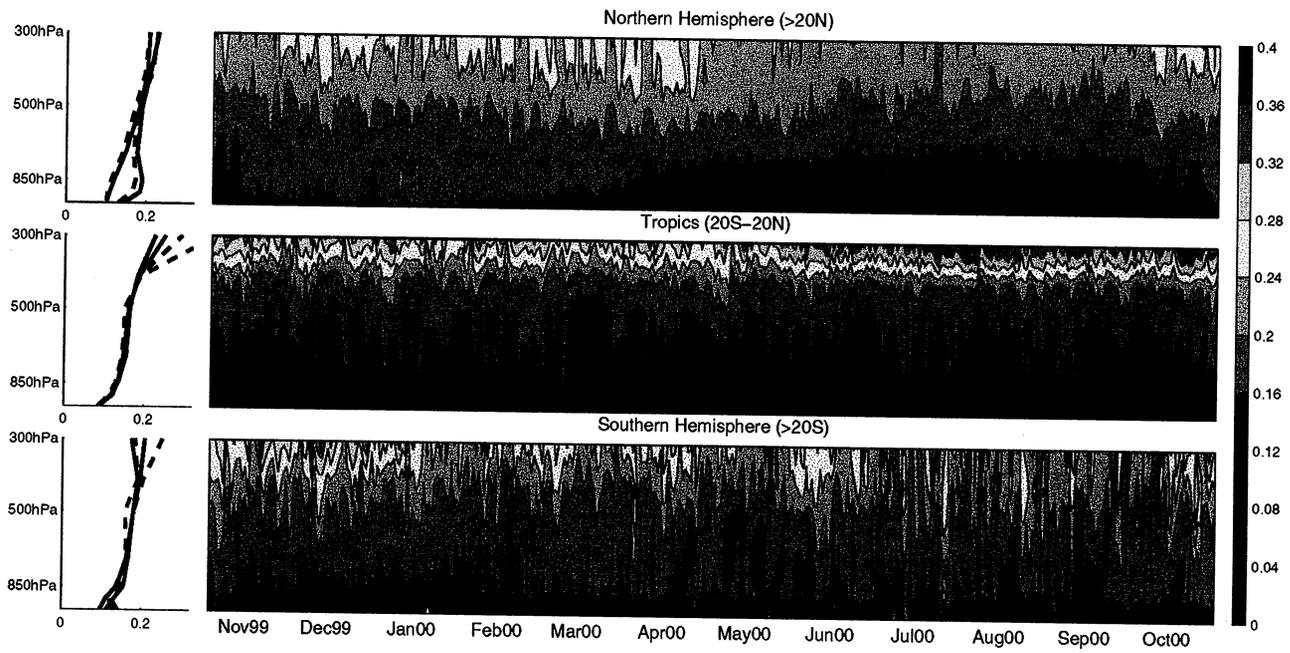
Figure 6: Mixing ratio analysis increments in a vertical column due to a single observation at 850, 500, 250, and 200 hPa, respectively, assuming vertically homogeneous and isotropic (in $\log p$) pseudo-relative humidity error covariances. Saturation mixing ratios are computed based on a temperature profile with a lapse rate of $8\,\mathrm{K\,km^{-1}}$ below 220 hPa, and constant temperature above. Temperature values are indicated along the vertical axis of the rightmost panel. The dotted lines in each panel mark the location and the value of the mixing ratio observation residual.
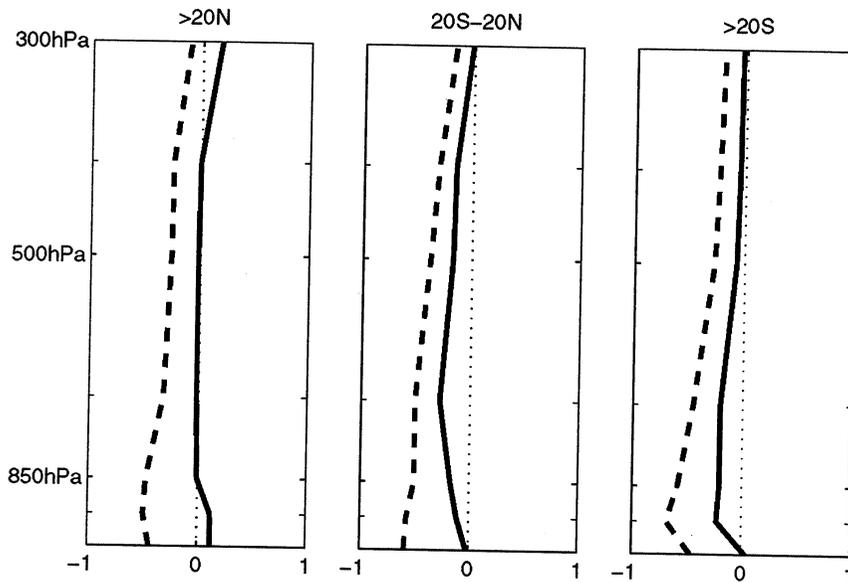
Figure 7: Average timeseries correlations for mixing ratio and temperature differences (solid) and for relative humidity and temperature differences (dashed), computed from December 1999 quality-controlled radiosonde observed-minus-forecast residuals. The correlations were computed for each station at each mandatory level from the residual timeseries, and then averaged over all stations in the Northern Hemisphere, Tropics, and Southern Hemisphere.

pseudo-rh + T analyses: implied rh increments — Contour interval is 3%

T analysis: T obs residuals + increments — Contour interval is 0.2 K

pseudo-rh analysis: w obs residuals + increments — Contour interval is 0.05 g/kg

Figure 8: Top panel: Mixing ratio observed-minus-background residuals (solid disks) and analysis increments (contours) obtained with a univariate pseudo-relative humidity analysis, for the eastern United States on 1 January 2002 at 0 UTC. Center panel: Temperature residuals and analysis increments. Bottom panel: Implied relative-humidity analysis increments. The contour interval for each panel is indicated along the left axis; blue/red shades represent negative/positive values, and the shading of the disks is consistent with that of the contours.

rh analysis: rh obs residuals + increments

Contour interval is 3%

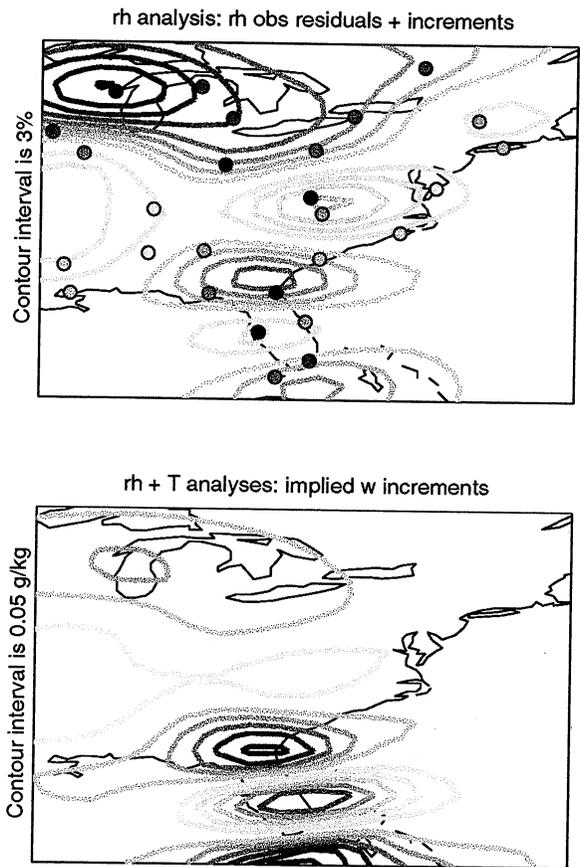rh + T analyses: implied w increments

Contour interval is 0.05 g/kg
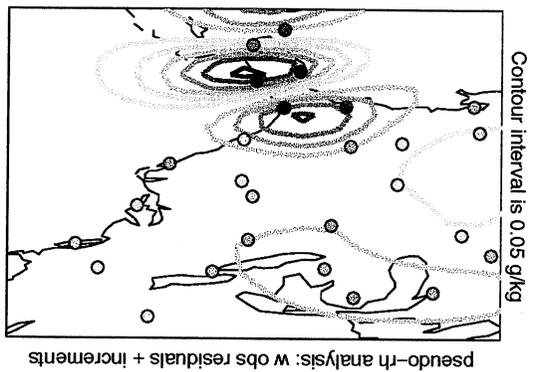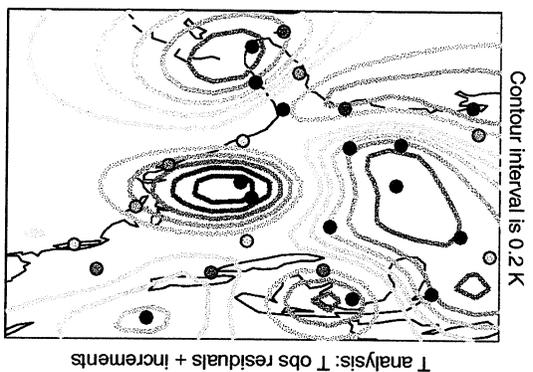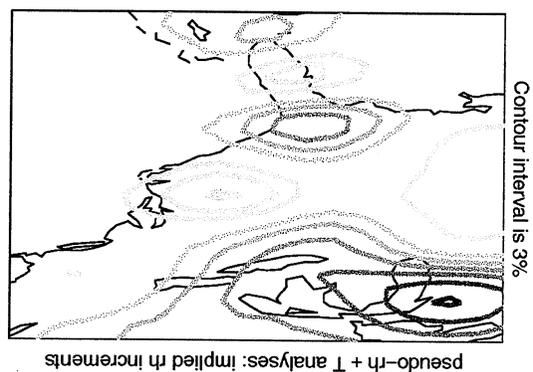
Figure 9: Top panel: Relative humidity observed-minus-background residuals (solid disks) and analysis increments (contours) obtained with a univariate relative humidity analysis, for the same region and time as Fig. 8. Bottom panel: Implied mixing ratio analysis increments. Color shading is identical with Fig. 8.

difference in rh increments

Contour interval is 3%

difference in w increments
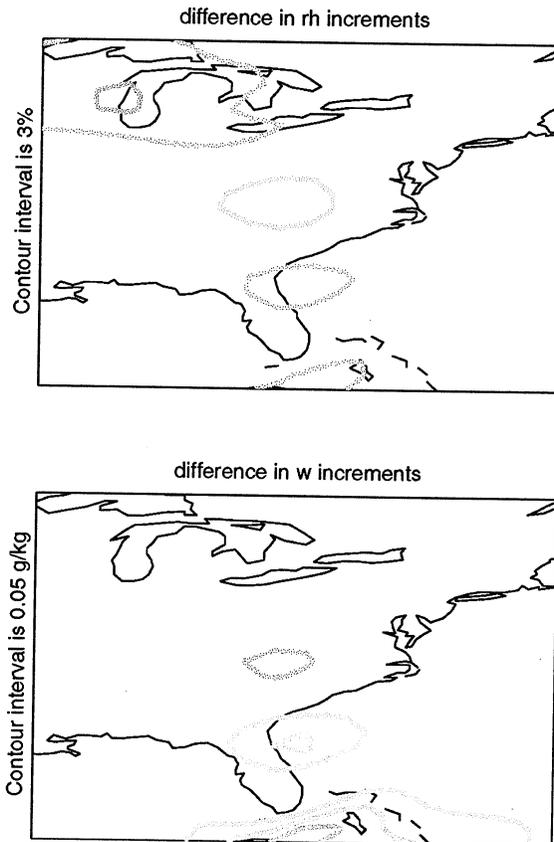
Contour interval is 0.05 g/kg

Figure 10: Top panel: Difference between the relative-humidity increments shown in the top panel of Fig. 8 and the implied relative-humidity analysis increments shown in the bottom panel of Fig. 9. Bottom panel: Difference between the implied mixing ratio analysis increments shown in the bottom panel of Fig. 8 and the mixing ratio analysis increments shown in the top panel of Fig. 9. Color shading is identical with Fig. 8.
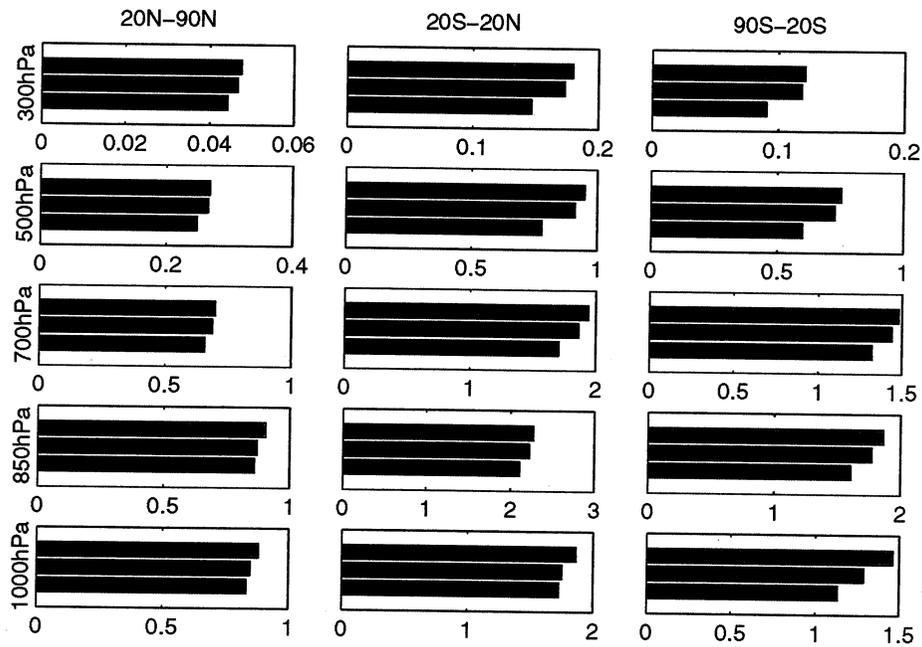
Figure 11: Root-mean-square mixing ratio background errors for the three experiments discussed in the text. Errors are calculated with respect to radiosonde mixing ratio observations in the Northern Hemisphere, Tropics, and Southern Hemisphere, at 1000, 850, 700, 500, and 300 hPa. The units along the horizontal axes are $g\,kg^{-1}$.