3.5

TESTING OF TECHNIQUES FOR ERROR ANALYSIS IN ASSIMILATION OF CLOUDY RADIANCES

Rosanne Polkinghorne*, Tomislava Vukicevic, K. Franklin Evans
University of Colorado, Boulder, Colorado

1. INTRODUCTION

Recent research results by Greenwald et al, (2004) and Vukicevic et al, (2005, 2006) show that it is possible to obtain skilled three dimensional cloudy atmosphere analysis by four-dimensional variational (4DVAR) assimilation of infrared (IR) satellite radiances into a cloud resolving model. In these assimilation studies the analysis and model simulation errors are evaluated against the observations. The observations used were the IR geostationary satellite radiances and ground based cloud radar reflectivity. The errors associated with the individual cloud state parameters were not estimated because direct observations of these quantities are not available, and the standard application of the variational data assimilation technique only produces the analysis but not the error estimates. In order to improve the error analysis to include an explicit estimate of errors in the quantities that are analyzed, it is necessary to perform additional computations which account for error sources in the inputs to the variational algorithm and produce estimates of the associated errors in the output (i.e. the analysis state quantities).

In this study, the focus is first on testing several approaches for the error estimates using a 1DVAR cloudy radiance assimilation algorithm. This algorithm is based on cloudy radiance observational models by Evans (2006) which are a revised and improved version of the equivalent models used in the 4DVAR retrievals in Vukicevic et al (2004, 2005). The error estimation techniques tested are: 1) background perturbations by Bauer et al (2005) where the perturbations are derived from a known background error covariance matrix, 2) perturbations added to the observation vector to produce an ensemble of retrievals from which the deviation is evaluated and 3) linear error analysis at the retrieved state. These techniques were tested in assimilation of cloudy radiances with simulated GOES (Geostationary Operational Environmental Satellites) observations.

2. MODEL AND ALGORITHM

Simulated radiances of GOES channels one, two, and four (0.63 μm, 3.92 μm, and 10.7 μm) are first produced by the spherical harmonic discrete ordinate plane parallel data assimilation (SHDOMPPDA) (Evans, 2006) forward model. This model uses profiles of pressure, temperature, water vapor, mass mixing ratio, and number concentration for a number of hydrometeor species to calculate the upwelling radiances (in reflectance units or brightness temperature) for each GOES channel.

For these tests, columns from an X-Z slice of a 3D cloud model simulation are used as the input to the radiative transfer model. The 3D cloud model used to generate the cloud fields is the Regional Atmospheric Modeling System (RAMS) using the two moment liquid cloud scheme (Saleeby and Cotton, 2004). The simulation is centered on the Atmospheric Radiation Measurement program central facility in Oklahoma on March 16, 2001. The clouds used here are from the end of a 6 hour run.

The parameters retrieved in the RAMS simulation are taken to be truth in these experiments. These parameters (pressure, temperature, water vapor, mass mixing ratio, and number concentration) are initially used in the forward radiative transfer model to calculate the true liquid water path and the upwelling radiances for each GOES channel.

The control vector is log cloud mixing ratio and log cloud concentration: **x**=ln(**q**) (**q** represents the vector containing both cloud mixing ratio and number concentration). The first 50 of the 84 RAMS levels are included in this vector, so the parameters are represented from the surface up to about 5 km. The pressure and temperature profiles are fixed to their true values for the specified column during the integration of the model.

The background mean x_b and variance σ_b are calculated using the 10,000 RAMS columns in the RAMS slice used in these simulations. The distribution is assumed to be multivariate Gaussian and the mean (M) and variance (V) of the cloud mixing ratio and number concentration when averaged over these columns is used in the calculation of the lognormal parameters:

$$\sigma_b=\sqrt{\log\left(1+\frac{V}{M^2}\right)} \quad ,$$

$$x_b=\log\left[M\exp(-\sigma_b^2/2)\right] \qquad (1)$$

The off-diagonal elements of the background covariance matrix are calculated assuming an exponential decrease in correlation with distance

**Corresponding author address:* Rosanne Polkinghorne, Department of Oceanic and Atmospheric Sciences, UCB 311, University of Colorado, Boulder, CO 80309-0311
email: rosanne.wilcox@colorado.edu

between two levels:

$$B_{ij}=\sigma_{b,i}\sigma_{b,j}\exp(-|z_i-z_j|/L_{corr}), \qquad (2)$$

where z_i is the height of the ith level and L_{corr} is a specified correlation length. The background covariance matrix B, like the control vector, contains values from both the cloud mixing ratio and number concentration. The sections of the matrix where there is cross-correlation between these two parameters are further multiplied by a constant cross-correlation $C_{r,N}$. Analysis of the vertical correlation function of log mixing ratio and log concentration for cloudy levels from the 10,000 RAMS columns led to the choice of L_{corr} = 0.25 km and $C_{r,N}$ = 0.92 (Evans, 2006).

The cost function to be minimized in this 1DVAR retrieval is:

$$J=\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x_b})^T B^{-1}(\boldsymbol{x}-\boldsymbol{x_b})+\frac{1}{2}[\boldsymbol{y_0}-H(\boldsymbol{q})]^T R^{-1}[\boldsymbol{y_0}-H(\boldsymbol{q})] \qquad (3)$$

where $\mathbf{y_o}$ represents the simulated observations (the truth in these experiments) and the forward model is H(**q**). The vector **q** is used with the forward model rather than **x** because the control vector must be exponentiated to obtain the cloud mixing ratio and number concentration in real space that is input into this model (Evans, 2006). The observation covariance matrix R is diagonal, with channel uncertainties σ_j of 0.02 for 0.63 µm reflectance and 2.0 K for 3.92 µm and 10.73 µm brightness temperatures. The gradient of the cost function is

$$\nabla J=B^{-1}(\boldsymbol{x}-\boldsymbol{x_b})-\boldsymbol{q}\,H(\boldsymbol{q})^T R^{-1}[\boldsymbol{y_0}-H(\boldsymbol{q})], \qquad (4)$$

where $H(\mathbf{q})^T$ represents the adjoint to the forward model.

The *a priori* used in these simulations, unless stated otherwise, is the background mean vector. The cost function is minimized with a conjugate-gradient algorithm. A line minimization is done at each conjugate-gradient iteration by bracketing the minimum and then performing a golden section search using only the forward model.

3. DESCRIPTION OF EXPERIMENTS

Four different error analysis techniques are used with this 1DVAR model.

3.1 *Experiment One*

Random Gaussian noise is added to the simulated radiances, and then a retrieval is performed. This is done sixty times each for several columns. The retrieval error is defined as

$$e^a=x^t-x^a \qquad (5)$$

where e^a is the retrieval error, x^t is the "true" atmosphere and x^a is the retrieved atmosphere. Error statistics such as the mean and median retrieval error, standard deviation of the error, and root mean square are computed for each column.

3.2 *Experiment Two*

The method used in this experiment uses the simulated radiances of neighboring columns, adding errors to the atmospheric profiles rather than to the radiance vector directly as is done in the previous experiment. A random number between -10 and 10 is generated at the start of each retrieval to determine from which neighboring column the simulated radiances will be calculated. The retrievals are again repeated 60 times for each column and the error statistics computed as in experiment one.

3.3 *Experiment Three*

The theoretical error is calculated from the results of experiment one. This is done by finding estimates of

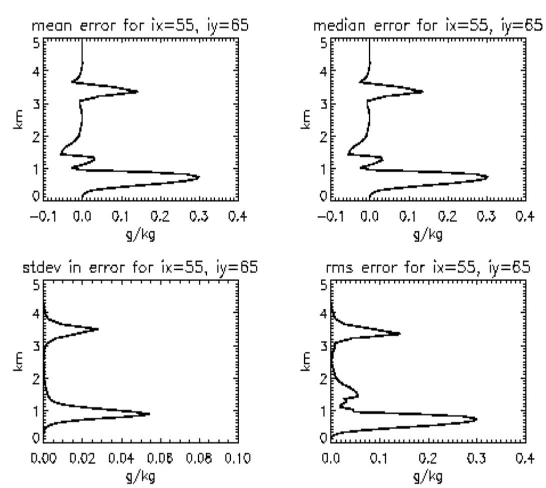

Figure 1: The error statistics for experiment one at the column ix=55, iy=65

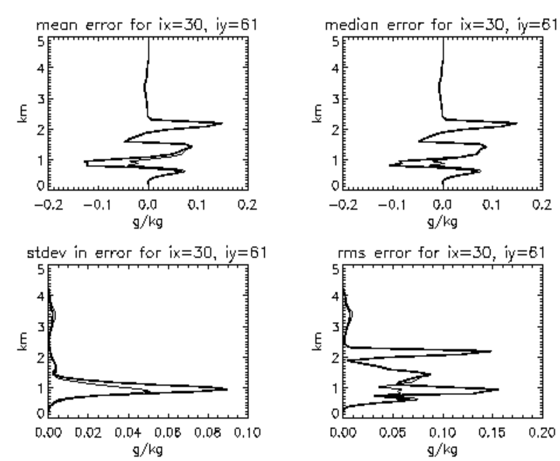

Figure 2: Error statistics for experiment one (thin solid line) and experiment two (heavy solid line) at the column ix=30, iy=61

the retrieval error covariance matrix (P^a) as defined by least square retrieval theory using the equation:

$$P^a = (H^T R^{-1} H + B^{-1})^{-1} \qquad (6)$$

where H is the observation model, R is the observation error covariance matrix, and B is the background error covariance matrix. This equation is derived assuming that the model is linear. The diagonal elements of the resulting matrix represent the theoretical variance of the column. This calculation is done independently for each of the 60 retrievals performed for each column, and then an average over the retrievals is computed.

3.4 *Experiment Four*

The ensemble approach is used to add error to the first guess state vector as in Bauer et al (2005). The eigenvalues are multiplied by their corresponding eigenvectors as well as a random Gaussian number. A sum is taken over all the eigenvalues and the resulting vector is added to the mean background state. This is then converted into real space and taken as the initial guess for the cloud mixing ratio and number concentration rather than having the mean background as the initial guess. As in experiments one and two, the retrievals were repeated 60 times for each column and the error statistics computed as in those experiments.

4. ANALYSIS OF RESULTS

4.1 *Experiment One*

The error statistics for the retrieval of one column are shown in figure 1. These statistics represent an average, median, etc over 60 retrievals. To test if 60 retrievals is sufficient to get a good representation of the statistics, these same statistics are also computed over 200 retrievals. The results of the two tests are nearly identical, indicating that the calculation of error statistics with 60 retrievals is a fairly accurate representation.

4.2 *Experiment Two*

As can be seen in figure 2, the method used in experiment two causes a greater error on average than the method used in experiment one. In particular, the standard deviation of the errors is greater in experiment two than in experiment one, as expected.

4.3 *Experiment Three*

When the theoretical error as calculated in experiment three is compared to the root mean square error as calculated in experiment one (see figure 3), it is seen that the root mean square of the error is significantly more than the theoretical standard deviation. This is in part due to the assumption of linearity made in this experiment. In addition, the background error covariance is poorly known, contributing to this discrepancy. The theoretical variances are also examined with no measurements. The variances in this case do not change much from the case with measurements, indicating that the mixing ratio and number concentration profiles depend mostly on the background rather than the simulated GOES channels.

4.4 *Experiment Four*

The results of this experiment and those of experiment one match fairly closely, as can be seen in figure 4. However, the standard deviation of the error in this retrieval is not negligible. The perturbations added to the initial guess have the shape of the background error covariance matrix, which reflects model errors in

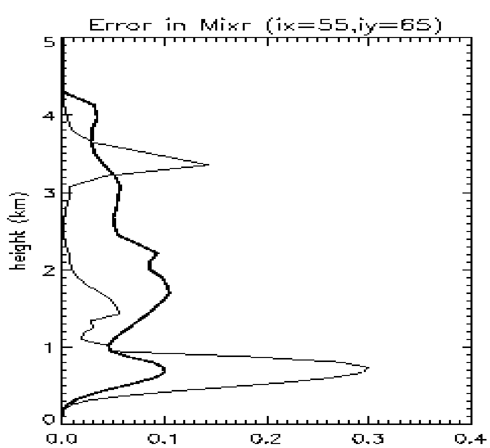

Figure 3: The comparison of the theoretical error (heavy solid line) with the root mean square error found in experiment one (thin solid line)

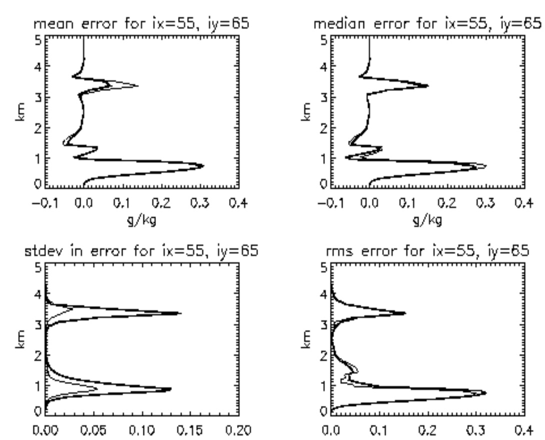

Figure 4: The comparison of the error statistics in experiments four (heavy solid line) and one (thin solid line)

RAMS since the background data were taken from the output of that model. This could bias the results towards the background and by so doing, contribute to the error observed in this experiment.

In theory, one needs to use only the set of eigenvalues and eigenvectors that represent about 90% of the total energy to save on computation time. In this case, it was found that the sum of the largest 30 eigenvalues (out of 100) represents about 90% of the total. When tested and compared to the results given when all of the eigenvectors are used to form the perturbation, the results are very similar, confirming that not all of the eigenvectors need to be used.

Figure 5 shows the eigenvectors that correspond to the ten largest eigenvalues of the background error covariance matrix. Since this model produces two moments, the cloud mixing ratio and number concentration, the pattern of each eigenvector repeats in the upper portion of the plot. Because the background decorrelation length is small, even the most dominant eigenvectors are still quite wiggly. When L_{corr} is set to be even smaller (0.05 km), the fine structure of the atmospheric state can be easily observed in the most dominant eigenvectors. On the other hand, when L_{corr} is large (2.5 km), the dominant eigenvectors become much smoother.

5. CONCLUSIONS AND FUTURE RESEARCH

The major conclusions from the preliminary 1DVAR experiments are:

- The errors in the background state cause larger errors in the retrieval than the observation errors, as expected.
- The results of experiment three show that the variance calculated by the linear theory is much smaller than the "true" variance caused by the observation errors only. This is due to the assumption of having a linear model, as well as using a poorly known background covariance matrix in the computation.
- The results of experiment four with the perturbations from the eigenvectors of the background covariance matrix also show that poor knowledge of the background error matrix could have a strong influence on the standard deviation of the resulting retrieval errors. In the case of a cloudy retrieval, these factors make the standard deviation of the retrieval error much larger than observed in experiment one, where random Gaussian perturbations are added to the simulated radiances.

In order to improve these calculations, a better model for the background error covariance should be developed. This could be achieved by using ensemble simulations with the cloud resolving model, or by a systematic comparison of the model forecast to available observations. The latter method will be applied first in future studies with both the 1DVAR and 4DVAR data assimilation experiments using the Atmospheric Radiation Measurements (ARM) project observations.

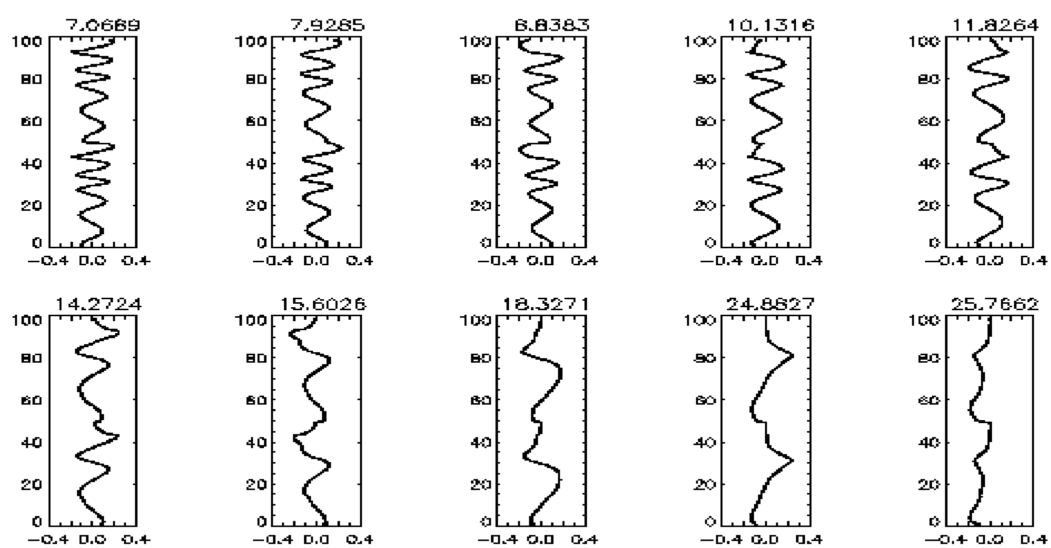

Figure 5: Ten eigenvectors corresponding to the 10 largest eigenvalues of the background covariance matrix. The eigenvalue appears on the top of the plot of its corresponding eigenvector.

REFERENCES

Bauer, P., E. Moreau, and S. Di Michele, 2005: Hydrometeor retrieval accuracy using microwave window and sounding channel observations. *J. of Applied Meteorology*, **44**, 1016-1032

Evans, K.F., 2006: SHDOMPPDA: a radiative transfer tool for cloudy sky data assimilation. *J. of the Atmospheric Sciences,* in press

Greenwald, T., T. Vukicevic**,** and L. Grasso, 2004: Adjoint Analysis of an Observational Operator for Visible and Infrared Cloudy-sky Radiance Assimilation*. Q. J. R. Met. Soc.,* **130**, 685-705

Saleeby, S.M., and W.R. Cotton, 2004: A large-droplet mode and prognostic number concentration of cloud droplets in the Colorado State University regional atmospheric modeling system (RAMS). Part I: Module descriptions and supercell test simulations. *J. of Applied Meteorology*, **43**, 182-195

Vukicevic**,** T., M. Sengupta, A. Jones and T. Vonder Haar, 2006: Cloud resolving data assimilation: information content of IR window observations and uncertainties in estimation. *J. Atmos. Sci.* **63,** 901-919

Vukicevic, T., T. Greenwald, M. Zupanski, D. Zupanski, T. Vonder Haar, and A. Jones, 2004: Mesoscale cloud state estimation from visible and infrared satellite radiances. *Mon. Wea. Rev.,* **132**, 3066-3077