A Nonlinear Physical Retrieval Algorithm—Its Application to the GOES-8/9 Sounder

XIA L. MA
Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin—Madison, Madison, Wisconsin

TIMOTHY J. SCHMIT
Advanced Satellite Products Team, Office of Research and Applications, NOAA/NESDIS, Madison, Wisconsin

WILLIAM L. SMITH*
Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, University of Wisconsin—Madison, Madison, Wisconsin

(Manuscript received 3 July 1997, in final form 9 July 1998)

ABSTRACT

A nonlinear physical retrieval algorithm is developed and applied to the GOES-8/9 sounder radiance observations. The algorithm utilizes Newtonian iteration in which the maximum probability solution for temperature and water vapor profiles is achieved through the inverse solution of the nonlinear radiative transfer equation. The nonlinear physical retrieval algorithm has been tested for one year. It has also been implemented operationally by the National Oceanic and Atmospheric Administration National Environmental Satellite, Data and Information Service during February 1997. Results show that the GOES retrievals of temperature and moisture obtained with the nonlinear algorithm more closely agree with collocated radiosondes than the National Centers for Environmental Prediction (NCEP) forecast temperature and moisture profile used as the initial profile for the solution. The root-mean-square error of the total water vapor from the solution first guess, which is the NCEP 12-h forecast (referred to as the “background”), is reduced approximately 20% over the conventional data-rich North American region with the largest changes being achieved in areas of sparse radiosonde data coverage.

1. Introduction

The Geostationary Operational Environmental Satellite (GOES)-8 three-axis stabilized geostationary satellite was successfully launched in April 1994. The GOES-9 satellite was launched in May 1995. Initially, a linear physical algorithm was used operationally at the National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellite, Data and Information Service (NESDIS) to retrieve the atmospheric temperature and water vapor profiles, as well as the surface skin temperature. Similar to the international Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) Processing Package 4 (ITPP4) (Smith and Woolf 1988) and the GOES Visible and Infrared Spin Scan Radiometer (VISSR) Atmospheric Sounder (VAS) sounding retrieval algorithm (Smith 1983), the method was a linearized, simultaneous solution for the perturbation for a first-guess temperature profile, moisture profile, and surface skin temperature estimate (Smith 1983; Hayden 1988). Improvements had been developed by the Advanced Satellite Products Team at the Cooperative Institute for Meteorological Satellite Studies (CIMSS) (Hayden 1994; Hayden and Schmit 1994, 1995, henceforth abbreviated as the linear version). In addition to the modifications needed to process the GOES channels, the major improvements to the algorithm included improving the cloud clearing, improved moisture basis functions, and an hourly binning of data for the radiance bias (i.e., “tuning”) adjustment.

To improve the accuracy of GOES retrievals, a nonlinear inversion algorithm, following Marquardt-Levenberg (Press et al. 1990), has been developed (hereafter referred to as the nonlinear version) and is described in this paper. The nonlinear physical retrieval algorithm is based on the formulation initiated by Rodgers (1976) and further developed by Purser (1984) and Eyre (1989a,b; Eyre et al. 1993); the algorithm employs a Newtonian iterative method that finds the maximum probability solution to the nonlinear inversion of the...
radiative transfer equation (i.e., the profile retrieval problem). In this paper, the characteristics of the GOES-8 sounder are briefly described. The nonlinear retrieval algorithm for processing GOES-8/9 data is described in detail. The nonlinear algorithm is applied in real time to GOES-8 and -9 observations. For GOES-8, statistical results are obtained to compare the accuracy of the new nonlinear method with the accuracy of the linear method initially used for profile retrieval. It is concluded that the nonlinear algorithm provides for closer agreement with radiosondes for both temperature and moisture profiles, which should provide better weather forecasts.

2. Sounder instrument

The GOES-8 and -9 sounders are filter wheel radiometers containing 18 thermal infrared channels plus a visible channel with 10-km linear resolution. The thermal infrared channels provide measurements of radiance from the earth’s surface, clouds, and atmospheric carbon dioxide (CO$_2$), moisture (H$_2$O), and ozone (O$_3$). Figure 1 shows the spectral positions of the GOES-8 channels and their filter functions. The spacecraft is designed to permit quasi-continuous viewing of the earth; thus, relatively large dwell times can be achieved to enhance signal-to-noise performance for atmospheric sounding. The instrument calibration is accomplished using one external blackbody and a space view reference (Weinreb et al. 1996; 1997). The GOES sounder channels were selected to permit atmospheric temperature and moisture profile retrieval with high spatial (10 km at nadir) and temporal (hourly) resolution. A more detailed description of the GOES sounder is given by Menzel and Purdom (1994).

3. Nonlinear retrieval algorithm

For a nonscattering atmosphere in local thermodynamic equilibrium, the radiative transfer equation may be expressed as

\[ R(v_j, \theta) = B(v_j, t)\tau(v_j, \theta, p) \]

\[ - \int_0^{\nu_j} B[v_j, t(p)] \frac{\partial \tau(v_j, \theta, p)}{\partial p} dp, \quad (1) \]

where \( R(v_j, \theta) \) is the mean spectral radiance measured in a channel whose mean effective wavenumber is \( v_j \) and the local zenith angle of observation is \( \theta \). \( B(v_j, T) \) is the Planck function of temperature, and \( \tau(v_j, \theta, p) \) is the transmittance from the pressure level \( p \) to the top of the atmosphere along the observation angle \( \theta \). The subscript \( s \) denotes surface values, either ground or cloud. In Eq. (1) it is assumed that the underlying surface radiates as a blackbody does. As shown in appendix A, which follows the derivation of Smith et al. (1991), Eq. (1) may be approximated in the numerical perturbation form of

\[ \delta t_s(j) = \delta t, K^s(j) - \sum_{i=1}^{3} \delta t(i) K'(i, j) + \sum_{i=1}^{3} \delta q(i) K^{s}(i, j). \]

(2)

where the perturbation \( \delta \) is with respect to an a priori
estimated or mean condition; \( t_b \) is a GOES channel brightness temperature vector; and \( K' \), \( K'' \), and \( K''' \) are the weighting functions of surface skin temperature, atmospheric temperature, and water vapor, respectively. The variable \( i \) is the quadrature pressure level, \( j \) denotes channel number, and \( ls \) is a quadrature level of the surface (or cloud) pressure. In its matrix form, Eq. (2) can be given as

\[
\delta y = K \delta x, \tag{3}
\]

where \( \delta y = \delta t_b \); the weighting function matrix \( K \) contains \( K' \), \( K'' \), and \( K''' \) and

\[
K'(i, j) = \frac{\partial [B(v_x, t(i))/\partial t_x]}{\partial R(v_x, \theta)/\partial t_b} (t_a - t_s) \tau_y(j), \tag{4}
\]

with

\[
\alpha(i, j) = -\frac{d \ln[\tau(i, j)]}{q(i) dp}. \tag{5}
\]

For surface skin temperature,

\[
K''(ls, j) = \frac{\partial [B(v_x, t(ls))/\partial t_x]}{\partial R(v_x, \theta)/\partial t_b} \tau_y(j), \tag{6}
\]

where \( t_a \) and \( t_s \) are surface air and skin temperature, respectively, and where \( \tau \) and \( \tau_y \) represent transmittance of the atmospheric column above level \( i \), and above surface \( s \), respectively, for all gases.

A “cost” function (Eyre 1989) is defined as

\[
J(x) = (x - x_0)^T C^{-1} (x - x_0) + [y_n - y(x)]^T E^{-1} [y_n - y(x)], \tag{7}
\]

where \( x_0 \) is the background (i.e., forecast) profile used as the first guess, \( C \) is the expected background error covariance matrix, \( y_n \) is the vector of radiance measurements, and \( y(x) \) is the vector of radiances calculated through the radiative transfer equation corresponding to the solution profile \( x \). The expected radiance error covariance \( E \) includes the random error of the measurement and the error produced by the forward model. Superscripts T and \( (−1) \) represent matrix transpose and inverse, respectively.

The nonlinear algorithm incorporates both the inverse Hessian method of solution,

\[
x_{n+1} = x_n - [\nabla^2 J(x_n)]^{-1} \nabla J(x_n), \tag{8}
\]

and the steepest descent method of solution,

\[
x_{n+1} = x_n - \gamma^{-1} \nabla J(x_n), \tag{9}
\]

to control the convergence process (Press et al. 1990; Rodgers 1996, personal communication). The method was originally developed by Marquardt (1963), following earlier work by Levenberg (1944). The combined solution, obtained by a weighted average of Eqs. (9) and (10), is

\[
x_{n+1} = x_n - [\nabla^2 J(x_n) + \gamma I]^{-1} \nabla J(x_n), \tag{10}
\]

where \( \nabla \) represents the gradient. The variable \( \gamma \), which absorbs the weights of Eqs. (9) and (10) incorporated into Eq. (11), is an experimental parameter to control convergence, and \( I \) is an identity matrix. If \( \gamma \to 0 \), Eq. (11) tends to the inverse Hessian method; if \( \gamma \to \infty \), Eq. (11) tends to the steepest descent method, with a smaller step size. The cost function \( J(x_n) \) is defined in Eq. (8) and \( x_n \) and \( x_{n+1} \) are the previous and current solutions, respectively. The most probable solution of Eq. (3) is obtained by setting the gradient of \( J(x) \) to zero, namely, minimizing the cost function with respect to the solution \( x \) (appendix B).

Matrix manipulation of Eq. (11) yields the following iterative solution,

\[
x_{n+1} = x_n + (C^{-1} + K''E^{-1} K_s + \gamma I)^{-1}
\times \{K''^T E^{-1} [\delta y_n + K_s (x_n - x_0)] + \gamma_n (x_n - x_0]\}, \tag{12}
\]

where \( C^{-1} \) is the inverse of a statistical representation of the covariance matrix representing the expected errors in the initial estimate \( x_0 \), \( \delta y_n \) is the difference of the observed and calculated brightness temperatures, and \( \gamma_n \) is allowed to vary with iteration as defined below.

It is computationally efficient to present the profile vector as a small series of eigenvectors of the temperature and water vapor profile (Smith and Woolf 1976). This efficiency reduces the number of unknown expansion coefficients to the same order as the number of measured radiances:

\[
\delta x = \sum_{j=1}^{M} f_j v_j = Vf, \tag{13}
\]

where \( v_i \) is the \( i \)th eigenvector, \( f_i \) is the \( i \)th expansion.
coefficient, and \( M \) denotes the number of terms. The terms \( \mathbf{V} \) and \( \mathbf{f} \) represent the eigenvector matrix and coefficient vector, respectively. The eigenvectors are derived from a statistical covariance matrix of a large sample of radiosonde temperature and moisture profiles. The inverse of the profile error covariance matrix, \( \mathbf{C}^{-1} \), in the eigenvector domain is \( \mathbf{C}^{-1} = \mathbf{V}^{T} \mathbf{C}^{-1} \mathbf{V} \). The matrix form of Eq. (3) becomes

\[
\delta y = \mathbf{K} \mathbf{\delta x} = \mathbf{K} \mathbf{f} = \mathbf{K} \mathbf{r}.
\]  

(14)

The corresponding cost function to Eq. (8) can then be written as

\[
\bar{J}(\mathbf{f}) = \mathbf{f}^{T} \hat{\mathbf{C}}^{-1} \mathbf{f} + [y^m - y(\mathbf{f})]^{T} \mathbf{E}^{-1} [y^m - y(\mathbf{f})].
\]  

(15)

The solution of Eq. (14) is given by

\[
\mathbf{f}_{n+1} = (\hat{\mathbf{C}}^{-1} + \hat{\mathbf{K}}_{T} \mathbf{E}^{-1} \hat{\mathbf{K}}_{n} + \gamma_{n} I)^{-1} \times (\hat{\mathbf{K}}_{T} \mathbf{E}^{-1} \delta y_{n} + \hat{\mathbf{K}}_{n} \mathbf{f}_{n} + \gamma_{n} \mathbf{f}_{n}).
\]  

(16)

Equation (16) is computationally a more efficient solution than Eq. (12). Therefore, the retrieval problem is reduced to finding a set of coefficients that may be applied to Eq. (13) with eigenvectors to update the meteorological variables.

At each iterative step, convergence tests are carried out. The purpose of the convergence tests is retrieval quality control. If one of the following tests at a sounding point fails, then a retrieval is rejected. There are two convergence criteria.

1) The expansion coefficient convergence test defines the coefficient distance, \( d_{n+1} \), as the difference between eigenvector amplitudes determined from one iteration to the next. Mathematically,

\[
d_{n+1} = (\mathbf{f}_{n+1} - \mathbf{f}_{n})^{T} (\hat{\mathbf{C}}^{-1} + \hat{\mathbf{K}}_{T} \mathbf{E}^{-1} \hat{\mathbf{K}}_{n} + \gamma_{n} I)^{-1} \times (\hat{\mathbf{K}}_{T} \mathbf{E}^{-1} \delta y_{n} + \hat{\mathbf{K}}_{n} \mathbf{f}_{n} + \gamma_{n} \mathbf{f}_{n}).
\]  

(17)

This should approach zero as the solution converges (i.e., \( \mathbf{f}_{n+1} \to \mathbf{f}_{n} \)). In the iteration, the norm of the coefficient distance \( d_{n+1} \) is calculated at each step and compared with the previous one. If the norm of \( d_{n+1} \) is less than or equal to the norm of \( d_{n} \), the iteration continues until the difference between the norm of \( d_{n+1} \) and the norm of \( d_{n} \) is less than an experimental threshold (0.1). If the calculated norm of \( d_{n+1} \) is greater than the norm of \( d_{n} \), \( \gamma_{n} \) is increased. The coefficient distance \( d_{n+1} \) is recalculated with the solution profile \( x_{n} \), where the \( n \)th iterative step remains unchanged. This procedure for changing \( \gamma_{n} \) is repeated until the norm of \( d_{n+1} \) is less than or equal to the norm of \( d_{n} \), at which point the iterative solution continues toward convergence. If the norm of \( d_{n+1} \) less than or equal to the norm of \( d_{n} \) is not met by the third increase in \( \gamma_{n} \), the retrieval process is considered to be nonconvergent and no further attempt is made to produce a sounding at this location. Fortunately, this nonconverging condition rarely occurs.

2) The brightness temperature residual test is the other convergence criteria. If the expansion coefficient convergence test is passed, the solution profile \( x_{n+1} \) can be obtained in terms of new coefficients \( \mathbf{f}_{n+1} \) and the eigenvector matrix \( \mathbf{V} \) using Eq. (13). New brightness temperatures can then be obtained by radiative transfer calculations using the retrieved solution \( x_{n} \) as input.

The rms radiance residual is defined as

\[
r_{n+1}^{2} = \sum_{k=1}^{n_{ch}} [y_{k}^{m} - y_{k}(x_{n+1})]^{2} / n_{ch},
\]  

(18)

where \( k \) is a channel number; \( n_{ch} \) is the total number of channels used in the retrieval; and \( y_{k}^{m} \) and \( y_{k} \) are, respectively, the measured and calculated brightness temperatures for channel \( k \). A convergence test is performed on the brightness temperature measurement residual. If \( r_{n+1} \leq r_{n} \), the iteration continues until \( r_{n+1} \) is acceptably small (i.e., less than that expected due to instrument noise). If convergence is not achieved, the retrieval is rejected. The retrieved coefficients and brightness temperature residual at each step are saved. The final solution of Eq. (3), constructed from a set of coefficients that minimizes the brightness temperature residual to the instrument noise level, is considered to be an optimum solution in the sense of minimizing a cost function.

4. Dataset

a. Historic data

The total of 2495 radiosonde samples used in the study were collected over North America from February to April 1995. An eigenvector decomposition technique is applied to this dataset containing temperature and water vapor mixing ratio profiles. Two sets of empirical orthogonal functions (EOFs) were calculated: one for temperature and one for the natural logarithm of the water vapor mixing ratio. As described in section 3, each atmospheric profile can be expanded in terms of a few EOFs to capture the atmospheric vertical structure information contained in the GOES sounding radiance observations.

b. Retrieval first-guess and surface observation

Any physical retrieval algorithm, linear or nonlinear, benefits from an accurate background to obtain a satisfactory final solution. The National Centers for Environmental Prediction (NCEP) 6- to 18-h forecast using the Nested Grid Model (NGM; Hoke et al. 1989) is adopted as background in both the linear and nonlinear physical retrieval (background refers to the first guess of the iterative solution).

Following the ITTP methodology of Smith and Woolf (1985; Smith et al. 1988), ancillary surface information is used as a boundary condition for the profile solutions.
The surface air temperature and water vapor mixing ratio are treated as additional “channels” of information to assist the lower atmospheric structure determination; the surface temperature and natural logarithm of the surface mixing ratio observations are used by including the equations

$$\delta t(\text{ls}) = \sum_{i=1}^{n_t} f_i^t v^i(\text{ls}) \quad (19)$$

and

$$\delta \ln[q(\text{ls})] = \sum_{i=1}^{n_w} f_i^w v^i(\text{ls}) \quad (20)$$

in Eq. (14) where $\delta t(\text{ls})$ and $\delta \ln[q(\text{ls})]$ are surface perturbations between “truth” and their initial estimation at the surface level $\text{ls}$; and $n_t$ and $n_w$ denote the number of temperature and water vapor mixing ratio eigenvectors used, respectively. The expansion coefficients $f_i^t$ and $f_i^w$ correspond to their eigenvectors $v^i_t$ and $v^i_w$, respectively. (The time difference between the synoptic surface observations and the sounding time may be as long as 30 min.)

Since the eigenvector expansion algorithm provides the most economical representation of the temperature or water vapor mixing ratio, the number of eigenvectors required to extract all the significant information in the GOES radiances is much less than the number of GOES radiances. Statistical analysis of the GOES radiance information content reveals that five temperature and three water vapor eigenvectors explain all the variance in the GOES radiances occurring above the instrument noise level. Thus, the sounding retrieval problem has been reduced to one of solving for nine unknowns (five temperature eigenvector coefficients, three water vapor eigenvector coefficients, plus surface skin temperature perturbation) from the 18 GOES radiances observations plus two surface observations. The initial $\gamma_2$ value is set to 1 and 5 for temperature and moisture, respectively. Choosing these initial $\gamma_2$ values for the solution is based on trial and error.

c. Error correlation matrix and bias correction

The physical retrieval process is conducted every hour at CIMSS. At 0000 and 1200 UTC, the retrieval results are compared with radiosonde (raob) data within a 1.0° box in location and 1 h in time. A retrieval/raob matched data file is generated and accumulated daily. This dataset contains all the necessary information needed for retrieval validation. The error correlation matrix $C$ is derived from a historic matched dataset containing radiosondes and NCEP forecasts. A correlation matrix rather than a covariance matrix is used to keep the same scale of the relationship between temperature and the natural logarithm of the water vapor mixing ratio. Due to the lack of water vapor radiances contributions above 100 hPa, the water vapor mixing ratio retrieval is limited to the lowest 20 quadrature levels (115–1000 hPa). The guess surface skin temperature is obtained from the GOES channel 8 brightness temperature observation using a radiative transfer solution in which the atmospheric contribution is calculated using the NCEP forecast profiles. The error of the surface skin temperature needed for the $C$ matrix is assumed to be random with a standard deviation equal to 3.0 K.

The correlation matrix $C$ is given as 61 rows by 61 columns (40 level temperature, 20 level water vapor plus surface skin temperature), which provide the interlevel correlation of errors. The inverse of the error correlation matrix plays a key role in the physical iterative procedure. Since no unique solution to Eq. (2) exists, the matrix $C^{-1}$, along with $\gamma_2$, constrains the solution to ensure a stable as well as a statistically optimal retrieval result.

Physical retrievals from satellite-measured radiances require that the calculated radiances be corrected for forward radiative transfer model error and errors in instrument calibration. This so-called tuning accounts for discrepancies between observed radiances and radiances calculated from collocated model profiles. These discrepancies can be classified into three categories: 1) instrument measurement error; 2) forward model error that originates from a number of sources including spectroscopic uncertainties and the numerical methods used to calculate the atmospheric transmittance spectra and the spectral radiances; and 3) other processing errors, such as calibration, cloud detection and cloud clearing, and scene nonhomogeneity. A regression method called “shrinkage estimation” is used to derive error estimates for GOES-8 radiances tuning (Schmit 1996). The major effect of the radiances tuning is to remove bias differences between observed and calculated radiances as needed to ensure a successful retrieval.

5. Experimental results

The nonlinear retrieval algorithm described in section 3 is applied in real time to GOES-8 and -9 observations. The first-guess profiles for the retrieval are provided by a time–space interpolation of the NCEP forecasts. The first-guess skin temperature is a statistical regression estimate using the GOES infrared window measurements. The GOES clear-sky observations for a $5 \times 5$ (10 km) field-of-view (FOV) averaged sample are used to derive a box average clear brightness temperature vector for a single retrieval. During daylight hours a solar correction is applied to the GOES shortwave channel measurements in the 4.3–μ region (Smith et al. 1974). Temperature and moisture retrievals from the GOES measurements require a knowledge of cloud information. A clear FOV selection algorithm is performed over the $5 \times 5$ array of observations (Hayden et al. 1998, unpublished manuscript). The average number of clear FOVs is 16 out of a possible 25. As the result of this process, each individual FOV within the
A retrieval is performed if at least 1 of the 25 FOVs is determined to be cloud free. In the following discussion, results corresponding to linear retrieval algorithm are referred to as version 1.5. Version 1.6 is the nonlinear retrieval algorithm as described here. Figure 2 shows the total precipitable water retrieval root-mean-square error (rmse) comparisons at 0000 UTC in April 1996. (An rmse is a root-mean-square difference between the satellite retrieval and the radiosonde measurement.) This difference is due to errors in the radiosonde, time and space discrepancies of the two measurements, as well as errors in the satellite retrieval. The radiosonde errors are on the order of 0.5 K for temperature and 10% for humidity (Schmidlin 1988; Pratt 1985; Wade 1994). The distance between a retrieval sounding location and a raob location is within 0.5° (i.e., approximately 60 km). The broken line with dots is the first-guess total precipitable water, the broken line and the solid line represent retrieved total precipitable water from the linear and nonlinear physical versions (the same convention is employed in the following figures). In Fig. 2 note that the total precipitable water retrieval rmse of the nonlinear version is almost always smaller than the rmse associated with the linear version. It is also evident that the nonlinear version generally has a smaller rmse than the first guess; the reduction of the rmse for the linear version is less consistent. The nonlinear version has better humidity retrieval performance than the linear version. Figure 3 shows the same type of results shown in Fig. 2 but for 1200 UTC.

They show that the nonlinear version can maintain the temperature retrieval accuracy of the good forecast first guess at radiosonde locations. The linear version temperature retrieval disagrees slightly more with respect to the radiosondes than the first guess above 850 hPa. There is better agreement between radiosondes of the nonlinear version than those associated with the linear version for moisture. The nonlinear version retrieval rmse is better than the guess results at 850 hPa where there is large moisture variation, both spatially and temporally. Figure 6 shows 0000 UTC verification statistics of total water vapor from April 1996 to March 1997. Both the linear version and nonlinear version more closely agree with the radiosondes as compared with the good first guess near radiosonde locations, although the nonlinear version shows more agreement during the winter months. The same relative performance for the linear and nonlinear versions occurs at 1200 UTC (not shown here).

The monthly rmse of the total precipitable water vapor and the sample sizes are listed in Table 1; the total number of samples is 14,793. The average rmse of the background is 3.32 mm, while the values for the retrieved versions 1.5 and 1.6 are 3.15 and 2.86 mm, respectively. The comparisons between the nonlinear retrieval and background first-guess differences with radiosondes shown in Figs. 7 and 8 are based on a 1-yr matched dataset. These matches were restricted to 0.25°, rather than 0.5°, to further minimize the collocation errors (the smaller match distance is possible due to the large sample size available). Figure 8 provides a clear demonstration of the capability of the nonlinear version to more closely agree with the radiosonde than the forecast even when the NCEP background is quite accurate as it usually is near radiosonde locations. The background dewpoint temperature rmse errors (Fig. 8) are significantly reduced above 850 hPa. At 1000 hPa, the retrieval error is about 0.15° worse than background error, and this may be a result of discrepancies between
conventional surface observations, used as a lower boundary condition for the GOES observations, and radiosonde observations just prior to ascent.

Table 2 provides moisture comparisons of total column, surface to 0.9 sigma (the ratio of pressure level to surface pressure), 0.9–0.7 sigma, and 0.7–0.3 sigma. The retrievals of the total water vapor show better agreement with radiosondes than the NCEP background by about 20% (2.72 versus 3.36 mm). The largest impact of the satellite data is in regions void of conventional observations. Over the radiosonde-rich continent, the first-guess profile is generally accurate in terms of tem-
perature; as a result, the GOES measurements rarely improve the temperature profile accuracy. However, water vapor is much more heterogeneous so that the first-guess profile based on forecasts initialized from coarsely spaced radiosondes is often not very accurate, particularly over oceans, where there are no radiosondes. (Note that the NCEP analyses over oceans benefit from the use of polar orbiting satellite TOVS sounding radiance information.) Total water vapor differences between the NGM 12-h forecast and the GOES-8 sounder retrieved values (with the forecast field as its background) are shown for 20 May 1997 (Fig. 9). Note the large differences away from the continental United States where the GOES radiances had a larger impact on the moisture. For example, near Cancun, Mexico, the retrieved total water vapor agrees more with the radiosondes than the background (from the NGM 12-h forecast) by 60% (4.6 versus 7.4 mm). Similar results have been obtained when the 12-h forecast from the Eta Model is used as the background (not shown).

The retrieved residuals of the GOES channels 1–16, except channel 9 (the ozone channel, which is not used in retrieval processing), are shown in Fig. 10. The residuals are defined to be the difference between the bias-corrected observed brightness temperature and the brightness temperature calculated from the retrieved temperature and moisture profiles. As shown, the residuals are comparable to the GOES instrument noise level. This demonstrates the nonlinear version is providing profiles that are physically consistent with GOES-measured radiances.

Numerical model assimilation provides another method to assess the quality of the retrievals. In the spring of 1996, NCEP ran a parallel test incorporating GOES-8/9 moisture data from the nonlinear retrievals into the Eta Model system; the operational model served as the control run and the experimental run consisted of the operational model plus retrieved moisture information (three-layer integrated water vapor) inserted every 3 h.

Table 1. Monthly total precipitable water vapor (mm) retrieval comparisons. (0000 UTC Apr 1996–Mar 1997, version 1.5 vs 1.6, matched distance = 0.5°, and rms is the root-mean-square error with the bias included.)

<table>
<thead>
<tr>
<th>Month</th>
<th>rmse (bgd)</th>
<th>rmse (1.5)</th>
<th>rmse (1.6)</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr</td>
<td>3.72</td>
<td>4.33</td>
<td>3.27</td>
<td>1224</td>
</tr>
<tr>
<td>May</td>
<td>3.77</td>
<td>3.72</td>
<td>3.44</td>
<td>2278</td>
</tr>
<tr>
<td>Jun</td>
<td>4.15</td>
<td>3.57</td>
<td>3.51</td>
<td>2230</td>
</tr>
<tr>
<td>Jul</td>
<td>4.25</td>
<td>3.45</td>
<td>3.50</td>
<td>1730</td>
</tr>
<tr>
<td>Aug</td>
<td>3.88</td>
<td>3.42</td>
<td>3.55</td>
<td>1552</td>
</tr>
<tr>
<td>Sep</td>
<td>3.52</td>
<td>3.09</td>
<td>2.95</td>
<td>974</td>
</tr>
<tr>
<td>Oct</td>
<td>3.51</td>
<td>2.93</td>
<td>2.77</td>
<td>1574</td>
</tr>
<tr>
<td>Nov</td>
<td>3.03</td>
<td>3.30</td>
<td>2.86</td>
<td>608</td>
</tr>
<tr>
<td>Dec</td>
<td>2.21</td>
<td>2.72</td>
<td>2.30</td>
<td>687</td>
</tr>
<tr>
<td>Jan</td>
<td>2.90</td>
<td>2.62</td>
<td>2.14</td>
<td>885</td>
</tr>
<tr>
<td>Feb</td>
<td>2.56</td>
<td>2.28</td>
<td>1.96</td>
<td>865</td>
</tr>
<tr>
<td>Mar</td>
<td>2.34</td>
<td>2.39</td>
<td>2.06</td>
<td>186</td>
</tr>
</tbody>
</table>
from the GOES-8/9 sounders. As in an earlier impact test (Lin et al. 1996), these data were assimilated using the Eta Data Assimilation System. Validation with respect to rain gauges and radiosondes revealed that the forecasts were generally improved by using the GOES data (Menzel et al. 1998). As a result of these findings, NCEP placed a high priority on the use of the GOES sounder moisture data in its numerical weather analysis/forecast operation. On 7 February 1997 the nonlinear version was implemented operationally by NOAA/NESDIS.

6. Future extensions

The following improvements to the current GOES retrieval procedure are under way. A fast transmittance model is currently used for the forward transfer radiance calculation (Eyre 1991). This model is an updated version of Weinreb et al. (1981). Unlike the model developed by Eyre and Woolf for microwave channels (1988), it provides a common version for both microwave and infrared channels. Currently, a new fast forward calculation algorithm is under development for the Atmospheric Infrared Sounder (AIRS), which is scheduled on the Earth Observation System PM platform (Hannon et al. 1996). There are two “high-resolution” AIRS forward model algorithms available for water vapor. One algorithm is referred as to pressure layer optical depth (PLOD) or pressure layer fast algorithm for atmospheric transmittances (PFAAST) and is based on atmospheric layers with fixed pressure and variable water amounts. Following the PLOD algorithm, a new fast transmittance model has been used for the GOES retrievals. Retrieval results show that total water vapor rmse can be reduced up to 15% using the new transmittance model in place of the current transmittance model. More detailed information will be presented in a subsequent paper.

As reported by Weinreb et al. (1996), both the imager and sounder exhibit an east–west radiance gradient due to reflection off the coating of the scan mirrors. This problem was essentially eliminated with ground processing in the imagers as of the spring and summer of 1995. The sounder instruments were corrected on 26 March 1996 for GOES-9 and 19 June 1996 for GOES-8 (Weinreb et al. 1997). This improvement in calibration had a direct positive effect on the temperature and moisture profile retrievals. Also, the improved calibration allowed for a simplified version of the radiance bias adjustment, one that does not depend on the time of day (or the scan mirror temperature). A bias-correction algorithm, which is similar to one used by Susskind (1993), has been developed and tested. Raob data are assumed as “truth,” and the bias correction is based on a linear regression with five predictors, including the channel brightness temperature observation, satellite zenith angle, and the brightness temperatures for the three water vapor channels. For the bias correction of channels 1 and 2, the moisture channel brightness temperatures are not used as predictors. The bias correction regression coefficients are updated once every three to six months.

It has been shown that the retrieval accuracy is dependent on the accuracy of the first guess (Hayden 1988). If the forecast is poor, it is desirable to improve

TABLE 2. Total precipitable water vapor (mm) retrieval comparisons. (0000 UTC Apr 1996–Mar 1997, version 1.6, 1637 samples, matched distance = 0.25°, rmse is the root-mean-square error, and rmsd is the root-mean-square difference.)

<table>
<thead>
<tr>
<th>Layer (Sigma)</th>
<th>Background Bias rmse</th>
<th>Retrieval Bias rmse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column</td>
<td>−0.29 3.34 3.36</td>
<td>−0.03 2.72 2.72</td>
</tr>
<tr>
<td>0.9–1.0</td>
<td>0.88 1.53 1.77</td>
<td>−0.62 1.46 1.59</td>
</tr>
<tr>
<td>0.7–0.9</td>
<td>0.24 2.08 2.09</td>
<td>0.29 1.74 1.76</td>
</tr>
<tr>
<td>0.3–0.7</td>
<td>0.31 1.38 1.41</td>
<td>0.26 1.10 1.13</td>
</tr>
</tbody>
</table>
this guess prior to the application of the physical retrieval algorithm. A technique has been developed to replace a poor forecast guess with an improved guess based on a linear combination of a statistical regression (between radiosonde profiles and GOES radiances) retrieval and the original forecast model guess profile. Regression coefficient datasets for sounder viewing angles, 0°, 15°, 30°, 45°, and 60°, are calculated using the historic dataset mentioned in section 4. A hybrid initial-guess profile to be used in the physical nonlinear retrieval algorithm is constructed by a linear combination of the NCEP forecast profiles and the regression retrieval. The weights for the linear combination are inversely proportional to the radiance discrepancies between the radiances calculated for each guess profile with the observed radiances. The initial profile with the smallest discrepancy is given the highest weight in the construction of the hybrid profile. The hybrid profile is used as a replacement to the NCEP forecast profile only if the radiance discrepancy for the hybrid initial-guess profile is less than the radiance discrepancy resulting from the NCEP profile.

Another important matrix in the retrieval is the GOES measurement error covariance matrix. Usually it is assumed that there is no correlation between channels, so the matrix \( \mathbf{E} \) is a diagonal matrix. In the physical retrieval processing, the error covariance diagonal elements are determined from instrument noise variance and the variance of the error in the local surface parameters. However, since this error should also include a forward model error, the correlation between off-diagonal elements (different channels of brightness temperature) may be significant. Therefore, the measurement error covariance is used to account for measurement errors, as well as errors in the forward model and data processing. This matrix is derived from the discrepancies between the bias-corrected GOES observations and brightness temperatures calculated from collocated raob profiles in the retrieval/raob matched dataset. More work needs to be done prior to incorporating this change; early tests have produced promising results.

7. Conclusions

This paper has described the nonlinear physical retrieval algorithm as it is being used for GOES sounding retrievals. The algorithm has been successfully applied in real time to GOES observations beginning in April 1996. Since 7 February 1997 the nonlinear version has been implemented operationally by NOAA/NESDIS. Statistical comparisons between the linear and nonlinear versions during a 1-yr period show that the linear version can degrade a high-accuracy NCEP forecast temperature profile achieved near radiosonde verification locations. However, the nonlinear version does retain the accuracy of the NCEP forecast temperature. For the moisture profile, GOES retrievals using both the linear and nonlinear algorithms achieve a major improvement; the nonlinear version providing much better moisture retrieval accuracy in the low atmosphere and during the dry months. The nonlinear version shows better agreement with radiosondes (of about 20%) for the total water vapor versus the NCEP forecast near radiosonde locations. Difference fields of retrieved minus background total precipitable water indicate much greater differences are achieved over oceanic regions where the accuracy of the NCEP forecast is degraded possibly as a result of the lack of radiosonde information. (The result has been verified using isolated radiosonde measurements in data-sparse regions.) During the spring of 1996, nonlinear retrievals from the GOES sounders were tested in the early Eta Model of NCEP and improved precipitation forecasts were achieved. Since the fall of 1997, NCEP has been using the GOES-8/9 sounder moisture data (total and three layers) in the operational Eta Model.

Acknowledgments. We are indebted to Dr. Christopher M. Hayden’s dedicated contributions to the GOES sounding retrieval program. Our sincere appreciation is also extended to Dr. Clive D. Rodgers for his valuable suggestion and to Dr. H. L. Huang for his extensive contributions in the implementation of the retrieval algorithm. Dr. Paul Van Delst provided Fig. 1. Hal Woolf did many of the radiative transmittance model calculations. This research was supported by NOAA Contracts 50WCNE-306075 and NA67EC0100.

APPENDIX A

Derivation of Eq. (2)

For simplicity, Eq. (1) can be rewritten as

\[
R = B_s \tau_s - \int_0^{\tau_s} B \, d\tau, \tag{A1}
\]

where

\[
R = R(v_s, \theta), \quad B = B(v_s, t), \quad \text{and} \quad \tau = \tau(v_s, \theta, p).
\]

The subscript \( s \) denotes either ground or cloud surface values.

The radiance spectrum corresponding to an assumed initial temperature and absorbing gas profiles is

\[
R^0 = B_{0}^s \tau_0^s - \int_0^{\tau_0} B_0^0 \, d\tau^0. \tag{A2}
\]

Subtracting Eq. (A2) from Eq. (A1) yields

\[
\delta R = R - R^0 = B_s \tau_s - B_{0}^s \tau_0^s - \int_0^{\tau_s} B \, d\tau + \int_0^{\tau_0} B_0^0 \, d\tau^0. \tag{A3}
\]

Employing the linear perturbation definitions
\[ \delta B = B - B^0 \quad \text{and} \quad \delta \tau = \tau - \tau^0, \]

Eq. (A3) becomes
\[ \delta R = \delta B \tau^0 + B \delta \tau - \int_0^\rho \delta B \, d\tau^0 - \int_0^\rho B \, d(\delta \tau). \]  

(A4)

Integrating the last integral on the right of Eq. (A4) by parts yields
\[ \delta R = \delta B \tau^0 + \int_0^\rho \delta B \, d\tau^0 + (B_x - B) \delta \tau_x \]
\[ + \int_0^\rho \frac{\delta B}{\delta p} \, dp. \]  

(A5)

where the subscript \( a \) represents the surface air value.

Assume that the total transmittance is given by the product of uniformly mixed gas, water vapor, and ozone transmittance components, that is,
\[ \tau = \tau_{\text{so}} \tau_{\text{w}}, \]  

(A6)

For the water vapor transmittance \( \tau_w \), it can be approximately expressed as
\[ \tau_w = \exp \left[ -\frac{1}{g} \int_0^\rho \left( k_w \, q_u \, q_u \right) \, dp \right]. \]  

(A7)

where \( g \) is the gravitational constant and \( k_w \) and \( k_w' \) are absorption line and continuum absorption coefficients. Thus,
\[ \delta \ln(\tau_w) = -\frac{1}{g} \int_0^\rho \left[ (k_w + k_w' q) q - (k_w + k_w' q^0) q^0 \right] \, dp, \]
\[ \delta \ln(\tau_w) = \frac{1}{g} \int_0^\rho \left( k_w + k_w' (q + q^0) \right) \, \delta q \, dp', \]  

and
\[ \delta \ln(\tau_w) = -\frac{1}{g} \int_0^\rho \left[ k_w + 2k_w' q^0 \right] \, \delta q \, dp'. \]

If we define
\[ \tau_{\text{dry}} = \tau_{\text{CO}_2} \tau_{\text{O}_2}, \]  

(A8)

then
\[ \delta \tau = \tau_{\text{dry}} \delta \tau_w = \frac{\delta \tau_w}{\tau_w} = \delta \ln(\tau_w) \]
\[ = -\tau^0 \int_0^\rho \frac{(k_w + 2k_w' q^0)}{g} \, \delta q \, dp'. \]  

(A9)

Let
\[ \alpha = \frac{(k_w + 2k_w' q^0)}{g}. \]  

(A10)

Substituting (A10) into (A9) and (A9) into (A5) yields
\[ \delta R = \delta B \tau^0 + \int_0^\rho \delta B \, d\tau^0 + (B_x - B) \tau_x \]
\[ + \int_0^\rho \frac{\delta B}{\delta p} \, dp \int_0^\rho \left( \alpha \delta q \right) \, dp. \]  

(A11)

Letting
\[ u = \int_0^\rho \alpha \delta q \, dp \]  

and
\[ v = \int_0^\rho \tau^0 \frac{\delta B}{\delta p} \, dp', \]

then
\[ du = \alpha \delta q \, dp \]  

and
\[ dv = \frac{\delta B}{\delta p} \tau^0 \, dp'. \]

Since
\[ \int_0^\rho u \, dv = uv - \int_0^\rho v \, du, \]

integrating the last term on the right of Eq. (A11) by parts yields
\[ \delta R = \delta B \tau^0 - \int_0^\rho \delta B \, d\tau^0 \]
\[ + \int_0^\rho \delta \left[ \left( B_x - B \right) \tau_x - \int_0^\rho \frac{\delta B}{\delta p} \, dp \right] \alpha \, dp. \]  

(A12)

Rewriting Eq. (A12) in the form of brightness temperature, we have
\[ \delta t_b = \left[ \frac{\partial B}{\partial \ln t_b} \right] \tau_x - \int_0^\rho \left[ \frac{\partial B}{\partial \ln t_b} \right] \tau_x \, dp \]
\[ + \int_0^\rho \left\{ \frac{\partial B}{\partial t} \left( t_x - t_b \right) \tau_x \right. \]
\[ \left. - \int_0^\rho \alpha \frac{\partial t}{\partial \ln t_b} \frac{\partial B}{\partial t} \, dp \right\} \delta q \, dp. \]  

(A13)

APPENDIX B

Derivation of Eq. (12)

Assuming that \( x \) is an \( n \)-dimensional column vector, \( a = a(x) \) and \( b = b(x) \) are \( m \)-dimensional column vector functions with respect to the vector \( x \), the following relation is held
\[ \frac{d(a^T b)}{d \mathbf{x}} = \frac{d a^T}{d \mathbf{x}} b + \frac{d b^T}{d \mathbf{x}} a, \]  

(B1)

where \( \tau \) denotes vector or matrix transpose.

Employing Eq. (B1), since \( \mathbf{C}^{-1} \) and \( \mathbf{E}^{-1} \) are symmetric matrices, we have
\[ \nabla J(x) = \frac{d(x - x_0)^T}{dx} C^{-1}(x - x_0) + \frac{d[C^{-1}(x - x_0)^T]}{dx}(x - x_0) + \frac{d[y^m - y(x)]^T}{dx} E[y^m - y(x)] + \frac{d[E^{-1}[y^m - y(x)]]^T}{dx}[y^m - y(x)] \]

\[ = 2C^{-1}(x - x_0) - 2\frac{d[y(x)]^T}{dx} E^{-1}[y^m - y(x)] \]

\[ = 2C^{-1}(x - x_0) - 2K(x)^T E^{-1}[y^m - y(x)], \quad (B2) \]

By differentiating Eq. (B2) and ignoring \( K(x) \) variations with respect to vector \( x \),

\[ \nabla^2 J(x) = 2C^{-1} + 2K(x)^T E^{-1}K(x). \quad (B3) \]

Substituting Eq. (B2) and (B3) into (11) yields

\[ x_{n+1} = x_n - \{ C^{-1} + K(x_n)^T E^{-1}K(x_n) + \gamma_n \}^{-1} \times \{ C^{-1}(x - x_0) - K(x_n)^T E^{-1}[y^m - y(x_n)] \}. \quad (B4) \]

By matrix manipulation of Eq. (B4), it becomes the following iterative form

\[ x_{n+1} = x_0 + \{ C^{-1} + K(x_n)^T E^{-1}K(x_n) + \gamma_n \}^{-1} \times \{ K(x_n)^T E^{-1}\hat{\delta} y^m + K(x_n)(x_n - x_0) + \gamma_n(x_n - x_0) \}. \quad (B5) \]

where \( \hat{\delta} y^m = y^m - y(x_n) \).

REFERENCES


M A E T A L.


