THE RELATIVE IMPORTANCE OF THE RETRIEVAL APPROACH AND THE FIRST GUESS IN THE NESDIS RETRIEVAL ALGORITHMS

Larry McMillin

National Oceanic and Atmospheric Administration
National Environmental Satellite, Data, and Information Service
Satellite Research Laboratory
Washington, D.C. 20233

1. INTRODUCTION

Satellite retrieval methods convert information obtained by satellites in the form of measured radiances to atmospheric parameters such as profiles of temperature and moisture. There is an inherent information content in the measurements and an information content in the final retrievals. A good retrieval method preserves the maximum amount of information in the final retrieval. Comparisons of retrievals using classification with those that don't have demonstrated that retrievals based on a classification guess are significantly more accurate than those based on a climatological guess. A probable explanation is that the conventional solutions include the physics of the radiative transfer equation, but fail to include meteorological constraints such as the knowledge that superadiabatic lapse rates should not occur. Classification contains meteorological information in empirical form. This suggests that the radiances contain information that is filtered out by the conventional physical and regression based retrieval methods now in use. The current retrieval situation at the National Satellite, Data, and Information Service (NESDIS) is reviewed, the auxiliary information is examined, and possible changes are discussed.

2. REVIEW OF RADIATIVE TRANSFER THEORY

The radiation emerging from the top of the atmosphere is given by the radiative transfer equation as

$$R(\nu) = B(\nu,T(s))\tau(\nu,s) + \int_{p(s)}^{0} B(\nu,T(p))(d\tau(\nu)/dp)dp \tag{1}$$

where $R(\nu)$ is the measured radiance at wavenumber ν, $B(\nu,T(s))$ is the blackbody radiance at wavenumber ν for an object at the surface temperature, $T(s)$, where s denotes the surface conditions, $\tau(\nu,s)$ is the transmittance of the atmosphere between the satellite and the surface at wavenumber ν, $B(\nu,T(p))$ is the radiance emitted at wavenumber ν by an object at temperature $T(p)$, $d\tau(\nu)/dp$ is the derivative of the transmittance at wavenumber ν with respect to pressure, p, and dp is the pressure increment of the layer. For soundings, the Planck function is often given by

$$B(\nu,T) = 1.191066\ 10^{-5}\nu^3 \ /\ (\exp[1.438833\ \nu/T]-1) \tag{2}$$

where $B(\nu,T)$ denotes the radiance in units of $mW/(m^2\ sr\ cm^{-1})$. For instruments used in soundings, $R(\nu)$ is restricted by filters to narrow ranges of wavenumbers. It is common to denote the radiance for a single channel is denoted as $R(\nu(i))$ where i identifies the channel an $\nu(i)$ is the average wavenumber for the particular filter.

It is also common practice to retrieve departures from an initial state. When this is done, eq. (1) can be written as

$$R - \bar{R} = B(s)\tau(s) - \bar{B}(s)\bar{\tau}(s) + \int_{p(s)}^{0} (Bd\tau/dp - \bar{B}d\bar{\tau}/dp)dp \tag{3}$$

where the barred quantities represent an average or some other initial guess and the dependencies on wavenumber and temperature that were explicitly noted in eq. (1) have been dropped to keep the notation simple. Eq. (3) can be simplified further by writing it as

$$\Delta R = \Delta B(s)\tau(s) + \int_{p(s)}^{0} \Delta B(d\tau/dp)dp. \tag{4}$$

Finally, the integral in eq. (4) is replaced with a summation over finite intervals to give

$$\Delta R = \Delta B(s)\tau(s) + \sum_{i=1}^{n} \Delta B(d\tau/dp)dp(i). \qquad (5)$$

In the retrieval process, values of ΔR are known and values of ΔB are desired. To start, the surface term in eq. (5) is assumed known and subtracted out. Strictly speaking, values of ΔR and ΔB in eq. (5) are for the wavelength of the particular channel. This makes ΔR a matrix, but for the retrieval, ΔR must be a function only of pressure. To remove the wavenumber dependence of ΔR, eq. (2) is used to scale values of ΔB and ΔR to either a brightness (equivalent) temperature or to an equivalent wavenumber and eq. (5) is written in matrix form as

$$r = At \qquad (6)$$

where r and t are column vectors of the departures of radiance and temperature from some initial condition, A is a matrix, and ' denotes the transpose. For more details, see Fleming (1972) or Rodgers (1976).

Although the elements of t in eq. (6) are the integral of the temperature over the over the layers in eq. 5, it is common practice to select a pressure scale in which layers change smoothly with height. When this is done, values of Δp change smoothly over adjacent levels and it is possible to drop the Δp by defining Δp as p(i+1)-p(i-1) and t as t(i). This avoids the need to interpolate the solution when physical retrievals are produced. However, at the bottom of the atmosphere, p(i+1) does not exist, so dp must be taken as 2(p(i)-p(i-1)). A similar situation occurs at the top of the atmosphere where p(i-1) does not exist. It needs to be emphasized that the simplification can be done only when the values of dp change smoothly with height. In particular, the simplification does not work for the 40 pressure levels (Table 1.) at which the NESDIS retrievals are calculated in the current regression approach because the spacing between levels is uneven. Physical retrievals at NESDIS are done at evenly spaced levels and interpolated to the unevenly spaced 40 levels.

When errors are included, eq. (6) becomes

$$r = At + e \qquad (7)$$

Table 1. Pressure levels for retrievals in the TOVS processing.

level	pressure (hPa)	level	pressure (hPa)	level	pressure (hPa)	level	pressure (hPa)
1	0.1	11	10	21	115	31	500
2	0.2	12	15	22	135	32	570
3	0.5	13	20	23	150	33	620
4	1.0	14	25	24	200	34	670
5	1.5	15	30	25	250	35	700
6	2.0	16	50	26	300	36	780
7	3.0	17	60	27	350	37	850
8	4.0	18	70	28	400	38	920
9	5.0	19	85	29	430	39	950
10	7.0	20	100	30	475	40	1000

where e denotes the errors. A solution of the form

$$t = Cr \tag{8}$$

is desired where C is the matrix of retrieval coefficients. Solving for C gives

$$C = TR'(RR')^{-1} \tag{9}$$

where T and R are matrices of r and t for a sample. Substituting from eq. (7) gives

$$C = TT'A'(ATT'A' + ee')^{-1} \tag{10}$$

which is commonly written as

$$C = S[T]A'(AS[T]A' + S[e])^{-1} \tag{11}$$

where S[T] and S[e] are the temperature and error covariance matrices, respectively. In the physical retrieval, S[T] is calculated from a set of radiosondes and S[e] is usually assumed to be a diagonal matrix whose elements are equal to the square of the noise for the given channel. Retrievals are also made by regression in which case a set of radiosondes collocated with satellite measurements are collected and eq. (9) is solved directly with the collocated data. Because there are always systematic differences between calculated and measured radiances, methods classified as physical use empirical methods to tune to observed data, thus blurring the distinction between the two methods.

A problem with both solutions arises from the nature of the matrix A. The rows of A are weighting functions which determine the regions of the atmosphere which contribute to r in eq. (7). The filters, which define the wavenumber regions for a particular channel, need to be broad enough to supply sufficient energy to the detector to be measured above the noise level. To be broad enough, they also cover several absorption lines. This broadens the region of the atmosphere represented by a single single channel. The atmosphere also has natural correlations between levels which tend to have the same effect as the broad weighting functions. As a result, the matrix that needs to be inverted in eq. (11) is nearly singular and the rows of C tend to have several coefficients greater than 1.0 with opposing signs, and the coefficient with the largest size often multiplies a channel that derives none of its signal from the layer being predicted. In the right situation, either one of these effects is sufficient to cause unrealistic retrievals to occur. Another result of the broad weighting functions and the correlations is the limited ability of the retrieval to recover fine scale features. Changes from the initial guess take the form of smooth changes which cover broad regions of the atmosphere.

3. THE CURRENT NESDIS RETRIEVAL APPROACH

Although NESDIS has plans to switch to a new retrieval approach in the near future, at the present time NESDIS is using an empirical approach. The retrieval has been described by Phillips et al. (1979), Smith et al. (1979), and McMillin and Dean (1982). The parts that are changing are the procedures for selecting the initial guess and the retrieval algorithm. The cloud clearing approach described by McMillin and Dean (1982) will remain in its current form.

Currently the initial guess can be described as a climatological approach, although it has some elements of a classification approach. Collocations of satellite and radiosonde data are collected for five latitude zones, -90 to -60,-60 to -30, -30 to +30, 30 to 60, and 60 to 90 degrees. Mean values and regression coefficients are generated for each of the five zones. When the coefficients are used, retrievals are made with coefficients which are a combination of the ones for the zone containing the sounding and the nearest neighboring zone. The measured value for MSU channel 4 is then used to interpolate between the zones. Details of the procedure are given in (Phillips et al. 1979).

The retrieval method is described by Smith and Woolf (1976). It is a regression approach, but differs from normal regression in that eigenvectors are generated for both the temperatures, which are the predictands, and the brightness temperatures, which are the predictors. When coefficients are generated, both the temperature vector and brightness temperature vector are converted to vectors containing the coefficients of the eigenvectors. The eigenvectors with eigenvalues judged to be below the noise level are discarded, and regression coefficients that predict the coefficients of the remaining temperature eigenvectors from the coefficients of the remaining brightness temperature eigenvectors are generated. The regression coefficients are then multiplied by the eigenvectors to produce coefficients that directly convert measured brightness temperatures to retrieved temperature profiles. Details of this procedure and equations are given in the paper by Smith and Woolf (1976).

4. THE NEW NESDIS APPROACH

As mentioned, NESDIS is planning to change to a new algorithm for the operational retrievals. The new NESDIS algorithm consists of two changes, one to the method of obtaining a first guess and one to the method of doing the retrieval.

4.1 The library search procedure

The climatological mean is being been replaced the library search procedure described by Goldberg et al. (1988). In their library search routine, a library of about 1000 recent and carefully selected radiosonde-satellite collocations is maintained. These are continually updated from a fixed set of 200 selected radiosonde stations. When a retrieval is made, 20 profiles are selected from the library and used to generate the initial guess for the retrieval method. The 20 are selected by scanning the library to find the 20 with radiances closest to the observed radiances where closeness is measured as

$$\rho = (r[m] - r[lib]S[lib](r[m] - r[lib])' \tag{12}$$

where r[m] is the vector of measured radiances, r[lib] is the vector of radiances for one of library members, and S[lib] is the covariance matrix or r[lib]. Because scanning the library is computationally intensive, a check is made before it is done. In the check, the average radiance for the 20 library members used for the last retrieval is obtained. Then the distance ρ between the r[m] and the average is determined from eq. (12) with r[lib] replaced by the average of the 20. This distance is compared to the average distance that was obtained the last time the sample of 20 was updated. A new search is initiated only if the new distance exceeds the average distance for the 20.

The use of the library search procedure results in significant increases in retrieval accuracy. Goldberg et al. (1988) show comparisons between the MVS solution with 27 bin means with the MVS solution with the library search. The 27 bin means are yet another form of climatological first guess. Comparisons are shown for three latitude bands, 60 - 90, 30 - 60, and -30 to +30 degrees. Improvements in accuracy of .1 to .3K are noticed near 300 hPa and near the surface in all three bands. Near 600 hPa, the results of the two methods are similar. The library search routine also produces improvements in the accuracy of the moisture retrievals. The biggest increase in accuracy occurs near the surface where most of the water vapor is concentrated. Improvements of 0.1 to 0.3 g/kg are observed.

4.2 Minimum Variance Simultaneous (MVS) retrieval

The second part of the new system consists of the change from a regression solution to the MVS retrieval. Because the MVS retrieval algorithm was developed over several years, a complete description of the method and its performance is scattered over several references, Fleming et al. (1988), Fleming et al. (1986a), Goldberg et al. (1986), Fleming et al. (1986b), and Smith and Woolf (1984). It should be noted that retrieval methods are subject to a large number of variations, and some of the particular variations discuss in the earlier papers are not used in the final solution. NESDIS plans to write a paper describing the details of the final version in the near future. To start our discussion, it is convenient to return to eq. (3) and follow the development given in Fleming et al. (1986a) which is a review of Smith and Wolf (1976). Adding and subtracting the term $B[s]\tau[s]$ outside the integral and $Bd\tau/dp$ inside the integral and collecting terms gives

$$\Delta R = (\Delta B[s])\tau[s] + \bar{B}[s](\Delta\tau[s]) + \int_{p(s)}^{0} \{(\Delta B)d\tau/dp + \bar{B}(\Delta d\tau/dp)\}dp. \tag{13}$$

When the right hand term in the integral is integrated by parts, eq. (13) becomes

$$\Delta R = (\Delta B[s])\tau[s] + \int_{p(s)}^{0} \{(\Delta B)d\tau/dp - (\Delta\tau)d\bar{B}/dp\}dp. \tag{14}$$

The second term in the integral in eq. (14) is the one used to bring water vapor into to the simultaneous solution. According the Fleming et al. (1986a), the Taylor expansion of $\Delta\tau$ in (14) is a one-term expansion about precipitable water vapor that can be inaccurate. They use the more elaborate expansion given in Fleming et al. (1986a). They also differ from Smith and Woolf (1984) in that they solve for ΔT and ΔU directly rather than use basis functions. As a result of these changes, the vector of ΔR in eq. (14) contains the surface temperature and the moisture as well as the temperature. Fleming et al. (1986a) use ΔV to represent the combined vector given by

$$\Delta V = (\Delta T[1], \ldots, \Delta T[n1], \Delta T[s], \Delta U[1], \ldots, \Delta U[n2])' \tag{15}$$

where n1 is the number of temperature levels and n2 is the number of water vapor levels. It follows that S[T] in eq. (11) is replaced by S[V] where S[V] is given by

$$S[V] = \Delta V \Delta V'. \tag{16}$$

Results of the MVS method are presented in Goldberg et al. (1986) and in Fleming et al. (1988). Goldberg et al. (1986), present a comparison between a theoretical version of the MVS retrieval method, the minimum variance method, and the operational regression. Results are shown for three latitude zones, -30 to +30, 30 to 60, and 60 to 90. For temperature, no one method was best at all levels in all zones, but there were some definite trends. Generally, the MVS and the minimum variance solution were similar with the MVS being slightly more accurate. In addition, the MVS and minimum variance solutions were more accurate than the operational retrieval except for the tropical cases where the regression was the most accurate. Differences varied with height and and ranged from 0K for some levels to .5K for others. For water vapor, the MVS was the most accurate in the lower levels (850 hPa to 1000 hPa), where water vapor is important, for all except one level in one latitude zone. A typical improvement in this region is .2 g/kg in RMS retrieval accuracy out of a total of 2 g/kg in the tropics, .2 g/kg out of 1.5 to 2 g/kg in the midlatitudes, and less than .1 g/kg out of .25 to .4 g/kg. in the polar regions. Fleming et al. (1988) present the results of a comparison between a theoretical approach and one that has been optimized for operational considerations and show that, while there is a general tendency for the more theoretical nonlinear operator to be more accurate than the more operational linearized operator, the difference in accuracy is small and the difference in computation time is large. The difference in accuracy is generally less than 0.1 K for temperature and .1 g/kg for water vapor. They recommend using the more efficient procedure because of the large saving in computation time. Other discussions of simultaneous retrieval methods are given in two recent parers, (Uddstrom 1988, and Hayden 1988).

5. FUTURE CHANGES

Changes are planned for the future, both as part of the new processing system designed for the Advanced Microwave Sounder Unit (AMSU) and as part of an ongoing retrieval program at NESDIS. The changes for the AMSU will be implemented in two stages, currently known as System 90 and System 92 although the dates change because of changes in delivery dates and planned launch dates. Changes will be made in two stages because the building of a new system is an opportunity to change features of the current operational system that are known to be deficient, but are so ingrained in the present system that they can't be changed without building a new processing system. The system 90 changes involve the changes that can be made with the current Microwave Sounder Unit (MSU) while the System 92 changes are those that depend on the presence of AMSU measurements. Changes for the initial System 90 have been specified so some the items to be discussed represent research that has occurred after the specification was completed. The successful approaches will be implemented at a later date.

5.1 The need for a meteorological retrieval

Before discussing changes, it is appropriate to review the information that is available when a retrieval is started, how various methods make use of the information, and examine ways of including more of the available information in the retrieval process. At the start we know that the solution must satisfy the measured radiances. We have some knowledge of the errors in the measurements, the relationship between measured radiances and retrieved temperatures, and the interrelationships between the various retrieval quantities in the form of a covariance matrix. Finally, the solution must provide a profile that satisfies meteorological physics. In particular, it must not provide a profile that is superadiabatic. Generally, current retrieval models include knowledge of the physics of radiative transfer, but leave room for improvement in their ability to account for meteorological physics. An evaluation of the current situation has led to several promising approaches for improving retrievals. Several of the more promising ones will now be discussed. Some of these approaches are the result of attempts to develop a meteorological retrieval that includes more physical meteorology in the retrieval process.

One way of evaluating the meteorological performance of retrieval methods is to compare the meteorological parameters derived from retrievals with the same parameters derived from radiosondes. Figures. (1 through 3) compare the frequency distributions of lapse rates derived from satellite retrievals with those derived from the collocated radiosondes. The distributions are shown for one degree intervals of lapse rate. The expected distribution is one that increases from left to right until the 10 K/km is reached. At 10 K/km the bar should be half the value at 9 K/km because only the half of the interval below the adiabatic limit of 9.8 K/km should be populated. The number of superadiabitic lapse rates should be low. In particular, the number of superadiabatic lapse rates for

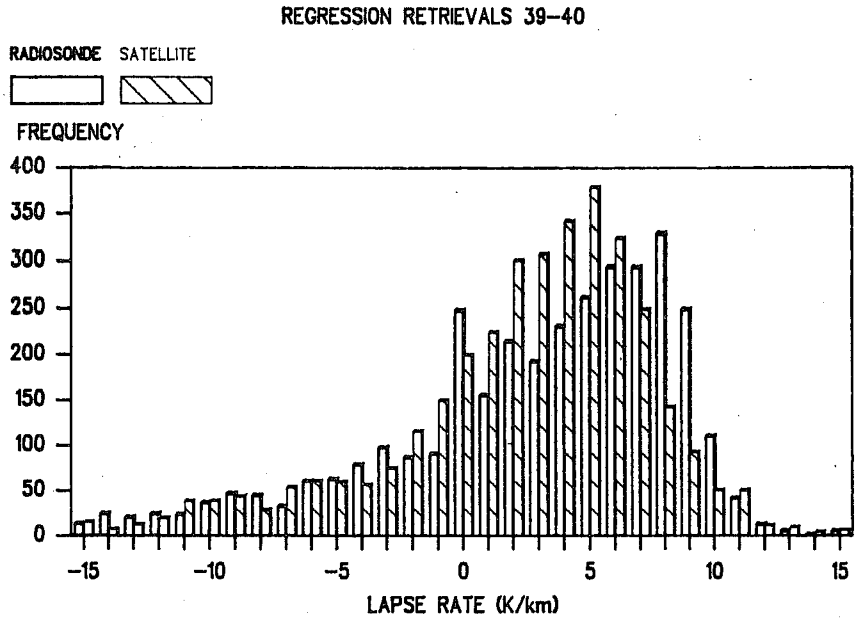

Figure 1. A histogram showing the frequency distribution of the lapse rate of temperature between levels 39 and 40 (950 hPa and 1000 hPa). The distribution for the operational retrievals is compared with the distribution for the collocated radiosondes.

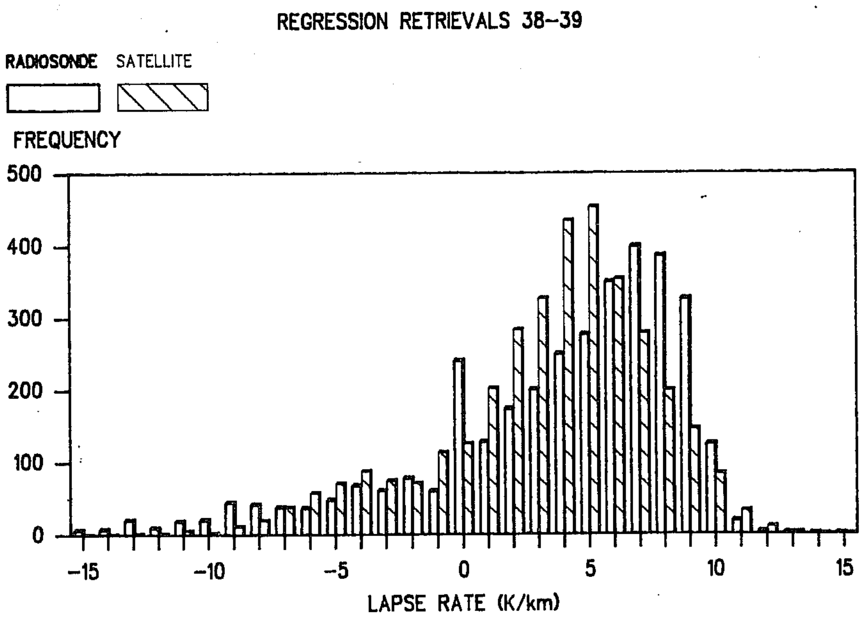

Figure 2. A histogram showing the frequency distribution of the lapse rate of temperature between levels 38 and 39 (920 hPa and 950 hPa). The distribution for the operational retrievals is compared with the distribution for the collocated radiosondes.

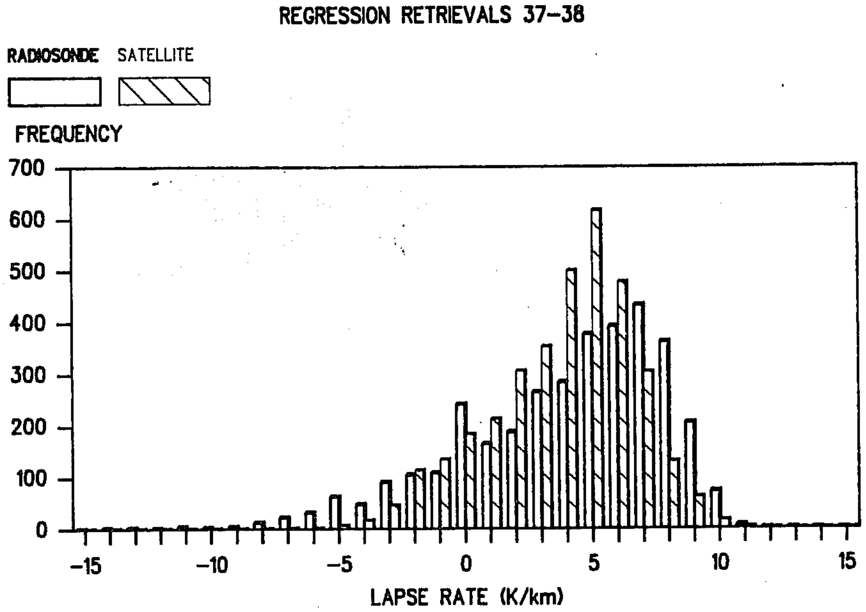

Figure 3. A histogram showing the frequency distribution of the lapse rate of temperature between levels 37 and 38 (850 hPa and 920 hPa). The distribution for the operational retrievals is compared with the distribution for the collocated radiosondes.

retrievals covering 400 km or more should be considerably less than the number for the radiosondes which represent a single location. The distribution for the radiosonde matches expectations well, with a peak at 8 K/km and a high population at the adiabatic limit and a low population beyond. The distribution for the satellite retrievals is more symmetric with a peak at 5 K/km. It is clear that the distributions for the retrievals and the radiosondes differ.

Similar effects are observed at other levels. The distribution for levels 920 to 950 hPa (levels 38 to 39) shows a depletion in the near adiabatic region, too many superadiabic cases, a shifted peak, and too few inversion cases (Fig. 2). At higher levels (Fig. 3), the distributions become more similar.

A feature occurs in the figures that is difficult to overlook. It is the single peak in the radiosonde lapse rates at 0 K/km. The radiosonde data are extracted from the NESDIS processing stream, so it is possible that the effect is due to some feature in the processing; but it is also possible that the feature is a result of procedures used to report radiosondes. We plan to investigate the cause of this effect more thoroughly, but it consistently appears in all the samples.

5.2 A meteorological coordinate for covariances

Many retrieval process use the covariance S[T] or S[V] as a constraint. Currently, retrievals used in the TOVS processing are interpolated to the 40 TOVS pressure levels, and covariances are generated for the temperature at one level relative to the temperature at another. When S[V] is used, correlations between the level temperatures and the other quantities are generated. This form of the covariance matrix is convenient, but not very meteorological. The problem is best illustrated at the surface where the atmosphere has a boundary layer. Because of the diurnal cycle, this layer has unique statistical properties. In general, the height of the boundary layer is about the same height over a surface at 1000 hPa as it is over a surface of 850 hPa. Near the surface, the covariance should be determined for fixed heights above the surface. It is particularly inappropriate to assume that the covariance between the surface and 800 hPa when the surface is at 850 hPa is the same as it is when the surface is at 1000 hPa. For meteorological purposes, it is more consistent to determine covariances at sigma levels (phillips 1957) defined as

$$\sigma[i] = p[i]/p[s] \tag{17}$$

where $\sigma[i]$ is a constant. To include this effect correctly requires a physical solution, because the best representation of the radiative transfer is in pressure levels and the best representation of the covariance matrix is in sigma levels. For physical retrievals, it would be necessary to interpolate between values of S[T] stored at sigma levels, and transmittances calculated at fixed pressures. The closest regression approach would be to predict temperatures at the sigma levels rather than at fixed pressures. An indication of the effect of using sigma levels can

Table 2. Results of a comparison between retrievals at fixed pressures with those at sigma levels given in term standard deviations (SD).

level	fixed pressure		sigma levels	
hPa	sample SD	retrieval SD	sample SD	retrieval SD
100	12.30	1.78	12.32	1.75
115	12.21	2.00	12.15	1.98
135	11.04	1.96	10.95	1.92
150	10.00	1.89	9.90	1.87
200	7.83	2.91	7.75	2.81
250	8.34	2.63	8.47	2.55
300	9.92	1.98	10.13	1.97
350	10.51	1.77	10.76	1.83
400	10.66	1.75	10.95	1.89
430	10.56	1.77	10.87	1.91
475	10.39	1.70	10.71	1.84
500	10.30	1.65	10.61	1.84
570	10.27	2.36	10.63	2.61
620	10.52	3.11	10.74	2.91
670	10.39	1.99	10.65	1.88
700	10.46	1.61	10.71	1.67
780	10.92	1.75	11.09	1.75
850	11.21	2.09	11.37	1.97
920	11.40	2.19	11.73	2.18
950	11.74	2.40	12.03	2.27
1000	12.37	2.72	12.75	2.84

be obtained by comparing the predictions of sigma level temperatures with those at fixed pressure levels even though regression is only approximately correct. The results of such a comparison are shown in Table 2.

At 1000 hPa, the sigma level retrieval errors are large because the diurnal cycle is a maximum at these levels, and the collocations contain large time differences. For a thorough test, the hourly surface observations should be used to remove the differences between measured and observed values due to the diurnal cycle. With the exception of this level, the sample variances are larger for the sigma levels near the ground while the retrieval errors are smaller. This means that the correlations between temperature and radiance are larger in sigma than in pressure coordinates. It is also noted that changes exist at all altitudes. At the higher altitudes there is pronounced tendency for the pressure coordinate to provide the more accurate results when both are accurate and for the sigma coordinate to provide the more accurate results when both are poor. These results are preliminary and obviously need to be examined, but they suggest that small but significant increases in retrieval accuracy near the ground can be obtained by using sigma coordinates. There is evidence that the removal of the diurnal cycle at the ground would lead to improved accuracy there as well. Davis and Tarpley (1983) used retrieved satellite temperatures to predict the shelter temperature, which is always a fixed height above the surface, to an accuracy 1.6 to 2.6 K for clear and partly cloudy retrievals, an accuracy that exceeds the accuracy of the satellite measurements at the surface. Gal-Chen (1988) has proposed that retrievals be made at constant heights rather than constant pressures.

5.2 Radiance adjustment procedures

In the retrieval process, it is important to match measured and calculated radiances if the retrievals are to be accurate. Due to uncertainties in a number of factors affecting the retrieval process, empirical adjustments are required to make calculated and observed radiances agree. One source of uncertainty that is not commonly recognized is that while carbon dioxide concentrations are relatively uniform in the southern hemisphere, they have an annual variation of about 5% in the northern hemisphere with strong latitudinal gradients at the peaks of the cycle (Gammon et al. 1985). This has some obvious implications for retrieval methods as well as retrieval accuracies in the two hemispheres. Retrieval techniques differ greatly in the approaches used to adjust the radiances. On the one extreme, the classification-regression technique employed by McMillin (1986) is an entirely empirical method that uses only radiosonde data collocated with satellite measurements. This method was developed to avoid the use of calculated radiances because of experience with the difficulties in obtaining the required agreement with calculated values. On the other extreme is the classical physical approach where calculated radiances are used for everything, including the initial guess. In between, is the approach used by Fleming et al. (1986b) in which the the initial guess depends on a set of radiances collocated with radiosondes to avoid the biases between calculated and observed radiances, but in which the retrieval coefficients depend on calculated radiances. A physical retrieval can only be as good as the accuracy to which measured radiances agree with calculated values. Differences in accuracy due to different correction methods are often larger than differences in due to differences in retrieval methods.

Attempts to force agreement between calculate and observed radiances have been presented by Weinreb (1979) and later by Susskind et al. (1983). Adjustments of this sort have been used in TOVS processing for some time and they reduce the uncertainty, but fail to bring the noise to acceptable levels, possibly because the correction makes inherent assumptions about the form of noise and can only adjust errors that conform to the assumptions. Uddstrom (1988) suggested using regression techniques to predict calculated radiances form observed values. This approach is more general than the one employed by Weinreb (1979), and can also be used to solve the problems caused by a radiosonde network consisting of two or more radiosonde types. A problem with this approach is that it can produce coefficients with some disturbing physical implications.\ Since the errors in the calculated values are generally small, the normal expectation is that the coefficient that predicts the calculated value of a channel from is measured value should be near 1.0 and all other coefficients for that channel should be near zero. In other words, the regression coefficients should look like a small perturbation about the identity matrix. Instead, the regression coefficients often have several elements with absolute values of 2 or 3 and often the measured channel with the largest coefficient measures a different region of the atmosphere from the one being predicted. McMillin et al. (1988) have developed a procedure for stabilizing the regression.

It is easiest to visualize the stabilizing method in two dimensions where the measured value of a given channel is used to predict its calculated value. For channels near the tropopause, the atmospheric variation can be small, and the signal to noise ratio becomes small, too. When this happens, the regression minimizes the the squared distance perpendicular to the line given by setting the calculated value equal to its mean. However, since the best initial estimate of the solution is that the calculated value is equal to the measured value, it is more appropriate to minimize squared distances perpendicular to that line instead of to the mean. This is accomplished by rotating the coordinate system 45 degrees counter clockwise so that the new x becomes the line calculated is equal to observed and the new y becomes calculated minus observed. Then regression coefficients are generated in the rotated coordinates and the resulting coefficients are rotated back to the original coordinate system. When the correlation is zero, the initial prediction is unchanged.

In the more general case where calculated radiances are predicted from measured ones, the solution is given by

$$r[c] = (I - D)^{-1} (I + D)\, r[m] \tag{18}$$

where r[c] is the vector of calculated values, r[m] is the vector of measured values, I is the identity matrix, and D is the matrix of regression coefficients in rotated space given by

$$\begin{aligned} D = {} & (R[m]R'[m] - R[c]R'[m] + R[m]R'[c] - R[c]R'[c]) \\ & (R[m]R'[m] + R[c]R'[m] + R[m]R'[c] + R[c]R'[c])^{-1}. \end{aligned} \tag{19}$$

McMillin et al. (1988) show that the regression coefficients obtained when normal regression is used range from -.224 to 1.182 for the diagonal terms, which are expected to be near 1.0, and from -1.005 to 1.081 for the off diagonal terms which are expected to be near 0.0. For normal ridge regression with a smoothing factor of 0.5 times the identity matrix added to the diagonal of RR' in eq.(9), they show coefficients ranging from .138 to .645 on the diagonal and from .176 to .600 off the diagonal. For their rotated regression with a smoothing factor of .5 they show coefficients ranging from .713 to 1.014 on the diagonal and from -.197 to .089 off the diagonal. The coefficients for the rotated regression are closer to the expected values. At the same time, the residual error on the dependent data set, which is by definition minimal for normal regression, increased only slightly from .350 to .426. When noise is taken into account, the expected error on an independent data set becomes more favorable to the rotated regression, because the average size of the coefficients is less. The result is an adjustment procedure that uses coefficients that satisfy physical expectations for a very slight penalty in accuracy. It is important to note that the rotated regression preserves the relationship between the radiance for a given channel and atmospheric levels that is expected by a physical retrieval. It solves one of the major problems with using either an historical data set, as suggested by Chedin et al. (1985), or a forecast for a first guesss rather than current collocated data.

5.3 Radiosonde errors

Radiosonde errors are an important consideration in the radiance correction process since radiosondes are used to judge the performance of retrievals and to adjust the retrieval process. Radiosondes are subject to errors in the temperature sensor, errors in the pressure sensor, and errors in coding procedures. The world radiosonde network is composed of instruments of different designs, built by different manufacturers, and processed by different countries. Systematic differences between radiosonde stations are often observed and procedures have been developed to make corrections (Uddstrom 1988). There is also an active program to intercompare different radiosondes (Schmidlin et al. 1986). However, the longwave radiation has been largely ignored in the design of the intercomparision experiments. The extremes occur when the temperature contrast between the radiosonde and the underlying surface is greatest. These conditions are found over a very cold surface, especially with stratospheric warming, and over a very hot surface with a cold tropopause. Existing radiosonde intercomparisons have been made in atmospheric conditions with more comfortable surface temperatures. Schmidlin (1988) is planning to take some measurements in the conditions where extremes can be expected.

The lead carbonate used as a coating on many instruments has a high reflectivity in the visible, but absorbs 86% of the incident infrared radiation according to Schmidlin (1986). In the same article, they give the emissivity of the aluminum paint used by Vaisalla as 22% although more recent measurements (Schmidlin 1988) indicate that the emissivity may be slightly higher. At the high emissivity values, our studies indicate that a typical radiosonde measures about 1/10 to 1/30 of the radiation striking the temperature sensor, the exact amount depending on factors such as the ventilation rate, the emissivity of the coating, and the density of the air. These values are consistent with those of Ney et al. (1961) who show that one type of sensor responds to about 1/20 of the temperature difference between the air and the radiation environment. It is worth noting that several vacuum deposited metals have very good reflective properties in both the visible and the infrared and that such a coating was recommended over 20 years ago (Ney et al. 1961). Any one of several coatings could reduce the radiation errors to acceptable amounts.

Another component of the error situation is the magnitude of the difference between the temperature of the air and the equilibrium temperature of the radiation environment. Gergen (1957) presented monthly averages of the difference for Minneapolis, Minnesota, that show the difference has a pronounced seasonal dependence at this location. In the summer, when the atmosphere is more tropical, the radiation temperature is over 10 K warmer than the air temperature at the tropopause. This means that radiosondes are subject to heating due to both longwave and shortwave radiation components. In winter, the radiation temperature is over 10 K cooler than the air for the same region. McMillin et al. (1988) found similar effects when satellite measurements were used to derive biases between two radiosonde types. Their results showed that the error due to longwave radiation can reach several degrees Kelvin, and that it varies with atmospheric conditions. Although errors of this magnitude are larger than the values reported by Schmidlin (1986), they were derived from atmospheres where the temperature difference between the radiosonde and the radiation environment were also larger, so the larger errors are expected.

These longwave radiation errors have two effects on satellite retrievals. Radiosondes are often used as a comparison to evaluate the accuracy of satellite retrievals. Any effect that causes an error, especially a profile dependent one, will complicate the radiosonde correction process and the remaining errors are usually attributed to the satellite rather than the radiosonde. It is particularly important to recognize that radiosondes in cloudy area are biased with respect to those in clear areas due to changes in the radiation environment of the radiosonde. The second effect is that attempts to use the difference between measured radiances and calculated values to infer information about errors in the transmittance calculations will produce limited success as long as the radiosondes have variable errors of this magnitude.

5.4 A meteorological view of classification

The temperature retrieval problem is considered to be ill-conditioned in the sense that solutions are not unique in a mathematical context. Because the weighting functions are so broad, it is theoretically possible to impose perturbations on the solution that are below the resolution of the instrument. From a mathematical point of view, these perturbations can be very large. However, Thompson et al. (1986) make the point that any solution that matches radiances of a given observation and occurs in the atmosphere is unique and that the problem is not ill-conditioned. They based their conclusions on an empirical study of 1600 radiosondes. Their argument is not an absolute proof because it can be argued that profiles that produce the same radiances yet differ in temperature can exist outside the data set used for their study. But it can be shown that their conclusions are based more than just empirical observations.

Any valid solution has an inherent lapse rate and any perturbation from a valid solution must move toward the superadiabatic solution for part of its cycle and away from it for the other part. The move toward and away must take place in less vertical distance than the minimum that can be resolved by the instrument, which is 2 - 4 km. If the slope of the perturbation is limited by the superadiabatic limit, and the length of the perturbation is limited by the vertical resolution, then the maximum temperature error caused by the perturbation is also limited. This type of physical limitation is not imposed on the present "physical" solutions, but it is contained in the solutions allowed by the classification. A meteorological view of classification is that it utilizes information about meteorological physics that is otherwise excluded from the retrieval process.

A strong argument for classification can be made by observing channels 10 and 8. Channel 10 is the lowest peaking water vapor channel and channel 8 senses the surface. Because water vapor concentration increases rapidly toward the ground, channel 10 has the sharpest weighting function, and thus is most sensitive to low level inversions. But since water vapor is so variable, the weighting function for channel 10 is unknown, and this makes it difficult to use this channel for retrieving temperature. Yet it is obvious that when channel 10 is warmer than channel 8, an inversion exists. McMillin (198a) has demonstrated that a classification approach is capable of recognizing these inversions and providing a guess with an inversion present. A physical retrieval that starts with a guess that does not include a shallow inversion has difficulty producing one while a solution based on a first guess that recognizes the existence of the inversion has a far better chance to retrieve the correct profile than one that doesn't.

Classification needs to be applied with caution. When retrieval coefficients are determined, a matrix with dimensions equal to the number of channels needs to be inverted. This means that as a minimum, the number of atmospheric profiles needs to equal the number of channels or the matrix will be singular. Unfortunately, there is often confusion between the number of atmospheres and the number of profiles. Multiple measurements on the same atmospheric condition are not multiple atmospheres in the sense required to stabilize the matrix inversion. As classification does a better job, the number of independent profiles in a class becomes smaller, and the relation of the atmospheric variance in the class relative to the noise becomes small. For this reason, blindly applying a classification method is not likely to be successful unless some attention is given to the matrix inversion problem. Several alternatives are available. Uddstrom (1988) does the inversion by adding one channel at a time and terminates the inversion when evidence of instability appears. McMillin (1986, 1988a) uses a large set of profiles to define the covariance matrix for the inversion, but multiple small sets to define class means. Fleming et al. (1988) use covariance matrices generated from a set that is far more diverse than the 20 soundings used to define the mean profile.

McMillin (1988a, 1986) has demonstrated substantial improvements in retrieval accuracy through the use of classification. In that approach, measured radiances are used to select a class and retrieval coefficients are obtained from a set of radiosondes collocated with satellite measurements. In the classification method, a covariance matrix of HIRS channels 1 - 8, 13 -16, and MSU channels 2 and 3 is generated from a representative sample of brightness temperatures. This sample doe not have to be matched with radiosondes. Eigenvectors of the covariance matrix are then generated using a standard statistical package that is designed to handle ill-conditioned matrices. Then the radiances in the sample are converted to eigenvector coefficients through the relationship

$$EC = EV\ BT \tag{20}$$

where EC is the matrix of the eigenvector coefficients, EV is the matrix of eigenvectors and BT is the matrix of brightness temperatures for the sample. Only the four eigenvectors corresponding to the largest eigenvalues are used. Then the values of EC are scanned to find the maximum value and the minimum value for each eigenvector to find the range. The range for each eigenvector is then divided into groups, 16 for the first eigenvector, and 4 each for the remaining three eigenvectors, and the values corresponding to the boundaries of the groups are stored. This completes the classification. When a retrieval is desired, the coefficients of the first four eigenvectors are determined and compared to the boundary values that define the groups.

For the retrieval, a set of radiosondes matched with satellite observations is required. All the pairs are classified and the averages of the measured brightness temperatures and radiosonde temperatures are calculated for each class. If the value for one or more of the eigenvectors exceeds the range, it is assigned to the closest class. For each class with a population greater than one, a covariance matrix was formed. Because some of the classes were sparsely populated and thus singualr, and most of the others were close to singular, the covariances for all classes were combined into a single covariance matrix which was used to generate retrieval coefficients using a modified stepwise regression procedure, giving a solution of the form

$$C_c = T_c R_c'(R_c R_c)^{-1} \tag{21}$$

where an element of T_c, t_c is the difference between the observed value t and the mean value, $\bar{t}_c$, for the selected class, and an element of R_c is the difference between the observed value r and the mean value, $\bar{r}_c$, for the selected class. In use, radiances are converted to brightness temperatures and then to eigenvector coefficients of the brightness temperatures. The eigenvector coefficients define a class for which the class means are known. A retrieval is then determined from the relationship

$$t = \bar{t}_c + C_c(r - r_c). \tag{22}$$

Results from the classification used by McMillin (1988 and 1986) are shown in Fig. (4). The classification used by McMillin (1986), resulted in a decrease of rms difference compared to radiosondes of 0.2 K over the operational result which varied from 1.63 to 2.33 K, depending on the particular level. Later, McMillin (1988) demonstrated an additional increase accuracy that resulted from the observation that the existence of a lapse or an inversion condition near the surface is correlated with the surface type and time of day. The profiles for nighttime over land were separated from the others on the expectation that the night cases over land would have nocturnal inversions. This step reduced the rms difference by another 0.2 K resulting in a total decrease in rms difference of about 0.4K. In practice, McMillin (1988a) recommended a separate classification based on channels sensing near the ground to detect surface inversions. The final retrieval would then be the result of two classifications, with the surface classification receiving a high weight near the ground and a decreasing weight with increasing height.

6. SUMMARY AND CONCLUSIONS

Retrievals based on linear departures from a climatological guess have been in use for some time. These retrievals are classified as being either physical approaches or regression approaches. Both have the characteristic of making smooth changes to an initial guess of the retrieved profile, but retaining in the retrieval the small scale vertical structure that is present in the initial guess. Because of the form of the equations, it is difficult to include meteorological knowledge, such as the adiabatic limit,

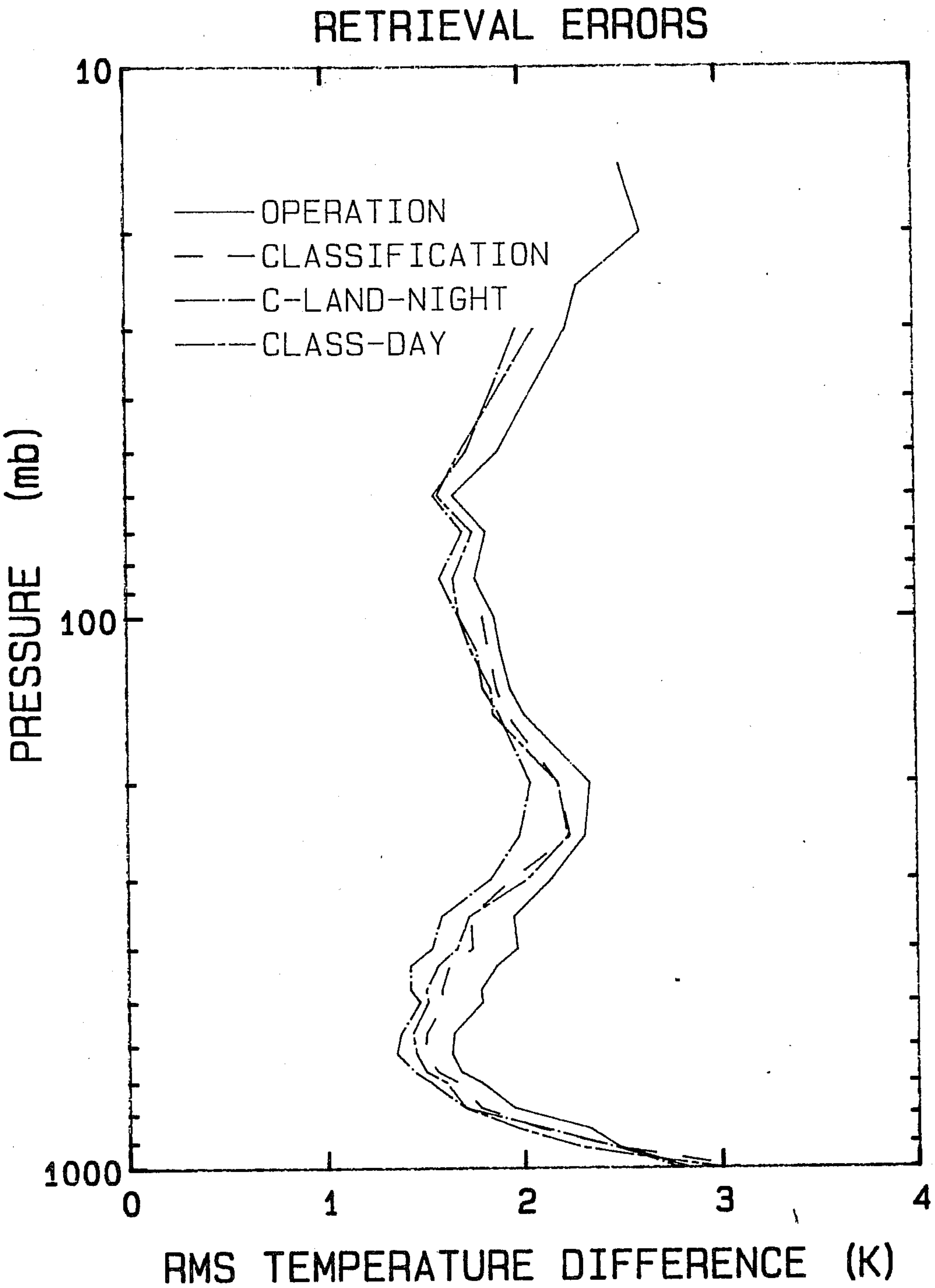

Figure 4. Root mean squared differences between retrievals and radiosondes for several retrieval algorithms. Operational retrievals are the standard TOVS retrievals produced by NESDIS. Classification retrievals are produced by the eigenvector classification method. C-LAND-NIGHT retrievals are produced by the same method on a sample which includes only soundings over land at night and thus are most likely to contain nocturnal inversions. CLASS-DAY soundings are produced by the same method on a sample consisting of day soundings over land and all soundings over oceans and thus are most likely to contain lapse conditions near the ground.

explicitly in to the retrieval equations. As a result, the physical solutions are physical only in the sense that they satisfy the physics of the radiative transfer equation. They do not at this time include meteorological constraints such as a limit on the lapse rate.

At the same time, it is clear that empirical approaches that seek to find a solution that satisfies the radiative transfer equation and belongs to the set of known atmospheric profiles, are not nearly as ill-conditioned as the approaches that depend entirely on linear departures from a climatological guess. Classification is one way to bring more meteorology into the retrieval process. Current studies show that improvements in the initial guess by using classification are capable of making substantial (0.3 to 0.4 K) improvements in the accuracy of current retrieval systems. Even after the new library search-physical retrieval is in place, the available data suggest an additional increase of 0.2K is possible by using the proven classification techniques. The best solutions are those given by obtaining a good initial guess followed by a good retrieval system. This indicates that there is information in the radiances that is not passed through the linear retrieval systems.

These findings must be considered when alternative first guesses are considered. For example, if a forecast is used as a first guess, then retrievals based on departures from the forecast radiances have a very large probability of including retrievals that never occur in the atmosphere. Going back to the philosophy that the best solution is the one that makes the best use of the most information, it seems that the best way to use forecast information is in combination with a classification approach. A retrieval based only on departures from a forecast is likely to ignore much of the available information.

Although classification is a way to incorporate additional meteorological knowledge into retrievals, it would be more satisfying to include meteorological physics directly into the retrieval process. Attempts to do so should be encouraged.

6.1 References

Davis, P.A., and J.D. Tarpley, 1983: Estimation of shelter temperatures from operational satellite sounder data. J. Climate Appl. Meteor., 22, 369-376.

Chedin, A., N.A. Scott, C. Wahiche, and P. Moulinier, 1985: The improved inversion method: a high resolution physical method for temperature retrievals from satellites of the TIORS-N series. *J. Climate Appl. Meteor., 24,* 128-143.

Hayden, C.M., 1988: GOES-VAS simultaneous temperature-moisture retrieval algorithm, *J. Appl. Meteor., 27,* 705-733.

Fleming, H.E., M.D. Goldberg, D.S. Crosby, 1986a: Minimum variance simultaneous retrieval of temperature and water vapor from satellite radiance measurements. *Second Conference on Satellite Meteorology/Remote Sensing and Applications.* Williamsburg, Va., 20-23.

Fleming, H.E., D.S. Crosby, and A.C. Neuendorffer, 1986b: Correction of satellite temperature retrieval errors due to errors in atmospheric transmittances. *J. Climate Appl. Meteor., 25,* 869-882.

Fleming, H.E., M.D. Goldberg, and D.S. Crosby, 1988: Operational implementation of the minimum variance simultaneous retrieval method. *Third Conference on Satellite Meteorology and Oceanography.* Anaheim, Calif., 16-19.

Gal-Chen, T., 1988: A theory for the retrievals of virtual temperature from remote measurements of horizontal winds and thermal radiation. *Mon. Wea. Rev., 116,* 1302-1319.

Gammon, R.H., E.T. Sundquist, and P.J. Fraser, 1985: History of carbon dioxide in the atmosphere. *Atmospheric Carbon Dioxide and the Global Carbon Cycle.* 27-62.

Gergen, J.L., 1957: Atmospheric infrared radiation over Minneapolis to 30 millibars. *J. Meteor., 14,* 495-504.

Goldberg, M.D., A. Swaroop, H.E. Fleming, and A. Reale, 1986: An analysis of temperature and water vapor retrievals using a minimum variance simultaneous method. *Second Conference on Satellite Meteorology/Remote Sensing and Applications.* Williamsburg, Va., 24-29.

Goldberg, M.D., J.M. Daniels, and H.E. Fleming, 1988: A method for obtaining an improved initial approximation for the temperature/moisture retrieval problem. *Third Conference on Satellite Meteorology and Oceanography.* Anaheim, Calif., 20-23.

McMillin, L.M. and C. Dean, 1982: Evaluation of a new operational technique for producing clear radiances. *J. App. Meteor., 21,* 1005-1014.

McMillin, L.M., 1986: The use of classification procedures to improve the accuracy of satellite soundings of temperature and moisture. *Second Conference on Satellite Meteorology/Remote Sensing and Applications.* Williamsburg, Va., 1-4.

McMillin, L.M., 1988a: Retrieval of air surface temperatures over oceans from satellite radiance measurements using stratification techniques. *Third Conference on Satellite Meteorology and Oceanography.* Anaheim, Calif., 10-12.

McMillin, L.M., M.E. Gelman, A. Sanyal, and M. Sylva, 1988b: A method for the use of satellite retrievals as a transfer standard to determine systematic radiosonde errors. *Mon. Wea. Rev. 116,* 1091-1102.

McMillin, L.M., L.J. Crone, and D.S. Crosby, 1988c: Adjusting satellite radiances by regression with an orthogonal transformation to a prior estimate. in preparation.

Ney, E.P., R.W. Maas, and W.F. Huch, 1961: The measurement of atmospheric temperature. *J. Meteor., 18,* 60-80.

Phillips, N.A., 1957: A coordinate system having some special advantages for numerical forecasting. *J. Meteor., 14,* 184-185.

Phillips, N., L. McMillin, A. Gruber, and D. Wark, 1979: An evaluation of early operational temperature soundings from TIROS-N. *Bull. Am. Meteor. Soc., 60,* 1188-1197.

Schmidlin, F.J., J.K. Luers, and P.D. Huffman, 1986: Preliminary estimates of radiosonde thermistor errors. *NASA Technical Paper 2637,* 15pp.

Schmidlin, F.J., 1988: Private communication. Goddard Space Flight Center, Wallops Flight Facility, Wallops Island, Va.

Smith, W.L., 1976: The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiometer observations. *J. Atmos. Sci., 33,* 1127-1140.

Smith, W.L., H.M. Woolf, C.M. Hayden, D.Q. Wark and L.M. McMillin, 1979: The TIROS-N operational vertical sounder. *Bull. Am. Meteor. Soc., 60,* 1177-1187.

Smith, W.L., and H.M.Woolf, 1984: Improved vertical soundings from amalgamation of polar and geostationary radiance observations. *Preprint Volume: Conference on Satellite Meteorology/Remote Sensing and Applications,* June 25-29, 1984, Clearwater, Florida.

Susskind, J., J. Rosenfield, and D. Reuter, 1983: An accurate radiative transfer model for use in the direct inversion of HIRS2 and MSU temperature sounding data. *J. Geophys. Res., 88C,* 8550-8568.

Thompson, O.E., M.D. Goldberg, and D.A. Dalzlich, 1985: Pattern recognition in the satellite temperature problem. *J. Climate & Appl. Meteor., 24,* 30-48.

Thompson, O.E., D.D. Daslich, and Yu-Tai Hou, 1986: The ill-posed nature of the satellite temperature retrieval problem and the limits of retrievability. *J. Atmos. Oceanic Technol., 3,* 643-649.

Uddstrom, M.J., 1988a: Retrieval of atmospheric profiles from satellite radiance data by typical shape function maximum a posteriori simultaneous retrieval estimators. *J. Appl. Meteor., 27,* 515-549.

Uddstrom, M.J., 1988b: The effect of collocation radiosonde errors on the assessment of the performance of a physical retrieval estimator. *The Technical Proceedings of the Fourth International TOVS Study Conference.* Cooperative Institute for Meteorological Studies, Space Science and Engineering Center, 1225 West Dayton Street, Madison, Wisconsin, 53706.

Weinreb, M.P., 1979: Atmospheric transmission for remote temperature sounding. *SPIE Vol. 195 Atmospheric Effects on Radiative Transfer(1979).*, 22-30.