A RETRIEVAL/FORECAST INTERACTIVE SYSTEM

Henry E. Fleming
NOAA/NESDIS/Satellite Research Laboratory
Washington, D.C., U.S.A.

Jaime M. Daniels and Mitchell D. Goldberg
ST Systems Corporation (STX)
Lanham, MD, U.S.A.

Bert Katz
Centel Federal Services Corporation
Reston, VA, U.S.A.

Wayman E. Baker and Dennis G. Deaven
NOAA/NWS/NMC/Development Division
Washington, D.C., U.S.A.

1. INTRODUCTION

A joint NESDIS/NMC research effort to build a retrieval/forecast interactive system recently was begun . Specifically, this system is a satellite retrieval/forecast model interactive assimilation system in which the NMC 6-hour forecast temperature and moisture data are used as the initial (first guess) profile by the satellite retrieval algorithm. This system is interactive in the sense that the output of the retrieval algorithm is assimilated by the forecast model for its initialization and the forecast of the model is assimilated in turn by the next cycle of retrievals. The purpose in doing this is to improve the accuracy of the satellite retrievals since they are inherently first guess dependent. Of course, any improvement in retrieval accuracy translates directly into improvement in forecast accuracy as well because of the interactive nature of the system.

Since the research effort is new, only the preliminary work has been completed. The purpose of this paper, therefore, is to report our early results and to discuss our plans for future work. The first topic discussed is the pressure slicing of the retrieved profile as compared to typical slicings of the NMC forecast model. They are incompatible near the surface and in the stratosphere, resulting an important problem of differing vertical resolutions between the two systems.

The retrieval algorithm requires not only a first guess temperature/moisture profile, but also a first guess radiance vector. Because the radiance first guess must be calculated from the forecast first guess temperature/moisture profile, the next three topics discussed concern the all-important forward problem, the method used to compensate for errors resulting from incomplete knowledge of the radiative transfer physics and mislocations between the radiosonde and the satellite measurements, and results from correcting the forward calculations. Next, results from a simplified version of our interactive system are discussed. Finally, our plans for a fully interactive system are outlined.

2. PRESSURE SLICINGS

Sigma coordinates are used for the vertical coordinate sytem of the forecast model. The forecast temperature/moisture data of the model are available to the retrieval algorithm only in the form of mean quantities over sigma layers. Conversion of these data to the pressure coordinate system is by way of the average pressure of the layer, which results in 18-element vectors of layer mean pressures that are quite variable globally and seasonally. However, if a crude average of these vectors of mean pressure is taken, the average pressure slicings shown in Fig. 1 are obtained.

Figure 1 is divided into two pressure columns with the essentially-tropospheric pressures on the left and the statospheric pressures on the right. The solid horizontal bars indicate the location on the pressure scale of the 18 pressures of the average of a sampling of forecast model layer mean pressure vectors. On the other hand, the dashed horizontal bars in Fig. 1 are the 18 pressures that are optimal, relative to the the retrieval algorithm, within the average 1013 to 20 hpa range of the model.

The retrieval pressure slicing is optimal in the following sense. A mathematical transformation of the pressure variable is chosen such that the peaks of the weighting function curves (i.e., the kernel functions) of the radiative transfer equation are more or less equally distributed along the vertical extent of that part of the atmosphere to which the satellite measurements are sensitive. This transformation is pressure to the two-sevenths power. Application of the transformation to the weighting function curves also causes them to have roughly equal half widths. The dashed lines in Fig. 1 are the mean pressures of the 18 equally-spaced layers, relative to the pressure to the power 2/7 slicing, in the 1000 to 20 hpa inter-

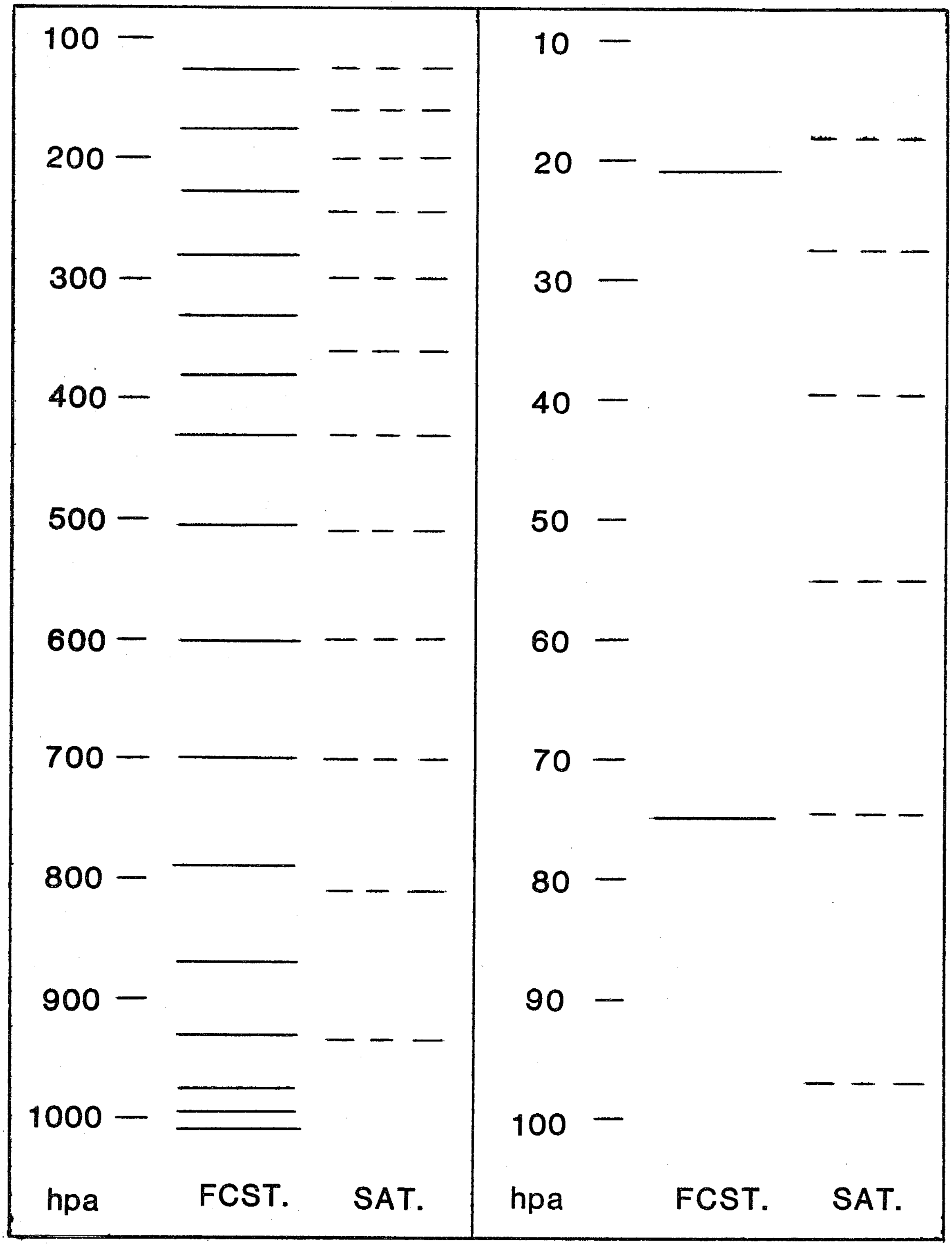

Fig. 1 Comparison of the pressure slicings for the forecast (solid lines) and for the satellite retrievals (dashed lines).

val.

Inspection of Fig. 1 yields the following observations: 1) the mean pressure points of the forecast model are clustered at the bottom of the atmosphere, particularly in the boundry layer, while those of the retrieval are very sparse there; 2) the densities of the two sets of pressure points are reversed in the stratosphere from what they are at the bottom of the troposphere ; and 3) the two sets of pressure slicings are very similar in the middle and upper troposphere. The slicings at the bottom of the atmosphere reinforce our knowledge of that region, namely that the satellite soundings have low vertical resolution there while the model resolves that region the best. Thus, the satellite retrievals should be helped there by the forecast first guess, while the reverse should be true in the stratosphere. On the other hand, the two systems complement each other in the middle and upper troposphere.

3. THE FORWARD PROBLEM

The term "forward problem" means that one applies a temperature profile to the Planck function (i.e., the source function) of the radiative transfer equation and performs the integration to determine a vector of radiances. It is a forward problem in contradistinction to an inverse problem in which one determines the source function from knowledge of the radiances. Note that the radiative transfer equation is nonlinear in that one also applies the temperature profile, along with the associated moisture profile, to the weighting functions.

The first step in our interactive research was to apply a global set of forecast temperature/moisture profiles for November 30, 1988, to the forward problem to produce a corresponding global set of radiances for the NOAA-10 TOVS instruments. To do this we interpolated the forecast to the locations of satellite measurements, giving us accurate spatial coincidence. However, time coincidence was no better than plus or minus three hours in that no time interpolation was used.

The purpose of the forward problem exercise was to compare the measured and calculated radiances, and in doing so we were attempting to answer two questions. First, is it true (as we expect it is) that the vast majority of the time the forecast is sufficiently accurate that the satellite data contribute little to what already is known from the forecast?. Are there places where there is substantial disagreement between the measured radiances and those calculated from the forecast, and if so, are they large contiguous areas that show definite patterns, or are they ill-defined and random?

The results from the experiment were displayed as color-coded global maps of differences between the measured and calculated brightness temperatures (not radiances). These maps will be shown via viewgraphs at the Workshop, but they cannot be reproduced for this paper; therefore, only the results are discussed here.

We compare brightness temperature difference maps of HIRS Channels 5 (longwave IR) and 15 (shortwave IR) and MSU Channel 2 (microwave), which all sense roughly around 500 hpa and, therefore should look quite similar. In fact, however, all three maps are different. The MSU Channel 2 map is the only one that appears as one would expect; about 80% of the map indicates temperature differences of less than plus or minus 1^o K with the rest of the map showing larger differences, both warm and cold, in well-defined areas.

On the other hand, the map for HIRS Channel 5 shows areas covering close to 50% of the globe in which the temperature differences exceed plus or minus 1^o K and these differences (with very few exceptions) have the forecast brightness temperatures colder than the satellite-observed ones. Finally, the Channel 15 map has fewer differences that exceed 1^o K than does the Channel 5 map, but the number of such areas is still excessive. However, the more important discrepancy is that the larger temperature differences for Channel 15 are mainly of one sign and in the opposite sense from those of the Channel 5 map. This exercise clearly reveals deficiences in the forward problem calculations.

4. CORRECTIONS TO THE FORWARD PROBLEM

To correct the deficiencies in the forward problem calculations, we use a regression procedure in which the measured brightness temperatures are predicted from the calculated ones. However, because each calculated brightness temperature that is being corrected receives its radiation from a unique portion of the atmosphere, by far the most important information about the correction should come from the measured brightness temperature of the same channel, since it receives its radiation from the same unique portion of the atmosphere. Its neighboring channels should be the next most important predictors, but with rapidly decreasing importance.

These underlying assumptions suggest that the "shrinkage estimator" procedure of Oman (1982) is more appropriate for our problem than is standard regression. Shrinkage estimation is a regression procedure based on the fact that if one has some prior knowledge of the regression coefficients, then it is more reasonable to shrink the estimator towards its projection onto this new subspace rather than towards the origin as standard regression does. Note that an alternative approach, but one quite similar to the shrinkage estimator procedure, is provided by the rotated coordinate regression method of McMillin, et al. (1989).

The motivation for using the shrinkage estimator is that the matrix of regression coefficients should have numbers very close to unity down the diagonal and that the off-diagonal terms in each row should fall off very rapidly in value toward zero. (Of course this assumes the channels have been sequentially ordered according to their vertical response in the atmosphere.) This is accomplished by constraining the regression coefficients matrix rows to unit basis vectors.

We now derive the version of the shrinkage estimator that we use. Let n be the size of our sample data sets of calculated and measured brightness temperatures, let m be the number of channels being corrected (i.e., number of predictands), and let p be the number of channels being used as predictors. Then m <= p. Furthermore, let C be the m x p matrix of regression coefficients, let S be the n x p sample matrix of simulated (i.e., calculated) brightness temperatures being used as predictors, and let M be the n x m sample matrix of measured brightness temperatures. The data in matrices M and S are assumed to be centered, i.e., the matrix elements are the deviations of the data from their sample means. Finally, let m_i and c_i be the ith columns of matrices M and C, respectively, i.e., they are vectors of dimension m and p, respectively.

The shrinkage estimation problem now reduces to that of minimizing the following penalty function f with respect to c_i:

$$f(c_i) = (Sc_i - m_i)^T(Sc_i - m_i) + \alpha(c_i - u_i)^T(c_i - u_i) \quad (1)$$

where the superscript T indicates vector transposition, α is a Lagrangian multiplier whose value (to be determined later) is chosen so as to stabilizes the solution, and u_i is the unit basis vector (with unit value in the ith position and zeros elsewere), which serves as the constraint vector.

The minimizer of f is

$$c_i = (S^TS + \alpha I)^{-1} (S^Tm_i + \alpha u_i) \quad (2)$$

where I is the m x m identity matrix. Since all the vectors in

(2) are column vectors, we can stack them from left to right and convert (2) to the matrix equation

$$C = (S^T S + \alpha I)^{-1} (S^T M + \alpha J) \qquad (3)$$

where J is the p x m rectangular matrix (recall that m <= p) with the upper m rows of the matrix being the identity matrix I and the remaining lower p-m rows being all zeros. Note that the mathematical structure of (3) allows one to use additional predictors that are quantities other than just brightness temperatures. The additional predictors can include quantities such as latitude and longitude, season, satellite scan angle, solar zenith angle, etc., provided scaling is used.

To better understand the implications of solution (3), consider the following limiting cases. In the limit as α approaches zero, we have

$$\lim_{\alpha \to 0} C = (S^T S)^{-1} (S^T M) \qquad (4)$$

which is the standard least squares solution. At the other extreme, as α approaches infinity, we have

$$\lim_{\alpha \to \infty} C = \lim_{\alpha \to \infty} [(1/\alpha) I] (\alpha I) = I \qquad (5)$$

which says that the predicted brightness temperatures are equal to the predictor temperatures. The coefficient matrix solution (3) clearly is a blend of these two extreme properties.

With the coefficient matrix C in hand, one obtains a vector of corrected brightness temperatures $\hat{m}$ from a vector of simulated brightness temperatures s by the regression equation

$$\hat{m} = \bar{m} + C^T (s - \bar{s}) \qquad (6)$$

where $\bar{m}$ and $\bar{s}$ are the sample mean vectors. Thus, (3) and (6) are the devices by which we make the calculated, or simulated, brightness temperatures look like measured ones.

5. RESULTS FROM THE CORRECTED FORWARD PROBLEM

The forward problem corrections of the previous section were applied to the brightness temperatures simulated from the November 30, 1988, data set described in Section 3. The coefficient matrix C was calculated from the initial analysis field, valid six hours prior to the forecast time, using the same interpolation scheme as was used for the forecast. This matrix and the simulated brightness temperatures of November 30, 1988, were applied to (6). The the resulting corrected brightness temperatures were subtracted from the corresponding measured ones and the differences mapped globally as before.

The purpose in comparing the measured brightness temperatures with simulated ones which are corrected so as to look like the measured ones, is again to attempt to answer the two questions asked in Section 3. Is the forecast sufficiently accurate the vast majority of the time that the satellite data contribute little to what already is known from the forecast? Are the places in which there is substantial disagreement between the measured radiances and those calculated from the forecast large contiguous areas that show definite patterns, or are they ill-defined and random?

Again, the maps of differences between the measured and corrected brightness temperatures will be shown via viewgraphs at the Worshop, but they cannot be reproduced for this paper; only the results are discussed here. This time when we compare brightness temperature difference maps of HIRS Channels 5 (longwave IR) and 15 (shortwave IR) and MSU Channel 2 (microwave), which all sense roughly around 500 hpa, the results are very different from those discussed in Section 3, except for Channel 2.

The corrected difference map for Channel 5 changed very dramatically from the original map in two respects. First, the total area in which the errors were less than plus or minus 1^{o} K increased from about 50% to much nearer 80%. Second, the regions for which the differences exceed 1^{o} K are now balanced so that there are almost equal total areas of warm and cold departures, as one would hope to find. The original discrepancies of Channel 15 where not as extreme as they were for Channel 5, and so one would not expect the changes to the map of corrected brightness temperature differences to be as extreme. This indeed is the case; nevertheless, the changes are precisely of the two kinds just described for Channel 5.

Now we consider how similar the corrected difference maps for Channels 2, 5, and 15 are to each other. The total area of brightness temperature differences exceeding 1^{o} K is the least for Channel 2 and is the largest for Channel 5. A few of the areas of large differences for Channel 2 are over cloudy areas where no tropospheric IR channels are used, suggesting difficulties with precipitation or excessive cloud liquid water in those areas. A few additional areas of large differences are over high ground where surface emissivity (which is not included in the calculations) affects the calculated Channel 2 brightness temperatures.

Channel 15 has more and larger areas of temperature differences exceeding 1^{o} K than does Channel 2, and the same situation exists for Channel 5, but even more so. These areas of large temperature differences coincide quite well for Channels 5 and 15, but coincide less well when compared to Channel 2. Overall, however, the three difference maps are reasonably similar when reasons for the disrepencies are taken into account. These reasons include the following differences among the three channels: 1) in how clouds are sensed, 2) in the size of the field of view of Channel 2 as compared to that of Channels 5 and 15, 3) in water vapor sensitivity, 4) in the accuracy of the forward problem computations, 5) in the nonlinear effects of the Planck function, 6) in the noise levels of the measured radiances, and most importantly, 7) in the shapes of the weighting functions, and even though they peak at roughly the same level in the atmosphere, their peak levels differ by as much as 50 hpa.

6. THE SEMI-INTERACTIVE SYSTEM

As the first step toward testing the interactive concept, a scaled-down version of the interactive system was run. The operational retrievals were used in this system, and the system was made semi-interactive by replacing the operational retrieval first guess temperature and moisture profiles with the forecast profiles. Also, the associated operational first guess radiance vector was replaced by one calculated from the forecast temperature and moisture profiles, and in addition, this calculated radiance vector was corrected in the manner described in Section 4.

The operational retrieval system is a minimum variance method which retrieves profiles of atmospheric temperature and moisture, along with surface temperature, simultaneously as a single solution vector. We call this approach the MVS method, which stands for minimum variance simultaneous method. In the operational mode the MVS first guess temperature and moisture profiles and associated first guess radiance vectors are obtained from an a priori library of coincident radiosonde temperature/moisture profiles and satellite radiance measurements. The library is searched for the best agreement between the radiance vectors in the library and the given, measured radiance vector. Then the 20 most closely matched vectors from the library are averaged, as are the associated 20 temperature/moisture profiles, and the average radiance vector and the average simultaneous temperature/moisture profile are used as the operational first guesses. Further details concerning the operational MVS method can be found in Dey, et al. (1989), Fleming, et al. (1986) and (1988), and Goldberg, et al. (1988).

The semi-interactive system is not fully interactive in that the retrieved profiles, although based on the forecast, are not used in the initialization of the next forecast cycle. Nevertheless, the test of the semi-interactive system proved to be very informative.

The evaluation of the semi-interactive test system during the Workshop talk once again will be via viewgraphs of color-coded global maps, which cannot be reproduced for this paper. Therefore, only descriptions of the highlights of these maps are given here. This series of maps is of layer mean virtual temperature differences, whose layer boundries are determined by the mandatory pressure levels. In the evaluations to follow, results over land areas are not considered because of large diurnal heating and cooling discrepancies at and near the surface due to the six-hour time window being used in matching the satellite and forecast data.

The first set of maps is of the mean virtual temperature differences between the interactive retrievals and the forecast, with separate maps for the layers 200-300 hpa, 500-700 hpa, 700-850 hpa, and 850-1000 hpa. The objective is to determine the degree to which the retrieval changes the forecast. The areas of disagreement (i.e., exceeding plus or minus 1^{o} C) over the oceans are small and limited in number , with the most disagreement occurring in the lowest 850-1000 hpa layer. This again shows that the forecasts are generally very accurate, but also that when necessary, the retrievals based on the forecast can make changes away from the forecast. Furthermore, these changes are significant in that the the few major ones that occur can be traced through more than one layer for which we have maps.

The next objective is to determine the accuracy of the few retrievals that moved away from the forecast by checking maps of the same layers, but of the mean virtual temperature differences between the interactive retrievals and the verifying analysis maps, valid at retrieval time. Unfortunately, the majority of the areas in which the retrieval changed the forecast, that change moved away from the verifying analysis as well. However, we also produced a map showing sky conditions which has different color codes for each of the following conditions: 1) clear areas, 2) partly cloudy areas where we are able to convert the cloud-contaminated radiances to equivalent cloud-free radiances, and 3) cloudy areas, where only microwave channels can be used to do the retrievals in the troposphere, and so the retrieval accuracy is degraded. It turns out that the areas where the retrievals differ from the verifying analysis are almost always cloudy, confirming the known degradation of the cloudy retrievals.

On the other hand, in the clear and partly cloudy areas over the oceans in which the retrieval changed the forecast first guess, the retrievals moved in the right direction. Furthermore, the interactive retrievals generally agreed with the verifying analysis to better than plus or minus 1^{o} C, even though the they disagreed with the forecast by more than 2^{o} C. This development bodes well for the interactive concept.

Comparing semi-interactive retrievals with the operational ones normally would also be of interest; however, over the open oceans the verifying analysis maps draw virtually all of their ground truth information from the operational retrievals. Consequently, this exercise is pointless because the operational retrievals are bound to look very good there. However, on the November 30, 1988 maps, there is a large area encompassing the East and South China Seas and the Philippine Sea that has numerous radiosonde stations which determine the ground truth, rather than the satellite data. The largest operational retrieval errors on the entire map occur in this area in the 700-850 hpa layer, and to a slightly lesser degree in the two adjacent layers. On the other hand, the semi-interactive retrievals show very good agreement with the ground truth, being less than 1^{o} C almost everywhere in that area in all three layers. This gives another indication of the potential positive impact of the full interactive system.

7. PLANS FOR THE FULL INTERACTIVE SYSTEM

Retrieved satellite products have always been made to look like radiosonde measurements. A significant feature of our planned full interactive system is that the retrieved products are tailored to the specific needs of the forecast model. For example, mean thicknesses and specfic humidities over the sigma layers of the forecast model are retrieved directly instead of retrieving temperatures and moisture at fixed pressure points and converting them to the desired quantities. This also means that the retrievals are allowed to terminate at a variable surface pressure. This differs from our present procedure of terminating our retrievals at a fixed surface pressure of 1000 hpa.

The full interactive system also allows us to improve the MVS retrieval system. The problem of the nonlinearity of the retrieval operator is largely circumvented by generating a new operator at each retrieval location based on the forecast parameters at that location. Also, the solution covariance matrix used by the operational retrieval operator is replaced by the expected forecast error covariance matrix, call it E, which is readily available from the optimum interpolation analysis.

Finally, MVS retrieval algorithm has to undergo some fundamental changes. First the regression coefficient matrix C of Section 4 is derived with the role of matrices M and S reversed in (1) through (3) and in (6). In other words, the measured radiances are made to look like calculated ones, rather than having the calculated radiances look like measured ones as required by our maps. Then using the mean vectors $\bar{m}$ and $\bar{s}$ of (6), we write the interactive, simultaneous temperature/moisture solution vector $\hat{t}$ in the form:

$$\hat{t} = f + E A^T (A E A^T + N)^{-1} [C^T (m - \bar{m}) - (s - \bar{s})] \quad (7)$$

where f is the first guess forecast temperature/moisture vector, m is the measured radiance vector, A is the matrix of weighting functions, s is the radiance vector calculated from f (i.e., s = A f), E is error covariance matrix discussed in the previous paragraph, and N is the expected error covariance of the fit of the vector $s\hat{}$ of (6) [with the roles of m and s reversed] to s. Note that the error covariance of the vector m is implicit in N.

8. CONCLUSIONS

We consider the procedure for correcting radiances simulated from the forecast to be effective and necessary for the success of the full interactive retrieval algorithm. The discussions of Section 5 verify this. Furthermore, the results from the semi-interactive system described in Section 6 are sufficiently promising that we are proceeding to build the full interactive system, according to the plan of Section 7, as a joint effort between NMC and NESDIS.

9. REFERENCES

Dey, C., R. Petersen, B. Ballish, P. Caplan, G. White, H. Fleming, A. Reale, D. Gray, M. Goldberg and J. Daniels, 1989: An examination of NESDIS TOVS physical retrievals using data impact studies. NMC Tech.Rep., to be published.

Fleming, H., M. Goldberg and D. Crosby, 1986: Minimum variance simultaneous retrieval of temperature and water vapor from satellite radiance measurements. Preprints, Second conf. on Satellite Meteorology/Remote Sensing and Applications, Amer. Meteor. Soc., Boston, 20-23.

Fleming, H., M. Goldberg and D. Crosby, 1988: Operational implementation of the minimum variance simultaneous retrieval method. Preprints, Third Conf. on Satellite Meteorology and Oceanography, Amer. Meteor. Soc., Boston, 16-19.

Goldberg, M., J. Daniels and H. Fleming, 1988: A method for obtaining an improved initial approximation for the temperature/moisture retrieval problem. Preprints, Third Conf. on Satellite Meteorology and Oceanography, Amer. Meteor. Soc., Boston, 20-23.

McMillin, L., L. Crone and D. Crosby, 1989: Adjusting satellite radiances by regression with an orthogonal transformation to a prior estimate. J.Appl.Meteor.,28, (to appear).

Oman, S., 1982: Shrinking towards subspaces in multiple linear regression. Technometrics,24,307-311.