SYNTHETIC AND REAL DATA APPLICATIONS OF THE TYPICAL SHAPE FUNCTION SIMULTANEOUS PHYSICAL RETRIEVAL ALGORITHM

Michael J. Uddstrom

New Zealand Meteorological Service
Wellington, New Zealand

1. INTRODUCTION

Recently a number of important advances have been made in the specification of linear measurement models for the radiative transfer equation (Smith and Woolf, 1984; Smith *et al.* 1985; Fleming *et al.* 1986). However, because the retrieval problem is fundamentally ill-posed, additional ancillary data must always be supplied to retrieval estimators. The problem of the specification of the *a priori* first guess profile and retrieval constraints has been addressed by a number of authors, including Westwater *et al.* (1983), Uddstrom and Wark (1985), Chedin *et al.* (1985), Thompson *et al.* (1985), Wark (1986) and McMillin (1986). Each of these papers suggest different methods for determining first guess profiles, and in some cases retrieval constraints, which are independent of numerical weather prediction models, yet appropriate to the sounded atmosphere.

The purpose of this paper is to outline the characteristics of a physical simultaneous retrieval algorithm which utilises the method of Typical Shape Function (TSF) classification (Uddstrom and Wark, 1985; henceforth abbreviated UW). The advances made in both the selection of the first guess profile and constraints, and measurement specification models, are incorporated in this new algorithm. The essential element of TSF classification is that it enables the development of a non-linear retrieval scheme, without recourse to relaxation or iterative techniques. Instead, local domains in both radiance and retrieval parameter space are defined, in which the constituent atmospheres are in some sense linearly related. Within each TSF domain, linear retrieval methods may be applied without the usual consequences associated with the application of linear perturbation theory to strongly non-linear problems. In addition to providing a good estimate of the first guess profile, the TSF classification method also provides covariance constraints which define the "shape" of each TSF atmospheric class. This additional information is important and may be utilized in an appropriately specified retrieval algorithm.

The full TSF retrieval algorithm, which includes *a priori* temperature classification, radiance discrimination, cloud detection and correction, and simultaneous maximum *a posteriori* retrieval estimators, is applied to synthetic data and two NOAA 7 passes over the New Zealand region in July 1984. The TSF retrieval results are subjectively compared with operational NWP analyses, and output from the New Zealand Meteorological Service's operational retrieval scheme, which is a version of the NOAA NESDIS Development Laboratory - Cooperative Institute for Meteorological Satellite Studies (CIMSS) "Export" regression retrieval software.

Complete details of the TSF simultaneous retrieval algorithm are given in Uddstrom (1986).

2. TYPICAL SHAPE FUNCTION CLASSIFICATION

The theoretical basis of the TSF classification algorithm is described in UW. Here the TSF classification scheme has been applied to a Southern Hemisphere sample (including data from the Indian Ocean, Australian, Pacific, New Zealand and Antarctic regions) of 2181 July radiosonde profiles from the years 1981 to 1983. The classification region is defined by the temperatures at the 16 pressure levels bounded by the 50hPa and 500hPa levels of the NESDIS TOVS pressure coordinate system.

Although the the TSF classification algorithm is an unsupervised classifier, the number of classes or patterns detected in the *a priori* data must be specified, since the amount of structure resolved in the tropopause region is ultimately constrained by the indeterminancy of the radiance discrimination problem. Here 13 TSF classes are selected from the July data. Some characteristics of these TSF class samples are listed in Table 1 and the mean temperature profiles for all 13 classes are plotted in Fig 1.

3. RADIANCE DISCRIMINATION

A Bayesian discriminant function is used to specify the TSF *a priori* class constraints for a retrieval, given the observed radiance measurements (UW). For this function the ith discriminant length measure $g_i(\mathbf{r})$ is;

$$g_i(\mathbf{r}) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{S}_{r_i}| + \log(P(\omega_i)) - \frac{1}{2}(\mathbf{r} - \bar{\mathbf{r}}_i)^T\mathbf{S}_{r_i}^{-1}(\mathbf{r} - \bar{\mathbf{r}}_i) \qquad \ldots(1)$$

where $\mathbf{r}$ is the vector of discriminator channel measurements (dimension d), $P(\omega_i)$ the prior probability for the ith class, $\bar{\mathbf{r}}_i$ the mean radiance profile of the ith TSF class and $\mathbf{S}_{ri}$ the associated covariance matrix. The measurement errors are assumed to be unbiased. The posterior probability for each class, in terms of the discriminant function $g_i(\mathbf{r})$ is just;

$$p(\omega_i|\mathbf{r}) = \exp(g_i(\mathbf{r})) \Big/ \sum_{j=1}^{N} \exp(g_j(\mathbf{r})) \qquad \ldots(2)$$

for N TSF classes.

Radiance discrimination equations utilizing channels 1, 2, 4, 23, 5 and 6 have been applied to an independent sample of 500 noise-contaminated (as specified in Table 7) July profiles of known TSF class. Posterior probability distributions for two classes of correctly discriminated radiance measurements are displayed in Fig 2. In the case of class 34, which has a TSF quite different from all remaining TSFs, the posterior probability distribution is narrow with a mean close to 1.0. Conversely, in the case of the class 11 profiles, for which other classes have similar TSFs, the posterior probability distribution is rather broader and has a mean value somewhat less than 1.0. However, in both cases, the mean of the posterior distributions is very much larger than the prior probabilities, which are of the order of 1/13 (see Table 1).

The consequences of misclassifications can be addressed by considering the distributions of the radiance discriminated classifications for all the profiles belonging to some particular class. Included in Fig 2, for both of the classes for which posterior probability distributions are plotted, are histograms of the discriminated class populations for the respective TSF class samples. When the TSF class in question is not similar to any of the other classes, as in the case of class 34, there are few misclassifications. For those TSF classes which are related to other classes in that the differences in their TSFs are rather subtle, misclassifications do occur, as is evident from the class 11 results. None of these misclassifications should lead to serious retrieval errors.

The accuracy of the radiance discriminant inverse mapping function is dependent upon the number of *a priori* TSF classes, their differing vertical features and expected covariances, the vertical resolving power of the satellite radiometer measurement kernels, the measurement noise statistics, and the number and selection of channels used in formulating the discriminant equations. The discriminant function is

sensitive to measurement noise through the last term in equation (3), the Mahalanobis distance. This term compares the relationships between the various measurement channels, with the expected covariances of each TSF class. Small measurement errors can lead to large differences in the contribution from this term if the radiance covariance matrix is nearly singular. However, this noise sensitivity can be suppressed through the choice of discriminator channels which are reasonably vertically independent, since in this circumstance the radiance covariance matrix is better conditioned.

4. CLOUD DETECTION

The TSF class of an atmosphere may be determined with considerable skill from within the radiance domain of the observations by choosing a discriminant function which is well conditioned and does not make use of measurements that could be badly cloud contaminated. Combining the direct measurements with the *a priori* information enlarges the set of "measurements". Taken together, these measurements define the domain in which the observed atmosphere falls, and this information can be used in a cloud detection and clearing algorithm which does not invoke any assumptions about horizontal homogeneity in the temperature fields.

A linear regression model has been developed, whereby within a particular TSF class, those channels unaffected by clouds are used to predict the radiances of other channels which might be so affected. The model equations are derived from the TSF class radiance samples using an eigenvector expansion method. The predictors may be the stratospheric and high tropospheric HIRS channels, together with the three highest MSU channels, and the predictands are the lower tropospheric HIRS channels. Separate equations are derived for each TSF class, and for land and sea surfaces.

Scatter diagrams of the predictand results of this model, applied to the independent sample of 500 noise contaminated July profiles are given in Fig 3. The discriminant channels are 1, 2, 4, 23, 5 and 6 and the cloud detection predictor channels are 1, 2, 4, 23 and 22. Excluding channels 11 and 12, it is evident that the regression equations are of nearly uniform quality over the entire dynamic range of the independent sample, thus substantiating the assumption of linearity within each TSF domain. The largest errors occur for profiles at the low temperature end of the sample, and there are few outliers. Accordingly, once the TSF class of the observation has been determined, the presence of cloud contamination is easily detected. Cloud contaminated channels are declouded through substitution of the regression model predictand radiance temperatures for cloud affected measurements.

This TSF cloud detection and correction algorithm, like the psi method of Chedin and Scott (1983), has the advantage of not requiring horizontal homogeneity in paired temperature fields, and consequently should work equally well in baroclinic and barotropic situations.

5. RETRIEVAL ESTIMATOR SPECIFICATION

Formally, the linearized model of the measurement process is given by equation (3)

$$\mathbf{y} = \mathbf{A}\mathbf{x}' + \boldsymbol{\varepsilon} \qquad \ldots(3)$$

where

$$\mathbf{y} = \Delta\mathbf{T}^*$$

$$\mathbf{A} = [\mathbf{H}, \mathbf{K}, \mathbf{G}] \qquad \ldots(4)$$

$$\mathbf{x}' = [\Delta T_s, \Delta\mathbf{T}, \Delta\mathbf{U}]^T$$

and ΔT^* is the radiance temperature signal; **H**, **K** and **G** are respectively the surface temperature, temperature, and precipitable water kernels; and ΔT_s, ΔT and ΔU are the unknown perturbations of the surface temperature, temperature, and precipitable water profiles. Two different "simultaneous" measurement models for the radiative transfer equation have been developed. One is linear in terms of U(p), the precipitable water profile whilst, the other is linear in terms of log(U(p)) (see Uddstrom 1986). Both models can be used in the TSF retrieval estimator with the log(U(p)) model being more appropriate to specification of the water vapour channels and the U(p) model to the temperature sensing channels.

The TSF maximum *a posteriori* (henceforth MAP) retrieval estimator is formally identical to that of a simple (non-simultaneous) linearization model. The MAP estimator and covariance of the estimate are (from UW):-

$$\hat{\mathbf{x}}' = \mathbf{S}_{xk}\mathbf{A}^T (\mathbf{A}\mathbf{S}_{xk}\mathbf{A}^T + \mathbf{S}_e)^{-1}\mathbf{y}' \qquad ...(5)$$

$$\hat{\mathbf{S}}_{xk} = \mathbf{S}_{xk} - \mathbf{S}_{xk}\mathbf{A}^T (\mathbf{A}\mathbf{S}_{xk}\mathbf{A}^T + \mathbf{S}_e)^{-1} \mathbf{A}\mathbf{S}_{xk} \qquad ...(6)$$

where k refers to the TSF class and $\mathbf{S}_{xk}$ is the measurement error covariance matrix. However under the simultaneous specification, the constraint matrices $\mathbf{S}_{xk}$ include both covariances and cross-covariances for all three data types T_s, T(p) and U(p), i.e.,

$$\mathbf{S}_{xk} = E\left\{ \begin{bmatrix} \Delta T_s^T \Delta T_s & & \\ \Delta T^T \Delta T_s & \Delta T^T \Delta T & \\ \Delta U^T \Delta T_s & \Delta U^T \Delta T & \Delta U^T \Delta U \end{bmatrix} \right\} \qquad ...(7)$$

where E{.} is the expected value operator, and only the lower triangular matrix values are specified. In this way all *a priori* information is used in the retrieval procedure, both directly through the *a priori* means and their covariances, and indirectly via the cross-covariance relationships.

In practice, the MAP estimator is utilized in a sequential mode and the retrieval estimate and error covariance matrix of the estimate are updated as each new measurement is added. With this method, troublesome collinearities in either the measurements or the constraint matrices may be identified. The MAP retrieval equations in sequential form are:-

$$\hat{\mathbf{x}}_i' = \hat{\mathbf{S}}_{xk_{i-1}} \mathbf{a}_i^T ((\, y_i - \mathbf{a}_i \mathbf{x}_{i-1}\,) / (\, \mathbf{a}_i \hat{\mathbf{S}}_{xk_{i-1}} \mathbf{a}_i^T + \sigma_i^2\,)) \qquad ...(8)$$

$$\hat{\mathbf{S}}_{xk_i} = \hat{\mathbf{S}}_{xk_{i-1}} - (\, \hat{\mathbf{S}}_{xk_{i-1}} \mathbf{a}_i^T \mathbf{a}_i \hat{\mathbf{S}}_{xk_{i-1}}\,) / (\, \mathbf{a}_i \hat{\mathbf{S}}_{xk_{i-1}} \mathbf{a}_i^T + \sigma_i^2\,) \qquad ...(9)$$

where $\mathbf{x}_{k0}$ and $\mathbf{S}_{xk0}$ are the *a priori* TSF sample mean and constraint matrices. The susceptibility of these equations to data collinearities, measured or *a priori*, is determined by the bracketed term of equation (8). The radiance collinearities result because the measurements are not independent, but the *a priori* collinearities are a result of the dynamics which govern the atmosphere; the constraint covariances are non-zero. When the denominator of equation (8) is small relative to the signal the estimator equations may be nearly singular and the resulting retrieval unstable. With a sequential estimator this condition is easily detected. Should an instability arise, a smoothing parameter is applied to the estimator in order to adaptively condition the equations and reduce the effects of collinearities. The resulting retrieval is, of course, biased.

Another advantage of the sequential estimator relates to solving problems which while specified by linear models are known *a priori* to have a non-linear character, e.g., the interaction between the retrieved water vapour distribution and the temperature profile in the lower troposphere. In utilizing the retrieval estimator in a sequential mode it is possible to choose the order in which the observations are applied, with the most non-linear being used first. The resulting modified guess profiles are better approximations of the unknown profiles, and the more nearly linear observations may be used within the perturbation domains in which they are most applicable.

A full description of the linearized measurement model of the radiative transfer equation and specification of the simultaneous retrieval equations used here is given in Uddstrom (1986).

6. RETRIEVAL VALIDATION

Finally, the performance of the retrieval algorithm is characterised by examining the inverse mapping from measurement to parameter space for a set of measurements computed synthetically via a forward mapping from parameter to measurement space. Here the forward mapping function is the radiative transfer equation, and the inverse mapping function the radiance discrimination function followed by the TSF simultaneous, adaptively conditioned, MAP sequential retrieval estimator.

The sample of independent data on which this validation analysis is carried out is identical to that used to test the radiance discrimination and cloud detection algorithms. The accuracy of the retrieved profile samples (P profiles) will be represented by two parameters, namely the vertically averaged rms "error" (armse) over some specified height range defined by q layers (equation (10)),

$$\text{armse} = \frac{1}{q} \sum_{i=1}^{q} \sqrt{\frac{1}{P} \sum_{j=1}^{P} (\hat{x}_{ij} - x_{ij})^2} \qquad \ldots(10)$$

and the associated vertically averaged coefficient of determination, aR^2, (equation (11)),

$$aR^2 = \frac{1}{q} \sum_{i=1}^{q} (1 - \sum_{j=1}^{P} (\hat{x}_{ij} - x_{ij})^2 / \sum_{j=1}^{P} (x_{ij} - \overline{x}_{ij})^2) \qquad \ldots(11)$$

(Kvalseth, 1985).

Noise contaminated synthetic measurements are used to indicate the susceptibility of the TSF retrieval algorithm to random measurement noise, but measurements from a "perfect" radiometer are employed to examine those features of the retrieval scheme which are not too sensitive to data noise. Some examples are different measurement model specification equations, the selection of the radiometer measurements used and the order in which they are applied, and errors arising from the radiance discrimination scheme.

Firstly consider retrieval results for "measurements" from a perfect radiometer. Complex multi-pass retrieval algorithms are possible since there is no requirement for either the measurement model or the channel selection to remain static between passes. Table 2 lists the retrieval error statistics for a selection of TSF MAP estimators.

Clearly, multi-pass estimators yield more accurate retrievals than single pass methods. Also, the advantages of using both linear and logarithmic precipitable water measurement models is demonstrated in the results labelled EN 3 and EN 4 (see Table 2). The largest improvements are evident in the water vapour retrievals. Addition of channel 11 measurements in the log(U(p)) model pass further improves the accuracy of the retrieval estimator. Comparing the error statistics for EN 5 and EN 6, it is apparent that there are important improvements in the water vapour retrievals and the accuracy of the temperature retrievals is not reduced. The channel 11 measurements enable better retrieval of the gradient of the precipitable water profiles, and hence more accurate mixing ratio profile retrievals.

The effect of radiance discrimination misclassifications on the TSF retrieval estimator results is demonstrated by the error statistics labelled EN 6 (radiance discrimination - DC) and EN 7 (perfect discrimination - TC). The rms error profiles for these two algorithms are plotted in Fig 4(a). There is a substantial improvement in the retrieval of the tropopause region with perfect discrimination, but the differences in the lower troposphere are small. Below 300hPa the temperature retrieval error (for DC) is little more than 1.0° C and the mixing ratio error is never more than 1.39 g/kg. The precipitable water rms error is always less than 0.16cm. The bias statistics in Fig 4(b) suggest that retrieval sample

biases are small at all levels and are similar for both TC and DC estimators.

The TSF MAP retrieval estimator provides an estimate of the accuracy of a retrieval (equation (9)), that is, the error covariance matrix of the estimate. Since this matrix is not diagonal the errors are correlated in the vertical. However, a simple although not precise estimate of the retrieval accuracy is afforded by the diagonal elements of this matrix, i.e., the variances of the estimate. The standard deviations of the sample error temperature and precipitable water estimates for the estimator of EN 6 are plotted in Figs 5(a) and 5(b). In this case, the estimated accuracy of the retrieval algorithm, over the whole sample, is close to the actual accuracy, and the vertical structure in the estimator error profile is similar to that in the "true" error profile. Fig 6 demonstrates the R^2 profiles for temperature, precipitable water and mixing ratio, and EN 6. It is most difficult to retrieve the gradient of the precipitable water profile in the height region where it changes most rapidly; i.e., between approximately 700 and 800hPa.

Finally, consider the retrieval algorithm error statistics for synthetic radiance measurements contaminated with Gaussian noise. Since the TSF retrieval algorithm is a two stage process - radiance discrimination followed by profile retrieval - there are two points at which data noise may affect the retrieval result. An incorrect TSF radiance classification could lead to a poor retrieval, as could noise sensitivity in the retrieval estimator.

The effect of measurement noise on the retrieval estimator can be isolated by comparing error statistics from two samples of retrievals where in each case perfect radiance discrimination is used followed by profile retrieval on samples of noise free and noise contaminated measurements. Results for the estimator of channel selection 5 (a four pass estimator (see Table 2); measurement models b,a,a,a), are presented in Table 3. Addition of noise to the measurements leads to only a small degradation in the performance of the retrieval estimator. The largest change in accuracy is for the mixing ratio retrievals. Apparently the TSF adaptively conditioned profile retrieval algorithm (as distinct from the radiance discrimination algorithm) is not particularly sensitive to measurement noise.

The susceptibility of the complete retrieval algorithm (including the radiance discrimination stage) to measurement noise is demonstrated by comparing the error statistics of retrieval samples generated from noise-free and noise-contaminated data. In these cases, the radiance discrimination algorithm is applied in order to determine the *a priori* TSF class of the sounded atmosphere. The errors due to misclassifications of the *a priori* TSF class arising from noise contamination of the radiance discrimination algorithm, and retrieval estimator noise sensitivity are accumulated and quantified by the sample error statistics. The results for three different retrieval estimators, applied to both noise-free and contaminated data, are given in Table 4. The same discrimination algorithm is used on both the noise-free and noise-contaminated radiance data.

For all three estimators, and the three profile types, there is a degradation in the accuracy of the retrieval algorithm when noise is added to the measurements. However the decrease in accuracy is not large, being of the order of 0.05°C per layer for temperature and 0.02 g/kg for water vapour mixing ratio. Considering the results of Tables 3 and 4 together it is clear that the radiance discrimination procedure is not particularly noise sensitive. The largest component of the error in the temperature retrievals arises from missclassification of the *a priori* TSF class due to the inherent indeterminancy of the discrimination problem.

The rms error profiles for one of these estimators, cs 5, for both noise-free and noise-contaminated measurements, are plotted in Fig 7. It is evident that the largest component of the difference in the error statistics arises from errors in retrieval of the tropopause region. The cause of such errors is missclassification of the *a priori* TSF class of the sounded atmosphere.

Some example retrievals from the TSF retrieval scheme and noise-contaminated data, are displayed in Fig 8. The only supplemental information used in these examples is the surface pressure. The retrieved profiles give accurate estimates of the temperature structure in the troposphere, especially in the region of the tropopause. Also, the retrieved dew point temperature profiles display a number of meteorological features of significance.

7. EXAMPLE NOAA 7 RETRIEVALS

The New Zealand Meteorological Service has a facility to receive HRPT data from TIROS-N polar orbiting weather satellites. The retrieval scheme presently in use is a version of the CIMSS regression software (Taylor, 1983). Owing to media storage overheads, the Meteorological Service does not archive radiance data. However, for a two week period in July 1984 some NOAA 7 satellite data were archived in support of the development of a new analysis algorithm. The stored data correspond to the final products from the regression software and are not the original TIROS Information Processor (TIP) data. Accordingly, the archived radiance data have been earth located, calibrated, corrected to nadir, adjusted for emissivity effects and corrected for solar radiation contamination, etc. Pre-processor and retrieval processor defined "bad" data have been removed. In this example, the TSF retrieval algorithms, which together incorporate: TSF *a priori* temperature classification, TSF cloud detection and correction, radiance discrimination, and retrieval by the TSF simultaneous adaptively conditioned MAP sequential estimators, are applied to two passes archived from this period. The purpose of this example is to give some indication of the type of results which might be expected from the TSF retrieval algorithms.

(a) Radiance temperature adjustments

The simulation results discussed in preceeding sections assume the TOVS transmittance functions are known accurately, and that the only errors present in the data are due to random measurement noise of mean zero. This assumption is in fact not correct (Fleming *et al.* 1983). The NESDIS provide transparency adjusting "gamma" correction parameters for transmittances computed by the fast transmittance model (Weinreb *et al.* 1981), but additional, "delta", radiance temperature corrections are required before computed and measured radiance temperatures can be brought into agreement. The delta corrections are generally computed from high quality samples of radiosonde and satellite colocations. Apparently, the delta values are not well known and are probably air mass dependent (Chedin and Scott, 1983).

The TSF retrieval equations were derived with delta values equal to zero; therefore, all measurements must be adjusted to equivalent "computed" radiance temperatures. Since no colocation data sample was available, the delta values have been estimated by another method. It is evident in Fig 1(a) that two of the TSF classes are associated with tropical atmospheres - classes 11 and 12 - with the second having the larger channel variances and, in both the temperature and radiance domains, the cooler surface. The variances of all channels for either of these classes is small, relative to all the other classes. These tropical classes define narrow peaks in the global spectrum of atmospheric radiances. An estimate of the delta values may be made by determining the differences between either or both of these radiance class mean vectors, and the mean of a sample of measured tropical radiances. From the two week archive of radiance data, mean radiance temperatures were computed for a sample constrained by the specifications that the measurements be cloud free (as defined by the CIMSS software), lie within nine IFOVs of nadir, and north of 21° S. The latter value was chosen so as to trade off sample size against the inclusion of non-tropical air, since the northern limit for HRPT data, received at Wellington, is near 17° S. Observations from both day and night time passes are included in the sample. The delta values are estimated from the class 12 synthetic radiance data since this class is assumed to be associated with the more southerly of the two tropical TSF classes. Table 5 lists the channel mean radiance temperatures and variances for class 12, the mean for the measured "tropical" radiance temperatures and their variances, the gamma values (supplied by the NESDIS) and the delta value estimates. The sample sizes are indicated.

The large delta value for channel 1 might result from the procedure used to extrapolate the radiosonde data to mesospheric heights, although the result for channel 2 does not substantiate this thesis. The channel 8 delta estimate may indicate a water vapour problem since the delta values for channels 18 and 19 (which are less sensitive to water vapour) are much smaller, although a residual component of reflected solar radiation may contaminate these channels. The correction for channel 16 is consistent with the computed transmittance being too transparent. Although the sample size for the measured data is 78, there are perhaps only 10 or so meteorologically independent observations, since

many of the sample measurements lie in the same air masses. Consequently these delta values and associated comments are necessarily tentative.

In the following example retrievals, the satellite measured radiance temperatures have been adjusted by the delta values given in Table 5. As a result of this approach those radiance measurements which are most sensitive to bias "noise", because they have the smallest signal, and their TSF classes the smallest hyper-volumes, are forced into agreement with the computed radiances. This strategy reduces the effect of bias errors which could result in serious misclassifications because of the sensitivity of the Mahalanobis term of the discriminant equations to data biases.

(b) Radiance discrimination and cloud detection

In the example discrimination classification of NOAA 7 radiance observations presented here, the radiance discriminator channels are 1, 2, 4, 23, 5 and 6, and TSF declouding predictor channels 1, 2, 4, 23 and 22. Channel 24 would be a useful addition to both the discrimination and declouding model equations, but because the weighting function is centered near 100hPa and has a high vertical resolution relative to channels 1 and 2, the information in this channel is quite sensitive to errors in both the gamma and delta correction parameters.

The results of TSF radiance discriminant classification to define a priori retrieval constraints for two July 1984 NOAA 7 passes over the New Zealand region are plotted in Fig 9. The associated AVHRR image for these same passes is presented in Fig 10. In Fig 9 the TSF class numbers, as defined in Fig 1, are printed for every third retrieval location, which is a 3 x 3 field of HIRS IFOVs, or when the TSF class changes. All classifications with posterior probabilities greater than 0.1 are plotted. It is evident that the progression in TSF class types from equatorial to polar regions is as might be expected. Also evident is the spatial consistency of the TSF classes in both the various airmass types and between orbits. Class 43 (double tropopause) *a priori* profiles are associated with surface low pressure features. In the anticyclone over New Zealand (see Fig 13) the selected TSF class has the expected 200hPa tropopause of the mid-latitude winter atmosphere. The transition to classes 41 and 21, which have lower mean tropopause heights (aproximately 250hPa) adjacent to the surface low near 40° S, is meteorologically realistic. The *a priori* retrieval constraint classifications are not zonal in nature, but appear to be related to different airmass types.

(c) TSF simultaneous physical retrievals for NOAA 7 passes

The radiance temperature observations, once discriminated to determine the *a priori* retrieval constraints, and declouded by the TSF stratified regression radiance prediction models, can be used in the TSF physical retrieval algorithm.

Table 6 indicates the channel selection used in the retrieval estimator. The HIRS shortwave window channels (18 and 19) are not used for day time passes since it is probable that these channels are contaminated by reflected solar radiation (Chedin and Scott, 1983). The size of the channel 16 delta correction suggests the possibility of some problem with this channel; accordingly it is not used in the retrieval algorithm. The measurement noise estimate applied in the retrieval algorithm to the information for any particular channel, is dependent upon whether the measurement has, or has not been, corrected for cloud contamination. The cloud-free measurement noise estimates together with the noise estimates of the declouding prediction equations, are given in Table 7.

The surface pressure assumed by the retrieval algorithm is set equal to the *a priori* TSF class value (see Table 1) except where the underlying surface is not at mean sea level in which case a new estimate for the surface pressure is computed. When the pressure at the surface is greater than 1000hPa the radiance temperatures are corrected for that part of the measured radiance which arises from the layer of atmosphere between 1000hPa and the surface. On the inclusion of each new measurement the above surface profile retrievals are checked for super-adiabatic lapse rates. However, the lapse rate between the surface and the lowest atmospheric retrieval level is permitted to be super-adiabatic since the satellite window channels measure the skin temperature. The gradient between the temperature of the underlying

surface and that a few meters above the surface can greatly exceed the dry adiabatic lapse rate (Taylor, 1973). No ancillary data are used by the retrieval algorithm other than the TSF *a priori* constraint information.

In Fig 11 a selection of TSF retrieval results for the two NOAA 7 passes is presented. Results from the Meteorological Service's operational regression retrieval software are given in Fig 12. The retrievals have been analysed onto a 41 x 29 grid with a Cressman analysis scheme. The sample sizes for both retrieval schemes are identical as are the locations of the retrieved profiles. The TSF retrievals have not been spatially filtered. Finally, in order to provide an almost independent assessment of the meteorological situation, the operational NWP analyses of surface pressure and 1000-500hPa thickness for the New Zealand region and times 0000 and 0600 GMT on 27/Jul/1984 are presented in Fig 13. GTS SATEM messages and a subset of the locally derived regression retrievals were used in constructing these analyses (which are on a much coarser grid than the satellite only analyses).

Considering the 1000-500hPa thickness fields first, it is evident that in broad detail the regression and TSF results are similar, with the exception of the trough in the TSF field near 45° S and 177° E. No strong bias is evident in either of the satellite-derived thickness fields relative to the NWP analyses. When considered in closer detail, the most noticable differences are in terms of the gradients in the main trough and ridge features. The TSF retrievals show stronger gradients than the regression retrievals - this observation is especially appropriate north of the cold pool near 175° W, and on the western edge of the ridge in the vicinity of 27° S and 153° E. During the period between the two passes there was considerable advection of warm air into the New Zealand region, which explains the apparent discontinuities in some thickness contours near the satellite pass boundaries. This advection is also evident in the 0000 and 0600 GMT analyses of Fig 13. The trough in the TSF retrievals at 177° E is difficult to explain. The AVHRR image indicates that there is low cloud generally in this region but does not reveal any significant trough related feature. Instead this trough might be evidence of a problem in the pre-processor limb correction algorithm since it occurs on the edge of the first pass.

In Fig 14, tephigrams of both regression and TSF retrievals are plotted for a selection of profiles from a cross-section along the trough axis and through the cold pool at 175° W. The retrieval locations are indicated on Fig 11(a). The results from the two retrieval algorithms show marked differences, especially in the tropopause region and near the surface. In particular it is evident that the regression temperature retrievals below 850hPa tend to have similar stability whereas the TSF retrievals display inversions as well as stable and dry adiabatic lapse rates in the boundary layer. The AVHRR imagery indicates convective activity is occurring along portions of this cross-section and therefore supports the low level vertical structure indicated in the TSF retrievals. With regard to the tropopause retrievals, the development in the TSF retrievals of a double tropopause structure in the vicinity of the cold pool and near the surface low at 165° W is reasonable, given the physical processes occurring. The 400hPa tropopause in the retrieval at 34.5° S near the centre of the trough is consistent with the 430hPa tropopause (an unexpectedly extreme value) measured by the Auckland (37° S, 175° E) 0000 GMT radiosonde on 26/Jul/84, when the trough was centred over Auckland. Further south along the trough axis the TSF retrieved tropopause rises to near 250hPa. The regression tropopause retrievals do not display structure of the type expected for this meteorological situation.

As might be expected from the preceding discussion, the 850hPa temperature and dewpoint temperatures, and the 500hPa temperature fields display some marked differences. From the 500hPa TSF temperature analysis it is evident that the trough in the TSF 1000-500hPa thickness analysis near 177° E is a low level feature since at 500hPa there is some evidence for a thermal ridge in this location. The presence of high cloud over the south of the South Island in the vicinity of this ridge is consistent with warm air advection aloft. The isotherms on the northern side of the cold pool feature are closely aligned with the lines of convection in the AVHRR image for both the regression and TSF retrievals. although the gradients are stronger for the TSF retrievals. At 850hPa, the gradients in the main trough are weaker for the TSF retrieval fields than they are for the regression fields, although the reverse is true for the ridge near 160° E. However, when the 850hPa dew point temperature fields are considered it is evident that the there is more structure in the TSF retrievals. Clearly the TSF retrievals are, in places, in error due to cloud contamination, especially in the cloud mass associated with the ridge in the region of 160° E. In the main trough, east of New Zealand, the TSF retrievals delineate clearly the moisture gradient between the open cell and adjacent area of subsidence (to the west). It is apparent

that for the regression algorithm the low level structure is contained principally in the temperature fields, with a flatter water vapour field, whereas, in this case, the reverse is true of the TSF retrieval results. This result is not unexpected since, unlike the regression scheme, the simultaneous retrieval algorithm utilizes the water vapour information in the temperature sensing channels directly, and therefore decouples the temperature and water vapour fields to a greater extent.

8. SUMMARY

Some of the characteristics of the TSF retrieval algorithm have been demonstrated for a winter case, and the results of a new type of TSF stratified, single field of view, cloud detection and clearing scheme presented.

Evidently the TSF simultaneous, adaptively conditioned, MAP retrieval estimator is, in so far as synthetic data experiments can prescribe, capable of producing accurate retrievals of temperature, precipitable water, and mixing ratio profiles. An additional advantage of the TSF retrieval algorithm is the specification of the expected accuracy of any particular retrieval, through recourse to the radiance discrimination posterior probability and the retrieval estimator error covariance matrix. Both pieces of information can be used by an optimum interpolation analysis algorithm.

In the case of the example satellite retrievals presented here, the TSF-retrieved 1000-500hPa analysed fields are similar, on the synoptic scale, to those produced from a regression retrieval scheme. However the gradients in the TSF results are greater than those for the regression scheme, and the TSF-retrieved profiles demonstrate greater variance than the equivalent regression retrievals. The differences between the TSF and regression retrievals would appear to be important and significant, but an objective analysis of the quality of individual TSF and regression retrievals must await full colocation and NWP analysis / forecast cycle experiments.

A number of problem areas can be identified for the TSF retrieval algorithm. The most important is the requirement for good estimates of the gamma transmittance, and delta radiance temperature correction parameters. The delta corrections should be calculated for each TSF *a priori* class. However once these problems, and those associated with various pre-retrieval processing algorithms, are resolved, it might be expected that the TSF retrieval accuracy will be close to that estimated from the validation analysis.

REFERENCES

Chedin, A. and N.A. Scott, 1983: Improved Initialization Inversion Procedure ("3I"). In *Technical proceedings of the first International TOVS Study Conference.* 14-79, 29 August-2 September, Igls, Austria, Pub CIMSS, University of Wisconsin.

Chedin, A., N.A. Scott, C. Waihiche and P. Moulinier, 1985: The Improved Initialisation Inversion Method: A high resolution physical method for temperature retrievals of the TIROS-N series. *J. Clim. Appl. Meteor.* **24**, 128-143.

Fleming, H.E., Crosby D.S. and A.C. Neuendorffer, 1983: Elimination of a major error component in satellite temperature retrievals. Preprint Volume, *Fifth Conference on Atmospheric Radiation,* October 31-4 November, Baltimore, Maryland.

Fleming, H.E., Goldberg M.D. and D.S. Crosby, 1986: Minimum variance simultaneous retrieval of temperature and water vapour from satellite radiance measurements. Preprint Volume, *Second Conference on Satellite Meteorology/Remote Sensing and Applications.* May 13-16, Williamsburg, Virginia.

Kvalseth, T.O., 1985: Cautionary note about R^2. *Amer. Stat.* 39, 279-285.

McMillin, L.M., 1986: The use of classification procedures to improve the accuracy of satellite soundings of temperature and moisture. Preprint Volume, *Second Conference on Satellite Meteorology/Remote Sensing and Applications.* May 13-16, Williamsburg, Virginia.

Smith, W.L and H.M. Woolf, 1984: Improved vertical soundings from an amalgamation of polar and geostationary radiance observations. In preprint volume to: *Conference on Satellite/Remote Sensing*

and Applications, June 25-29 1984, Clearwater Beach, Fla. Pub AMS.

Smith, W.L, H.M.Woolf, C.M.Hayden and A.J.Schreiner, 1985: The simultaneous retrieval export package. In *Technical proceedings of the second International TOVS Study Conference.* 224-253, February 18-22, Igls, Austria, Pub CIMSS, University of Wisconsin.

Taylor, B.F., 1973: The maritime convective layer in middle latitudes. 150pp, *PhD Thesis*, University of London

Taylor, B.F., 1983: The first International TOVS Study Conference case study report. In *Technical proceedings of the first International TOVS Study Conference.* 304-315, 29 August - 2 September, Igls, Austria, Pub CIMSS, University of Wisconsin.

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

Uddstrom, M.J. and D.Q.Wark, 1985: A classification scheme for satellite temperature retrievals. *J. Climate Appl. Meteor.* **24**, 16-29.

Uddstrom, M.J., 1986: The retrieval of atmospheric profiles with Typical Shape Function simultaneous physical retrieval estimators. In preparation.

Wark, D.Q., 1986: Temperature profiles derived from a restricted range of solutions using satellite measurements. *Advances in Remote Sensing Retrieval Methods* (eds. A. Deepak, H. Fleming, and M. Chahine) A.Deepak Publishing, Hampton, Virginia, USA.

Weinreb, M.P., H.E.Fleming, L.M.McMillin and A.C.Neuendorffer, 1981: Transmittances for the TIROS operational vertical sounder. *NOAA Tech. Rep.* NESS 85, 60pp.

Westwater, E.R., M.T. Decker, A. Zachs and K.S. Gage, 1983: Ground- based remote sensing of temperature profiles by a combination of microwave radiometry and radar. *J. Clim. Appl. Meteor.*, **23**, 689-703.

TABLE 1: Characteristics of July TSF class samples.

Class name	Sample size	A priori probability	Trace (50hPa - 1000hPa) (K)	Mean sfc pressure* (hPa)
11	459	0.192	292.5	1010.1 (6.0)
12	412	0.173	298.3	1007.8 (13.9)
21	201	0.084	324.7	1012.9 (10.5)
22	116	0.049	456.2	1014.3 (12.9)
23	130	0.054	221.1	1015.4 (9.9)
24	146	0.061	514.7	1013.2 (10.5)
31	174	0.079	729.4	1006.7 (14.7)
32	71	0.030	1356.7	1002.5 (13.8)
33	66	0.028	2284.4	998.8 (15.1)
34	109	0.046	568.5	980.3 (10.7)
41	228	0.096	296.2	1011.4 (12.2)
42	125	0.052	463.5	1008.2 (15.1)
43	151	0.063	1124.7	1002.1 (13.9)

* Class standard deviation in brackets

TABLE 2: Multi-pass retrieval armse error statistics, (aR^2 in brackets) for indicated channel selections and measurement specification models. TC indicates perfect discrimination and DC indicates radiance discrimination. Sample size: 500.

EN'	Channel Selection*	Measurment models+	Temperature (100-1000hPa) K	Precipitable Water (500-1000hPa) cm	Mixing Ratio (500-1000hPa) g/kg
1	1 DC	a	1.4197 (0.9748)	0.1342 (0.9128)	0.9029 (0.8419)
2	2 DC	b,a	1.3522 (0.9765)	0.1062 (0.9453)	0.8645 (0.8590)
3	3 DC	a,a,a	1.2770 (0.9790)	0.1013 (0.9528)	0.8602 (0.8643)
4	3 DC	b,a,a	1.2676 (0.9794)	0.0983 (0.9547)	0.8323 (0.8709)
5	4 DC	b,a,a,a	1.2480 (0.9801)	0.0935 (0.9600)	0.8243 (0.8754)
6	5 DC	b,a,a,a	1.2466 (0.9800)	0.0866 (0.9649)	0.7934 (0.8835)
7	5 TC	b,a,a,a	1.1172 (0.9848)	0.0832 (0.9672)	0.7863 (0.8862)

' Estimator Number (i.e., EN) identifier as used in text.
+ Measurement model for each pass:- a : linear U(p); b : log(U(p))
* Channel selections and radiance discrimination algorithm used, where [.] defines a single pass
1 is: [18,12,10,1,2,3,4,5,6,7,8,16,15,14,13]
2 is: [18,12,10], [18,12,10,1,2,3,4,5,6,7,8,16,15,14,13]
3 is: cs 2, [18,12,10,1,2,3,4,5,6,7,8,16,15,14,13]
4 is: cs 3, [18,12,10,1,2,3,4,5,6,7,8,16,15,14,13]
5 is: [18,12,11,10], 3 x [18,12,10,1,2,3,4,5,6,7,8,16,15,14,13]

TABLE 3: Retrieval armse error statistics and aR^2 values for noise free and noise contaminated radiances. (TC, cs 5)

Noise	Temperature (100-1000hPa) K	Precipitable Water (500-1000hPa) cm	Mixing Ratio (500-1000hPa) g/kg
Yes	1.1493 (0.9838)	0.0898 (0.9668)	0.8154 (0.8801)
No	1.1172 (0.9848)	0.0832 (0.9672)	0.7863 (0.8862)

TABLE 4: Retrieval armse error statistics and aR^2 values for noise free and noise contaminated radiances. (DC)

Channel selection and models+	Noise*	Temperature (100-1000hPa) K	Precipitable Water (500-1000hPa) cm	Mixing Ratio (500-1000hPa) g/kg
1 a	yes	1.4590 (0.9729)	0.1360 (0.9119)	0.9211 (0.8359)
1 a	no	1.4197 (0.9748)	0.1342 (0.9128)	0.9029 (0.8419)
6 a,a,a,a	yes	1.3593 (0.9760)	0.0976 (0.9588)	0.8606 (0.8681)
6 a,a,a,a	no	1.2908 (0.9791)	0.0916 (0.9619)	0.8342 (0.8731)
5 b,a,a,a	yes	1.3006 (0.9776)	0.0912 (0.9639)	0.8092 (0.8785)
5 b,a,a,a	no	1.2466 (0.9800)	0.0866 (0.9649)	0.7934 (0.8835)

+ Channel selection and measurement models (see Table 9), and 6 is: [cs 1], [cs 1], [cs 1], [cs 1].
* Assumes 9 IFOVs for noise contaminated observations.

TABLE 5: Tropical atmosphere radiances, gamma transmittance adjustment and estimated delta radiance temperature adjustment values.

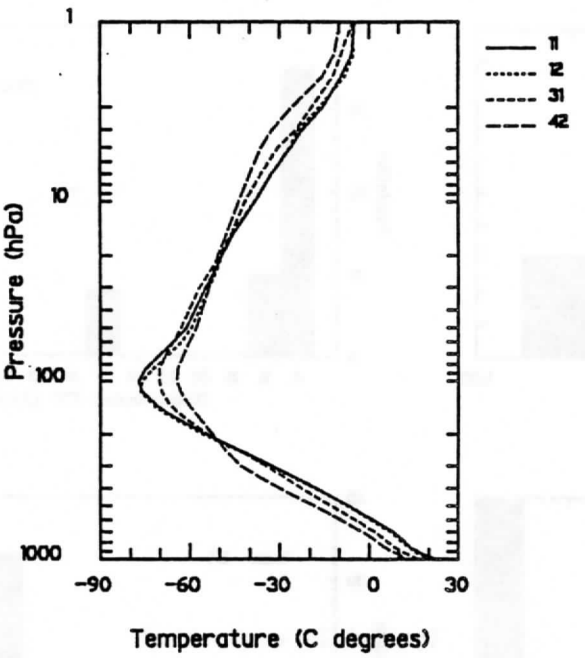
Channel No.	Class 12 Radiance temperatures (n=412) mean (K)	std dev	Measured Radiance temperatures (n=78) mean (K)	std dev	gamma	delta (K)
1	234.01	1.057	230.40	0.185	1.069	3.604
2	219.92	0.992	219.05	0.243	0.942	0.870
3	220.44	0.851	218.91	0.279	1.108	1.525
4	233.90	1.070	233.18	0.254	1.031	0.728
5	249.39	1.615	248.34	0.538	0.972	1.054
6	260.90	2.058	260.89	0.899	0.965	0.008
7	273.32	2.625	273.43	1.529	0.909	0.113
8	291.71	4.963	288.50	2.970	1.000	3.208
10	286.81	4.074	285.67	2.496	1.120	1.136
11	263.47	4.578	264.26	1.805	1.032	0.789
12	245.38	6.628	246.93	2.100	1.035	1.551
13	282.25	4.293	279.98	2.153	1.115	2.266
14	269.13	3.374	267.51	1.344	1.028	1.615
15	252.32	2.349	250.87	0.636	1.057	1.456
16	232.17	1.442	228.95	0.277	1.002	3.219
18	293.17	6.155	293.65	3.986	1.000	-0.488
19	293.12	5.883	293.70	2.817	1.000	-0.578
22	257.29	2.332	256.00	0.243	1.010	1.297
23	226.78	0.882	227.01	0.111	1.010	-0.231
24	209.10	1.784	209.90	0.256	1.070	-0.804

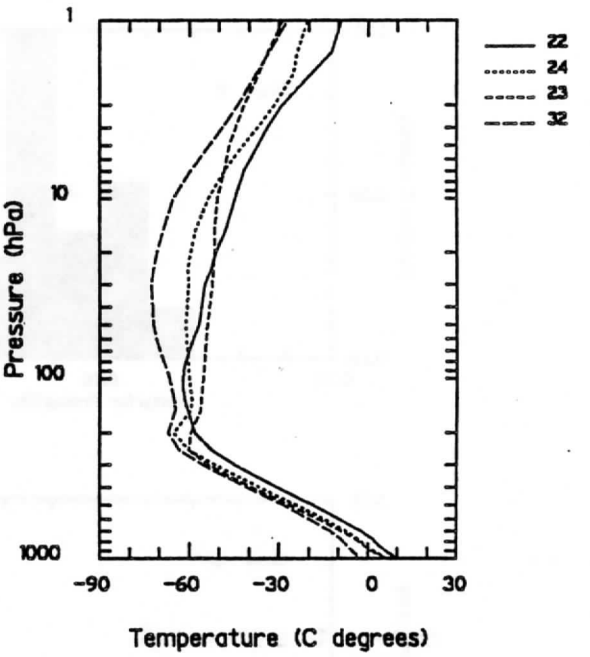
TABLE 6: NOAA 7 TSF retrieval channel and model specifications

Orbit time	TSF Retrieval Pass No.	PCW model	Channel selection
03GMT (day)	1	log	8,12,11,10
	2	lin	8,12,10,1,2,3,4,5,6,7,15,14,13
	3	lin	8,12,10,1,2,3,4,5,6,7,15,14,13
	4	lin	8,12,10,1,2,3,4,5,6,7,15,14,13

TABLE 7: Assumed noise characteristics for cloud free and TSF stratified regression declouded channels.

Channel Number	Measurement noise (K)	Declouding noise (K)
1	1.0	-
2	0.6	-
3	0.4	-
4	0.3	-
5	0.2	0.28
6	0.1	0.57
7	0.1	1.26
8	0.1	2.29
10	0.1	2.03
11	0.2	3.57
12	0.3	5.58
13	0.1	1.51
14	0.1	0.71
15	0.2	0.23
16	0.1	0.33
18	0.1	2.26
22	0.3	-
23	0.3	-
24	0.3	-





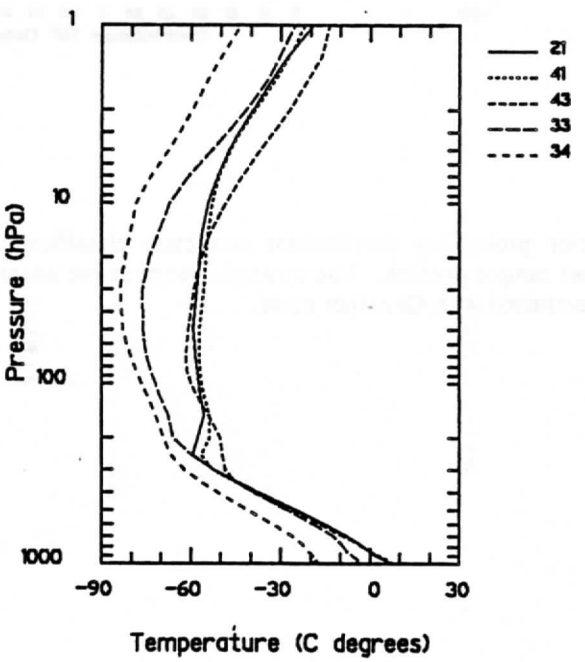


Fig 1. TSF class mean temperature profiles for July Southern hemisphere extrapolated radiosonde data. The class name identifiers are as given in table 1.

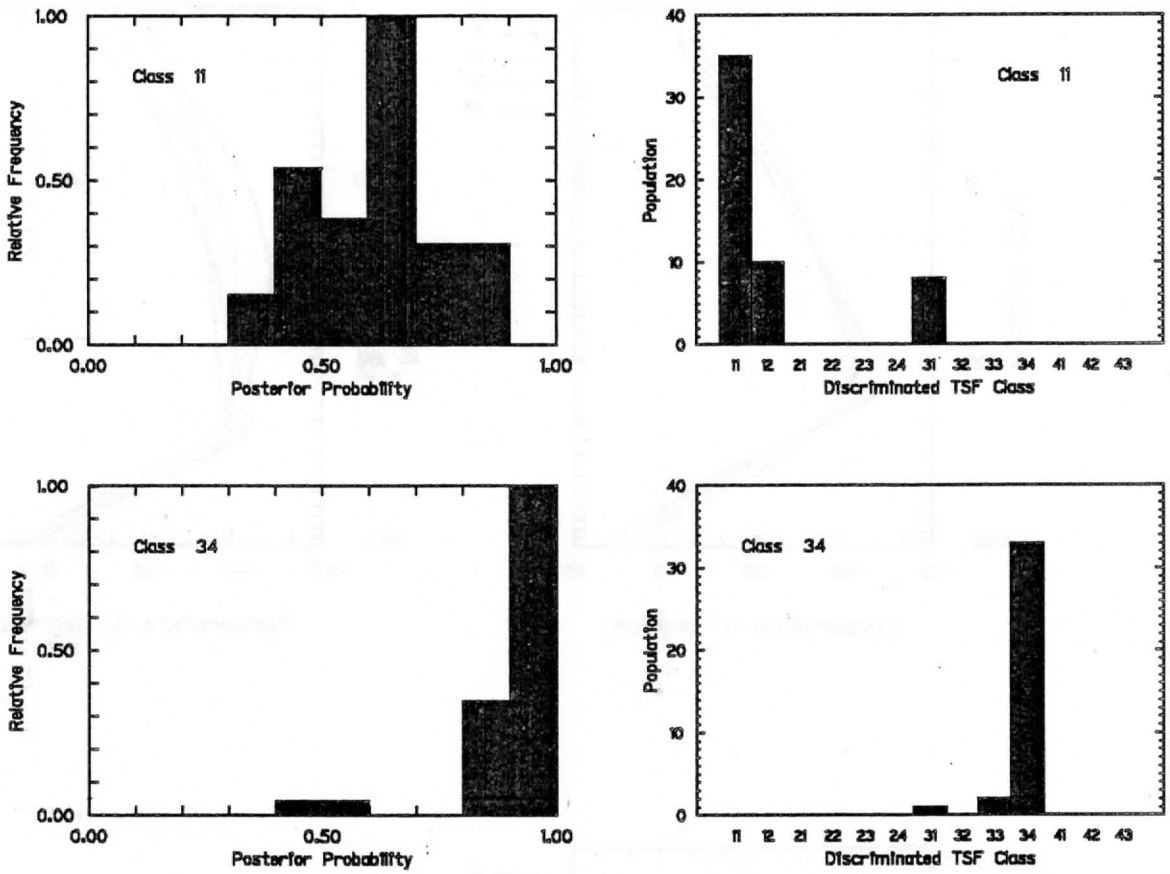


Fig 2. Radiance discriminant posterior probability distributions and class classification histograms for the TSF class 11 and 34 independent sample profiles. The synthetic radiance measurements used in the discrimination equations were contaminated with Gaussian noise.

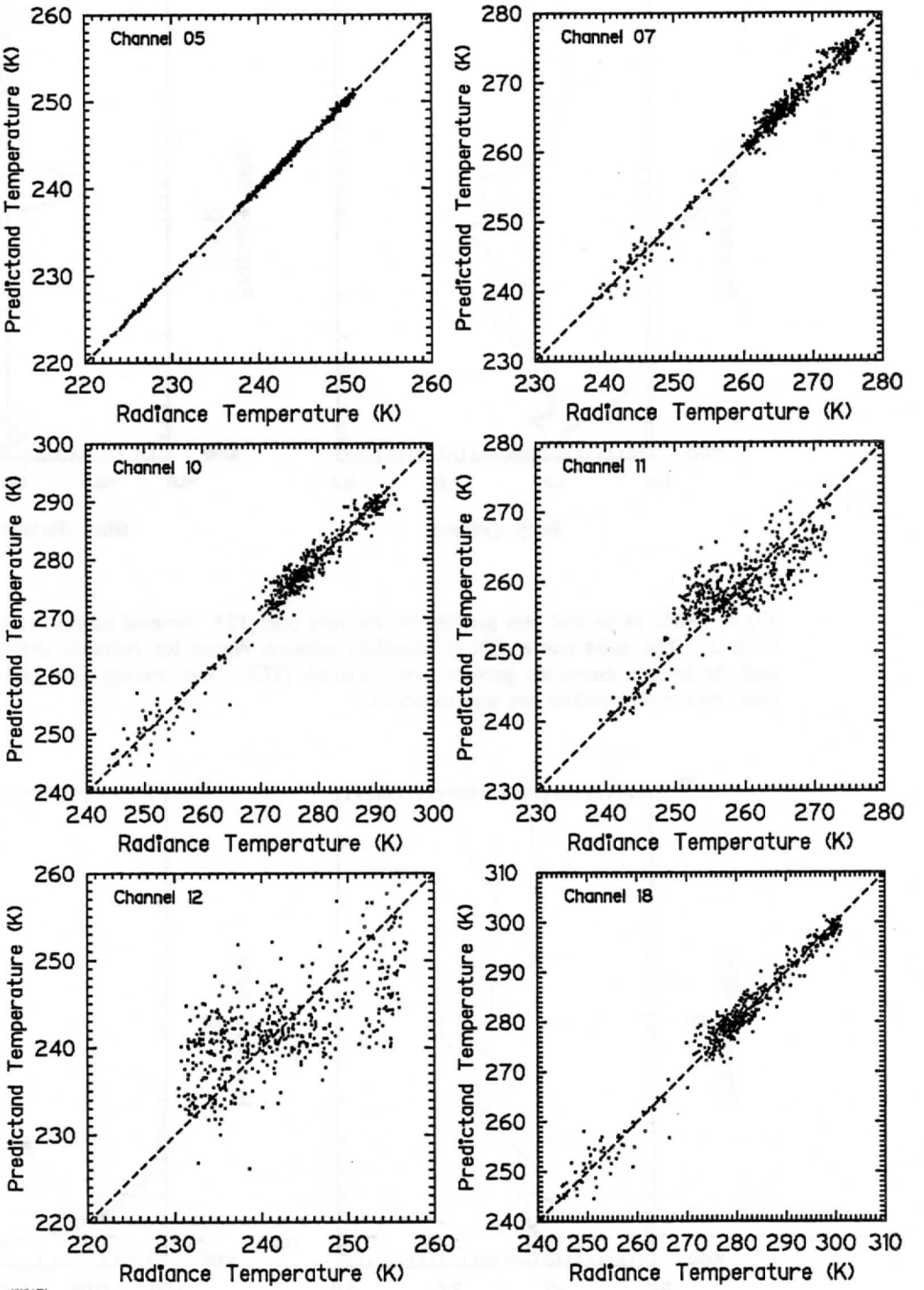


Fig 3. Scatter plots for the TSF cloud decontamination algorithm. The sample size is 500, divided into 13 TSF classes. The synthetic radiance measurements used in the discrimination classification were contaminated with Gaussian noise.

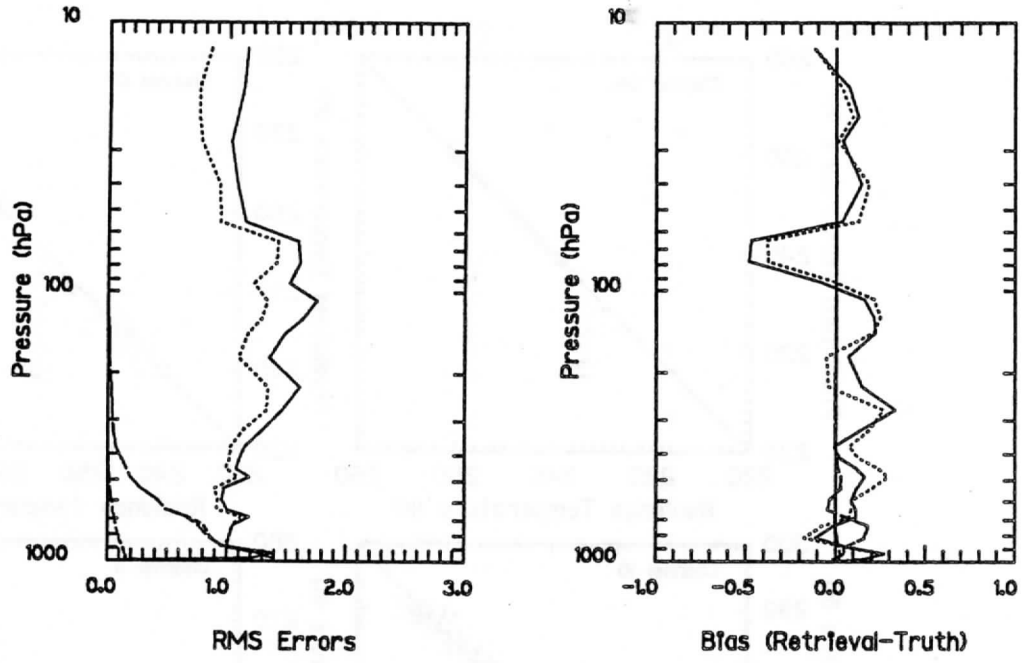


Fig 4. RMS error and bias profiles for the four pass TSF retrieval estimators labeled EN 6 and EN 7 in table 2. The solid curve (EN 6) specifies retrieval errors for radiance discrimination retrievals (DC) and the broken curve for perfect discrimination (TC). The mixing ratio (g/kg) and precipitable water (cm) rms error profiles are also indicated.

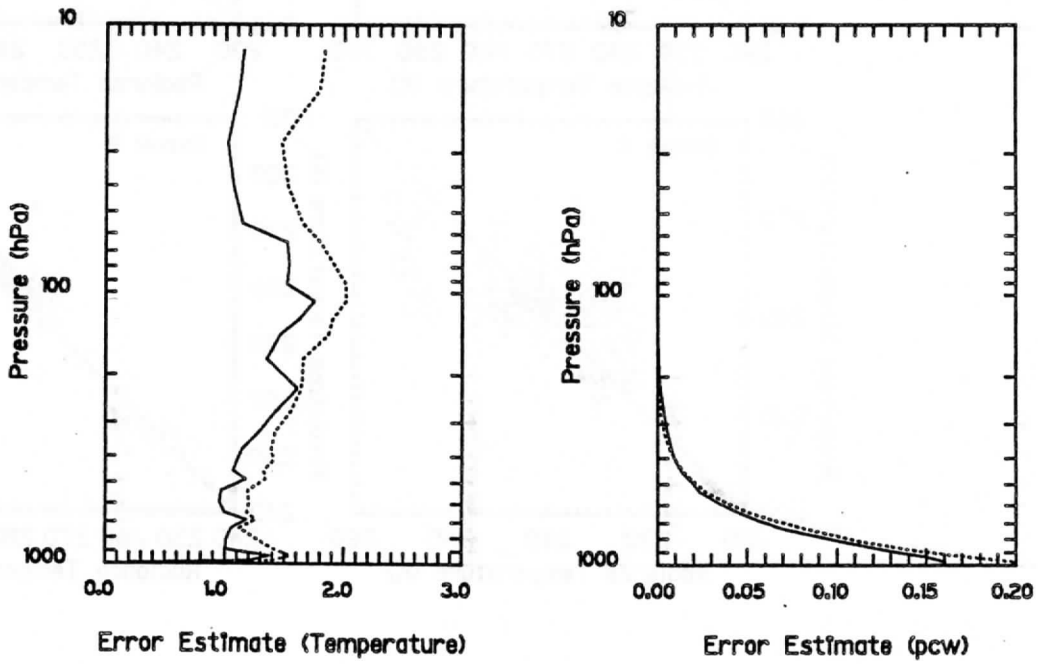


Fig 5. TSF MAP estimated error profiles for temperature and precipitable water (cm) for the TSF retrieval estimator labeled EN 6 in table 2. The solid curves indicate the "true" rms error profiles and the broken curves the estimator error profiles.

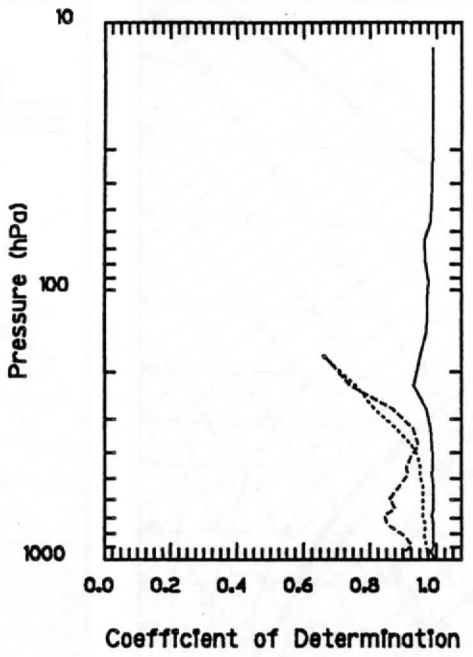


Fig 6. R^2 profiles for EN 6 TSF retrievals. The solid curve refers to temperature, the dotted curve to precipitable water and the dashed curve to mixing ratio.

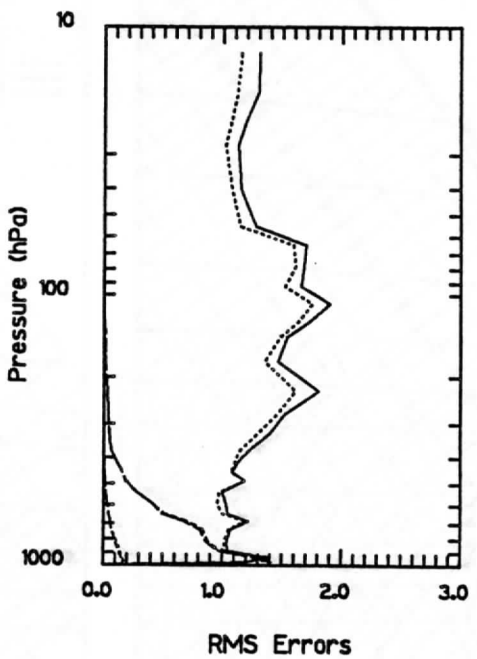


Fig 7. RMS error curves for the TSF retrieval algorithm for noise free and noise contaminated synthetic measurements. A four pass estimator has been used in both cases with cs 5 (see table 4). The solid curve refers to the results for the noise contaminated measurements and the dotted curve for the noise free measurements. Mixing ratio (g/kg) and precipitable water (cm) error profiles are also given.

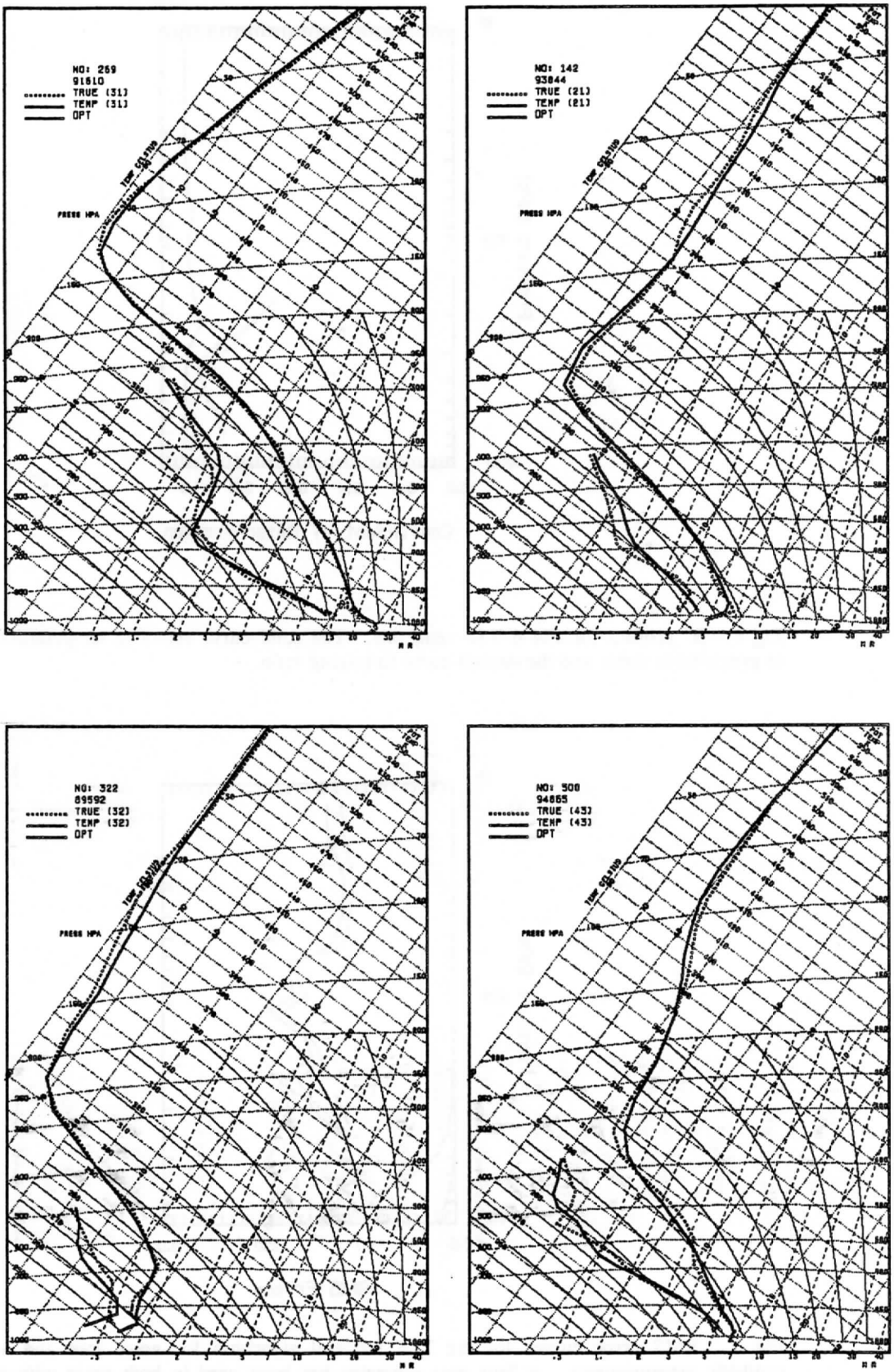


Fig 8. Tephigram plots of temperature and dew point temperture for four representative synthetic data retrievals. The true profiles are the dashed curves and the solid curves the TSF retrieved profiles. The TSF retrieval estimator utilized cs 5 data and the radiance temperature measurements were contaminated with Gaussian noise.

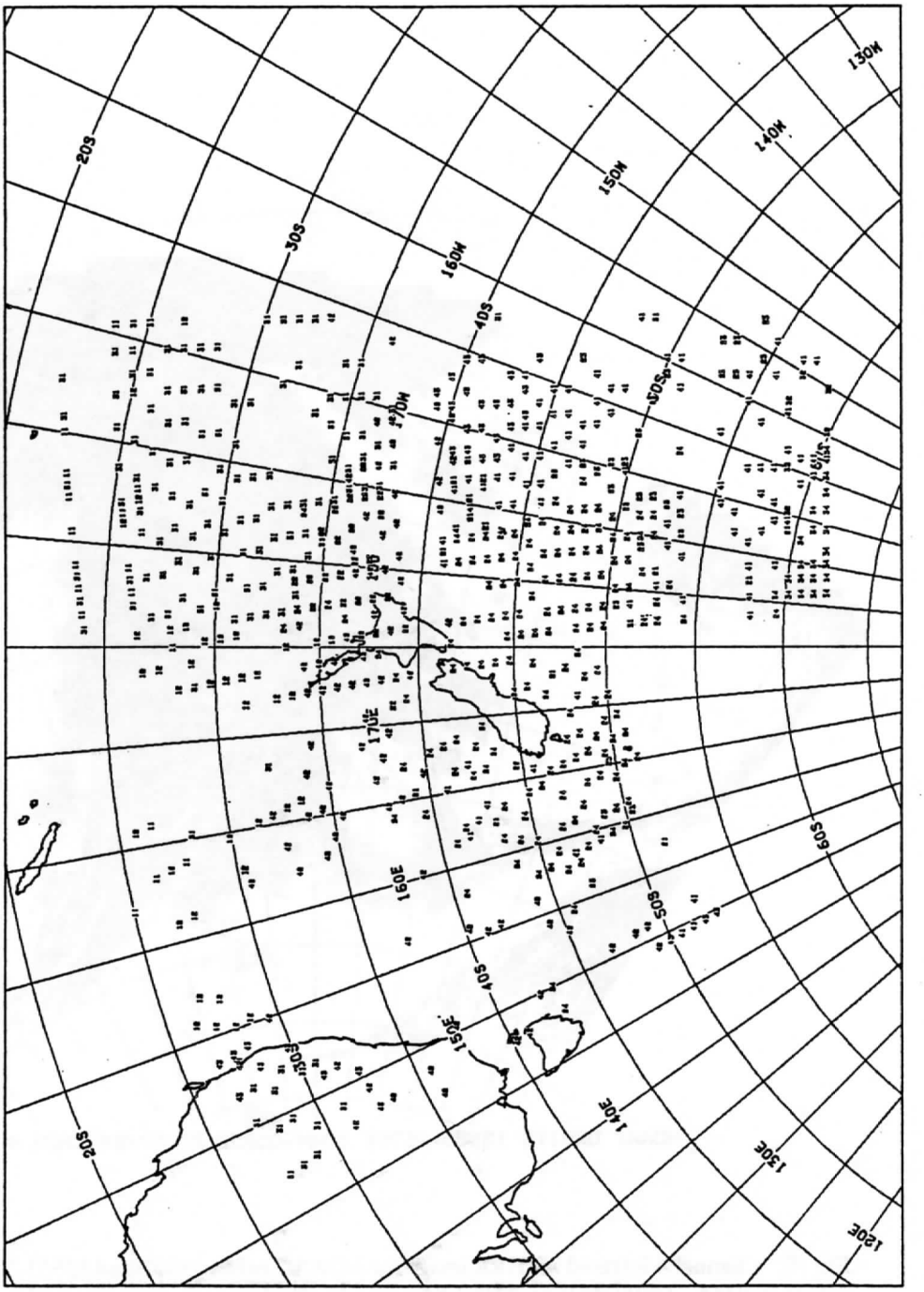


Fig 9. Radiance discrimination classes for NOAA 7 orbits 15950 and 15951. The class identifiers refer to figure 1 and table 1. The discrimination function utilized channels 1, 2, 4, 23, 5 and 6 and the declouding algorithm, channels 1, 2, 4, 23 and 22.

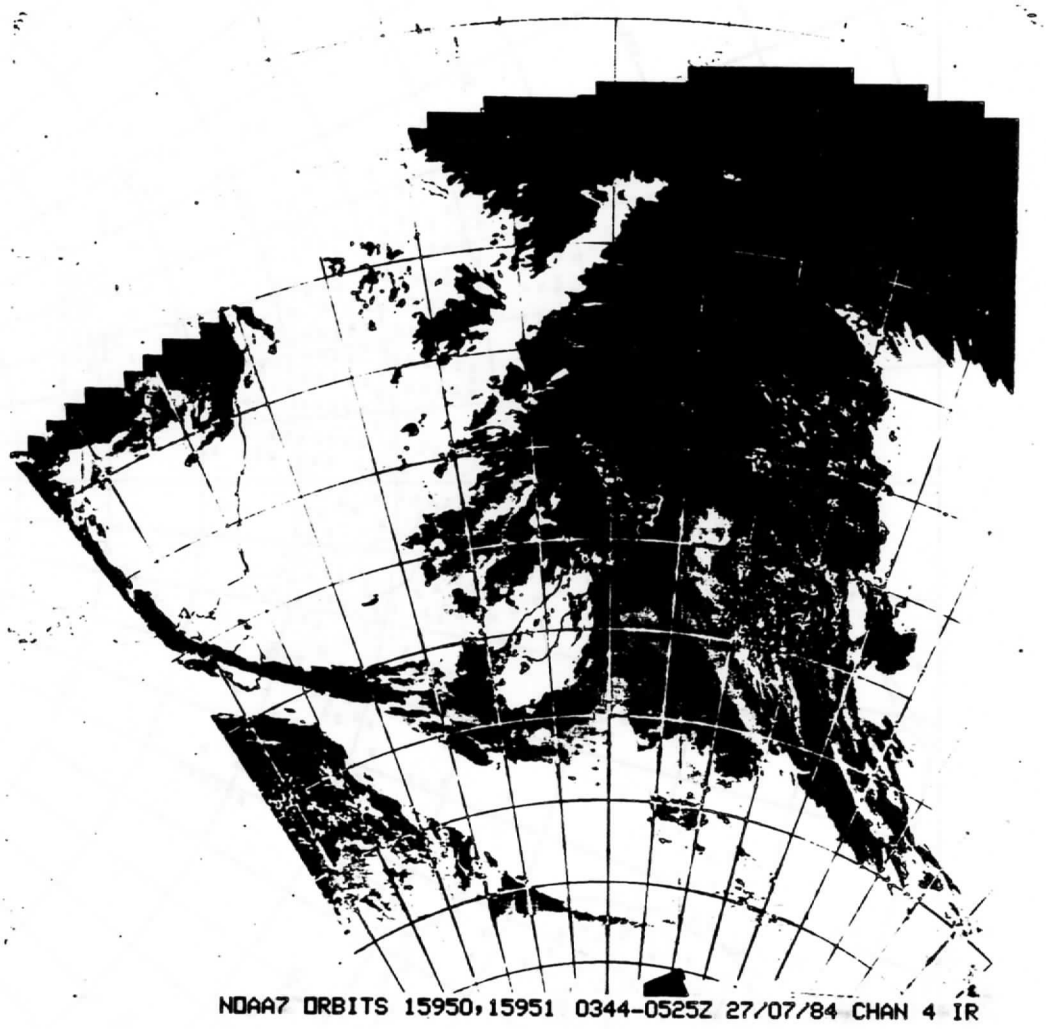
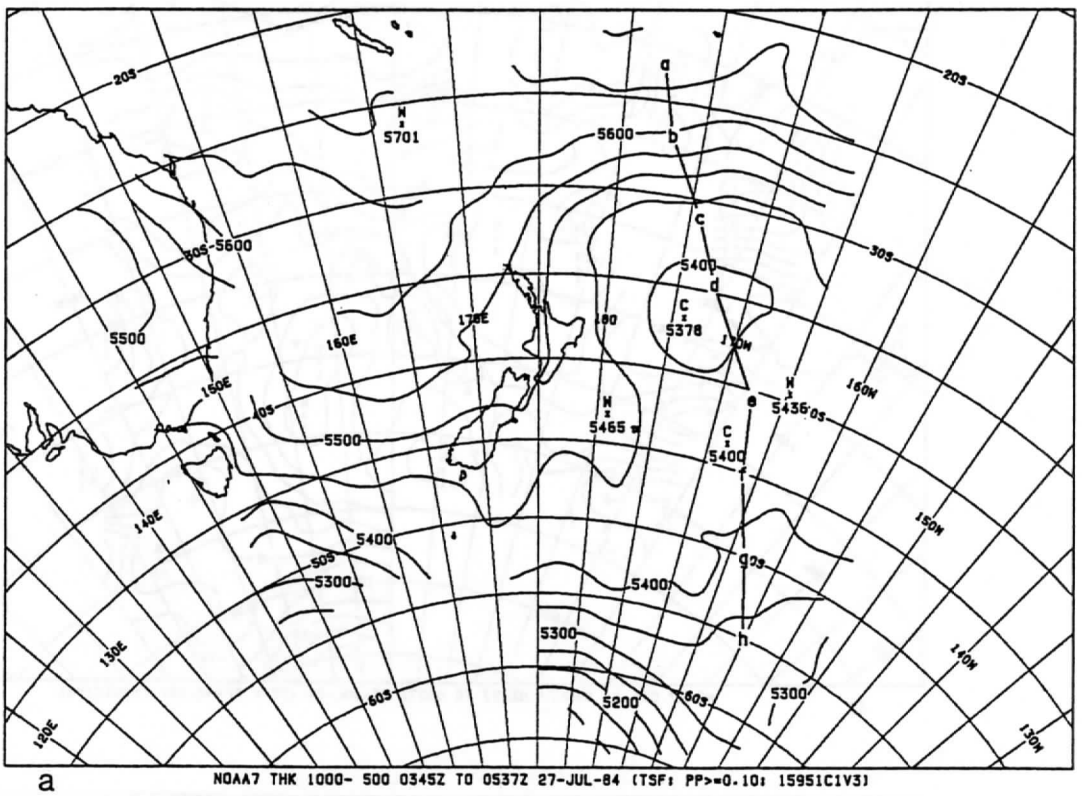


Fig 10. Channel 4 infrared AVHRR image for NOAA 7 orbits 15950 and 15951. The time span of the two passes is 0344 - 0525GMT on 27/Jul/84.



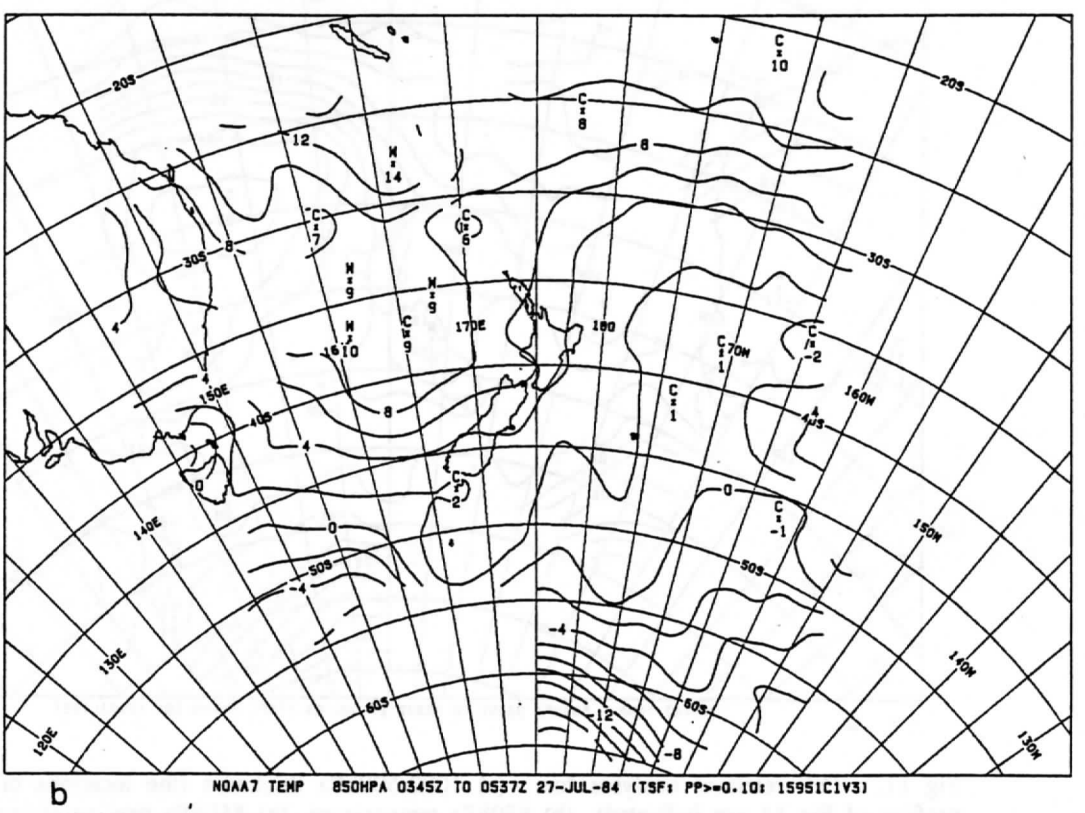
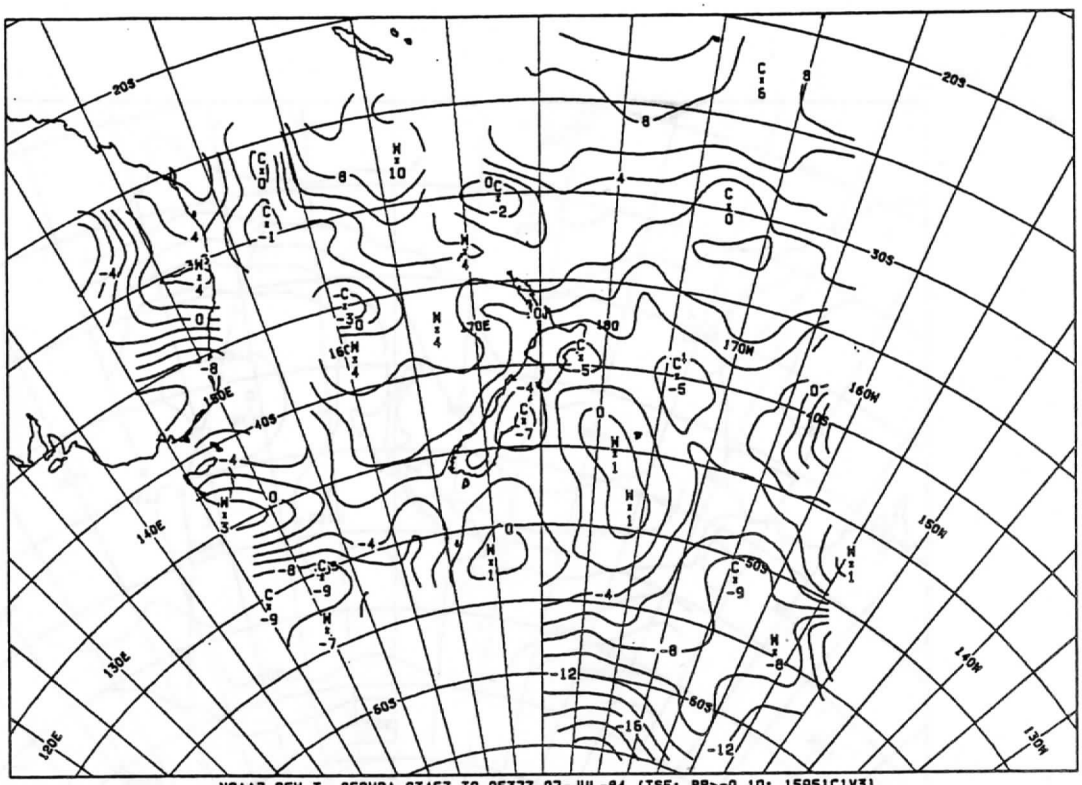
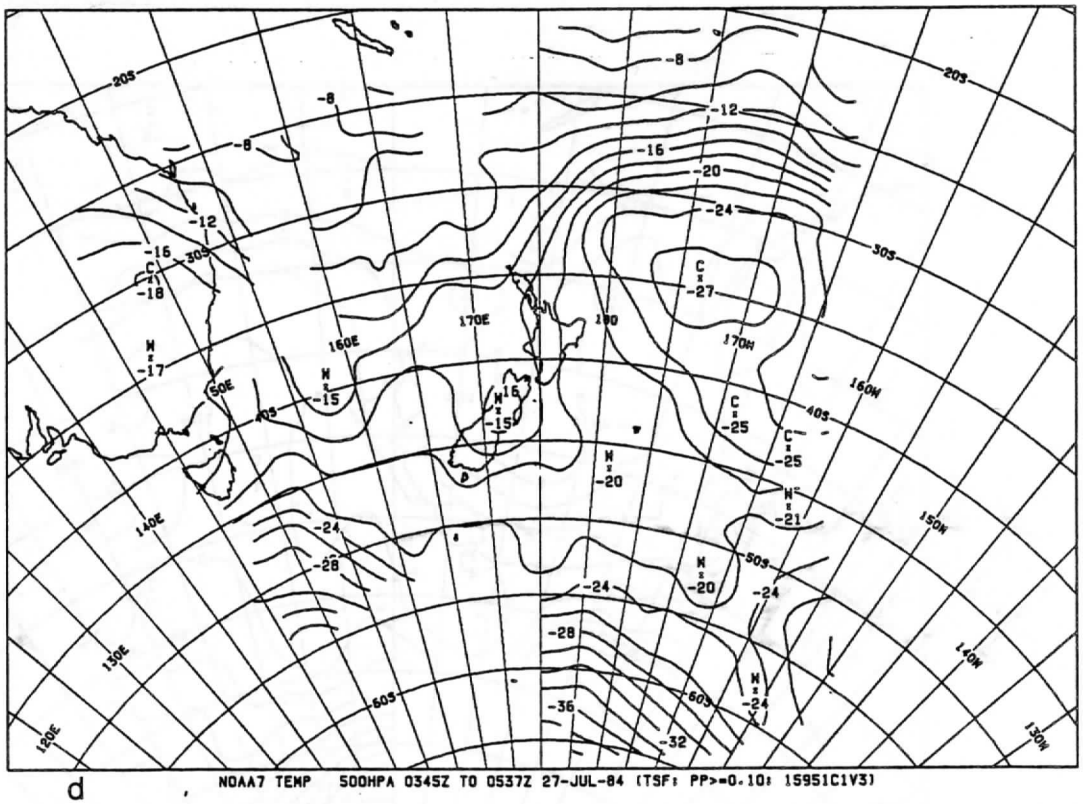


Fig 11. TSF retrieval analyses for; (a) 1000-500hPa thickness (the locations of the cross-section profiles of Fig 14 are indicated), (b) 850hPa temperature, (c) 850hPa dew point temperature, and (d) 500hPa temperature fields, for NOAA 7 orbits 15950 and 15951.

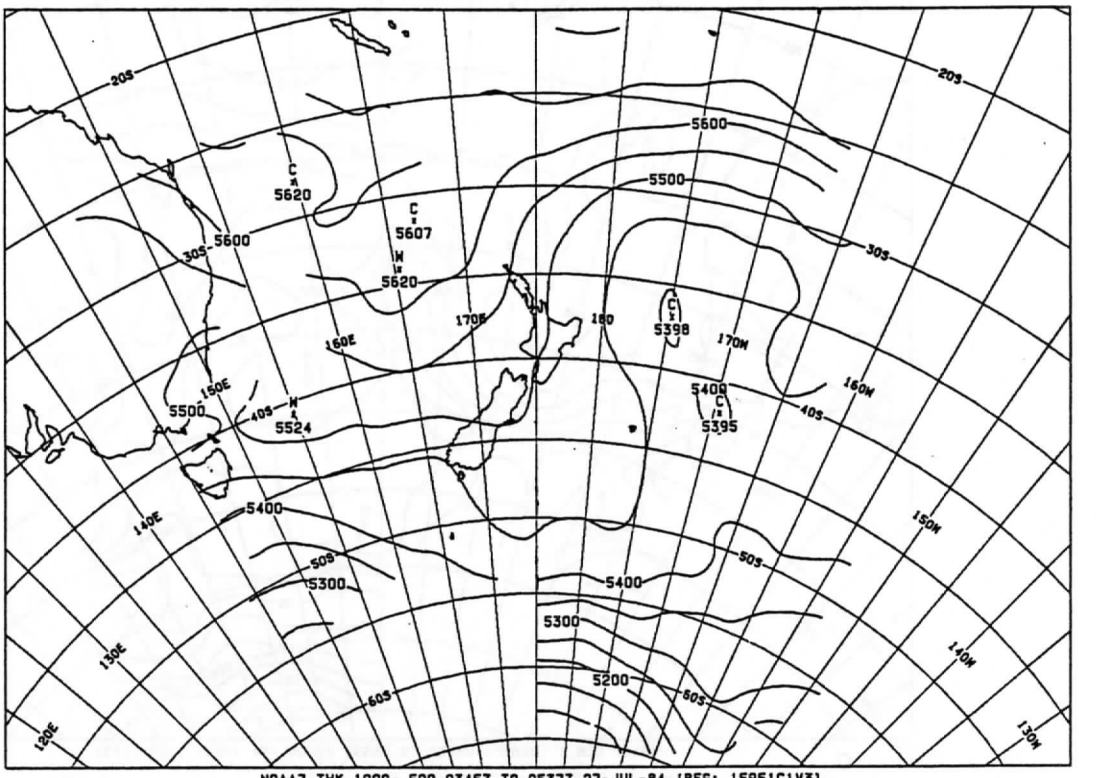


c

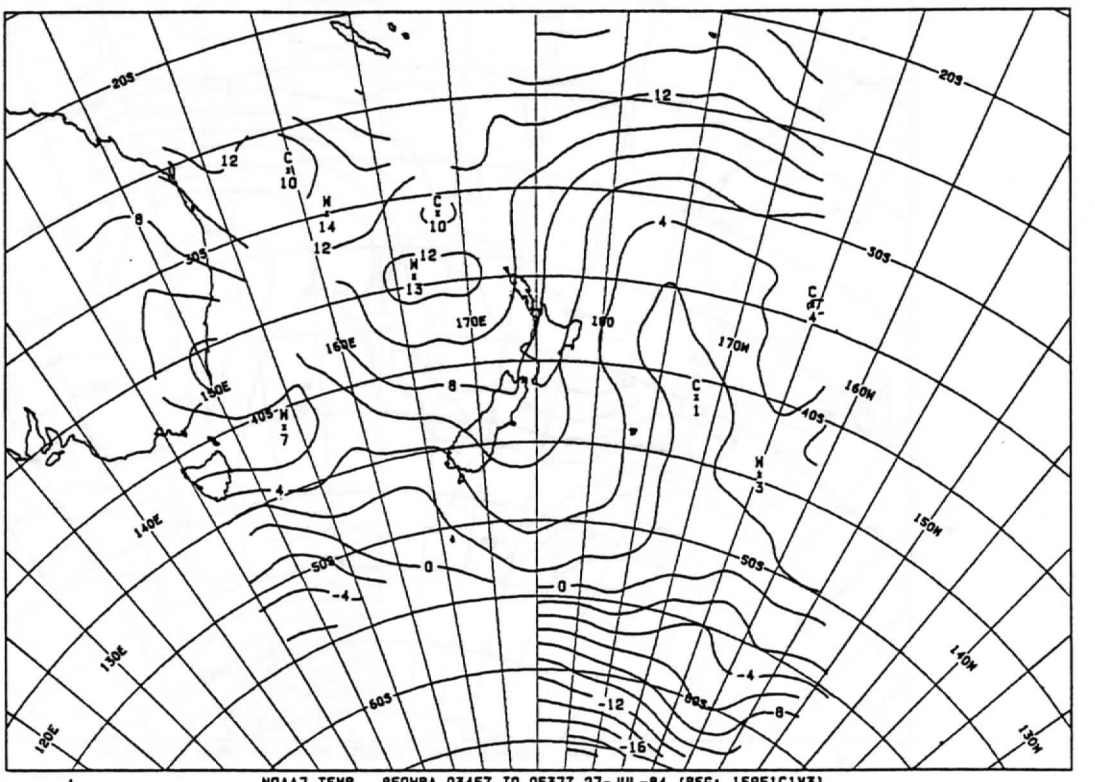


d

Fig 11. TSF retrieval analyses for; (a) 1000-500hPa thickness (the locations of the cross-section profiles of Fig 14 are indicated), (b) 850hPa temperature, (c) 850hPa dew point temperature, and (d) 500hPa temperature fields, for NOAA 7 orbits 15950 and 15951.

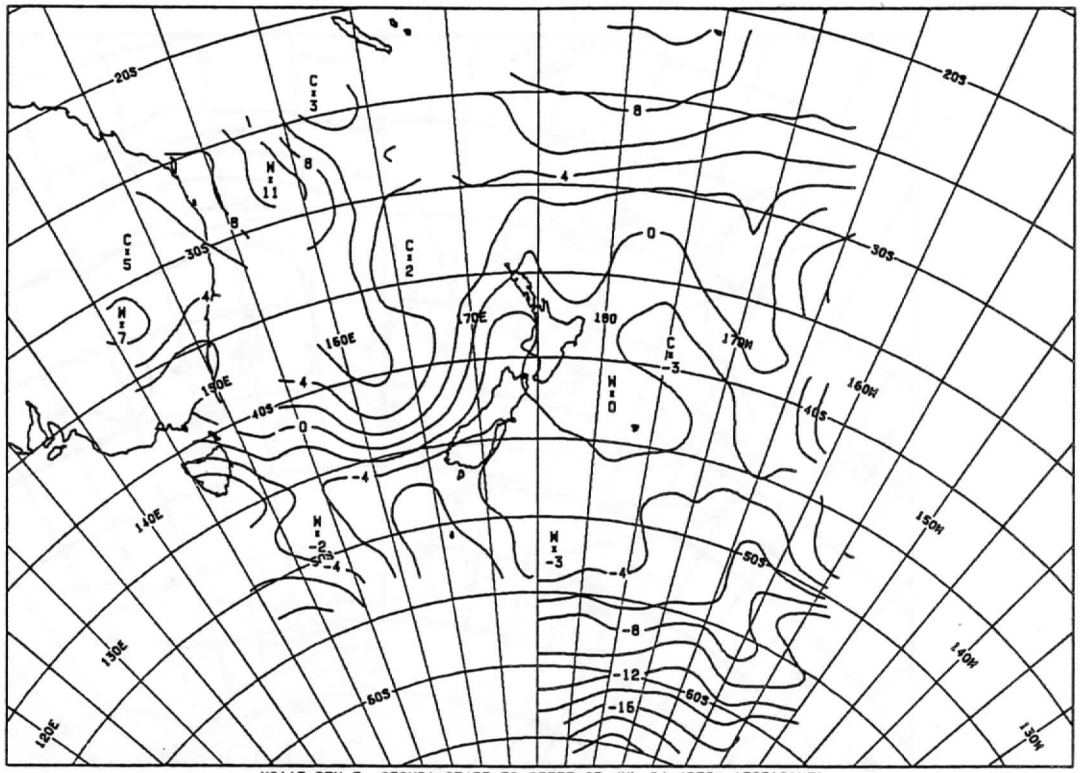


a

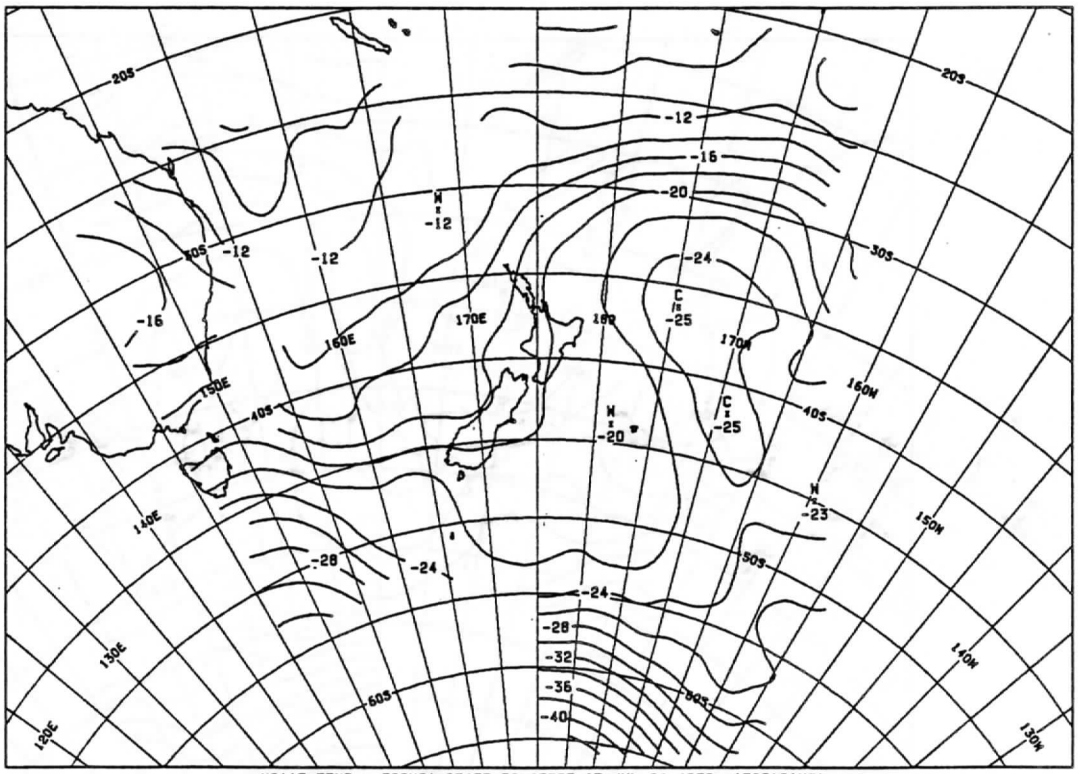


b

Fig 12. Regression retrieval analyses for; (a) 1000-500hPa thickness, (b) 850hPa temperature, (c) 850hPa dew point temperature, and (d) 500hPa temperature fields, for NOAA 7 orbits 15950 and 15951.

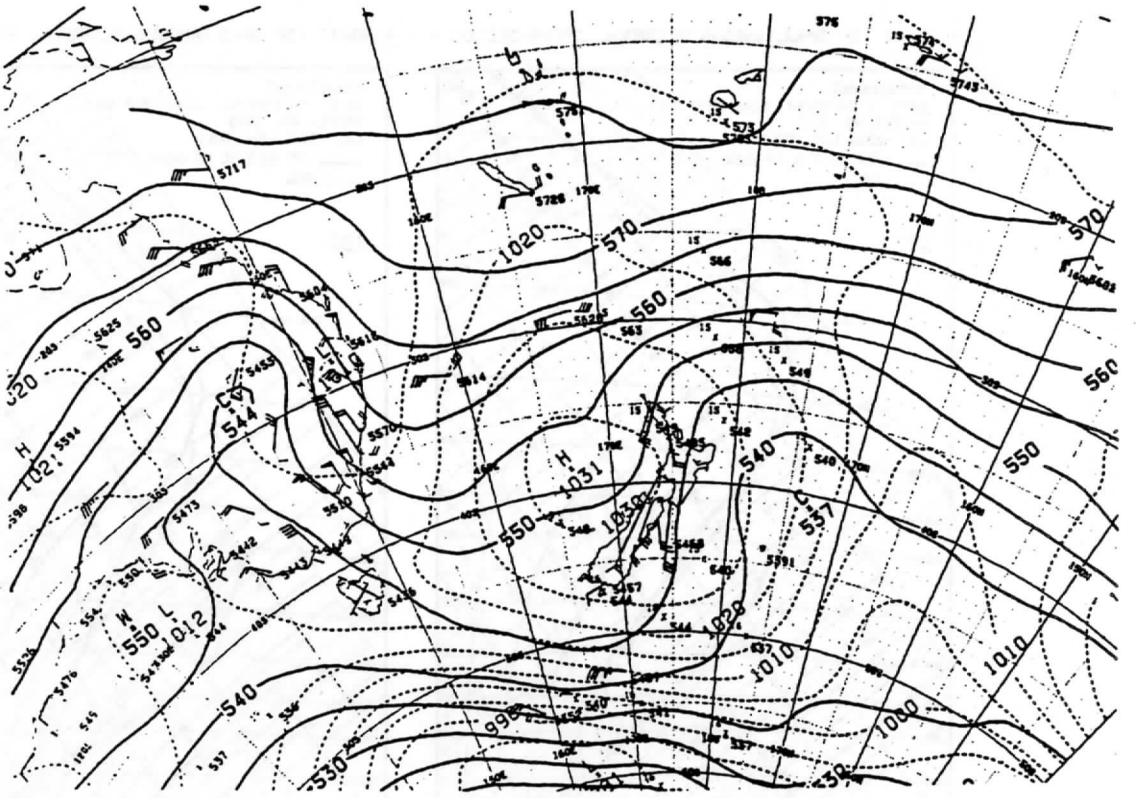


c



d

Fig 12. Regression retrieval analyses for; (a) 1000-500hPa thickness, (b) 850hPa temperature, (c) 850hPa dew point temperature, and (d) 500hPa temperature fields, for NOAA 7 orbits 15950 and 15951.



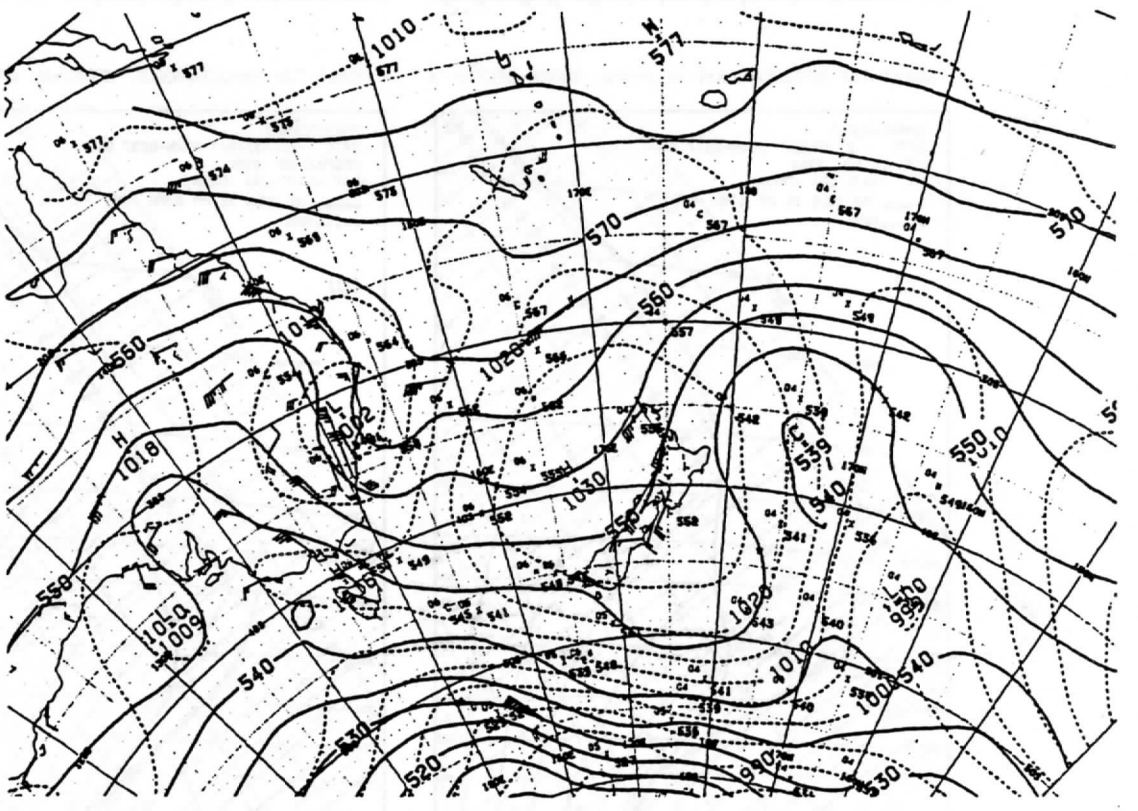


Fig 13. NWP analyses of 1000-500hPa thickness (solid curves) and surface pressure (dashed curves) fields at 0000 (a) and 0600 GMT (b) 27/Jul/84. Subsets of the GTS SATEM (marked by a cross) and locally derived SATEM observations are indicated (c refers to a clear retrieval, * to a cloudy retrieval, and M to a microwave only retrieval).

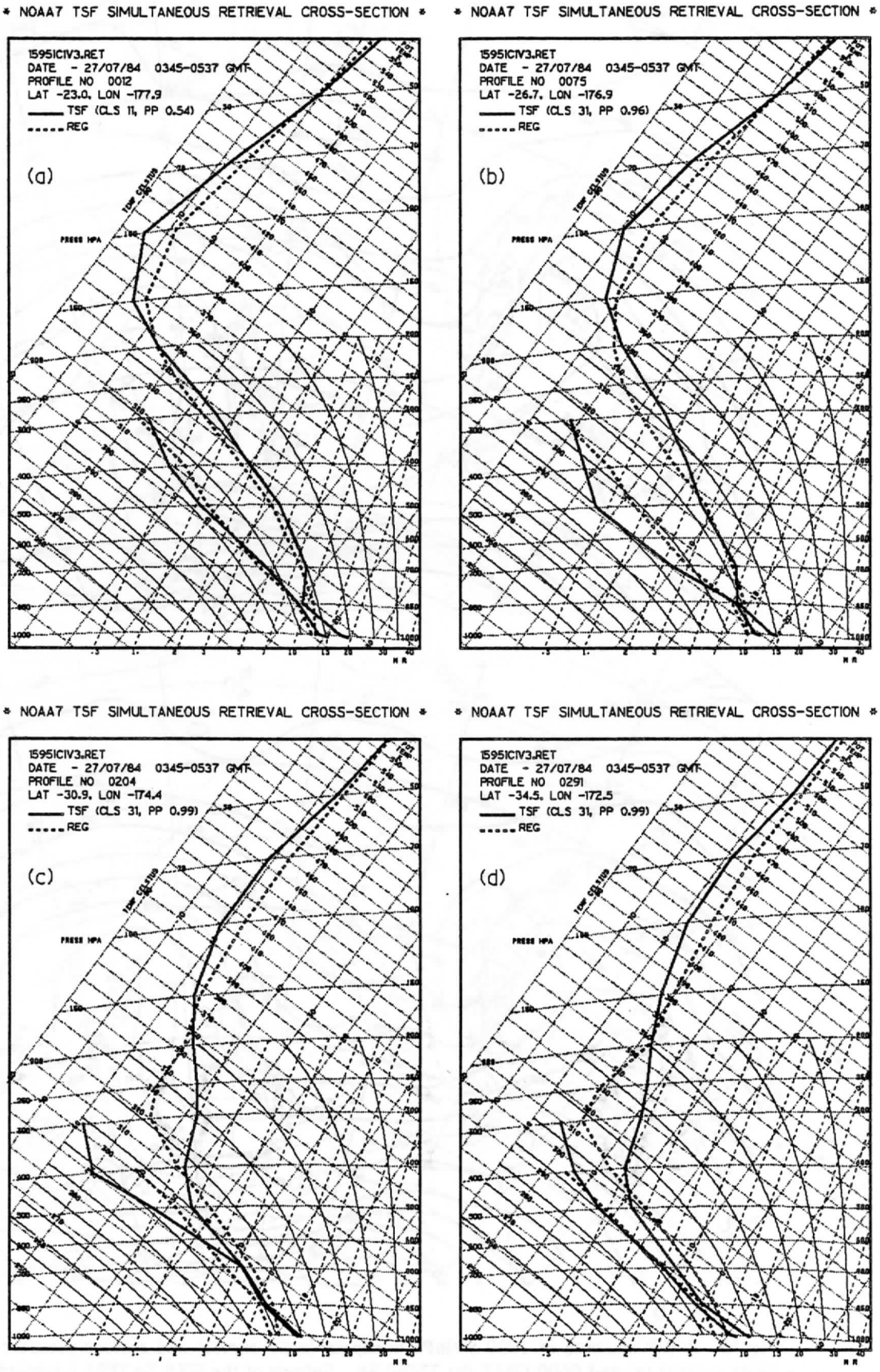


Fig 14. Tephigrams of TSF (solid curves) and Regression (dashed curves) retrievals for the cross-section indicated in figure 11(a).

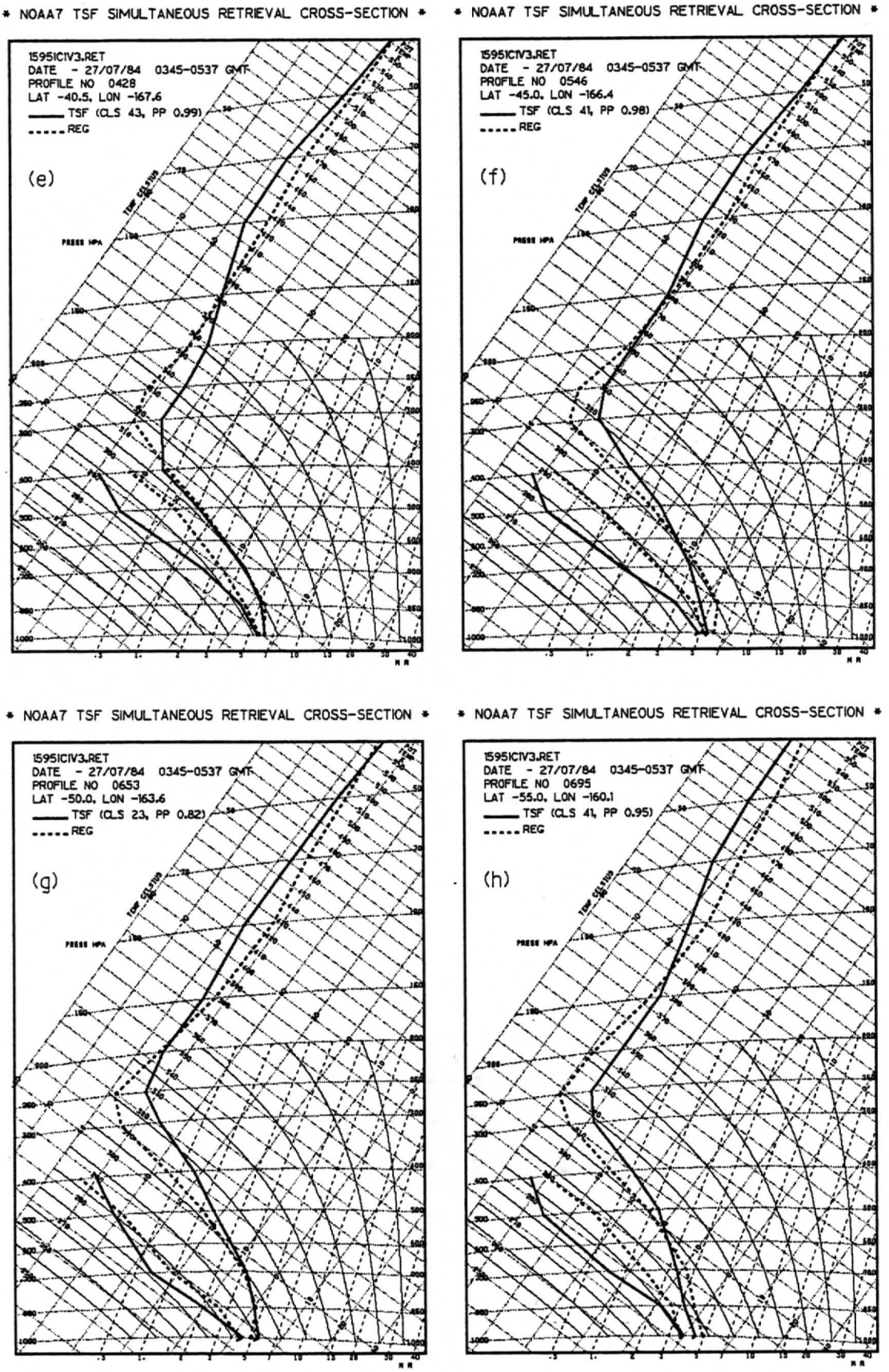


Fig 14. Tephigrams of TSF (solid curves) and Regression (dashed curves) retrievals for the cross-section indicated in figure 11(a).

The Technical Proceedings of

The Third International TOVS Study Conference

Madison, Wisconsin

August 13 - 19, 1986

Edited by

W. P. Menzel

Cooperative Institute for Meteorological Satellite Studies
Space Science and Engineering Center
University of Wisconsin
1225 West Dayton Street
Madison, Wisconsin 53706
(608)262-0544

November 1986