

A Statistical Algorithm for Clear-sky Temperature and Humidity Retrievals from TOVS Radiance

Jing-Shan Hong^{1, *} and Dongsoo Kim²

(Manuscript received 12 February 1999, in final form 17 December 1999)

ABSTRACT

In this study a regression algorithm was developed to retrieve the atmospheric parameters from TOVS (TIROS Operational Vertical Sounder) clear-sky radiances. The development of this algorithm requires calculations, at RAOB (Radiosonde Observation) sites, of the cross-covariances among fields obtained from rawinsonde observations, satellite radiances, and forecast products. The data were collected in every assimilation cycle for three months from May to August 1997.

The retrievals were verified statistically with the co-located RAOB measurements for two periods of experiments. The results show that the bias is comparable to the first guess. Compared with the forecast, the root mean square errors of the retrievals in the troposphere were better by about 1~1.5 K in temperature and 0.2~0.4 g/Kg in the mixing ratio. The root mean square error of retrievals at surface showed an improvement of 0.5~1 K in temperature and 1.5~2 g/Kg in the mixing ratio. The results during the August period were noticeably better than in the June period. This implies that the retrieved profiles were closer to the RAOB profiles than the 12-hour forecast profiles were.

The contour field of the retrievals showed a slight correction on the first guess field both in temperature and the mixing ratio except in the subtropical area. The results are also encouraging in that the retrievals in this case had a local coherence both in the horizontal and vertical.

(Key words: TOVS radiance, Regression retrieval)

1. INTRODUCTION

Satellite data now play an increasingly important role in Numerical Weather Prediction

¹Central Weather Bureau, Taipei, Taiwan, ROC

²Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado and NOAA, Forecast Systems Laboratory, Boulder, Colorado, USA

* *Corresponding author address:* Dr. Jing-Shan Hong, Computer Center, Central Weather Bureau, 64 Kung-Yuan Road, Taipei, Taiwan, ROC; E-mail: rfs14@cwb.gov.tw

(NWP). For example, the SATEMs (satellite temperature and humidity profile) generated by NOAA/NESDIS (National Environmental Satellite, Data and Information Service) are used in many operational NWP centers, such as the Central Weather Bureau (CWB) in Taiwan. Though investigations have shown that the SATEMs have a positive impact on the southern hemisphere and a neutral or negative impact on the northern hemisphere, the data show a systematic, air mass dependent bias arising from the included statistical or climatological component in the SATEMs (Andersson et al., 1991). Hong (1999) has recently presented the impact of SATEMs on the operational limited area model in the CWB and has concluded two major points. First, the objective analysis which includes the SATEM shows a low correlation at levels below 500 hPa. This is due to the low resolution of the SATEMs in the lower troposphere. Second, there is significant analysis increment around cloud areas, which may be due to the systematic bias of SATEM data in cloudy regions.

Recent progress in satellite data assimilation is due to two distinct features from earlier approaches. First, the analysis equation is written in nonlinear form, which must rely on intensive computer power (Eyre et al., 1993, Chou 1998). Second, the measured radiances are used instead of retrieved profiles, thereby ensuring error covariances used in the assimilation are being managed properly (Andersson et al., 1994, Derber and Wu 1998). However, the regression retrieval scheme has the advantage of simplicity, and the retrieved profiles are convenient as they have the same physical unit of predictive variables as the NWP model.

In general, for the regression approach, the desired solution can be written in finite difference form as:

$$\mathbf{T} = \mathbf{C}\mathbf{x},$$

where \mathbf{T} is a matrix whose columns are the temperature differences from the mean for one particular profile out of a large sample; \mathbf{x} is a matrix whose columns represent the brightness temperature differences for a given profile; and \mathbf{C} is the matrix of solution coefficients which are given from a large sample of observed brightness temperatures and initial guess and which can be expressed as:

$$\mathbf{C} = \mathbf{S}_{\mathbf{T}\mathbf{x}} \mathbf{S}_{\mathbf{x}}^{-1}, \quad (1)$$

where $\mathbf{S}_{\mathbf{T}\mathbf{x}}$ and $\mathbf{S}_{\mathbf{x}}$ stand for sample cross-covariance and covariance matrices. In general, regression retrieval methods can be considered as being made up of two primary components: a method of selecting the initial guess, and a retrieval algorithm which modifies the initial guess. For example, several techniques based on constrained regression have been developed and discussed for satellite meteorology application (Crone 1996). With regard to the initial guess, it has been common practice to use climatology (Smith et al., 1979). However, the initial guess profiles used in the retrieval are spatially correlated and the correlated errors can propagate to the retrieved profiles, which means isolating retrieval errors is made impossible. Thus, the NWP forecast results, which provide a better estimate than climate or a library of profiles, are recommended as the initial guess fields (Hayden 1988, Eyre et al., 1993).

In this paper, a retrieval algorithm that takes advantage of the existing analysis system as well as the real-time ingestion of radiance data from the NOAA's polar orbiting TIROS-N series satellites is presented. The retrieval scheme proposed by Kim (1993) is based on statistical regression and requires information from the cross-covariances among the observed and

computed radiance, temperature/humidity profiles from rawinsonde as well as forecasts. The computed radiances are simulated by using a radiative transfer model (called RTTOV) developed by Eyre (1991), and the first guess profile is the 12-hour forecast products from the operational Limited Area Forecast System (LAFS, Yeh et al., 1994) in the CWB.

This paper mainly focuses on the evaluation of the performance of the regression retrieval system in the CWB LAFS. A description of the algorithm, data processing, and the statistics for retrievals versus co-located RAOB (RADiosonde OBServation) measurements are presented in the following sections.

2. METHODOLOGY AND DATA PROCESSING

Figure 1 presents the flow chart of the regression retrieval procedure from TOVS radiance in the LAFS. The procedure is composed of three major components. First, the computed radiances are simulated by the radiative transfer model based on the atmospheric profile provided by the LAFS's 12-hours forecast. The radiative transfer model used here was the European Center for Medium-Range Weather Forecasts' (ECMWF) forward model called RTTOV (Eyre 1991). The information of the cloud fraction used in the forward model was estimated by the observed AVHRR (Advanced Very High Resolution Radiometer) pixels within the HIRS (High resolution Infrared Radiation Sounder) field of view (fov). Since only clear cases are presented in this paper, the complexity of the estimation of cloud information has been greatly reduced. Second, the difference between the observed radiances and computed radiances were taken as the observation increment. Third, the observation increment in radiances was applied to retrieve the profile increment from the first guess. The term of first guess was defined as the LAFS's 12-hours forecast here.

The use of locally ingested radiance data provides important advantages in three ways: (1) The NWP center can understand error sources attributed to satellite data processing procedures, such as cloud clearing, limb correction and bias corrections. (2) The NWP center can exploit full resolution AVHRR data in screening cloud affected HIRS sounder data. (3) The NWP center can monitor radiance data in the region and easily check its quality.

2.1 Algorithm for Regression Retrieval

Here we used the constrained least-square estimation proposed by Kim (1993) to retrieve the profile increment (the details are shown in Appendix):

$$T' = M(y) + S_{yx} [\lambda E + S_{xx}]^{-1} [x - M(x)], \quad (2)$$

where T' is the retrieved profile increment from the first guess which is defined as the LAFS's 12-hours forecast; x is the observation increments in brightness temperature (HIRS channel 3 - 8 and 10 - 15), which is the difference between the observed brightness temperature and the computed brightness temperature from first guess; y is the increment in atmosphere profiles, which is the difference between RAOB and the first guess.

There are 18 elements in the profile increment (T' and y) vectors which are composed of

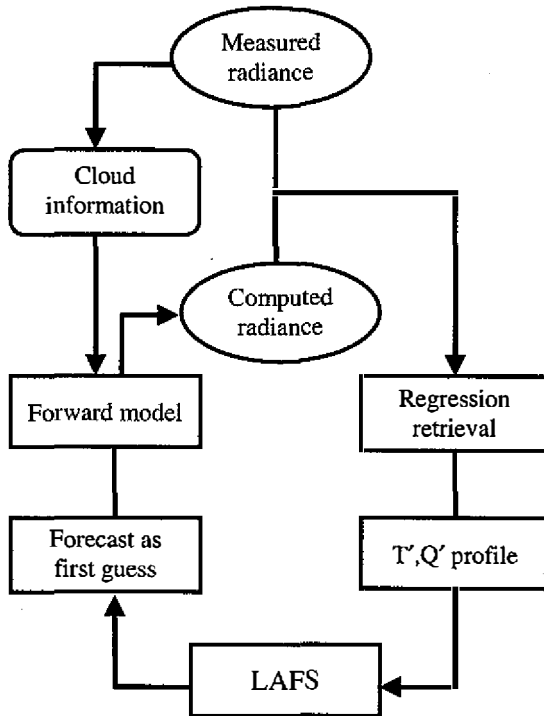


Fig. 1. Procedure of the regression retrieval from the TOVS radiance in the LAFS System.

9 mandatory pressure levels in temperature (1000 - 200 hPa), 7 levels in the mixing ratio (1000 - 300 hPa), surface temperature, and the surface mixing ratio. Thus, (2) provides a simultaneous retrieval of the temperature and humidity profile as well as surface information. The matrices in (2) are defined as follows:

$$\begin{aligned}
 \mathbf{M}(y) &= \langle \mathbf{P}_r - \mathbf{P}_b \rangle \\
 \mathbf{M}(x) &= \langle \mathbf{R}_o - \mathbf{R}_b \rangle \\
 \mathbf{E} &= \text{cov}(\mathbf{R}_r - \mathbf{R}_b) \\
 \mathbf{S}_{xx} &= \text{cov}(\mathbf{R}_o - \mathbf{R}_b) \\
 \mathbf{S}_{yx} &= \text{cov}(\mathbf{P}_r - \mathbf{P}_b, \mathbf{R}_o - \mathbf{R}_b)
 \end{aligned}$$

where

- R_o : measured brightness temperature;
- R_b : computed brightness temperature from the NWP profiles;
- R_r : computed brightness temperature from the RAOB profiles;
- P_b : the NWP profiles of T and q ;
- P_r : the RAOB profiles of T and q ;

$\langle \rangle$ stands for the sample mean; $\text{cov}(a)$ represents the autocovariance matrix of a ; and $\text{cov}(a,b)$ is the cross-covariance of variable a and b for a large sample collocated at the RAOB site. Since the matrix E is the covariance between the computed radiance from the RAOB and first guess, it represents errors contained in the radiative transfer model and the NWP forecast. The scalar λ in (2) is the weighting of E relative to S_{xx} . In the case of $\lambda=0$, the number of covari-

ance matrices in (2) is reduced to two, and then the coefficient matrix in (2) is similar to (1). This implies that the errors produced by the radiative transfer model and the NWP forecast were ignored. The value of λ can be determined statistically. Kim (1993) suggested that λ could be obtained via cluster analysis. We set $\lambda=1$ in this paper. The value was somewhat chosen arbitrarily

Obviously, there are many weaknesses in the statistical regression algorithm.

- (a) The linear form in (2) is unable to resolve the nonlinear characteristics of the atmosphere.
- (b) The least-square solution tends to underestimate deviations in the predictands from their mean when the predictors have measurement errors.
- (c) As the regression coefficient is determined in a certain period and area, the retrievals may be constrained by the specific weather pattern and may therefore be difficult to correct when such bias exists. Thus, it is necessary to update the error covariances to improve the results.
- (d) A long period of time is required to collect the necessary samples because the construction of the covariance matrix in (2) relies on a large number of RAOB measurements.

On the other hand, the regression method in (2) also has several advantages.

- (a) E and S_{xx} are symmetric and positive definite such that $S_{xx} + \lambda E$ is non-singular and its inverse can be solved efficiently.
- (b) Since the calculation in (2) is only the multiplication of the matrices, high computing performance can be reached once the covariance matrices are determined.
- (c) (2) uses the NWP forecast as the initial guess, which can be expected to provide a better estimate than climate or a library of profiles (Eyre 1988).
- (d) The rawinsonde data, which represents the ground truth, were used in (2) to construct the covariance matrices. Thus, the retrieved profile can be statistically expected to be closest to the rawinsonde (truth).
- (e) E and S_{xx} implicitly contain a systematic error in the NWP forecast as well as in the forward radiative transfer model. This is especially important due to the great uncertainty in the surface information.
- (f) Since the solution in (2) is the retrieved profile increment and has the same physical unit of predictive variables, these are the most suitable to ingest to the OI (Optimal Interpolation) module as an independent measurement, which is able to avoid the correlation between the background and observation error.

2.2 The NWP Model

One of the important components in (2) is the first guess supported by the NWP model. The model used here was the Limited Area Forecast System (LAFS) which is operated daily at the Central Weather Bureau (CWB) in Taiwan. The LAFS is a hydrostatic PE model incorporating various physical parameterizations: the Kuo scheme (Kuo 1974) for cumulus param-

eterization, long wave and short wave radiative transfer calculations following Harshvardhan et. al. (1987) and the turbulent kinetic energy (TKE) dissipation scheme for planetary boundary layer (PBL) parameterization (Detering and Etling 1985). The vertical coordinate, defined by the terrain pressure at 20 levels, is used in the model. The governing equation is horizontally differentiated on the Arakawa-C grid system with a coarse mesh domain of 161×121 and nested fine mesh size of 91×91 . The resolution is 60 km in the coarse domain and 20 km in the fine domain. Thus, the entire model domain covers most of Asia and the Western Pacific Ocean, as shown in Fig. 2.

The objective analysis in the LAFS is the OI scheme with an update cycle of 12 hours. Most of the upper air observations in the LAFS are RAOB data. The distribution of the RAOB sites within the LAFS domain is shown in Fig. 2. The figure also shows the rarity of RAOB measurements over the ocean area. In some circumstances, the subtropical High over the Pa-

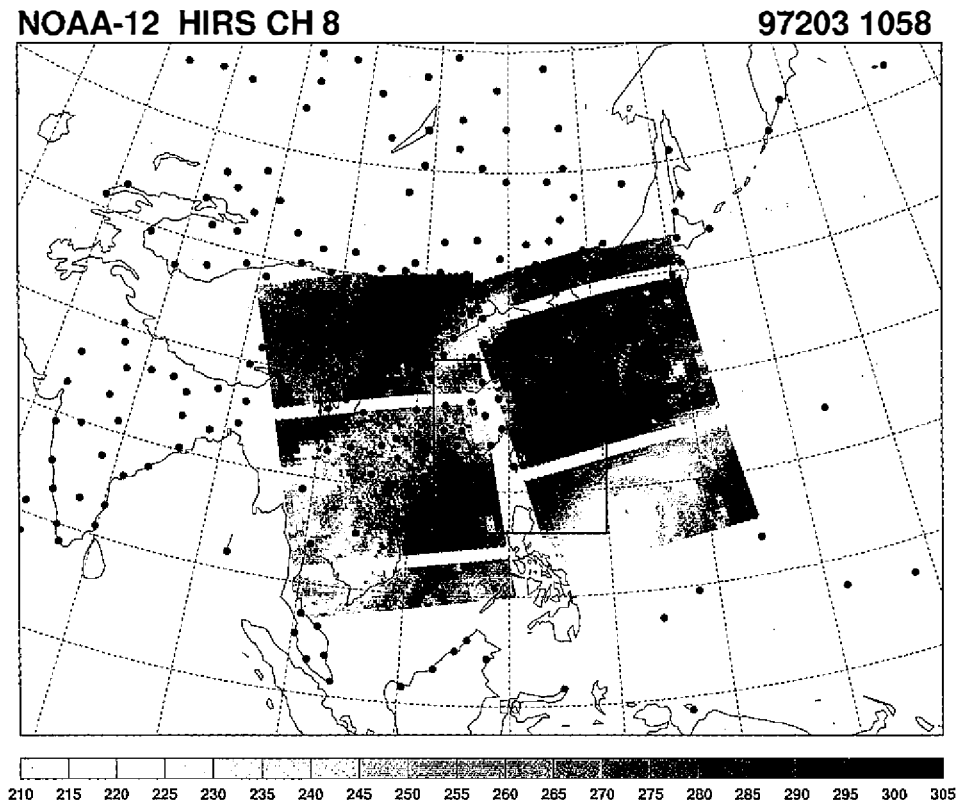


Fig. 2. HIRS channel 8 brightness temperatures (K) in two consecutive paths of the NOAA-12 satellite when it passed over Taiwan in 0918 LST and 1058 LST June 22 (Julian Day 203 as shown in the upper-right corner), 1997. The outer and inner boxes are the coarse and fine domains in the LAFS, respectively. The heavy dots are the RAOB site used in the LAFS assimilation system.

cific Ocean play an important role to in determining the weather system around the East Asian area, for example, the movement or evolutions of a Mei-Yu front system during the early summer, a typhoon track and even thermal thunderstorm activities in summertime etc. However, it is clear that there are not enough observations over the ocean area to correct forecast errors. Figure 2 also shows the image of the observed radiance of HIRS channel 8 in two typical consecutive paths of the NOAA-12 polar orbiting satellite when it passed over Taiwan.

In order to construct the covariance matrix in (2), it was necessary to collect the measured radiances and profiles from NWP forecasts and RAOB measurements long enough to ensure statistical stability. Since the error covariance matrices in (2) were constructed at the RAOB site, the NWP forecasts were interpolated from the model grid to the RAOB station. The representative brightness temperatures of TOVS observations at the RAOB site are the average of the possible clear *fovs* which are located within the area of 0.5° centered at the RAOB site. In the meantime, not only did these data have to be collocated correctly, but also they were subject to quality control. The pre-processing of the data set is discussed in greater detail in the following sub-section.

2.3 Measured Radiance

The TOVS and AVHRR data were extracted from the raw High-Resolution Picture Transmission (HRPT) data that are operationally received by the Meteorological Satellite Center in the CWB. The available satellite paths over Taiwan were chosen within a time window of ± 3 hours at synoptic times, 00Z and 12Z. Multi-channel measurements were required in (2) to retrieve the profile increment. Table 1 presents the descriptions of the characteristics of the TOVS channels. In this study, only HIRS channels 3-8 and 10-15 were considered.

One of the most important concerns in satellite measurement is the effect of clouds. Of particular concern in this study was determining which HIRS pixels were cloud-free. The HIRS instrument resolves a circular area that is 17.4 km in diameter at the satellite subpoint, whereas the AVHRR resolves a circular area that is 1.1 km in diameter at the satellite subpoint. Thus, each HIRS *fov* can collect about 300 - 450 AVHRR pixels depending on the viewing angle. The statistics of the AVHRR pixels within each HIRS *fov* can then provide the opportunity to estimate the cloud information of that HIRS *fov*.

In this paper, we adopted the method proposed by Aoki (1980) which involves matching and extracting the AVHRR pixels in each HIRS *fov*. The methodology to estimate the cloud information followed that of Kim and Chou (1995) and Kim (1996). Spatial cloud distribution was not of interest here, but the fraction of cloud coverage was. The original design in Kim (1996) involves computing the smoothed density function of the AVHRR pixels in each HIRS *fov* and finding the local minimum as a threshold in order to separate one cloud mass from another. Thus each cloud mass has a unique representative value which implies a unique cloud-top height. The simulation results in Kim (1996) show that the method is capable of estimating the proportion of clouds with different cloud-top heights within a HIRS *fov*. Following Kim's algorithm, the density function of the AVHRR pixels should exhibit a single peak pattern when the atmosphere is cloud-free or overcast. Figure 3 is an example of the density function of the AVHRR pixels with a single peak in a TOVS *fov*. A total of 324 AVHRR pixels were

Table 1. Characteristics of HIRS channels (adapted from Smith et al., 1979). The shading indicates the channel used to retrieve the atmospheric sounding.

HIRS channel	Central wavelength (μm)	Principal absorbing constituents	Level of peak energy contribution	Purpose of the radiance observation
1	15.00	CO ₂	30 hPa	Temperature sounding
2	14.70	CO ₂	60 hPa	
3	14.50	CO ₂	100 hPa	
4	14.20	CO ₂	400 hPa	
5	14.00	CO ₂	600 hPa	
6	13.70	CO ₂ /H ₂ O	800 hPa	
7	13.40	CO ₂ /H ₂ O	900 hPa	
8	11.10	Window	Surface	Surface temperature
9	9.70	O ₃	25 hPa	Total Ozone
10	8.30	H ₂ O	900 hPa	Water vapor sounding
11	7.30	H ₂ O	700 hPa	
12	6.70	H ₂ O	500 hPa	
13	4.57	N ₂ O	1000 hPa	Temperature sounding
14	4.52	N ₂ O	950 hPa	
15	4.46	CO ₂ /N ₂ O	700 hPa	
16	4.40	CO ₂ /N ₂ O	400 hPa	
17	4.24	CO ₂	5 hPa	Surface temperature
18	4.00	Window	Surface	
19	3.70	Window	Surface	
20	0.70	Window	Cloud	Cloud detection

collected within that TOVS *fov*. The figure shows the homogeneity of the AVHRR brightness temperatures with the mean of 292.49 K and the standard deviation of 0.43K. However, it was necessary to further distinguish whether the HIRS *fov* was a clear or overcast case when the density function of the AVHRR pixels exhibited as a single peak pattern. The HIRS *fov* could be regarded as overcast if the difference between the observed and computed (from the NWP forecast) brightness temperatures in HIRS channel 8 was greater than 4 K. Figure 4 is the distribution of clear TOVS *fov* sifted according to the above algorithm in the case of Fig. 2. The figure shows most of the cloudy area was in fact removed.

2.4 Computed Radiance

The forward model used in this study was the RTTOV that was first designed for a satellite sounding system (Eyre 1991). This model requires an input profile vector of 67 elements,

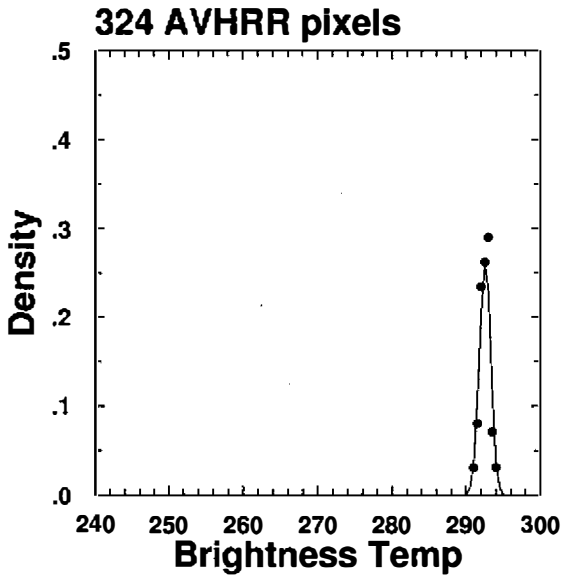


Fig. 3. Density plot of the AVHRR channel 4 brightness temperatures in a HIRS field of view. The horizontal and vertical axes are the brightness temperature (K) and density, respectively. A total of 324 AVHRR pixels were collocated in the HIRS field of view.

Clear FOV

972031058

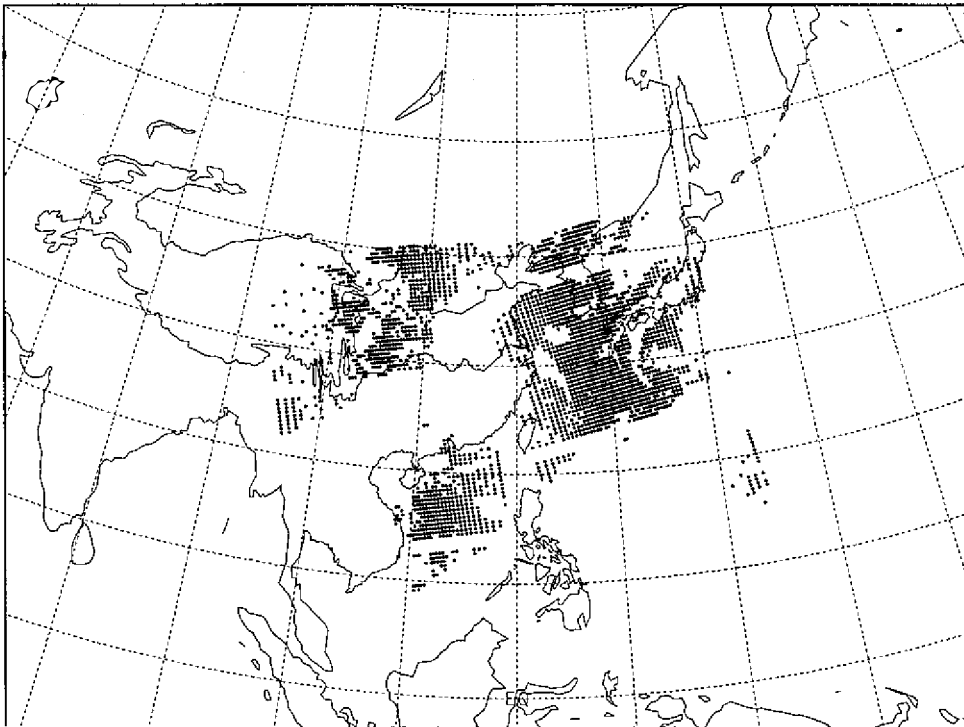


Fig. 4. Distribution of clear-sky HIRS pixels as determined by the procedure described in Fig. 1.

including temperature at 40 fixed pressure levels, mixing ratios at 15 fixed levels, skin temperature, surface temperature, surface mixing ratio, and surface pressure. These input profile vectors were interpolated or extrapolated from the NWP grid data or RAOB profile.

Figure 5 is the scatter diagram of the measured versus computed brightness temperature which was calculated from the RAOB profile during 20 June to 15 August 1997. HIRS channels 3 - 8 and 10 - 15 in Fig. 5 were used in the study (the characteristics of the TOVS channels are described in Table 1). The exclusion of HIRS channels 1 and 2 was based on the fact their peak level of the weighting function was far from the retrieval levels of interest (lower than 100 hPa) in (2). Channels 18 and 19 were also excluded because the reflection of the sunlight in daytime had a large impact on the observed radiances. A bit of scatter in channels 11 and 12 (water vapor channel), as shown in Fig. 5, was probably due to the inappropriate treatment of the humidity profile. The small bias and root mean square error shown in Fig. 5 not only implies a good performance on the radiative transfer calculations but also provides a good basis for statistical regression retrieval.

3. RESULTS

Two series of experiments were designed to demonstrate the performance of the regression retrieval. In the first experiment, hereafter referred to as the June experiment, the necessary data from a total of 248 clear samples were collected at the RAOB site from 15 May to 20 June 1997 to construct the error covariance matrices. The statistics of the error covariance matrix became stable as the sample size became greater than 150 (not shown). Using the covariance constructed in the June experiment, the regression retrieval experiment collocated at the RAOB site was processed from 20 June to 1 July 1997 and verified by the RAOB measurements. In the second experiment, hereafter referred to as the August experiment, the covariance matrices were constructed using the data collected from 20 June to 1 August 1997 (239 cases), and the retrieval results were verified from 1 August to 15 August 1997. The two experiments represent two kinds of weather regimes. The June experiment was sampled during the Mei-Yu season, which is the seasonal transition period and the atmosphere is more convectively unstable (Chen 1983,1991). The August experiment was sampled during summer time that was mainly dominated by the subtropical High and tended to be much more stable. Figure 6 is the regression coefficient: $C_{yx} = S_{yx}^{-1} [\lambda E + S_{xx}]^{-1}$, for the June and August experiments, where the sub-script x denotes the channel index (HIRS channel 3 - 8, 10 -15) and y is the profile index (temperature and mixing ratio for mandatory and surface levels). The figure shows that the coefficients in the June experiment either at peak value or the square summation were larger than those in the August experiment. The distribution of the coefficient was rather smooth for the August experiment. Because of the weather-regime dependence of the regression retrieval, the two experiments provided the opportunity to highlight the differences in the regression coefficients under various weather regimes.

Figure 7 is the bias of the 12-hour forecast error (defined as the RAOB minus the first guess) and the retrieved profile (defined as the RAOB minus retrievals) for the temperature and mixing ratio in the June experiment (86 samples). Figure 8 is the August experiment (117 samples). Figures 7 and 8 show that the LAFS has a cold/wet bias through the troposphere, a

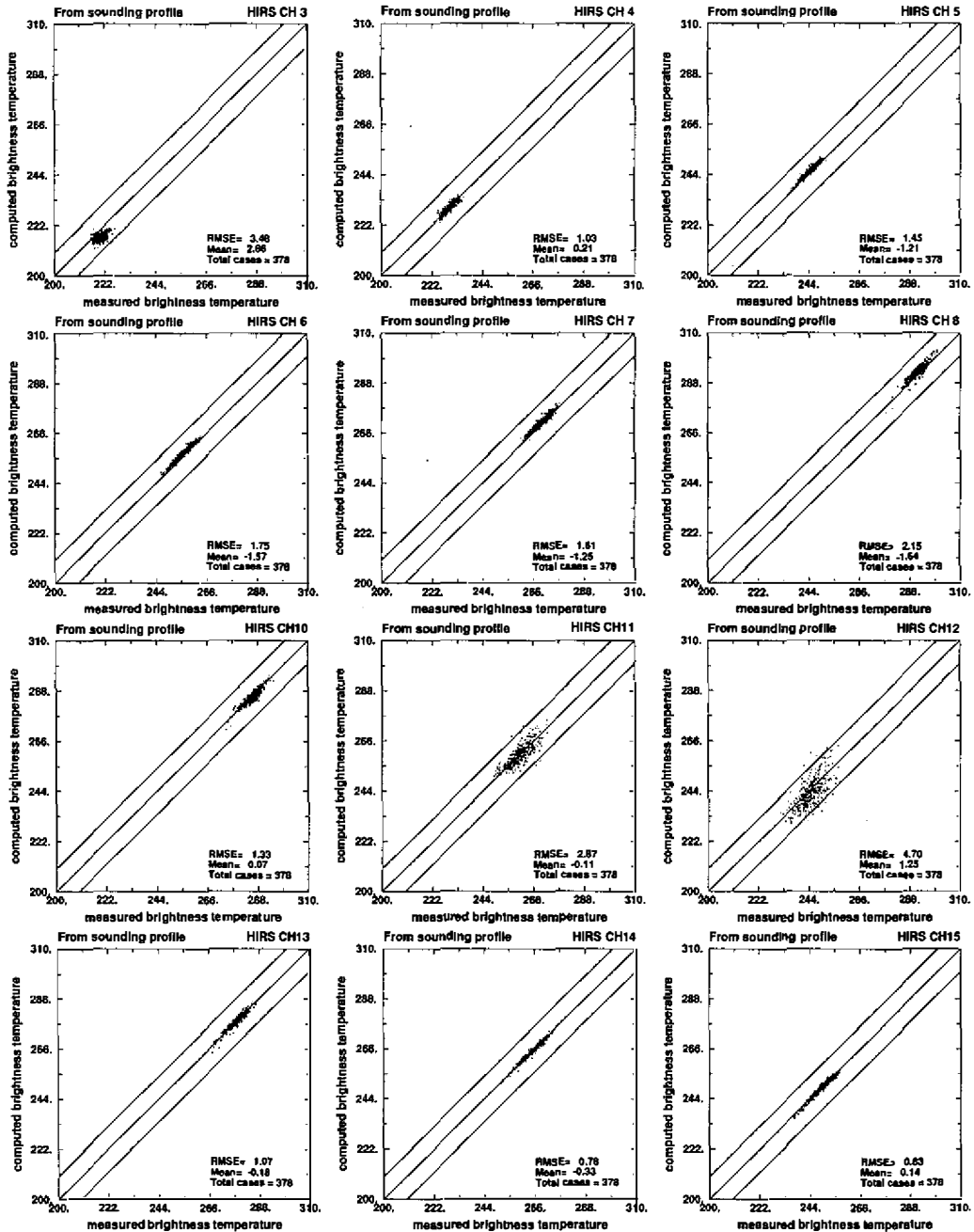


Fig. 5. Scatter diagrams of the measured HIRS brightness temperature (K) versus the computed brightness temperature (K) which were calculated from the RAOB profile. The sample number, bias (stands for mean), and root mean square errors (stands for RMSE) are shown in the bottom right corner of each panel.

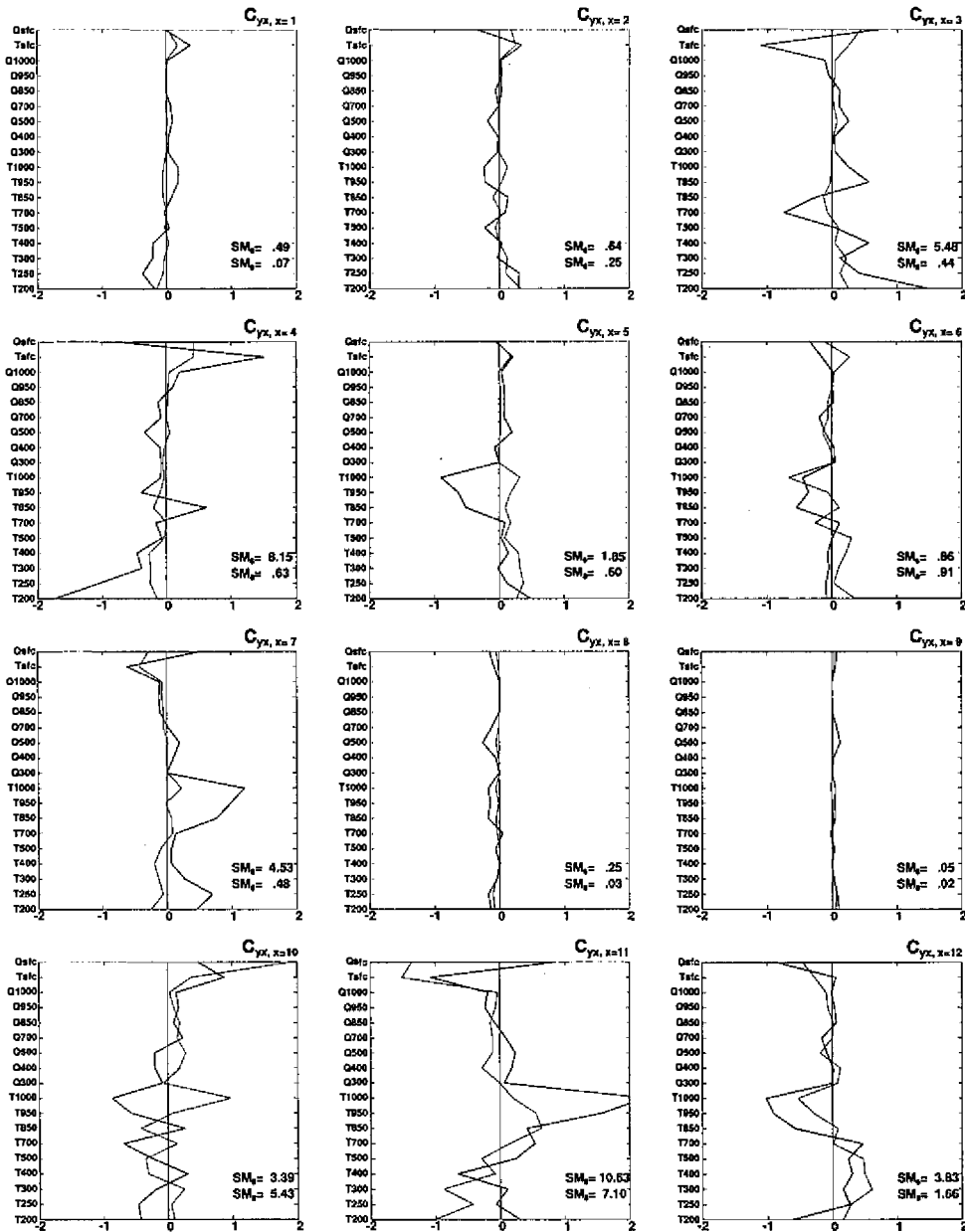


Fig. 6. Distributions of regression coefficients, $C_{yx} = S_{yx}^{-1} [\lambda E + S_{xx}]^{-1}$, for the 12 channels (x , x is the channel index) with respect to the components in the profile (y , y indicates the profile index, see details in context). Thick line: June experiment. Thin line: August experiment. The horizontal axis is the value of the coefficient and the vertical axis is the profile index. SM_y and SM_y are the sum of the squares of the coefficients for the y index in the June and August experiments, respectively.

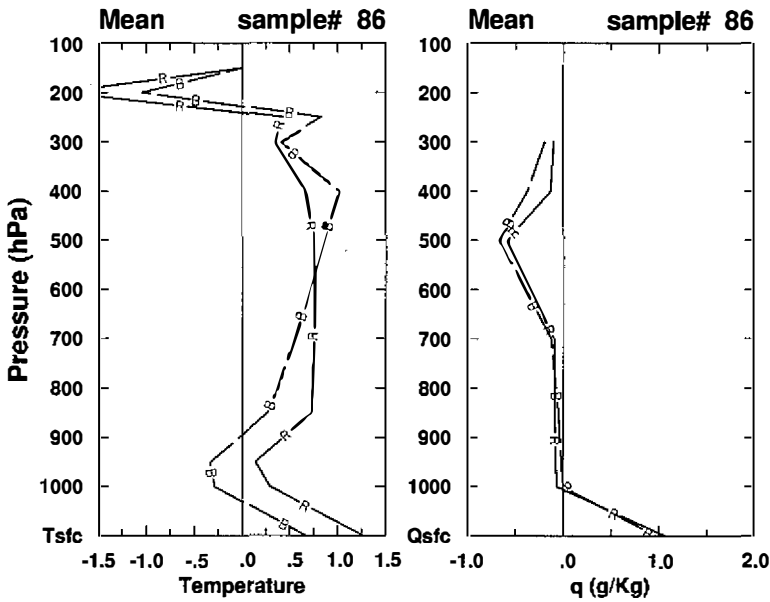


Fig. 7. Bias (June experiment) of the 12-hour forecast error defined as the RAOB minus the first guess (dashed line with B) and the retrieved profile defined as the RAOB minus retrievals (solid line with R) for the temperature (K) and mixing ratio (g/Kg). A total of 86 samples were collected from 20 June to 1 July 1997.

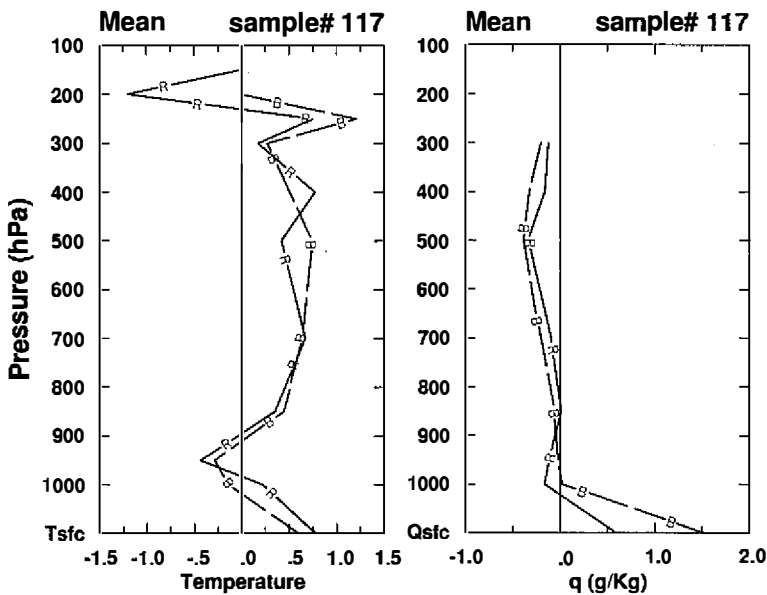


Fig. 8. Same as Fig. 7, but for the August experiment, A total of 117 samples were collected from 1 August to 15 August.

warm bias in the near-surface layer, and a cold/dry bias at the surface layer. The bias of the retrieved temperature profile is larger than the forecast errors below 500 hPa in Fig. 7 but closer to the forecast errors through the entire atmosphere in Fig. 8. The bias of the retrieved mixing ratio profile is quite close to the forecast errors in Figs. 7 and 8. That the retrieval bias is in the “vicinity” of the first guess is not surprising because the retrieval procedure is first guess dependent.

Figures 9 and 10 are the root mean square errors of the June and August experiments. The reference ground truth is supposed to be the RAOB measurements. The figures show that the root mean square error of the forecast temperature error is about 1~1.5 K within the troposphere and up to 2.1 K at the surface. However, the retrieved profile is only 1 K through the troposphere and 1.4 K at the surface, which is significantly smaller than the forecast error. The retrieved mixing ratio profile increment also has an improvement of 0.2~0.4 g/Kg in the troposphere and up to 1.2 g/Kg at the surface. This improvement implies that the retrieved profiles are closer to the RAOB profiles (as truth) than the 12-hour forecast profiles.

Because the weather system during summer is primarily dominated by the sub-tropical High, which is more steady and stable, the root mean square error of the forecast errors in Fig. 10 is reduced substantially, especially the mixing ratio in the mid-troposphere. The figure shows that the smaller root mean square error of the retrievals up to 0.5 K in temperature and less than 0.4 g/Kg in the mixing ratio is achieved. The root mean square error of retrievals at surface has an improvement of 1 K in temperature and 2 g/Kg in the mixing ratio. The results are noticeably better than those in the June experiment.

Figure 11 shows the contour field of the retrieved temperature on 700 and 500 hPa for every 3 HIRS fov on 0928 LST, 27 Aug 1997. As expected, the patterns of both the retrieved

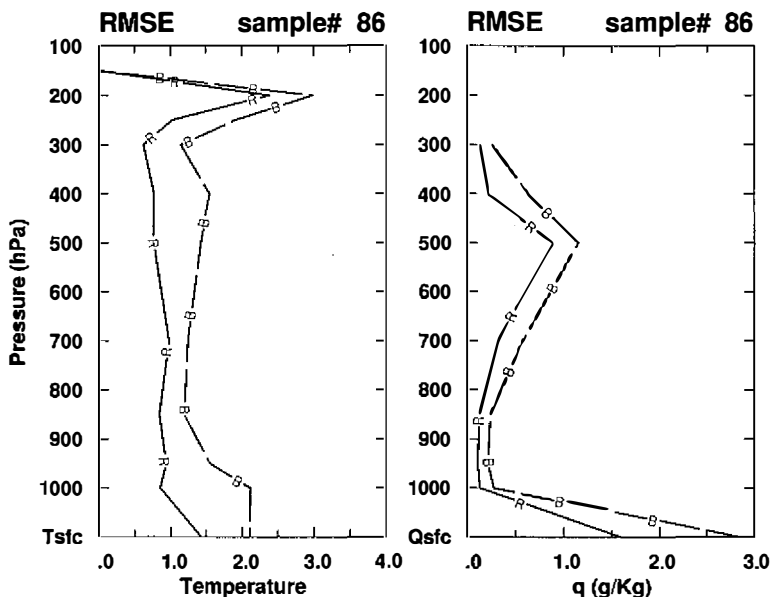


Fig. 9. Same as Fig. 7, but for the root mean square error.

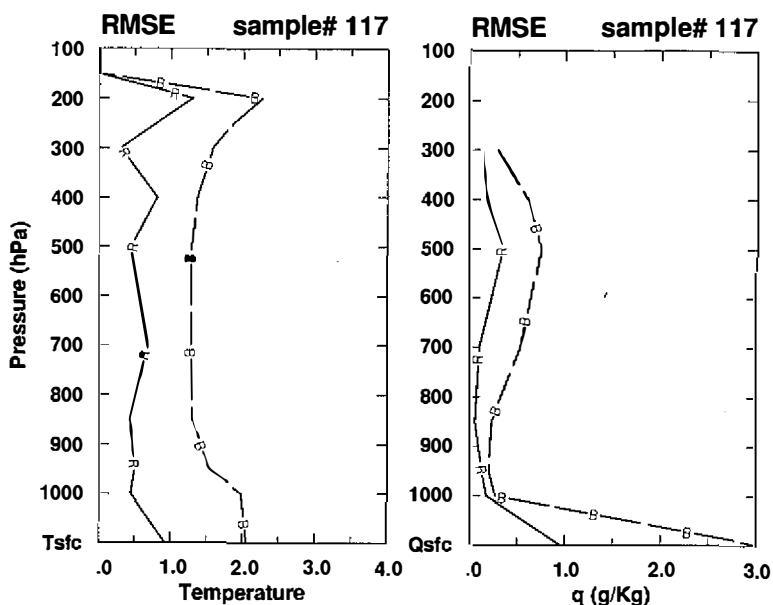


Fig. 10. Same as Fig. 8, but for the root mean square error.

field (solid line) and the background field (12-hour model forecast, dashed line) are quite similar. It is reasonable that the modification of the retrievals relative to the background is limited if the forecast error is small (Eyre 1989). In spite of the limited correction, the figure shows that the retrieved field at 700 hPa enhances the temperature perturbation near Japan and Korea. This figure also shows that there exists a larger difference between the retrieval and background fields in the sub-tropical area (south of 30°) on 700 hPa and 500 hPa. This is common in other cases (not shown here). Such a geographic dependence on the retrievals may be due to the different air mass characteristics between the mid-latitude and subtropical areas.

The differences in the mixing ratio field (Fig. 12) on 500 hPa are more prominent than 700 hPa. This figure also shows that the retrieval field tends to weaken the moist tongue near N through 700~500 hPa. The results from this case study are encouraging in that the regression retrieval algorithm exhibits characteristics which are locally coherent in the horizontal and vertical. Such characteristics are important to ingest the retrieval products in the analysis module.

4. SUMMARY AND CONCLUSIONS

In this study a regression algorithm was developed to retrieve the atmospheric parameters from the TOVS (TIROS Operational Vertical Sounder) clear-sky radiances. To construct the covariance matrix, this retrieval system requires a database, which collects the measured radiance, the profiles from the NWP forecast and RAOB observations. The RTTOV radiative transfer model used in the European Center for Medium Range Weather Forecasts (ECMWF) was chosen to simulate the computed radiance. The regression retrieval system and data pro-

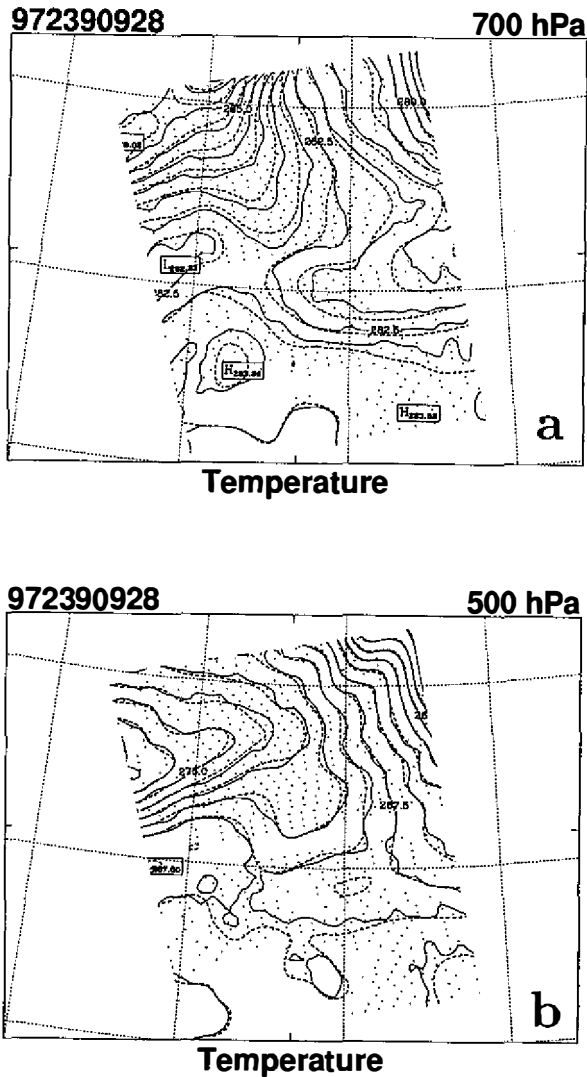
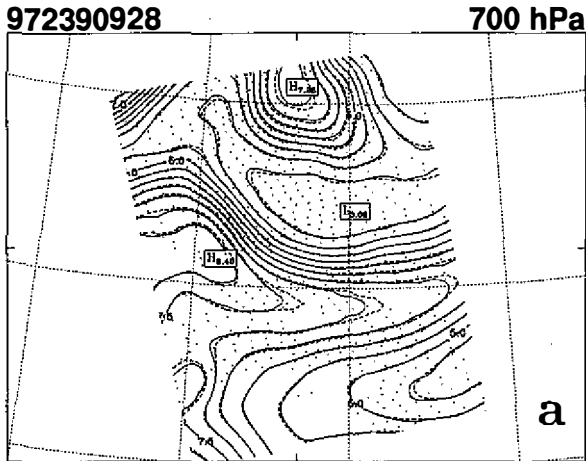


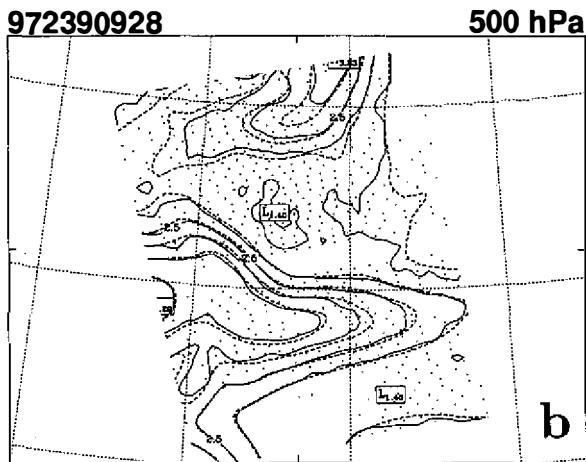
Fig. 11. Temperature field of 700 hPa (a) and 500 hPa (b) for the path on 0928 LST, 27 Aug 1997. The dashed line is the 12-hour forecast field and the solid lines are the retrieved field. The contour interval is 0.5K. The dots represent every three HIRS fovs available in the path.

cessing were closely related to the NWP model, the Limited Area Forecast System (LAFS), which is operated daily at the Central Weather Bureau (CWB) in Taiwan.

Clouds are an important factor in the radiative transfer processes in the atmosphere. In this paper, the authors combined the AVHRR and HIRS data to estimate the cloud information in a HIRS *fov* and determine which HIRS *fov* is "cloud-free". Two experiments were designed to show the retrieval performance, and were statistically compared with the co-located RAOB measurements. The results show that the bias was comparable to the first guess. This is not surprising because the retrieval is first guess dependent. The root mean square errors of the retrievals in the troposphere were better than the forecast by about 1~1.5 K in temperature and 0.2~0.4 g/Kg in the mixing ratio. The root mean square error of retrievals at the surface had an improvement of 0.5~1 K in temperature and 1.5~2 g/Kg in the mixing ratio. The results of the



Mixing ratio (g/Kg)



Mixing ratio (g/Kg)

Fig. 12. Same as Fig. 11, but for mixing ratio. The contour interval is 0.5 g/Kg.

August experiment were noticeably better than those in the June experiment. This improvement implies that the retrieved profiles were closer to the RAOB profiles than the 12-hour forecast profiles were.

The contour field of the retrievals for the case on 0928 LST 27 Aug 1997 shows that the temperature enhanced the thermal ridge and weakened the moisture tongue. The differences between the retrieval and background fields in the sub-tropical area (south of 30°) were larger than in the mid-latitude area. This phenomenon has also been noted in other cases and therefore not a special case here. The reason may be due to the different air mass characteristics between the mid-latitude and subtropical area.

This paper focuses mainly on the evaluation of the performance of the regression retrieval system in CWB LAFS. The pre-processing of cloud-cleared radiances has also been discussed.

However, many refinements still have to be made to the current retrieval framework. For example, we did not discuss further the possible impact due to the effects of the reflected solar radiation, though we excluded HIRS channels 18 and 19, which are most sensitive to sunlight reflection. Second, the retrieval performance under different weather regimes must be studied further. How to classify the different weather regimes or the so-called "seasons" is also an important issue in regression retrieval. Third, to include cloud information into the regression retrieval procedure and the inverse cloudy profile ought to be the next step. Finally, the impact of assimilating the retrieval products into the forecast model is currently under investigation.

Appendix

Without loss of generality, we consider the unbiased case:

$$\langle P_r - P_b \rangle = \langle R_r - R_b \rangle = \langle R_o - R_b \rangle = 0,$$

where $\langle \rangle$ stands for the sample mean. The basic form of the linear regression equation can be written as:

$$P^k - P_b^k = \sum_{i=1}^N w_{ki} [R_o^i - R_b^i] + \varepsilon_a^k. \quad (\text{A.1})$$

The definition of the symbols is the same as the descriptions in Section 2.1. The superscripts k and i indicate the levels of profile and HIRS channels, respectively. N is the total number of HIRS channels. ε_a^k represents the errors. w_{ki} is the underdetermined weighting function indicating the weight given to each observation increment in radiance in the retrieved profile increment. The minimum variance estimation procedure was used to minimize the expected error variance and satisfy the weak constraint:

$$\begin{aligned} J &= \left\langle (\varepsilon_a^k)^2 \right\rangle + \lambda \left\langle \left(\sum_{i=1}^N w_{ki} (R_r^i - R_b^i) \right)^2 \right\rangle \\ &= \left\langle \left((P^k - P_b^k) - \sum_{i=1}^N w_{ki} (R_o^i - R_b^i) \right)^2 \right\rangle + \lambda \left\langle \left(\sum_{i=1}^N w_{ki} (R_r^i - R_b^i) \right)^2 \right\rangle \\ &= \left\langle (P^k - P_b^k)^2 \right\rangle - 2 \sum_{i=1}^N w_{ki} \langle (P^k - P_b^k)(R_o^i - R_b^i) \rangle + \\ &\quad \sum_{i=1}^N \sum_{j=1}^N w_{ki} w_{kj} \langle (R_o^i - R_b^i)(R_o^j - R_b^j) \rangle + \lambda \sum_{i=1}^N \sum_{j=1}^N w_{ki} w_{kj} \langle (R_r^i - R_b^i)(R_r^j - R_b^j) \rangle \end{aligned} \quad (\text{A.2})$$

Here λ is a pre-specified weight called the Lagrange multiplier. If we want to satisfy the condition

$$\sum_{i=1}^N w_{ki} (R_r^i - R_b^i) = 0 \quad (\text{A.3})$$

very precisely, then we should designate λ as a very large value. Conversely, the constraint (A.3) is only approximately satisfied if λ is small. (A.3) can interpret the errors contained in the radiative transfer model and NWP forecast. The problem now is to find the weights w_{ki} that minimize (A.2) by using a large number of rawinsonde observations, satellite radiance, and forecast products. Differentiating (A.2) with respect to each of the weights w_{ki} , gives:

$$0 = \frac{\partial J}{\partial w_{ki}} = -2 \left\langle (P_r^k - P_b^k) (R_o^i - R_b^i) \right\rangle + 2 \sum_{j=i}^N w_{kj} \left[\left\langle (R_o^i - R_b^i) (R_o^j - R_b^j) \right\rangle + \lambda \left\langle (R_r^i - R_b^i) (R_r^j - R_b^j) \right\rangle \right] \quad (\text{A.4})$$

Applying the matrix form defined in Section 2.1, (A.1) can be rewritten as

$$T' = Cx, \quad (\text{A.5})$$

where C is the coefficient matrix. C can be obtained from (A.3) and presented as

$$S_{yx} = C(S_{xx} + \lambda E) \quad (\text{A.6})$$

Inserting (A.5) into (A.4), yields

$$T' = S_{yx} (S_{xx} + \lambda E)^{-1} x \quad (\text{A.7})$$

Acknowledgments The authors wish to thank Dr Kuang-Hua Wang and Mr. Kung-Lu Chang for support with the satellite data. Mr. Chien-Ben Chou and Mr. Shyu-Jui Wu are especially acknowledged for their discussions and help with data processing. Finally, we would like to thank two anonymous reviewers for their very helpful comments that have substantially improved the quality of this manuscript. All the computations were supported by the Computer Center of the Central Weather Bureau.

REFERENCES

- Andersson, E., A. Hollingsworth, G. A. Kelly, P. J. Pailleux, and Z. Zhang, 1991: Global observing system experiments on operational statistical retrievals of satellite sounding data. *Mon. Wea. Rev.*, **119**, 1851-1864.
- Andersson, E., J. Pailleux, J.-N. Eyre, A. P. McNally, G. A. Kelly, and P. Courtier, 1994: Use of cloud-cleared radiances in three/four dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **120**, 627-653.
- Aoki, T., 1980: A method for matching the HIRS/2 and AVHRR pictures of TIROS-N satellites. Tech. Note No. 2, Met. Sat. Center, Japan, 15-26.
- Chen, G. T. J., 1983: Observational aspects of the Mei-Yu phenomena in subtropical China. *J. Meteor. Soc. Japan.*, **61**, 306-312.

- Chen, G. T. J., 1991: Mesoscale features observed in the Taiwan Mei-Yu season. *J. Meteor. Soc. Japan.*, **70**, 497-516.
- Chou, C. B., 1998: Application of AVHRR data to one dimensional variational retrieval scheme for TOVS data. *Atmos. Sci.*, **26**, 343-362. (Chinese with English abstract)
- Crone, L. J., L. M. McMillin, and D. S. Crosby, 1996: Constrained regression in satellite meteorology. *J. Appl. Meteor.*, **35**, 2023-2035.
- Derber, J. C., and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287-2299.
- Detering, H. W., and D. Etling, 1985: Application of the E - (turbulence model to the atmosphere boundary layer. *Bound. Layer Meteor.*, **33**, 113-133.
- Eyre, J. R., 1989: Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. I: Theory and simulation for TOVS. *Quart. J. Roy. Meteor. Soc.*, **115**, 1001-1026.
- Eyre, J. R., 1991: A fast radiative transfer model for satellite sounding system. ECMWF Tech. Memo., 176, 28pp.
- Eyre, J. R., G. A. Kelly, A. P. McNally, E. Andersson, and A. Persson, 1993: Assimilation of TOVS radiance information through one dimensional variational analysis. *Quart. J. Roy. Meteor. Soc.*, **119**, 1427-1463.
- Hayden, C. M., 1988: GOES-VAS simultaneous temperature moisture retrieval algorithm. *J. Appl. Meteor.*, **27**, 705-733.
- Hong, J. S., 1999: The impact study of the satellite sounding (SATEM) in the objective analysis of Limited Area Forecast System (LAFS) in Central Weather Bureau. *Atmos. Sci.* **27**, 61-73. (Chinese with English abstract)
- Kim, D., 1993: Application of cluster analyses of radiance data measured by satellite and computed from forecast profiles. Technical Proceedings of the Seventh International TOVS Study Conference, 10-16 Feb, Igl, Austria, 209-220.
- Kim, D., and J. Chou, 1995: Multi-level cloud parameter estimation using smoothed density of AVHRR data and HIRS data. Preprints, Eighth International TOVS Study Conference, 5-11 Apr, Queenstown, New Zealand.
- Kim, D., 1996: Multilevel cloud parameter estimation combining HIRS and AVHRR data - asimulation study. Preprints, Eighth Conf. on Satellite Meteorology and Oceanography, 28 Jan - 2 Feb, Atlanta, GA.
- Kuo, H. L., 1974: Further studies of the parameterization of the effect of cumulus convection on large scale flow. *J. Atmos. Sci.*, **31**, 2132-1240.
- Harshvardhan, R. Davies, D. A. Randall, and T. G. Corsetti, 1987: A fast radiative parameterization for atmospheric circulation models. *J. Geophys. Res.*, **92**, 1009-1016.
- Smith, W. L., H. M. Woolf, C. M. Hayden, D. O. Wark, and L. M. McMillin, 1979: The TIROS-N Operational Vertical Sounder. *Bull. Amer. Meteor. Soc.*, **60**, 1177-1187.
- Yeh, T. C., S. C. Lin, S. W. Li, W. M. Chen, S. W. Chang, and M. S. Peng, 1994: The second generation Limited Area Forecast System at the Central Weather Bureau in Taiwan. Conf. on Weather Analysis and Forecasting, 24 - 26 Nov 1994, Taipei, Taiwan, 11-20. (In Chinese)