TECHNICAL MEMORANDUM

158

Quality control of NESDIS physical retrievals of TOVS satellite data

G. Kelly, E. Andersson,
A. Hollingsworth, P. Lönnberg,
J. Pailleux and Z. Zhang

Research Department

October 1989

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

ABSTRACT

Recent work has identified serious errors and biases in the operational statistical retrievals produced by NESDIS in 1987. Similar errors and biases are found in the physical retrievals produced operationally since September 1988. We report on experiments to design quality control algorithms to deal with the errors in the data. We document the quality control changes implemented in the ECMWF system in January 1989, and evaluate the performance of the changes.

1. INTRODUCTION

Recent developments in the ECMWF analysis/forecast system have lead to a significant increase of the sensitivity of the forecast to initial data (Lönnberg, 1988). A recent Observing System Experiment (OSE) found that the statistical SATEM retrievals of TOVS satellite data in the ECMWF analysis forecast system (as configured in late July 1988) had a negative impact on analyses and forecasts (Andersson et al., 1989) in the Northern Hemisphere.

Synoptic and statistical investigation of the SATEM data for the study period (30 January to 14 February, 1987) showed serious defects in the statistical retrievals. The defects included important air-mass dependent biases in the data. These lead to geographically fixed biases of substantial magnitude. The random component of the SATEM observation error for the gross tropospheric static stability was larger than the errors of the first-guess for the same field. Similar results were noted even for the 1000-300 hPa layer-mean temperature.

Andersson et al. noted that the quality control procedures in the assimilation system (called OPS-JUL88) did not exclude the worst of the SATEM data. Given the enhanced sensitivity of the OPS-JUL88 system to data, and therefore a enhanced vulnerability to bad data, and given also the magnitude of the errors in the SATEM data, it was not surprising to see a negative impact of the SATEM data on the forecast skill.

In September 1988 NESDIS changed from a statistical procedure to a physical procedure for their operational SATEM retrievals (Fleming et al., 1986). The new retrieval scheme is based on a search through a library of atmospheric profiles and radiances. The library entry closest to the observed radiances is chosen as the initial profile for the inversion of radiances to temperatures. The final temperature profile is obtained with the radiative transfer equation, linearized around the initial guess. Cloud clearing is still done as before.

In Section 2 we examine the quality of the new SATEM data as received during the winter of 1988-89. We demonstrate that the new physical retrievals have much the same problems of bias and noise that were noted in the statistical retrievals. We also report on a variety of experiments to improve the quality control of the SATEM data, either by tightening the acceptance criteria for data (Section 3), or by deleting certain types of data altogether (Section 4). We conclude by documenting the quality control changes implemented in the ECMWF system in January 1989.

2. QUALITY OF THE NESDIS PHYSICAL RETRIEVALS IN LATE 1988

We performed synoptic studies and data monitoring studies on the operational physical retrievals for November, December 1988 and January 1989 to determine if the error structure of the SATEMs had changed as a result of the new retrieval scheme.

2.1 Synoptic Study

Detailed synoptic evaluation of the operational NESDIS physical retrievals, concentrating on the synoptic systems which develop along the polar front, found that in these areas the (physical retrieval) operational SATEMs are affected by several problems including:

a) Large errors in the lower layer 1000-700 hPa which tend to smooth the horizontal gradients near the fronts. SATEMs are too warm in the cold air and too cold in the warm air.

b) Large errors were observed in the static stability.

Both types of error found in the physical retrievals have been documented already by Andersson et al. for the statistical retrievals.

There are generally very large and coherent observation errors in the cloudy areas of the mid-latitudes. The organisation, relative to the synoptic pattern, of the increments (OBS-FG) is striking. Figs.1a and 1b show an example (12 January 1989, 12 UTC) of the SATEMs near a developing extra-tropical cyclone which, with a deepening rate of 45 hPa over the following 24 hours, falls in the 'bomb' category.

The two figures illustrate the large discrepancy in the tropospheric stability between model and observations. The deviations from the first-guess were large enough for the data to be rejected by the operational analysis quality control, partly because the first-guess is considered fairly accurate in the western Atlantic. The forecast from the resulting analysis was very successful, which justifies the decisions of the automatic quality control. The same pattern of observation errors can be seen on any daily chart in the winter months, and there are several cases where there are radiosondes to confirm the accuracy of the first-guess.

2.2 Statistics of the Departures of the Physical Retrievals from the First-guess

Global collocations with radiosondes showed that the bias, SD (Standard Deviation) and RMS (Root Mean Square error) of the new retrievals were quite similar to those for the old retrieval scheme. On the global scale there were very small biases, generally below 0.5K, with SD and RMS around 2K in the mid troposphere and 3.5K near the surface. When the data were partitioned into air-mass classes we found large positive biases in the lowest layers in the Tropics and Subtropics, Figs.2a and 2b, and large negative biases in the polar air-masses, Figs.2c and 2d. These biases were often compensated aloft by biases in the upper troposphere of the opposite sign.

The bias of the physical retrievals relative to the first-guess shows a strong regional dependence, particularly in the lowest layer. Fig.3 shows the 1000-850 hPa NOAA-10 SATEM minus first-guess for clear soundings, Fig.3a, and for fully cloudy soundings, Fig.3b, for December 1989. See the large positive biases off the east coast of N America and Asia, which are very similar to what was seen with the statistical retrievals.

For this thin layer, the bias in the retrievals has positive maxima in the Subtropical highs, Figs.3a and 3b. It is largest for the fully cloudy soundings, exceeding 4K in the South Indian Ocean, Fig.3b. Only a small part of the bias comes from the first-guess. The first-guess is known to be almost unbiased in the Subtropics although it has a small positive bias in the deep Tropics.

Fig.3c shows the bias for all soundings in the lowest analysis layer, 1000-700 hPa, for December 1988. This layer is about twice as thick as the 1000-850 hPa layer. There are large biases off the coast of Asia, and somewhat smaller biases off the coast of N America. This is a reflection of a systematic under-estimation of the low-level temperature in cold-air out-breaks.

There is a compensation of bias between the 1000-850 and 850-700 hPa layers in some areas of the sub-tropics, but there is still a large positive bias in the Indian Ocean. The SATEMs seem to misrepresent the cold air below the Subtropical (trade wind) inversion.

Examination of the tropospheric stability index, S

$$S = T_V(1000\text{-}700) - T_V(500\text{-}300)$$

reported by SATEMs has proven to be particularly fruitful. The SATEM minus first-guess biases for the physical retrievals in December 1988 (not shown) are very similar to those shown by Andersson et al. (1989) for the statistical retrievals in February 1987. The largest biases in the physical SATEM retrievals, relative to the first-guess, are found along the Northern Hemisphere storm tracks, especially with the fully cloudy soundings. In the mean, and relative to the first-guess, the SATEM observations are up to 7K less stable in the area off the East Asian coast. The western part of the North Atlantic also has a large positive (less stable) bias, whereas the Norwegian Sea bias is up to 4K on the negative (too stable) side. The bias patterns for NOAA-10 and NOAA-11 are almost identical.

The standard deviation of the stability deviation from the first-guess is very large North of 30° North for both satellites and generally between 3K and 6K.

In the mid-latitudes there is a strong tendency for the deviations from the first-guess to compensate within the depth of the troposphere. This can be seen from maps of correlations of the first-guess departures between the two layers used in the definition of S (Strauss, 1989).

2.3 Forecast sensitivity to SATEM data in November 1988

On 3 November 1988, 12 UTC the operational forecast error was unusually large over the Pacific and North America. The SATEM data quality appeared poor over the Pacific. A NOSATEM assimilation was run and gave a significant improvement of the forecast scores. This confirms the sensitivity of the ECMWF system to the SATEMs noted earlier by Andersson et al. (1989).

The NOSATEM forecast run for the previous day (2 November 1988, 12 UTC) was also sensitive to the SATEM data. In this case however, there was a strong positive impact of SATEM data. We have not investigated in detail the reasons for the difference in sensitivity to SATEMs in these two forecasts. Flobert et al. (1989) have shown that intermittent problems in radiance calibration can sometimes occur, and can adversely affect the forecast skill if they are not detected.

3. QUALITY CONTROL AND ANALYSIS OF SATEM DATA

The results of forecast experiments and the synoptic studies described by Andersson et al. (1989), which were confirmed in the last section, have identified two important problems with the SATEMs. They are 'air-mass biases' and the errors in static stability in the mid-latitude baroclinic zones.

These shortcomings of the data have a strong effect on the analyzed fields. The analysis system of ECMWF assumes unbiased observations as well as an unbiased first-guess. In the absence of independent data the analysis system cannot correct or filter biases; clusters of data with uniform biases in the horizontal will be drawn for to the extent illustrated in Fig.4, provided the data passes the quality control procedures.

Clusters of data with the same type of biases in the vertical will also be drawn for to the extent illustrated in Fig.4, provided the data passes the quality control procedures. Current OI methods of using the information on vertical structure in the SATEM data are unsatisfactory.

3.1 Filtering in the OI System

Kelly and Pailleux (1988) changed previous practice at ECMWF in the use of SATEM data by only extracting information on thick layers (specifically 1000-700-500-300-100-50-30-10 hPa layers) from the operational SATEM bulletins. They argued that this choice of layers corresponded to the most likely information content in the original radiance measurements, and that any attempt to extract more information from the measurements would only introduce climatological information at best, and noise at worst.

Lönnberg (1989) has examined the response of the OPS-JUL88 ECMWF system to SATEM data. In this system the vertical correlation functions are fairly narrow in the vertical over most oceans, but are rather broad over the East Pacific. Lönnberg shows that the response of the analysis system to SATEM data alone is much more satisfactory when the broader vertical correlations are used than when the narrower functions are used. The use of the narrow functions can lead to peculiar responses when the half-width of the structure functions is comparable with the spacing of the observations.

The vertical correlation between the layers are currently prescribed by one correlation matrix which does not allow for regional and air mass variations. The matrix for SATEM observation eor, D, is described in Kelly and Pailleux (1988). The off-diagonal terms of this matrix provide an important control on the response of the analysis to different components of the SATEM data.

Strauss (1989) shows that the correlation of the departures from the analysis of T_V(1000-700) with T_V(500-300) is strongly negative in most regions of the Northern Hemisphere mid-latitudes. Given the relative magnitude of observation error and analysis error, Strauss' results will be dominated by the observation error. Use of such a feature in D would lead to a response which gave more weight to a tropospheric mean temperature than to a tropospheric gross static stability. However when the data used for the correlation calculation is screened by the quality control procedures described below, then the correlation of the T_V(1000-700) and T_V(500-300) departures from the first-guess is almost zero in the active regions of the Northern Hemisphere mid-latitudes. The matrix specified by Kelly and Pailleux (1988) does in fact have small off-diagonal terms, and so agrees with the calculations on screened data by Strauss (1989).

The current specification of SATEM observation error and first-guess error is such that the ECMWF OI system gives a better response to gross tropospheric static stability than to a tropospheric layer-mean temperature (Hollingsworth and Lönnberg, in preparation). This result seems counter-intuitive given the problems in the observations. However the filtering properties of the OI system are well grounded in screened empirical data. The question of whether it is better to use all the data or just the screened data is discussed further in Section 6.2.2, after some further results have been presented.

3.2 Discussion

Alternative approaches to reliance on OI to filter the analysis response to the SATEM data would be:

- Pre-filter the SATEM data itself with empirical corrections, or
- Pre-filter the data through selective filtering in the vertical, or
- Pre-filter the data through data rejection in quality control procedures.

Correction of biased satellite data by ECMWF does not seem feasible because tuning is carried out continuously by the data producers. Filtering in the vertical might, through the strongly non-linear properties of the radiative transfer equation, result in a profile which no longer satisfies the measured radiances and produce a filtered profile which is incorrect. The more practical solutions appear to be either to eliminate whole categories of SATEM data (Section 4) or to develop tailored quality control procedures (Section 5).

4. EXPERIMENTATION ON EXCLUDING CLASSES OF SATEM DATA

When it became clear that in the Northern Hemisphere the NOSATEM experiment reported in Andersson et al. (1989) produced as good as or better forecasts than OPS-JUL88, an experiment was set up to reduce the number of SATEMs used in the analysis, eliminating those with the worst error characteristics.

4.1 Experiment where Three Categories of SATEM Data are Not Used: All SATEMs over Ice, Cloudy and Partly Cloudy SATEMs North of 20° South

Both our collocation and synoptic studies in the Northern Hemisphere, and in the tropics, show higher RMS errors and larger biases for the cloudy and partly cloudy SATEM soundings than for the clear ones. Since clouds are nearly black bodies in the infra-red, the atmosphere within and below clouds cannot be observed by the HIRS instrument and the retrieval has to rely on the MSU channels, which are accurate but have very broad weighting functions. The cloud clearing for the partly cloudy soundings also introduces further errors. We therefore experimented with the elimination of the partly-cloudy and cloudy retrievals north of 20° South.

We are also aware that satellite soundings are less accurate over ice than over sea. Cloud-detection is more difficult over ice. The very sharp temperature gradients in the boundary layer over ice make temperature retrievals more inaccurate. We therefore extended our experiment to exclude SATEM data over ice. In order not to use any TOVS over ice we excluded all soundings where the surface temperature was less than 2K.

We made a three and a half day data assimilation (called MOD-NOSAT1) from 30 January 1987 to 2 February 1987. Three forecasts were run from 12 UTC data on 31 January through 2 February 1987. As controls for this experiment we had the OPS-JUL88 and NOSATEM-JUL88 assimilations and forecasts for the same period. In the Northern Hemisphere the OPS-JUL88 and NOSATEM-JUL88 forecast scores show the same relationship to each other during this short period as was found in the more extended comparison discussed by Andersson et al. (1989). The forecast scores for the experimental MOD-NOSAT1 forecasts were worse than either of the two sets of control forecasts, Fig.5a. We do not have a satisfactory explanation for this result. Further synoptic investigation is needed to explain why the exclusive use of clear SATEMs in areas where they are thought to be better than the cloudy or partly cloudy retrievals should give a result for the scores that is worse than the result got when no SATEM data whatever is used, or when all of it is used.

The scores for the MOD-NOSAT1 forecasts in the Southern Hemisphere are better than the NOSATEM scores but not as good as the scores when the SATEM data was used over ice, Fig.5b. Fig.6 shows the Southern Hemisphere day-3 forecast errors for one control forecast (OPS-JUL88, where all the SATEM data is used) and the MOD-NOSAT1 experiment which did not use SATEM data over ice. The errors in the experimental forecast are clearly larger in the South Pacific. Detailed investigation showed that larger errors originated over the ice-covered areas. The forecast errors appear to be reduced by the satellite data in an annulus around to the Antarctic continent. North of 50° South, there is little difference in the forecast errors up to day 3.

4.2 Discussion

Given the value of the SATEM data in the Southern Hemisphere, and the known problems with the fully cloudy retrievals in the Northern Hemisphere the experiments just discussed were the obvious follow-up experiments to the work of Andersson et al. (1989). Though the sample size is small, the MOD-NOSAT1 experiments indicated that there are some features of the use of SATEM data which we do not yet understand, and which need clarification.

Given the need to improve the use of SATEM data, we then explored the possibility of a different approach to the use of SATEM data.

5. THE JANUARY 1989 QUALITY CONTROL MODIFICATIONS

Instead of excluding whole categories of SATEM data we next explored the possibility of doing a much tighter quality control on all SATEM data regardless of retrieval path. This appeared to produce better results, as outlined in this section.

5.1 Quality Control Procedures Implemented in January 1989

The following enhanced quality control procedures on SATEM temperature soundings were tested and implemented in the operational suite on 31 January 1989. The results of the tests are described in the next section.

5.1.1 Revised OI-check

The OPS-JUL88 system (Lönnberg, 1988) contained a modification of the OI-check for acceptance of data. Since erroneous SATEMs tend to occur in clusters they support each other in a conventional OI-check, which makes it very inefficient for satellite data. The revised procedure checks SATEMs without using neighbouring data of the same type, provided there are at least two observations from other data sources in the vicinity. Together with a decrease in the OI-rejection limit this gives a more efficient check in areas with a sufficient number of conventional data.

In addition, a multi-level summary of the OI-check decisions was introduced; the whole tropospheric or stratospheric part of a sounding is rejected if there are several suspect layers or one large error within the report.

5.1.2 Tighter check on SATEM minus First-guess departures at a Single Level

The OPS-JUL87 first-guess check seems to be sufficient for the middle layers but very few rejections are observed in the lowest layer and in the top two stratospheric layers. This is due to the higher FG error variance for those layers. In terms of standard deviation of 'normalized departure' ((OBS-FG)/SD of FG) we have implemented a reduction from 3.0 at all levels to 1.2, 2.1, 2.75, 2.75, 2.75, 2.6, 2.3, listed from 1000-700 to 30-10 hPa. In absolute terms, this approximately corresponds to a rejection limit of 4K for the 1000-700 hPa layer-mean virtual temperature in the eastern part of the North Pacific.

5.1.3 Stability check

A further method to identify incorrect data is to compare observed and first-guess stabilities. Large errors in the lowest layer tend to be compensated aloft by errors of opposite sign. The soundings differ from the first-guess mainly because of their limited vertical resolution in overcast situations. From comparisons with ocean stations and weather ships (Andersson et al., 1989, and Fig.10 below) in six oceanic regions, it is clear that the most noisy tropospheric SATEM stability is the difference of temperature between the two layers 1000-700 and 500-300 hPa. The satellites (in the form of SATEMs) have almost no skill at all in measuring this stability index S. From these scatter diagrams we chose to reject all soundings in our test assimilations where the value of S in the SATEM differed from S in the first-guess by more than 4.5K. However when we implemented the check in operations, the limit for rejection was tightened to 3.5K.

5.2 Experiments with the Pre-FEB89 quality control

The tighter quality control of Section 5.1 was introduced in operations on 31 January 1989 after tests in the February 1987 period. The stability check had a threshold of 4.5K in the tests but it was decreased to 3.5K for operations. For clarity of presentation we shall speak of the system with the 4.5K cut-off on the stability index S as the Pre-FEB89 system, and the operationally implemented system with the 3.5K cut-off as the OPS-FEB89 system.

To examine the effect of the January 1989 changes we studied in detail the SATEM data statistics from the OPS-JUL88 assimilation on the period 30 January 1987 to 4 February 1987. Fig.4 shows summary plots of the data volumes and departures of the SATEM data from the first-guess, analysis and initialised analysis for NOAA-10. The volume of accepted SATEM data has declined substantially as a result of the tightening of the quality control.

Fig.4a shows the vertical distribution of the deviation statistics and average data volumes for all NOAA-10 data received over the Atlantic in the period, with separate plots for each retrieval path. Fig.4b shows the corresponding plots for all NOAA-10 data accepted and used in an assimilation for the period with the OPS-JUL88 system. Fig.4c shows the corresponding plot for the data that would have passed the 3.5K limit of the stability first-guess check of the OPS-FEB89 system. The data counts give the average number of data per analysis cycle. These results show that as a result of the quality control changes, about 20% of the clear retrievals and about 40% of the fully cloudy retrievals are rejected in the period. Similar results are found for NOAA-9. In February 1989, after the more stringent quality control had been made operational, 20% to 40% of the available SATEMs were rejected to the north of 20° North in any given analysis.

Figs.4b and c also show that for the data used by the assimilation, the tighter quality control has been very effective in reducing the bias in the stability in the north Atlantic, (and indeed in all Northern Hemisphere regions), particularly the cloudy retrievals.

Fig.7 shows more detail on the effect in the Northern Hemisphere of the quality control changes on the NOAA-10 data used by the OPS-JUL88 and OPS-FEB89 assimilations of the period 30 January 1987 to 4 February 1987. The figure shows histograms of NOAA-10 Tv(1000-700) departures from the first-guess for the fully cloudy retrievals. The bias and standard deviation of the departures in the accepted data are roughly half the values in the unscreened data. The bias in the accepted data in the order of 1K over the large oceans. The spread of the histograms is considerably reduced in the accepted data, as expected.

The two changes to the FG-check led to a large increase in the number of rejections. In a test with the Pre-FEB89 system with data from the period 30 January 1987, 00 UTC to 2 February 1987, 12 UTC the increase in SATEM rejections compared to the assimilation with OPS-JUL88 was from 85 to 240 on average per data assimilation cycle. The rejections were in the desired areas and led to more coherent analyses.

5.3 Tests with the OPS-FEB89 system

The OPS-FEB89 change has been tested in a data assimilation started on 24 January 12 UTC with forecasts run from the 25, 26 and 27 January, all at 12 UTC. As a control for this run we used the real-time operational assimilations, which used the OPS-JUL88 assimilation system. The period was chosen because the operational forecasts verified unusually badly beyond day 5, particularly the forecasts from 25 and 27 January. In addition, we ran a third assimilation MOD-NOSAT2, which excluded all SATEM data in the troposphere (below 100 hPa) between 20° South and 70° North. The differences between the OPS-JUL88 and OPS-FEB89 analyses are large, particularly in the temperature over the North Pacific. Fig.8 illustrates the analysis differences in the lower and middle troposphere at 12 UTC 25 January 1989 and together they show large differences in the analyzed static stability index of up to 6K, mainly due to SATEM rejections.

The effect on the forecast skill of the change from OPS-JUL88 to OPS-FEB89, or even to MOD-NOSAT2, was small. The cases of January 25 and 27 turned out to be very insensitive to changes in the use of satellite data; the very large forecast errors still developed from the Pacific regardless of how satellite data had been used in the assimilation. The impact on average forecast scores was neutral in both hemispheres.

In the Tropics, the effect of leaving out the SATEMs in the troposphere in the MOD-NOSAT2 assimilation was mostly a cooling of the lowest layer and a warming between 700 and 300 hPa i.e. initially a convectively less active atmosphere.

6. EVALUATION OF THE OPS-FEB89 QUALITY CONTROL CHANGES

6.1 Synoptic Aspects of the SATEM Data Rejections

The operational assimilations in February 1989 were run with the OPS-FEB89 system described above. The rejected satellite data were typically clustered in rather large groups associated with areas of warm advection or cold advection in mid-latitude synoptic systems. The rejections occur in the frontal zone near Japan, in the large amplitude trough in the eastern North Pacific, in the cold air out-break over the Gulf of Mexico, and along the front in the Atlantic. The number of rejections in the Southern Hemisphere is comparatively small due mainly to seasonal effects (not shown).

6.2 Verification of the OPS-FEB89 First-guess against Radiosondes

Perhaps the most convincing evidence for the problems with the statistical TOVS retrievals presented by Andersson et al. (1989) was a series of scatter plots of the value of S (the stability index) in the first-guess compared against the value of S measured by radiosondes in isolated locations, or measured by the SATEMs in the same locations. In this section we show similar plots to demonstrate that the major retrieval problems have been unaffected by the change from statistical to physical retrievals. All the scatter plots to be shown in this section are based on the operational 12 UTC analyses during February 1989, and so were generated with the OPS-FEB89 assimilation system. On the plots accepted data are shown with full circles, and rejected data are shown with open circles.

6.2.1 North Atlantic

Fig.9a shows the scatter plot of radiosonde reports for S from the three Atlantic weather ships (C, L, M) against the first-guess value for S. Reports are only plotted when the surface pressure is 1000 hPa or larger. The maximum deviations are about 5K. None of the radiosonde data is rejected.

Fig.9b shows the corresponding scatter plot for S from the SATEMs (physical retrievals) in the area of the Atlantic bounded by 40° North, 50° North, 40° West and 5° East. The largest deviations between SATEM and first-guess is of order 15K. The scatter is much larger than in Fig.9, and indicates that the first-guess for S in this area is considerably more accurate than the SATEMs.

Substantial quantities of the SATEM data have been rejected in the course of the month. If we assume that open dots lying more than 3.5K from the diagonal have been rejected by the check against the first-guess, and that open circles lying closer to the diagonal have been rejected by the main analysis check, then it is evident that the check against the first-guess is causing most of the rejections.

6.2.2 Error Characteristics of the Screened SATEM Data

It is important to know how to calculate the error characteristics the SATEM data used by the analysis system. Should one use all the data (i.e. black and white dots) in collocation studies with radiosondes to determine the error of the SATEMs?, or should one use only the accepted data?

The results of Strauss (1989) imply that there are important differences between the results one will get for off-diagonal entries of the vertical covariance matrix for SATEM error, depending on which data set is used. The OI system assumes there is no correlation between the forecast error and the observation error. One could ignore all first-guess information in the calculation of the observation error covariance, and one would use all the data, whether rejected or not, in the estimation of the observation error covariances. One then assumes that even the screened data are typical of a population which is capable of having an error in the observation of S as large as 15K, and so is rather inaccurate. In this approach, the screening against the first-guess is regarded as a safety device which has no implications for the intrinsic error of the SATEM data.

The alternative approach would say that the SATEM data used in the analysis really represent a combination of satellite information and first-guess information, and the screened data (i.e. the black dots) have quite different error characteristics from the original SATEM data. The error of the screened data is much lower than the error of the unscreened data (black and white dots). However the errors of the screened observations must have important correlations with the error of the first-guess. Such correlations then need to be taken into account in the OI correlation.

Neither of these approaches is fully satisfactory, but the first approach with a crude quality control is simpler.

6.2.3 Southern Part of Japan

Fig.9c shows the scatter in S of the first-guess compared with the radiosondes. The first-guess tends to be too stable by slightly over 1K. The largest deviation around this bias is about 1.5K. None of the radiosonde data is rejected.

Fig.9d shows the corresponding scatter in the SATEMs versus the first-guess in the area 25-35° North, 130-140° East. There is an enormous bias in the SATEM data, as well as a huge scatter. In cases of large static stability (low values of S) the bias in S is a large as

10K. Largest deviations are about 15K. Much of the SATEM data is quite properly rejected by the first-guess check on S. The first-guess here is much more accurate than the SATEM data.

6.2.4 Islands in the Sub-tropical Mid-Pacific

Fig.9e shows the scatter in S at the radiosondes on Midway Island and the Hawaiian Islands, compared with the first-guess for S at these points. The largest deviations are 3 to 4K, and none of the data is rejected.

Fig.9f shows the corresponding data for the scatter of SATEM measurements against the first-guess. The largest deviations are 8 to 9K. Even in a data sparse area such as the sub-tropical mid-Pacific, it would appear that the first-guess for S is more accurate than the SATEM estimates.

6.2.5 The Extra-tropical South Atlantic

Fig.9g shows the scatter in S for the radiosonde at Gough Island (in the extra-tropical South Atlantic), plotted against the first-guess. There is evidence of a bias in the first-guess of order 1K, with a small scatter. Largest deviations between radiosonde and first-guess are about 5K, with typical deviations of order 2 to 3 K. Allowing for the seasonal differences between north and south Atlantic in February, the performance of the first-guess in the south Atlantic (Fig.9g) is not very different from the performance in the north Atlantic (Fig.9a).

Fig.9h shows the corresponding scatter plot for S for the SATEMs in the area 35-45° South, 0-20° West. The largest deviations are about 8K. The largest deviations occur because the range of variability in the first-guess is larger than the range of variability in the SATEMs. Judged on the radiosonde evidence, the range of variability in the first-guess is quite reasonable. Hence the retrieval procedure in this area does not reproduce the full meteorological variability.

6.2.6 The Southern Ocean, South of New Zealand

Fig.9i shows the scatter in S at Macquarie and Campbell Islands (in the Southern Ocean South of New Zealand). The performance of the first-guess here is of about the same quality as in the South Atlantic.

Fig.9j shows the corresponding scatter plot for the SATEMS against the first-guess in the area 45-55° South, 150-170° East. The performance of the SATEMs in this area seems to be better than in the south Atlantic, in that the range of variability is more realistic. Rather few SATEM data are rejected in the area.

6.3 Discussion

All the available evidence indicates that the quality control modifications introduced as a result of the present work perform well, and correctly reject large volumes of suspect SATEM data in the winter extra-tropics of the Northern Hemisphere. We need to examine the need for a tighter quality control in other seasons, and also in the tropical area, where there is evidence that the errors in the SATEM data are large relative to the climatological variability.

7. CONCLUSIONS AND FURTHER DEVELOPMENTS

The extensive experimentation in Andersson et al. (1989) highlighted the sensitivity of the ECMWF analysis/forecast system to SATEM data, and demonstrate d serious quality problems in the SATEMs produced by stistical retrievals of TOVS data. NESDIS changed their retrieval procedure to a physical retrieval in September 1988. Global collocation statistics with radiosonde data do not indicate any change in overall performance between the statistical and physical retrievals although the new retrieval technique appears to have eliminated some lateral inconsistencies experienced earlier between adjacent satellite tracks. In this paper we have demonstrated that the SATEMs produced by the physical retrieval procedure (introduced by NESDIS in September 1988) have just as serious errors and biases as the statistical retrievals. These errors in the SATEMs have an adverse effect on analysis and forecast quality.

We explored two approaches to quality controlling the SATEM data in the Northern Hemisphere, where the problems are most serious. In our first experiments we retained only the best (clear) SATEM retrievals in the Northern Hemisphere and tropical troposphere, and excluded all SATEM data over ice. Although we only made limited tests of this approach, we found the paradoxical result that to retain only the best of the SATEM data, without any further changes in quality control, gave worse forecast scores in both hemispheres than using either all of the SATEM data or none of it. The reasons for this result need further investigation.

Our second approach was to develop a revised set of quality control tests for all SATEM data. These tests tightened existing tests against the first-guess, and introduced a new test on a stability index related to the gross tropospheric static stability. SATEMs which depart too far from the first-guess in this index are discarded. A good understanding of the SATEM data errors and of the forecast guess errors, together with detailed synoptic studies, are important in quality controlling the data.

Following the introduction of the analysis changes on 31 January 1989, up to 40% of the SATEMs in the Northern Hemisphere oceans are rejected, and there is now a much closer fit of SATEMs to the first-guess. The vertical stability check is mostly responsible for the rejections. Routine monitoring against radiosondes has confirmed that the rejected SATEMs have large errors.

Work is under way to improve the stability check by introducing a geographical dependence of the threshold and by using estimates of both observation and forecast errors. Other changes under investigation include the use of SATEMs over ice and land, and the use of DMSP soundings to quality control SATEMs.

Further work is also underway to make use of the forecast first guess to quality control the cloud-cleared radiances following work of Flobert et al. (1989). It is not uncommon to have small regions within satellite orbits effected particulary for the microwave channels.

Longer term approaches to improving the use of satellite data include the use of different retrieval methods (Flobert et al., 1989), and the use of variational retrieval methods. It is now becoming clear that if satellite retrievals methods do no make use of additional information, such as a six hour forecast, it is not always possible to produce a solution that improves on the first guess and in certain air-masses the results are extremely poor.

References

Andersson, E., A. Hollingsworth, G. Kelly, P. Lönnberg, J. Pailleux and Z. Zhang, 1989: Observing system experiments on NESDIS statistical retrievals of TOVS satellite data using the 1988 ECMWF data assimilation system. ECMWF Tech.Memo. 157.

Fleming, H.E., M.D. Goldberg and D.S. Crosby, 1986: Minimum variance simultaneous retrieval of temperature and water vapour from satellite radiance measurements. Second conference on satellite meteorology/remote sensing applications. May 13-16, 1986, Williamsburg. Published by the American Meteorological Society.

Flobert, J.F., E. Andersson, A. Chédin, G. Kelly, J. Pailleux and N. Scott, 1989: Data assimilation and forecast experiments at ECMWF using the "3I" inversion scheme for satellite soundings. ECMWF Techn.Memo. 156.

Kelly, G. and J. Pailleux, 1988: Use of satellite vertical sounder data in the ECMWF analysis system. ECMWF Technical Memorandum No. 143.

Lönnberg, P., 1988: Developments in the ECMWF analysis system. Proceedings of the ECMWF Seminar on data assimilation and the use of satellite data. Reading, September 1988.

Lönnberg, P., 1989: Quality control and filtering of satellite data. Proceedings of ECMWF/EUMETSAT workshop on use of satellite data in operational numerical weather prediction: 1989-1993. Reading, 8-12 May 1989.

Strauss, B., 1989: Monitoring of TOVS/DMSP data. Proceedings of ECMWF/EUMETSAT workshop on use of satellite data in operational numerical weather prediction: 1989-1993. Reading, 8-12 May 1989.

Zhang, Z. and J. Pailleux, 1989: Evaluation and development of the ECMWF humidity analysis. ECMWF Tech.Memo. 160.

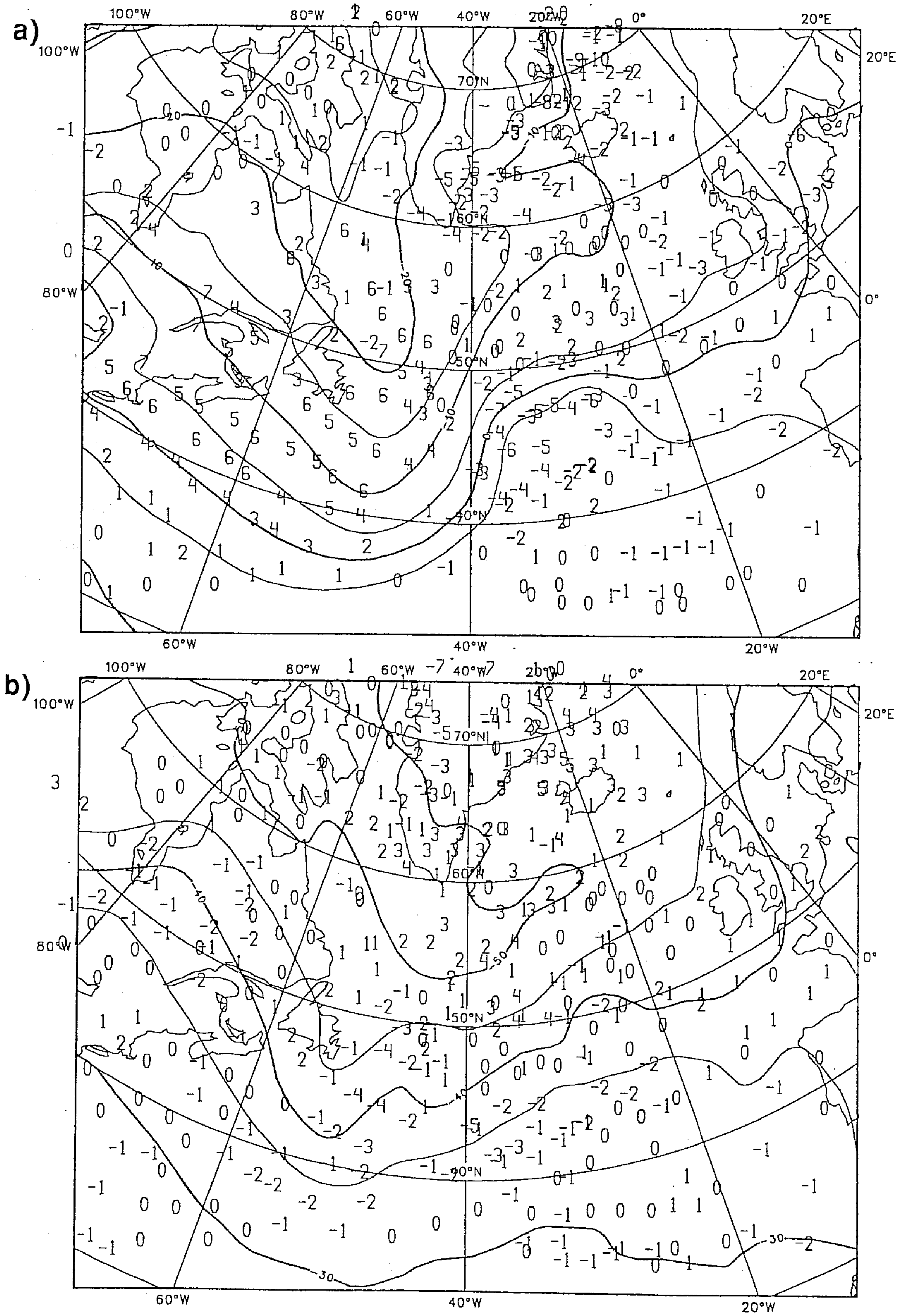


Fig.1 Observed SATEM layer-mean virtual temperature deviations from first-guess (K), 12 January 1989, 12 UTC, for the North Atlantic. The first-guess field is contoured with an interval of 5K. a) is 1000-700 hPa and b) is 500-300 hPa.

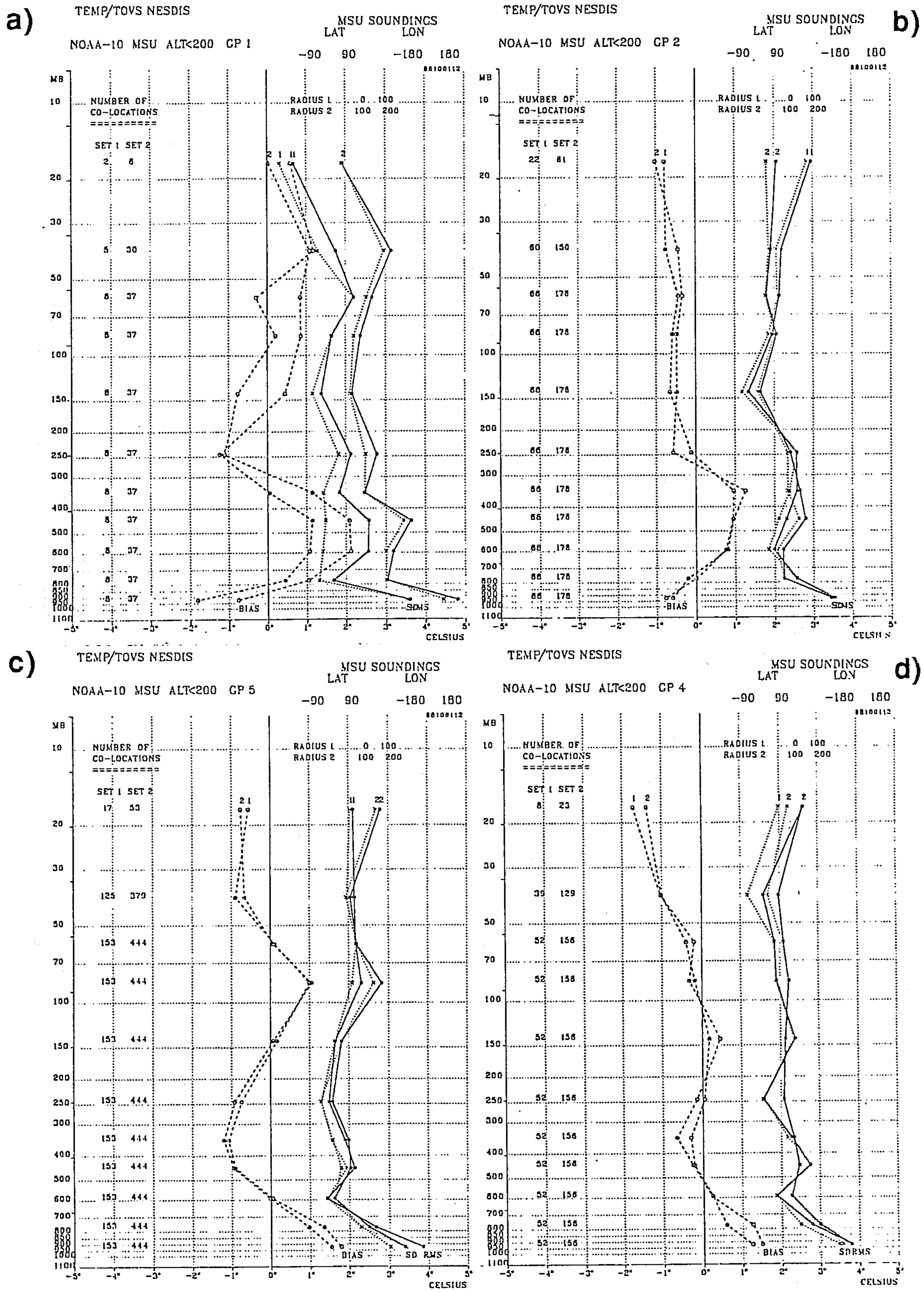


Fig.2 Statistics of the difference between NOAA-10 SATEMs and collocated radiosonde temperature profiles for various air-masses. The distance difference is up to 100 km for the curves marked (1) and between 100 and 200 km for the curves marked (2). Solid lines are RMS, dotted lines are standard deviation and dashed lines are bias. The different air-masses are: (a) Tropical, (b) Subtropical, (c) Near-Polar and (d) Polar, from a separation into five air-mass classes by discriminant analysis of temperature profiles.

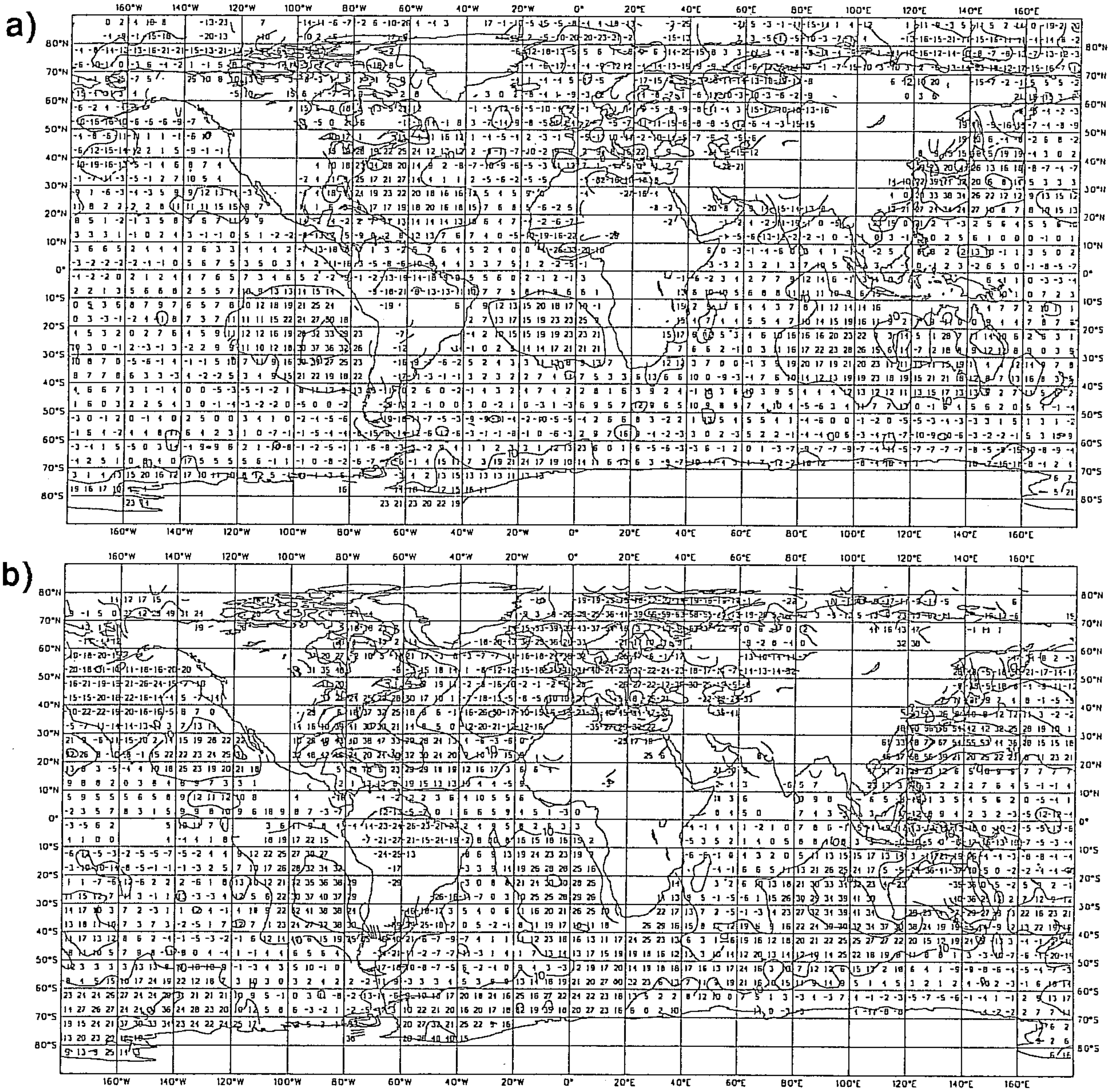


Fig.3 Monthly mean bias of NOAA-10 observed layer-mean virtual temperature deviation from the first-guess, December 1988. (a) shows clear soundings 1000-850 hPa, (b) MSU (cloudy) soundings, also 1000-850 hPa and (c) all soundings (clear, partly cloudy and cloudy) for the 1000-700 hPa layer.

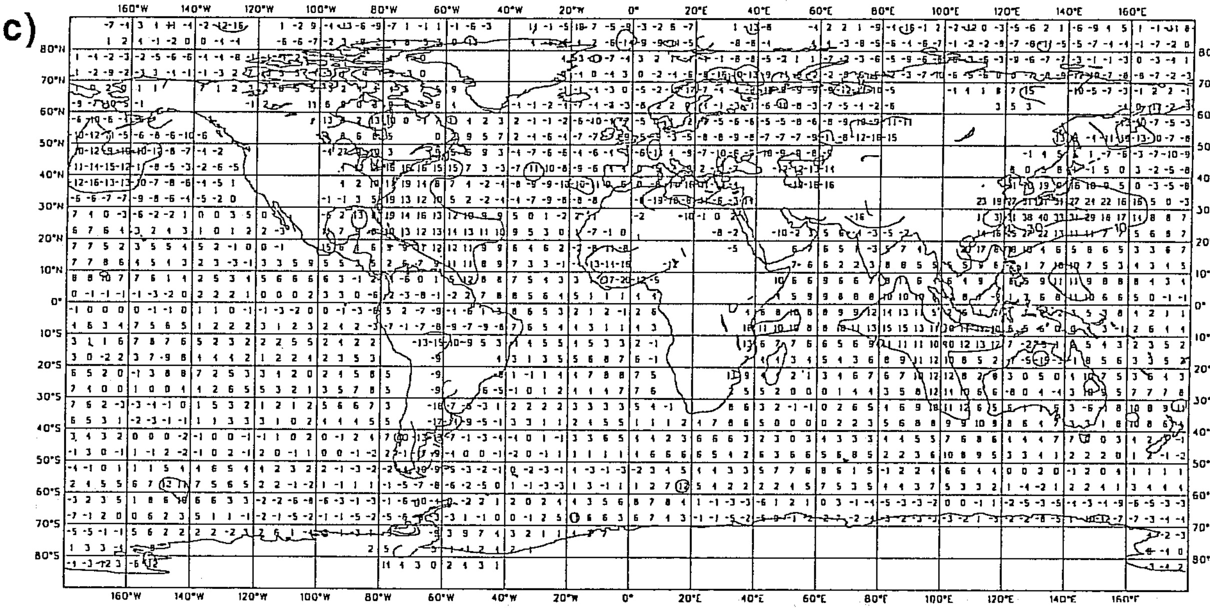


Fig.3 Continued

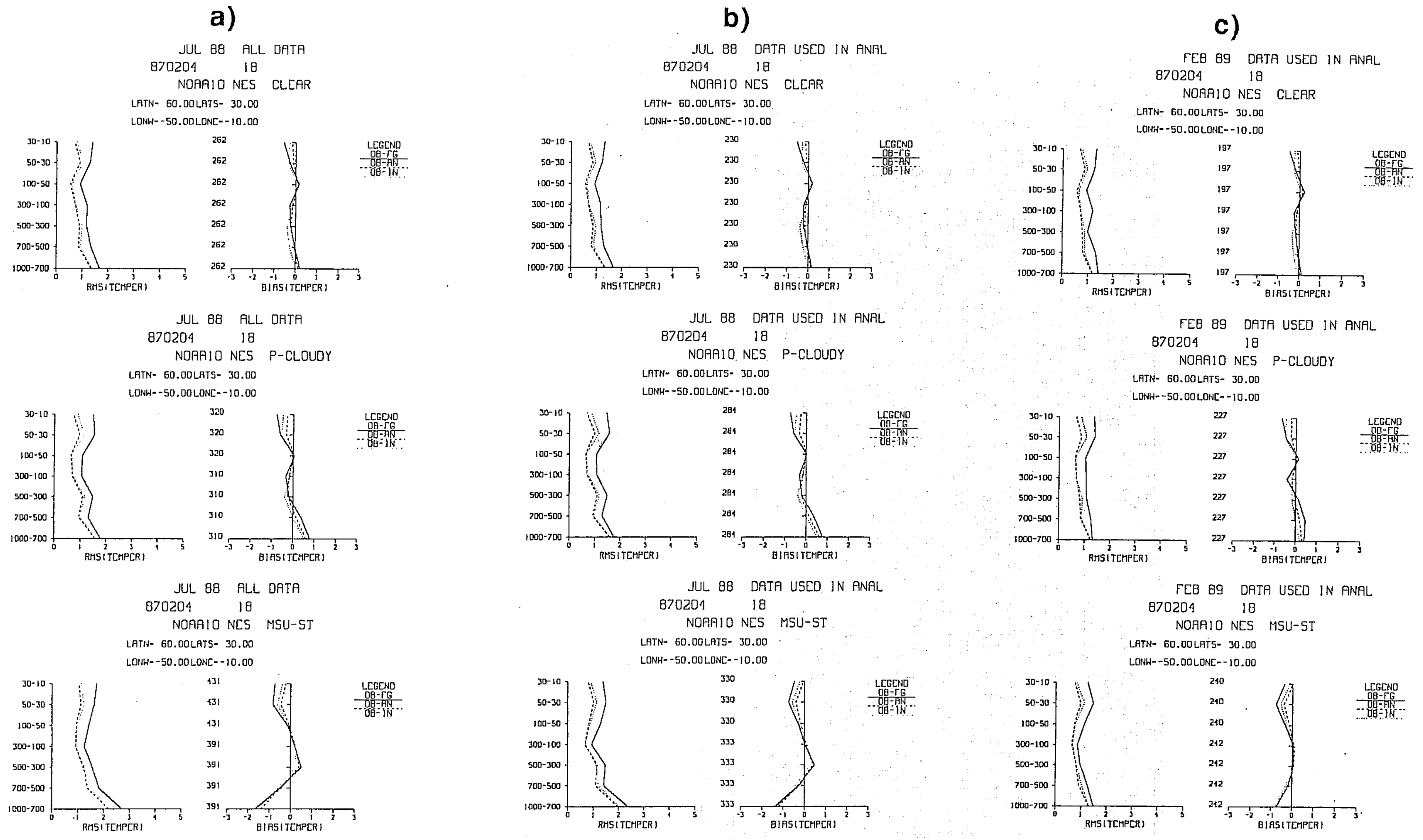


Fig.4 Mean and standard deviations of temperature differences between NOAA-10 SATEMs and first-guess (full lines), analysis (dashed) and initialized analysis (dotted), for the assimilation of the period from 30 January 1987 to 4 February 1987, using the OPS-JUL88 system. The top row is for clear retrievals, the second row is for partly cloudy retrievals, and the third row is for cloudy retrievals. (a) shows all data received in the North Atlantic, (b) the data accepted by the OPS-JUL88 analysis quality control and (c) the data accepted by the revised (tightened) quality control of the OPS-FEB89 system.

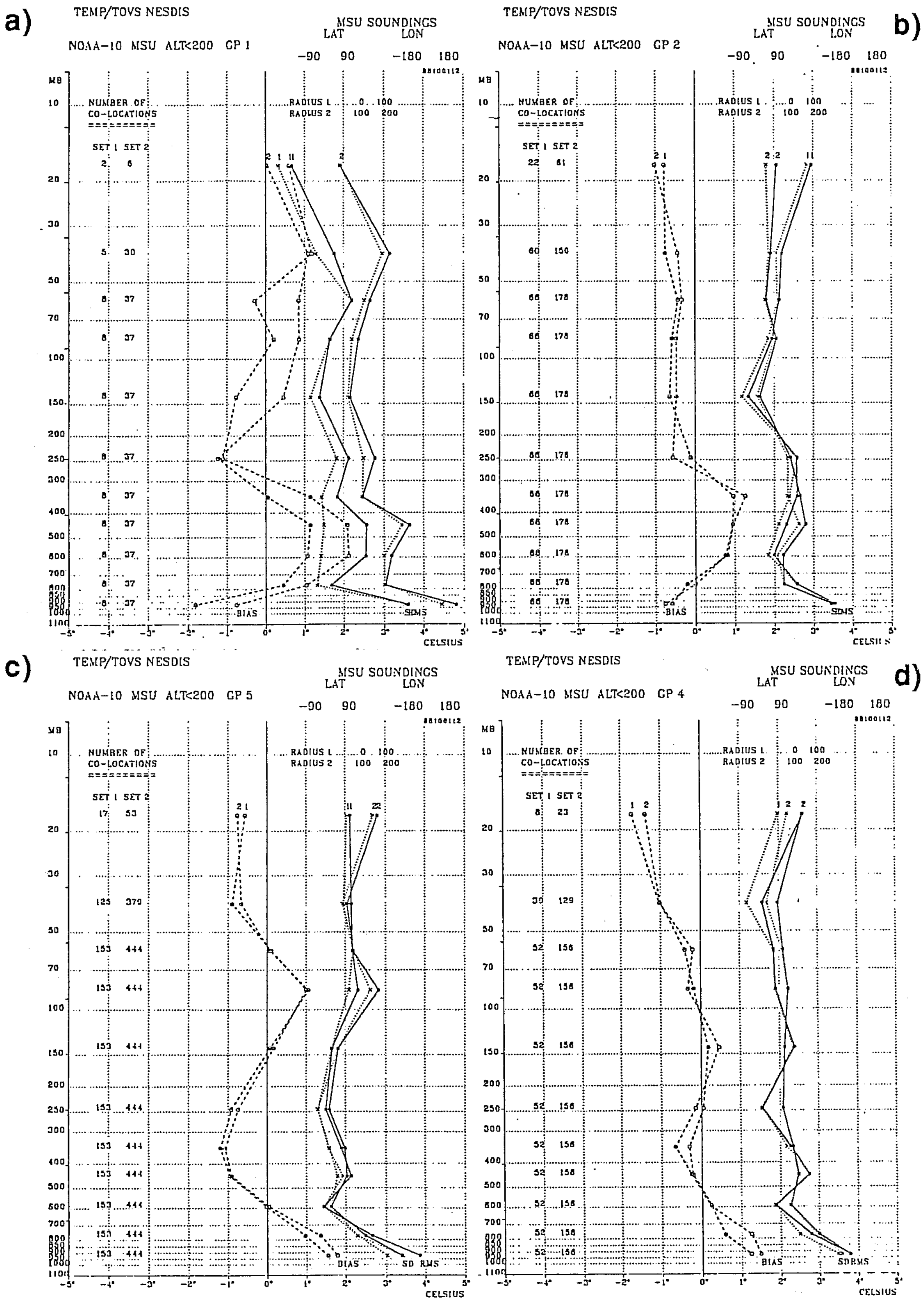


Fig.2 Statistics of the difference between NOAA-10 SATEMs and collocated radiosonde temperature profiles for various air-masses. The distance difference is up to 100 km for the curves marked (1) and between 100 and 200 km for the curves marked (2). Solid lines are RMS, dotted lines are standard deviation and dashed lines are bias. The different air-masses are: (a) Tropical, (b) Subtropical, (c) Near-Polar and (d) Polar, from a separation into five air-mass classes by discriminant analysis of temperature profiles.

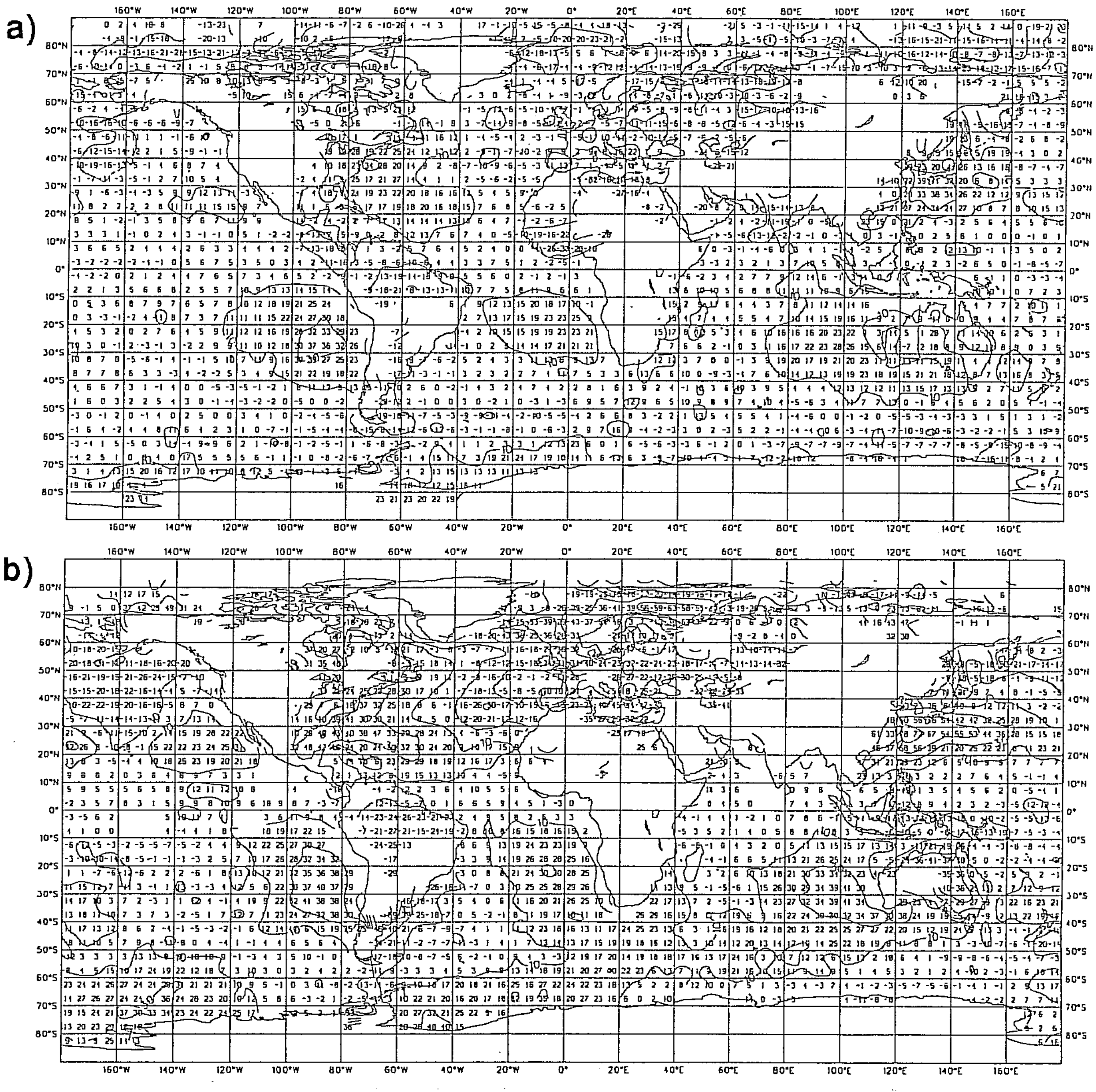


Fig.3 Monthly mean bias of NOAA-10 observed layer-mean virtual temperature deviation from the first-guess, December 1988. (a) shows clear soundings 1000-850 hPa, (b) MSU (cloudy) soundings, also 1000-850 hPa and (c) all soundings (clear, partly cloudy and cloudy) for the 1000-700 hPa layer.

c)

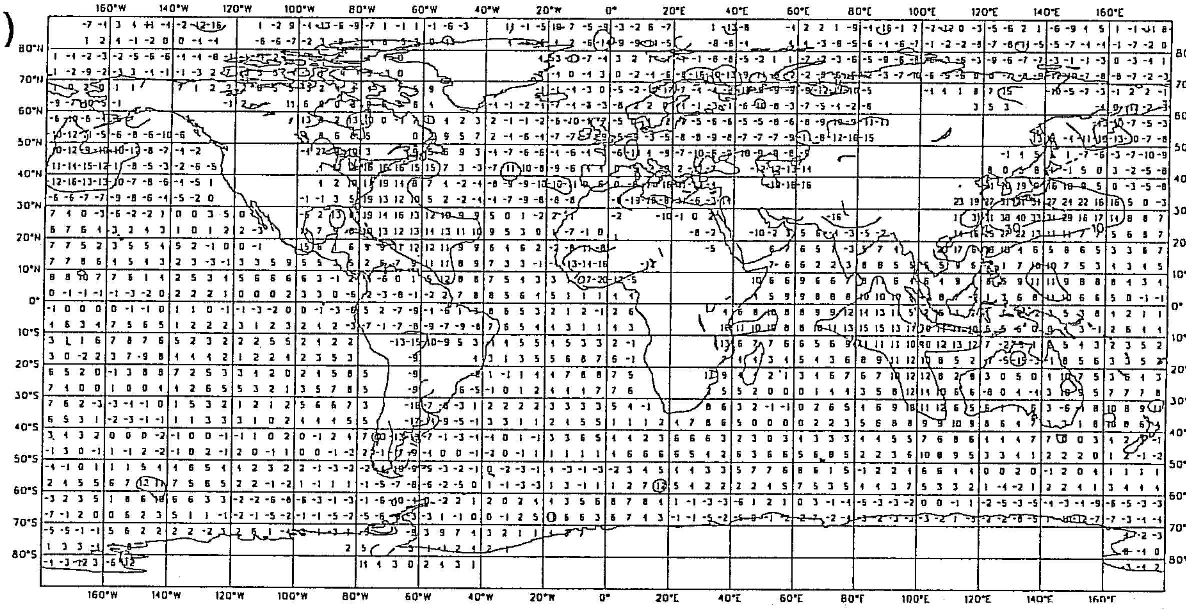


Fig.3 Continued

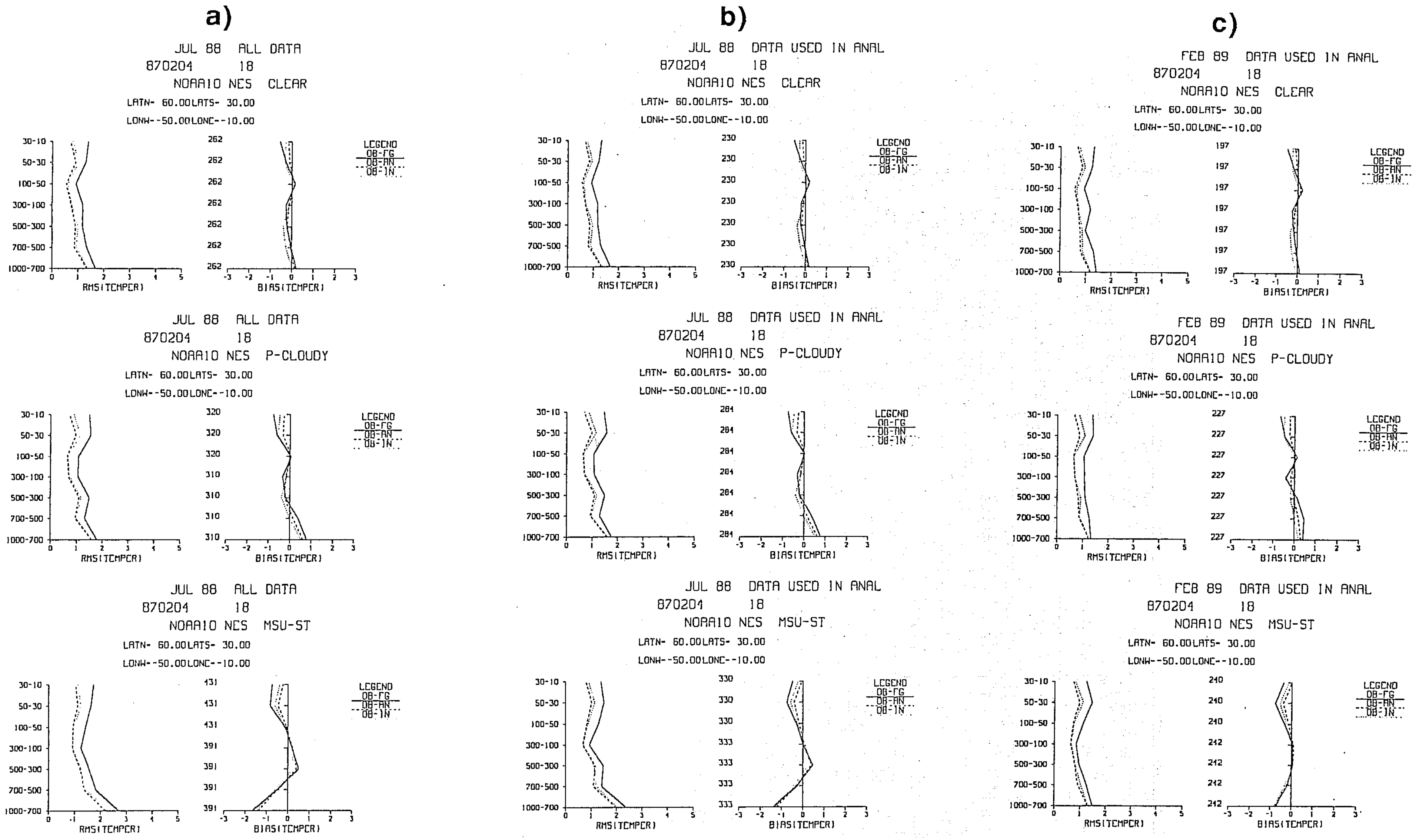


Fig.4 Mean and standard deviations of temperature differences between NOAA-10 SATEMs and first-guess (full lines), analysis (dashed) and initialized analysis (dotted), for the assimilation of the period from 30 January 1987 to 4 February 1987, using the OPS-JUL88 system. The top row is for clear retrievals, the second row is for partly cloudy retrievals, and the third row is for cloudy retrievals. (a) shows all data received in the North Atlantic, (b) the data accepted by the OPS-JUL88 analysis quality control and (c) the data accepted by the revised (tightened) quality control of the OPS-FEB89 system.

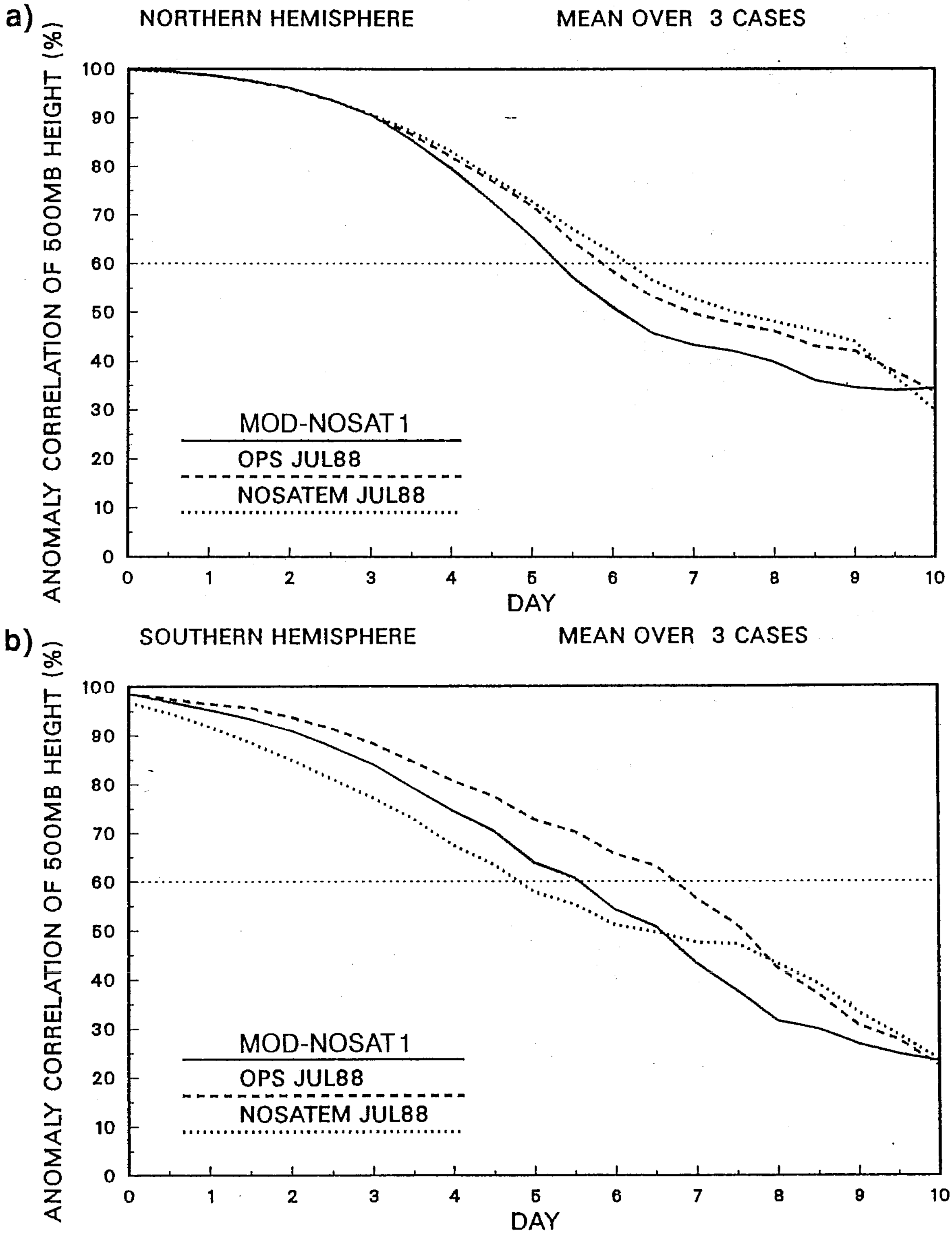


Fig.5 500 hPa anomaly correlations for (a) Northern Hemisphere and (b) Southern Hemisphere forecasts in the MOD-NOSAT1 experiment (solid) and for the two sets of control forecasts: OPS-JUL88 (dashed) and NOSATEM (dotted).

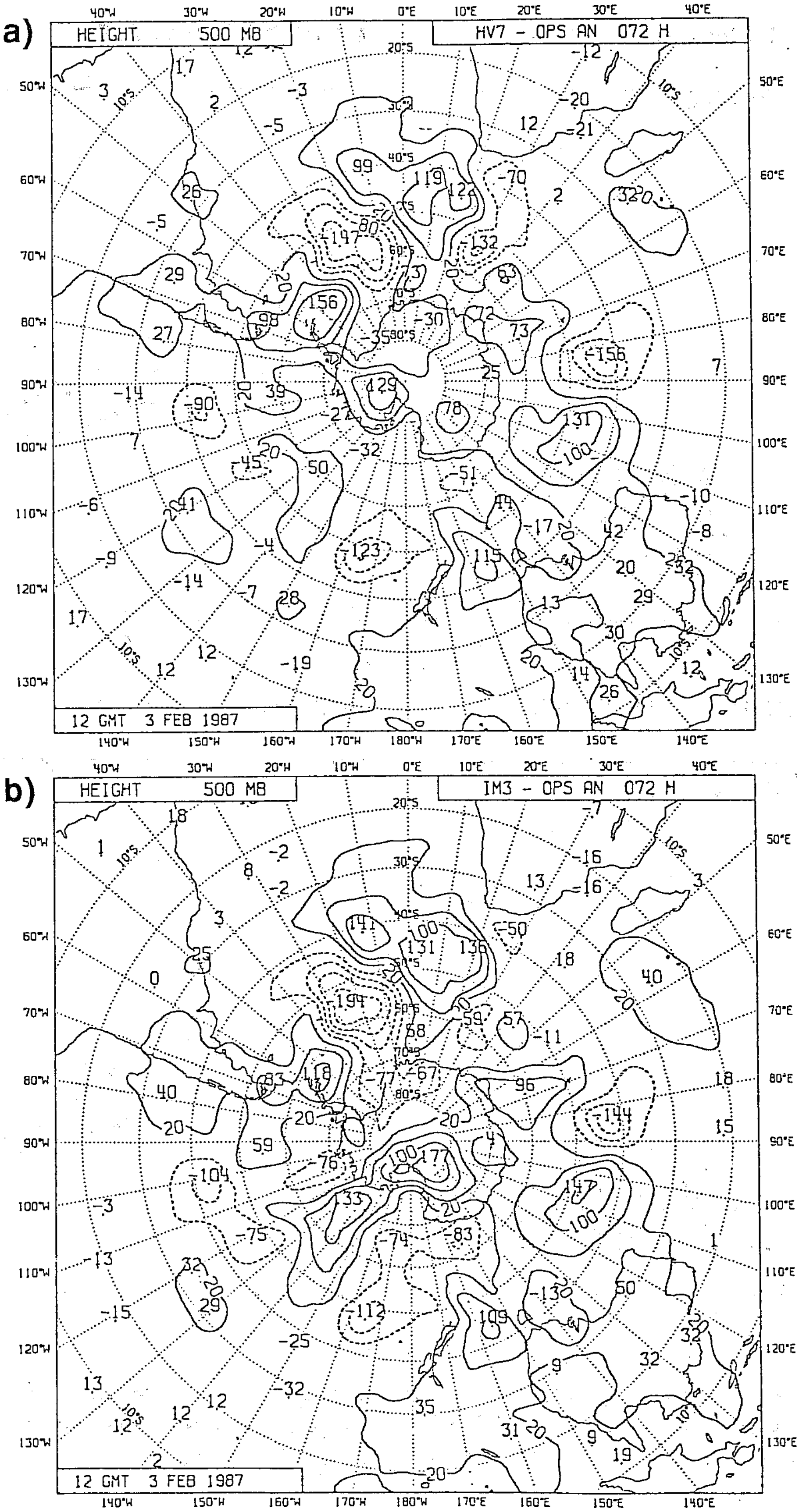


Fig.6 500 hPa 72-hour Southern Hemisphere forecast errors in forecasts from (a) the OPS-JUL88 and (b) the MOD-NOSAT1 assimilation at 31 January 1987. Contour interval is 40m starting at 20m; positive contours are solid, negative contours are dashed. The verifying analysis is the operational analysis for 3 February 1987.

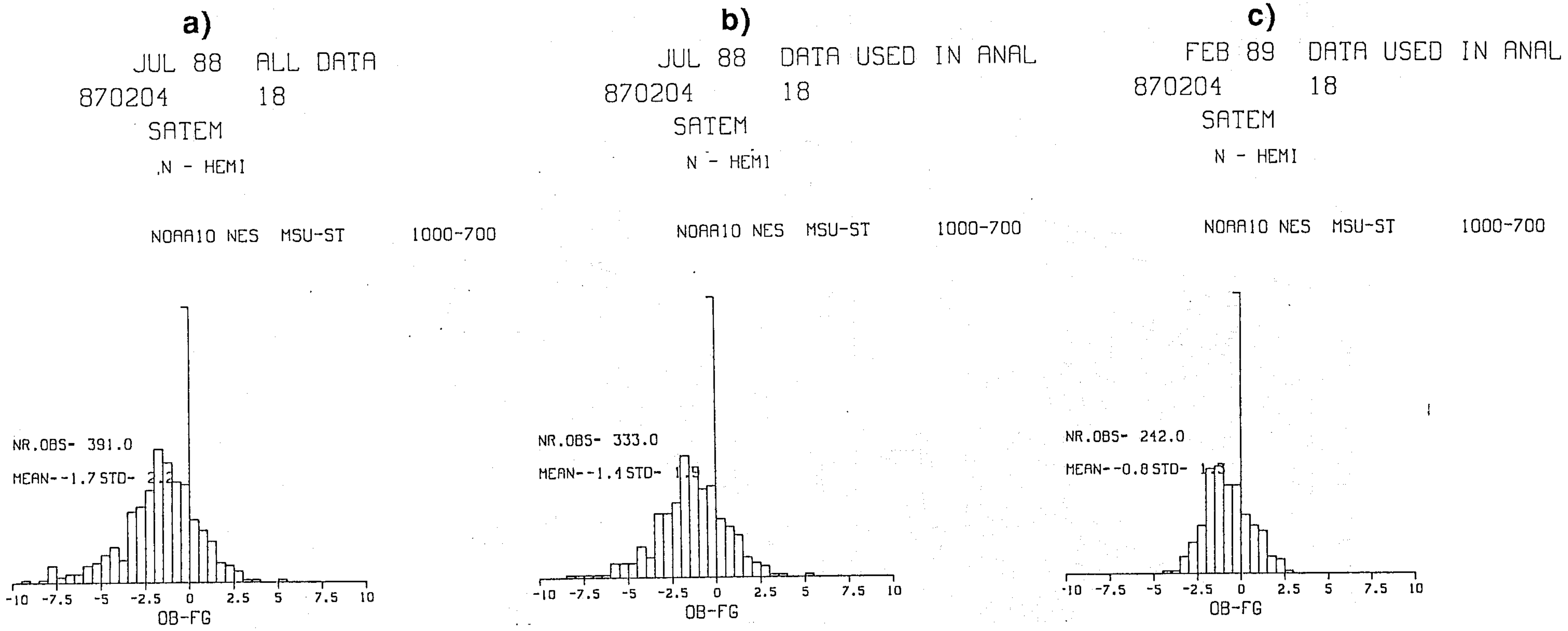
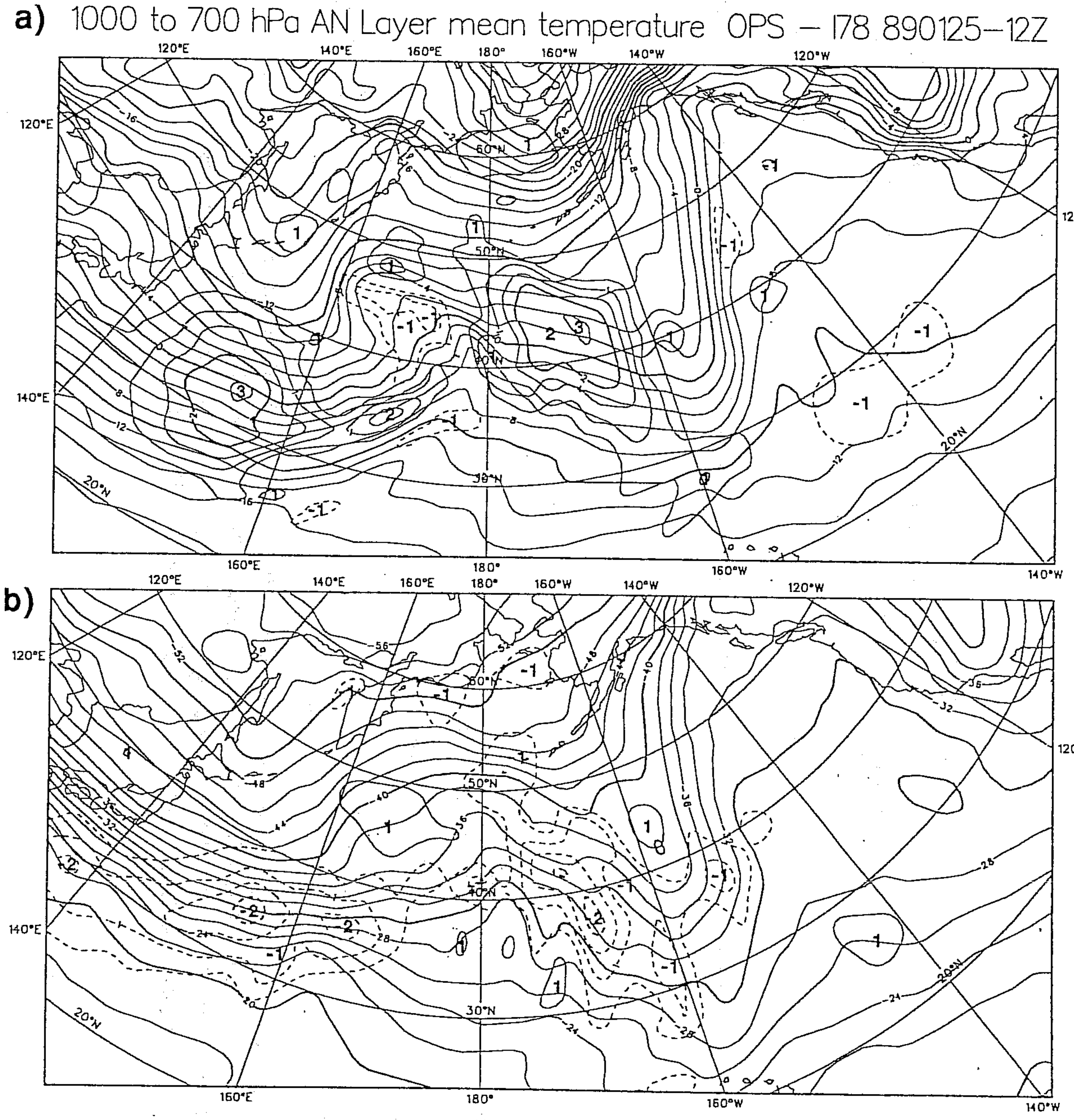


Fig.7 Accumulated histogram statistics in the Atlantic of SATEM departures from first-guess, for NOAA-10 MSU (cloudy soundings) in the 1000-700 hPa layer. (a) is all data, (b) is the data accepted by the OPS-JUL88 analysis quality control and (c) is the data accepted by the tightened quality control of OPS-FEB89



a) 1000 to 700 hPa AN Layer mean temperature OPS – 178 890125–12Z

b)

Fig.8 Layer-mean virtual temperature difference (K) between two analysis with different quality control on SATEMs: OPS-JUL88 minus OPS-FEB89, the latter with the tougher quality control, 25 January 1989, 12 UTC for the North Pacific. Contour interval is 0.5K negative differences are dashed. Also contoured is OPS-JUL88 analysis with a contour interval of 2K a) is 1000-700 hPa and b) is 500-300 hPa.

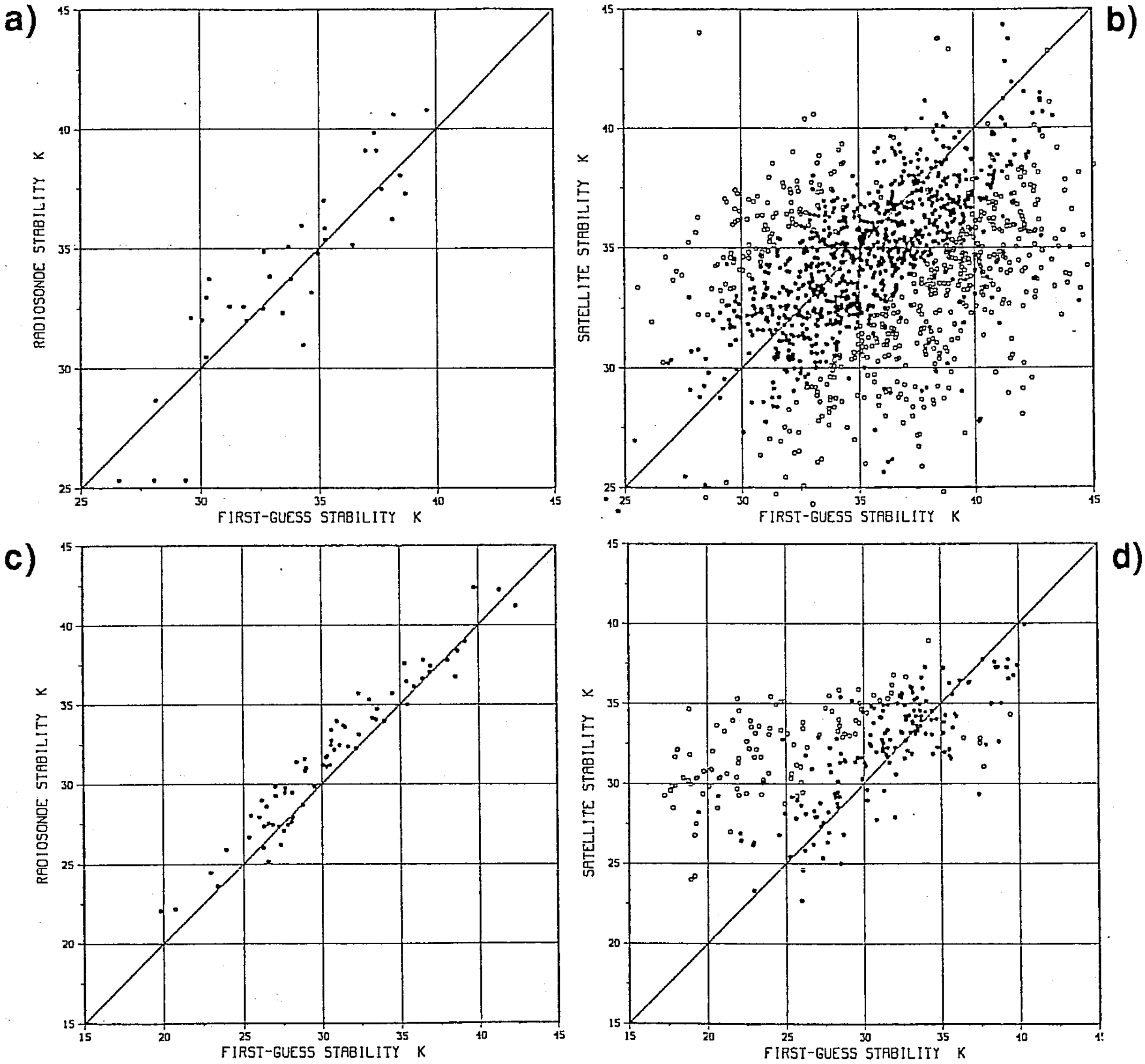


Fig.9 Scatter diagrams of observed tropospheric stability, S=T(1000-700) - T(500-300), from radiosonde observations (vertical axis) versus first-guess stability (horizontal axis), February 1989, for five different areas of the globe. The open circles represent SATEMs rejected by the analysis quality control. (a) is Atlantic Weather Ships (C, L, M) and (b) all SATEMs in the Atlantic area (50-68°N, 40°W-5°E). (c) shows five radiosondes stations along the south coast of Japan and (d) SATEMs for an area around southern Japan (25-35°N,130-140°E).

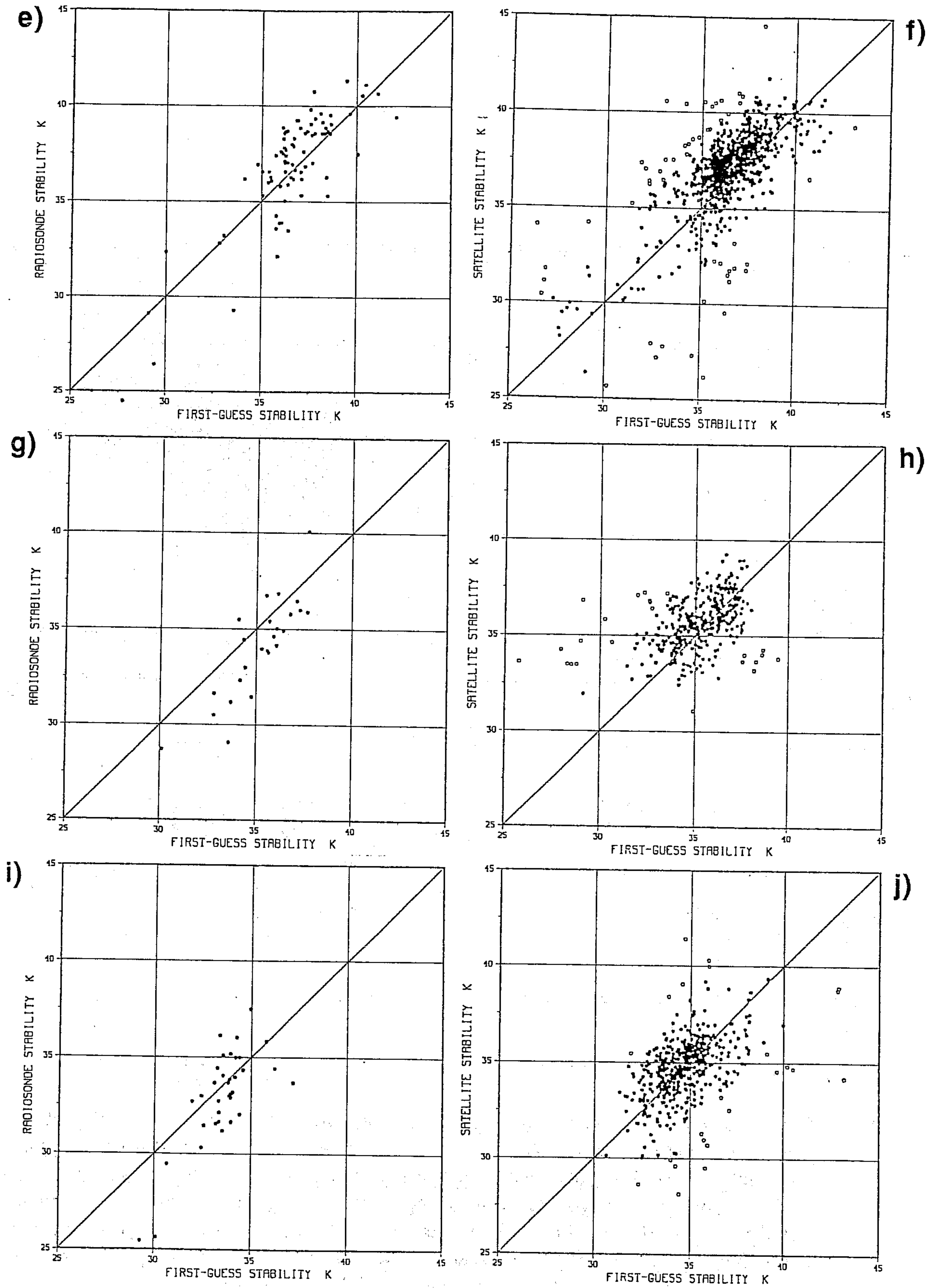


Fig.9 Continued. (e) shows the radiosondes of Midway and Hawaii and (f) SATEMs for the mid-Pacific (19-29°N,155-180°W), (g) shows the radiosonde station at Gough Island in the South Atlantic and (h) SATEMs around Gough Island (35-45°S,0-20°W), (i) shows radiosonde observations from Macquarie and Campbell Islands to the south of New Zealand and finally (j) all SATEMS in the region to the south of New Zealand (45-55°S,150-170°E).