Quality Control of Operational Physical Retrievals of Satellite Sounding Data

G. KELLY, E. ANDERSSON, A. HOLLINGSWORTH, P. LÖNNBERG, J. PAILLEUX AND Z. ZHANG

European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, United Kingdom

(Manuscript received 30 April 1990, in final form 28 February 1991)

ABSTRACT

Earlier work identified serious errors and biases in the operational temperature and moisture satellite retrievals produced by statistical methods. We show that similar errors and biases are found in the physical retrievals produced operationally since September 1988. We report experiments on quality control algorithms to deal with the errors in the satellite data. The quality control changes resulting from this work were implemented in the European Centre for Medium Range Weather Forecasts (ECMWF) system in January 1989. The performance of the quality control changes in the period after the change has been satisfactory.

1. Introduction

Developments in the ECMWF analysis–forecast system have led to a significant increase of the sensitivity of the forecast to initial data (Lönnberg 1988). An observing system experiment (OSE) by Andersson et al. (1991) found that the operational statistical temperature and moisture satellite retrievals (SATEMs) produced by the National Environmental Satellite, Data, and Information Service (NESDIS) of the National Oceanic and Atmospheric Administration (NOAA) had a negative impact on analyses and forecasts in the Northern Hemisphere when used in the ECMWF analysis forecast system as configured in late July 1988 (Andersson et al. 1991).

Synoptic and statistical investigation of the SATEM data for the 15.5-day study period of 0000 UTC 30 January–1200 UTC 14 February 1987 showed serious defects in the statistical retrievals. The defects included important airmass-dependent biases in the data. These led 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 virtual temperature.

Andersson et al. (1991) noted that the quality control procedures in the assimilation system that became operational in late July 1988 (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 an 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. The new retrieval scheme is based on a search through a library of atmospheric profiles and radiances. The library entries closest to the observed radiances are averaged to form the initial profile for the inversion of radiances to temperatures. The final temperature and humidity profile is obtained with the minimum variance simultaneous method developed by Fleming et al. (1986). Cloud clearing is done as before, following McMillin and Dean (1982), as is the statistical zenith angle correction. Handling the effects of clouds and satellite zenith angle on the observations is a very important step that can introduce larger errors in the radiances than any of the other approximations used in either the physical or the statistical method.

We examine the quality of the new SATEM physical retrievals 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 (section 2). We 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. We document the quality control changes implemented in the ECMWF system in January 1989 to remove the undesirable SATEM data (section 4) and discuss the performance of these changes (section 5). An important reason for documenting the changes is that the ECMWF analyses are widely used in diagnostic studies. Table 1 summarizes the various experiments discussed in this paper. We conclude with suggestions for further work.

2. Quality of the NESDIS physical retrievals in late 1988

We performed synoptic studies and data monitoring studies on the operational physical retrievals for No-
November, December 1988 and January 1989 to determine if the error structure of the SATEM had changed as a result of the new retrieval scheme.

In all the versions of the assimilation system discussed in this paper, the SATEM data used are mean-layer virtual temperature $T_v$ for thick layers, specifically the 1000–700, 700–500, 500–300, 300–100, 100–50, 50–30, 30–10 hPa layers from the operational SATEM bulletins. Kelly and Pailleux (1988) argue 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.

\textbf{a. 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 SATEM are affected by several problems including, inter alia:

- large errors in the lower layer (1000–700 hPa) that tend to smooth the horizontal gradients near the fronts, as the SATEM are too warm in the cold air and too cold in the warm air;
- large errors in the gross static stability of the troposphere.

Both types of error found in the physical retrievals have been found in the statistical retrievals discussed by Andersson et al. (1991). A fundamental problem with SATEM is the lack of vertical resolution in the satellite radiance measurements. This is more true of cloudy regions and, in particular, in midlatitude fronts. In such cases the microwave channels of the MSU instrument contain the only information on the tropospheric temperature and this is not enough to resolve fine vertical temperature structures in these active areas. The errors in the retrievals are therefore strongly related to the synoptic situation, which makes the errors air mass dependent and geographic in nature.

There are generally very large and coherent SATEM observation errors in the cloudy areas of the midlatitudes. There is striking organization, relative to the synoptic pattern, in the differences between the 6-h forecast (or “first guess,” FG) and the SATEM (OBS).

As an example, Figs. 1a and 1b show the first-guess $T_v$ fields for the 1000–700 hPa and 500–300 hPa layers at 1200 UTC 12 January 1989 over the North Atlantic. The figures also show the differences between the physical retrieval SATEM and the first guess. The largest deviations in the lower level in the western Atlantic exceed 6°C. They occur near a developing extratropical cyclone that deepened by 45 hPa over the following 24 h, and so falls in the “bomb” category. The figures illustrate the large discrepancy in the tropospheric stability as estimated by the 6-h forecast and the SATEM observations; negative deviations in the lower layer often coincide with positive deviations in the upper layer.

The retrieval type of the soundings are plotted in Fig. 1c (where “1” is clear, “2” is partly cloudy and “3” is fully cloudy). It is apparent that the largest deviations of the retrievals from the first guess occur in the cloudy areas with the strongest horizontal temperature gradients. (The 1000–700 hPa $T_v$ is plotted in Fig. 1a.) In this case the deviations from the first guess were so large that the data were rejected by the operational analysis quality control. The data rejection occurred in part because the first guess is known to be fairly accurate in the western Atlantic, based on comparisons with radiosonde data.

The forecast from the resulting analysis was very successful, which justified the decisions of the automatic quality control. However, the same pattern of observation errors can be seen on any daily chart in the winter months, and the suspect SATEM data were not always rejected.

\textbf{b. 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 physical retrievals were quite similar to those for the old statistical retrieval scheme. On the global scale there were very small biases, generally below 0.5°C, with SD and rms around 2°C in the midtroposphere and 3.5°C near the surface. When
compensated by biases of the opposite sign in the upper troposphere.

Figure 3 shows the 1000–850 hPa $T_v$ difference for NOAA-10 SATEMs minus first guess for clear soundings (Fig. 3a) and for fully cloudy soundings (Fig. 3b) for December 1989. The bias of the physical retrievals relative to the first guess shows a strong regional dependence, particularly in the lowest layer. For example, large positive biases in the physical retrievals occur off the east coast of North America (2.5°C/5.0°C) and Asia (4.2°C/7.9°C, for clear/cloudy soundings, respectively). These features are very similar to what was seen with the statistical retrievals in Andersson et al. (1991) and presumably arise for similar reasons.

Besides the extratropical problems, Figs. 3a and 3b show large positive biases in the 1000–850 hPa layer in the subtropical highs. The largest biases occur for the fully cloudy soundings and exceed 4°C in the south Indian Ocean (Fig. 3b). Only a small part of the bias comes from the first guess. The first guess is almost unbiased when verified against radiosondes in the subtropics, although it has a small positive bias in the deep tropics.

Figure 3c shows the temperature bias for all soundings in the 1000–700 hPa layer 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 (maximum 4°C), and somewhat smaller biases off the coast of North America (1.9°C). This is a reflection of a systematic underestimation of the temperature lapse rate in cold-air outbreaks.

In the subtropics the biases for the 1000–700 hPa layer are smaller than those for the 1000–850 hPa layer, so there is a compensation of bias between the 1000–850 and 850–700 hPa layers in some areas. Nevertheless, there is still a large area with 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 $S$:

$$S = T_v (1000–700) - T_v (500–300)$$

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. (1991) 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 7°C 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 4°C 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...
Fig. 2. Statistics of layer-mean virtual temperature difference between NOAA-10 fully cloudy SATEMs and collocated radiosonde temperature profiles for various air masses. The distance difference is up to 100 km. Solid lines are rms and dashed lines are bias (°C). The different air masses are: (a) tropical, (b) subtropical, (c) near-polar, and (d) polar. Pressure bounds of the plotted layers are 1000, 850, 700, 500, 400, 300, 200, 100, 70, 50, 30, and 10 hPa.
from the first guess is very large north of 30°N for both satellites and generally lies between 3° and 6°C.

In the midlatitudes 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).

**c. Case studies of forecast sensitivity to physical retrievals in November 1988**

Extensive experimentation is needed to document the overall impact of SATEM data on analyses and forecasts. There can be large variations from case to case, presumably depending on a case-to-case variability of the SATEM data. Two examples will illustrate the point:

- on 1200 UTC 3 November 1988 the operational forecast error was unusually large over the Pacific and North America. The SATEM data quality for the initial time appeared to be rather poor over the Pacific. An assimilation without SATEMs (NoSATEM) was run and gave a significant improvement of the forecast scores.

- the forecast run for the previous day (1200 UTC 2 November 1988) 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.

These examples illustrate the sensitivity of the ECMWF system to the physical retrievals, a feature noted earlier for statistical retrievals. The results of this section show that overall, the physical retrievals suffer from very similar defects to the statistical retrievals studied by Andersson et al. (1991).

**3. Quality control and analysis of SATEM data**

The forecast experiments and synoptic studies of physical retrievals in the last section found results very similar to those described by Andersson et al. (1991) for statistical retrievals. Serious problems have been identified with both the statistical retrievals and physical retrievals: there are large airmass-dependent biases in both types of retrievals, with associated errors in static stability in the midlatitude baroclinic zones. It is important to understand how these biases may affect the analyzed fields.

**a. The effect of data bias on the analyzed fields**

The analysis system used at ECMWF assumes that both the observations and the first guess are unbiased. In the absence of independent data the analysis system cannot compensate for data biases; clusters of data with uniform biases will influence the analyses, provided the data pass the quality control procedures. Thus, we expect that the shortcomings in the SATEM data will have a strong effect on the analyzed fields, both in the horizontal and in the vertical.

Figure 4a shows statistics on the differences in the North Atlantic in the period 30 January–4 February 1987 between three types of satellite retrievals (clear, partly cloudy, cloudy) and the three fields (first guess, analysis, initialized analysis) used in a trial assimilation with the OPS-Jul88 system, in terms of bias and standard deviation. Biases in the data relative to the first guess are clearly evident. The bias is largest in the fully cloudy retrievals and is of opposite sign to the bias in the other types of retrievals. Figure 4b shows the same set of statistics for the data actually used in the assimilation, about 10%–15% of the data were rejected. Both Fig. 4a and 4b show that most of the bias in the stratosphere has been assimilated by the analysis. In the lower troposphere the conflict in the sign of the bias between the cloudy retrievals on the one hand, and the clear or partly cloudy retrievals on the other, has prevented the analyses from being influenced too much by either bias. However, synoptic studies show that in conditions of extensive cloud cover, where only cloudy retrievals are available, the analyses become biased because of the biased SATEM data.

**b. Discussion**

The filtering properties of the optimal interpolation (OI) analysis used at ECMWF are discussed in section 5b. In the absence of independent data, biased observations will result in biased analyses. From a practical point of view, the alternative approaches to reliance on OI to filter the SATEM data bias would be to:

- correct the SATEM data itself with empirical corrections, or
- prefILTER the data through scale-selective filtering in the vertical, or
- remove the bias through data rejection in quality control procedures.

Correction of biased satellite retrievals by a user such as ECMWF does not seem feasible because of the tun-
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 to 4 February 1987, in the North Atlantic (30°–60°N, 10°–50°W). 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, (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. The numbers plotted on a vertical axis between diagrams indicate sample size.
ing that is carried out continuously by the data producers. Prefiltering in the vertical might, through the strongly nonlinear properties of the radiative transfer equation, result in a profile that no longer satisfies the measured radiances. The more practical short-term solutions appear to be either to eliminate whole categories of SATEM data or to develop tailored quality control procedures.

4. The January 1989 quality control modifications

To quality control the data we examined the possibility of excluding certain categories of SATEM data, e.g., cloudy and partly cloudy retrievals. This did not work well, so instead of excluding whole categories of SATEM data we 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.

a. 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.

1) REVISED OI CHECK

Since erroneous SATEMs tend to occur in clusters they provide spurious support for each other in a conventional OI quality check. The OPS-Jul88 system (Lönngberg 1988) contained a quality control modification that checks SATEMs without using neighboring 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 multilevel 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.

2) TIGHTER CHECK ON SATEM MINUS FIRST- GUESS DEPARTURES AT A SINGLE LEVEL

The OPS-Jul88 first-guess check performs well for the midtropospheric layers but it generates very few rejections in the lowest layer and in the top two strato- spherical layers. This is due to the higher FG error variance for those layers. The threshold for rejection is specified in terms of standard deviation of the "normalized departure" [(OBS-FG)/SD of FG]. We have implemented a reduction from a threshold of 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 4°C for the 1000–700 hPa mean-layer virtual temperature in the eastern part of the North Pacific.

3) STABILITY CHECK

A useful method to identify incorrect SATEM 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, particularly in overcast situations. From comparisons with ocean stations and weather ships (Andersson et al. 1991, and section 5 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 satellite data (in the form of SATEMs) have almost no skill in measuring this stability index S. Based on scatter diagrams like those in section 5 we chose to reject soundings in our test assimilations where the value of S in the SATEM differed from S in the first guess by more than 4.5°C.

b. Experiments with the pre-Feb89 quality control

The tighter quality control of section 4a was introduced in operations on 31 January 1989 after tests in the February 1987 period. The stability check had a threshold of 4.5°C in the tests but it was decreased to 3.5°C for operations. For clarity of presentation we shall speak of the system with the 4.5°C cutoff on the stability index S as the pre-Feb89 system, and the operationally implemented system with the 3.5°C cutoff as the OPS-Feb89 system.

To examine the effect of the pre-Feb89 changes we studied in detail the SATEM data statistics from the OPS-Jul88 assimilation on the period 30 January 1987 to 4 February 1987. Figure 4 shows summary plots of the data volumes and departures of the SATEM data from the first guess, analysis, and initialized analysis for NOAA-10. As already discussed, 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. Figure 4b shows the corresponding plots for all NOAA-10 data accepted and used in an assimilation for the period with the OPS-Jul88 system. Figure 4c shows the corresponding plot for the data that would have passed the 3.5°C 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 by the OPS-Feb89 system 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 SATEM departures were rejected to the north of 20°N in any given analysis.

Figures 4b and 4c 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 in the cloudy retrievals.

Figure 5 shows more detail on the effect in the North Atlantic of the quality control changes on the NOAA-10 data used by the OPS-Jul88 and pre-Feb89 assimilations of the period 30 January 1987–4 February 1987. The figure shows histograms of NOAA-10 $T_e$ (1000–700) departures from the first guess for the fully cloudy retrievals. The corresponding results for NOAA-9 were similar and are not shown. Both in the Atlantic and in the Pacific (not shown) 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 is of order 1°C over the large oceans. The spread of the histograms is considerably reduced in the accepted data, as expected.

c. Tests with the OPS-Feb89 system

The OPS-Feb89 change has been tested in a data assimilation started on 1200 UTC 24 January 1989 with forecasts run from the 25, 26, and 27 January, all at 1200 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.

The differences between the OPS-Jul88 and OPS-Feb89 analyses are large, particularly in the temperature over the North Pacific. Figure 6 illustrates the analysis differences in the lower (Fig. 6a) and middle (Fig. 6b) troposphere at 1200 UTC 25 January 1989. There are large differences in the analyzed static stability index of up to 5°C [near (35°N, 155°E) and (40°N, 168°W)], mainly due to additional SATEM rejections.

The effect on the forecast skill of the change from OPS-Jul88 to OPS-Feb89 was small. The cases of 25 and 27 January 1989 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.

To summarize, the effect of the tougher quality control on the short set of test forecasts was small. Given the weight of the statistical and synoptic evidence, the decision was nevertheless taken to implement the OPS-Feb89 change.

5. Evaluation of the OPS-Feb89 quality control changes

The operational assimilations in February 1989 were run with the OPS-Feb89 system described above. The
AN Layer mean temperature

(a) 1000 to 700 hPa

(b) 500 to 300 hPa

Fig. 6. OPS-Jul88 North Pacific layer-mean temperature analysis for (a) the 1000–700 hPa and (b) the 500–300 hPa layers for 1200 UTC 25 January 1989 (contour interval 2°C). Superimposed on the plots are the layer-mean virtual temperature differences between the OPS-Jul88 analysis and the OPS-Feb89 analysis which has tougher quality control on SATEMs (OPS-Jul88 minus OPS-Feb89). The contour interval for the differences is 0.5°C, with negative values dashed.

Rejected satellite data were typically clustered in rather large groups associated with areas of warm advection or cold advection in midlatitude 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 outbreak over the Gulf of Mexico, and along the front in the Atlantic. The number of rejections in the Southern Hemisphere is comparatively small due to the larger first-guess error there and to seasonal effects.

a. Radiosonde-based cross validation of the first guess and the SATEM data

Perhaps the most convincing evidence for the problems with the statistical TOVS retrievals presented by
Andersson et al. (1991) was a series of scatterplots 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 study similar plots to demonstrate that (a) the major retrieval problems have been unaffected by the change from statistical to physical retrievals and (b) the quality control changes described in the previous section perform well.

The scatterplots shown in this section are based on the operational 1200 UTC analyses during February 1989 and were generated with the OPS-Feb89 assimilation system. In order to demonstrate the performance of the OPS-Feb89 quality control, we show the accepted data with full circles, and we show the rejected data with crosses.

1) NORTH ATLANTIC

Figure 7a shows the scatterplot of radiosonde reports for \( S \) from the three Atlantic weather ships (C, L, and 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 5°C. None of the radiosonde data are rejected.

Figure 7b shows the corresponding scatterplot for \( S \) from the SATEMs (physical retrievals) in the area of the Atlantic bounded by 40°N, 50°N, 40°W, and 5°E. The largest deviations between SATEM and first guess are of order 15°C. The scatter is much larger than in Fig. 7a, 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 crosses lying more than 3.5°C from the diagonal have been rejected by the check against the first guess, and that crosses lying closer to the diagonal have been rejected by the main analysis check, then it is evident that the check against the first guess causes most of the rejections.

2) SOUTHERN JAPAN AND ADJACENT SEAS

Figure 7c shows the scatter in \( S \) of the first guess compared with the radiosondes for a group of five sondes along the south coast of Japan. The first guess tends to be too stable by slightly over 1°C. The largest deviation around this bias is about 1.5°C. None of the radiosonde data are rejected.

Figure 7d shows the corresponding scatter in the SATEMs versus the first guess in the same area (25°-35°N, 130°-140°E). There is an enormous bias in the SATEM data, as well as a huge scatter. For cases of large static stability (low values of \( S \)) in the first guess, the bias in the SATEM value of \( S \) is as large as 10°C. The largest deviations are about 15°C. Much of the SATEM data are quite properly rejected by the first-guess check on \( S \). The first guess here is much more accurate than the SATEM data.

3) ISLANDS IN THE SUBTROPICAL MID-PACIFIC

Figure 7e 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°-4°C, and none of the data are rejected.

Figure 7f shows the corresponding data for the scatter of SATEM measurements against the first guess. The largest deviations are 8°-9°C. Even in a data sparse area such as the subtropical mid-Pacific, it would appear that the first guess for \( S \) is more accurate than the SATEM estimates.

4) THE EXTRATROPICAL SOUTH ATLANTIC

Figure 7g shows the scatter in \( S \) for the radiosonde at Gough Island (in the extratropical South Atlantic), plotted against the first guess. There is evidence of a bias in the first guess of order 1°C, with a small scatter. Largest deviations between radiosonde and first guess are about 5°C, with typical deviations of order 2° to 3°C. Allowing for the seasonal differences between North and South Atlantic in February, the performance of the first guess for \( S \) in the South Atlantic (Fig. 7g) is not very different from the performance in the North Atlantic (Fig. 7a).

Figure 7h shows the corresponding scatterplot for \( S \) for the SATEMS in the area 35°-45°S, 0°-20°W. The largest deviations are about 8°C. 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.

5) THE SOUTHERN OCEAN, SOUTH OF NEW ZEALAND

Figure 7i 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.

Figure 7j shows the corresponding scatterplot for the SATEMS against the first guess in the area 45°-55°S, 150°-170°E. 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 this area.

6) SUMMARY

The results of this section demonstrate that the first-guess static stability in most parts of the world is rather accurate when verified against radiosonde data. The results also show that the physical retrieval SATEM data on the static stability is very noisy and very biased in some areas. A similar result was found in earlier work on the statistical retrievals. It is therefore reason-
able to use the first-guess static stability to screen out
the largest errors in the SATEM retrievals. The current
performance of the quality control algorithms seems
satisfactory for present purposes.

b. Filtering in the optimal interpolation (OI) system,
and the vertical correlation of observation errors

The scatterplots just shown have important impli-
cations for specifications of the filtering properties of
the OI analysis (Lorenz 1981; Hollingsworth 1987).
Strauss (1989) shows that the correlation of the SA-
TEM 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 that
observation error is larger than analysis error, Strauss'
results suggest that the observation error correlation
matrix $D$ should have a strong negative correlation be-
tween the 1000–700 hPa and 500–300 hPa layers. Use
of such a feature in $D$ would lead to a response that
gave more weight to a tropospheric mean temperature
than to a tropospheric gross static stability (Lönnberg
1989).

When the data used for the correlation calculation
are screened by the quality control procedures de-
scribed above, 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 agreed with the calculations on
screened data.

It is important to know what data to use to estimate
the error characteristics of the SATEM data used by
the analysis system. Should one use all the data (i.e.,
dots and crosses in Fig. 7) in collocation studies with
radiosondes to determine the error of the SATEMs, or
should one use only the accepted data (i.e., the dots
only)? 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 dataset
is used.

The OI system assumes there is no correlation be-
tween the forecast error and the observation error. One
approach to estimating the error correlations is to ig-
nore 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 would then
assume that even the screened data are typical of a
population that is capable of having an error in the
observation of $S$ as large as 15°C and so is rather in-
accurate. 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.

An alternative approach would be to assume that
the SATEM data used in the analysis really represent
a combination of satellite information and first-guess
information, and the screened data 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. However, the errors of
the screened observations must have important cor-
relations with the error of the first guess. Such corre-
lations 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 sim-
pler. It would probably result in significantly negative
vertical correlations for SATEM error between the up-
per- and lower-tropospheric layers. This would then
have the effect of giving more weight to information
in the SATEM data on the mean temperature of the
troposphere compared to information on the gross
static stability.

c. 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 ex-
amine 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.

There is still an unresolved question about whether
it is best to use all the data, or only the screened data
in calculations aimed at specifying the observation error
characteristics of the SATEM data. Each approach has
limitations and difficulties, and further experimenta-
tion is needed.

6. Conclusions and further developments

The extensive experimentation in Andersson et al.
(1991) highlighted the sensitivity of the ECMWF analysis-
forecast system to SATEM data and demonstrated
serious quality problems in the SATEMs produced by
statistical retrievals of TOVS data. NESDIS changed
their retrieval procedure to a physical retrieval in Sep-
tember 1988. In this paper we have demonstrated that
the SATEMs produced by the physical retrieval pro-
cedure contain errors and biases that are as serious as
those contained in the statistical retrievals. These errors
in the SATEMs have an adverse effect on analysis and
forecast quality. On the plus side, the physical retrieval
technique appears to have eliminated some lateral in-
consistencies between adjacent satellite tracks that were
seen in the earlier statistical retrievals.

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

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.

Our basic assumption in the work on quality control has been that if use of all soundings produces a worse forecast than use of no soundings, inclusion of just those retrievals close to the first guess will not degrade the forecast as much, or even help it. A problem with this approach, of course, is that in areas where the first guess is poor, good retrievals may be rejected. This is precisely where the satellite information would be most important if it were accurate and reliable. The use of a tighter quality control (tolerance of less difference from the first-guess in the rejection criteria) in a short test assimilation (section 4c), rejected a significant number of retrievals but did little to improve forecast skill. Nevertheless, the new tightened criteria were made operational. Part of the justification of using tightened criteria are figures such as Figs. 4c, 5, and 7, which
show that the accepted retrievals agree closer statistically to the first guess and to the radiosondes when a tighter quality control is used.

Work is underway 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 under way to make use of the forecast first guess to quality control the cloud-cleared radiances following work of Flobert et al. (1991).

We have not yet fully resolved the issue of how best to estimate the vertical correlation of SATEM observation error. Further experimentation is needed in this area.

Longer term approaches to improving the use of satellite data include the use of different retrieval methods (Flobert et al. 1991), and the use of variational retrieval methods. It is now becoming clear that if satellite retrieval methods do not make use of additional information, such as a 6-h 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.

Acknowledgments. We thank Mr. R. Hine and Mrs. J. Williams for their work on the graphics and Mrs. M. Simpson for typing the manuscript.

REFERENCES


