DATA MONITORING AT ECMWF

A. M. Radford

European Centre for Medium-Range Weather Forecasts Shinfield Park, Reading, U.K.

1. INTRODUCTION

Much progress has been made in recent years in numerical weather forecasting. Models and analysis techniques have been developed and refined while the rapidly increasing power of supercomputers has actually enabled the scientists to implement their schemes. However, numerical forecasting is not only a modelling but also an initial data problem. At a recent meeting on data for global models (ECMWF, 1987) the need for an improved data coverage and better quality data was stressed. Much can be gained from improvements in the Global Observing System (GOS), but very little has happened in the last few years. Instead of an improvement there are signs of a degradation of the system, in particular in the conventional surface based observations.

In order to manifest the deficiencies in the availability and quality of the observations, ECMWF undertakes regular data monitoring and has on various occasions made the results available to the WMO, data producers and other GDPS centres. WMO/CBS has endorsed such monitoring activities and recommended the exchange of results between monitoring centres (WMO, CBS Ext(85)).

The diagnostic facilities for the data monitoring activities at ECMWF are provided by the data assimilation system described by Lorenc (1981), with recent revisions documented by Shaw, Lönnberg and Hollingsworth (1984) and Lönnberg, Pailleux and Hollingsworth (1986).

This paper describes some of the different tools that ECMWF has at its disposal to monitor the performance of the GOS. Some recent results concerning both the availability and quality of the data are presented with reference to three different timescales - daily, monthly and annually.

2. DAILY MONITORING

Daily, or real-time, monitoring involves the day to day survey of the availability and quality of the observations as received via the Global Telecommunications System (GTS) at ECMWF and is essential in order to react quickly to any problems that may occur.

2.1 Availability

In order to check that observations are reaching ECMWF and are flowing into the reports database, data coverage maps of each observation type are produced four times a day for the four analysis cycle periods (00 UTC, 06 UTC, 12 UTC and 18 UTC). An example of one such map for radiosondes is shown in figure 1. A symbol is plotted for every report received. Complementing these maps are graphs displaying the total number of observations received (for each type) over the last 7 days for both hemispheres. Any deficiency in the data coverage is shown up immediately and is investigated.

2.2 Quality

Observations presented to the data assimilation undergo several automatic quality control checks, including:

- comparison with climatological values;
- hydrostatic check (for radiosondes);
- comparison with the first-guess field.

All data that fail any of the checks are plotted on a global rejection map, an example of which is shown in figure 2. Each symbol corresponds to a rejected observation, with different symbols representing the various data types. Any abnormal rejection patterns, such as several observations being rejected in a small area, are investigated further and are normally indicative of deficiencies in either the data, the first-guess or the analysis.

Once the observations have passed the automatic checks they may be used in the data assimilation system. They are used in conjunction with the

first-guess field (the 6-hour forecast from the previous analysis) in order to produce a new analysis. The impact of the data is represented by the difference between the analysis and the first-guess fields (the analysis increments). Excessive increments can indicate problems with either the data or the first-guess and are investigated immediately. For example, figure 3 shows the 200 hPa height increment chart over North America at 00 UTC on 28 December 1986. The magnitudes of the increments in this case were considerably larger than normal, and further investigation revealed that they were caused by the duplication (on the GTS) of the U.S. radiosonde reports from 24 hours earlier. Figure 4 shows the erroneous reports superimposed on the first-guess field.

Another tool at the disposal of the meteorological analyst is a plot of vertical profiles of collocation statistics between radiosondes and polar orbiting satellites. On 1 January 1987 an 'earth location' problem at NOAA/NESDIS caused the orbits of the retrieved NOAA-9 soundings to be displaced by 90 degrees longitude. The routine plot of collocation statistics immediately highlighted the problem and prompt action was taken to inform NESDIS and eliminate the data from subsequent analyses. Figure 5 shows how the problem was reflected in the vertical profiles. Collocation statistics are shown for two radii (0 - 100 km and 100 - 200 km) around the location of each radiosonde station, sample sizes are given on the left. Small bias values and a tight data fit in the RMS sense indicate good agreement between satellite and radiosonde soundings. Note the degradation in the data on 1 January.

3. MONTHLY MONITORING

Day to day monitoring is essential in order to react quickly to problems but does not provide a good understanding of the systematic deficiencies in the availability and quality of the data. An accumulation of statistics over a longer period of time, such as a month, is needed. It should be noted that the monthly availability of data reflects the number of observations received via the GTS and successfully decoded at ECMWF.

3.1 Availability

In this section two aspects of data availability are examined, namely timeliness of arrival, and reception rate.

At the end of every month statistics are compiled on the timeliness of arrival of individual SYNOP, TEMP and PILOT stations. An extract from the TEMP/PILOT statistics for November 1987 is given in figure 6. For each station the numbers of occasions during the month that the observation was received within 3 hours, and between 3 and 6 hours, of the nominal observation time are displayed. These statistics have been supplied on a monthly basis to two of the regions within WMO for monitoring purposes.

Monthly reception rates, of the global radiosonde network for example, may also be displayed pictorially. Figure 7 shows the reception rate of 100 hPa geopotential height at 00 UTC during the month of November 1987. The information is presented in symbolic form, each symbol representing a category of reception percentage as defined in the top right corner of the figure. A symbol is plotted for each radiosonde station which should report at 00 UTC, as specified in WMO Publication No. 9 Vol. A. The figure shows that the 100 hPa height was received more than 90% of the time over much of south-east Asia and Europe, whereas over India, Africa and South America the reception was mostly less than 30% and often zero. Poor reception rates at 100 hPa can be due to either the radiosonde ascents not reaching that level, or few ascents being made, or data not reaching ECMWF because of GTS problems.

3.2 Quality

Hollingsworth et. al. (1986) provided the rationale for using modern data assimilation systems as the appropriate tools for monitoring the quality of observations. There is good evidence that in areas where an adequate observational network ensures a sufficient data coverage, the 6-hour forecast error is quite low and allows the evaluation of data quality by comparison with the first-guess. In data sparse areas, however, a more cautious approach is required before any conclusions on the data quality are drawn from the comparison with the first-guess alone, as the model errors could be dominating.

Only if additional independent comparisons, such as results from collocation statistics of radiosonde ascents and atmospheric soundings derived from satellite radiance measurements, corroborate the data versus first-quess findings may they be accepted with confidence.

Data quality monitoring at ECMWF is based on the existence of the 'analysis statistics files'. A file is created at the time of each analysis (4 times per day) containing, for each observation, the departures from the first-guess, uninitialised analysis, and initialised analysis fields interpolated to the observation point. Also stored are the flags assigned by the automatic quality control procedures and the analysis. At the end of each month two accumulated files are created for 00 and 12 UTC, and it is these that form the basis for the monitoring of data quality. The system is planned to be extended to include monthly files for 06 and 18 UTC.

In this section various tools available within the ECMWF data monitoring system will be used to examine the quality of several data types, namely radiosondes and pilot balloons, satellite temperature soundings, satellite cloud-vector winds, drifting buoys and aircraft reports.

3.2.1 Radiosondes and Pilot Balloons (TEMP/PILOT)

In order to obtain an idea of how the quality of temperature and wind observations from an upper-air station varies with height, vertical profiles of observed minus first-guess and analysis differences may be used. Figure 8 shows an example for station 72694 using statistics averaged over the month of June 1987. The differences between height and wind observations and the ECMWF first-guess and analysis are presented for each of the standard pressure levels. The figures give standard deviation (left) and mean (right) of the differences for u-component wind (top), v-component wind (centre) and geopotential height (bottom) in the units ms-1 for wind and m for height. The numbers in the centre indicate the number of observations used for calculations (TEMP/PILOT for the wind components). Dashed lines denote deviations from the uninitialised analysis, solid lines deviations from the first-guess fields. The mean observed minus first-guess difference is henceforth referred to as the bias.

The character of the geopotential bias profiles in figure 8 is typical of that for US radiosonde stations with the 00 UTC bias positive increasing with height, while the 12 UTC bias behaves similarly but with a negative sign. The main effect is in the stratosphere. It is interesting to note that the difference between the first-guess profile (solid line) and the analysis profile (dashed line) is significant indicating that the data assimilation is adjusting the geopotential fields to fit the observations from this station.

3.2.2 Satellite Temperature Soundings (SATEM)

Figure 9 shows the bias of the 200-300 hPa layer temperatures measured by the NOAA-9 polar-orbiting satellite for 00 UTC in January 1987. The statistics are derived from all data valid between 2101 and 0300 UTC and are averaged over 5-degree squares. The information is presented in symbolic form, each symbol representing a category of bias as defined in the top right corner of the figure (values are in tenths of Celsius).

For the greater part of the atmosphere at these levels the values of layer mean temperature derived from the satellite measurements are on average within 1 Celsius of the ECMWF first-guess fields. However it is clear that over the Mediterranean and south-eastern Europe there are much larger differences. These differences are accentuated in the latitude band between 40 and 50 degrees north around the Black Sea and the Caspian Sea, where the bias changes abruptly from cold (north of 45 degrees north) to warm (south of 45 degrees north). These errors were due to the data supplier (NOAA/NESDIS) having problems with the regression coefficients used to calculate the temperature soundings during January 1987.

3.2.3 Satellite Cloud-Vector Winds (SATOB)

Figure 10 shows the bias of the cloud-vector winds between 250 and 150 hPa measured by the geostationary satellite METEOSAT for 00 UTC in October 1987. The statistics are averaged over 5-degree squares and the information is presented in the form of wind arrows. A reference value is given in the top right corner of the figure. An arrow is only plotted if 5 or more observations are used in that 5-degree square.

In general the bias vectors are small, although certain patterns are discernible. In particular in the latitude band around 30 degrees south, where the westerly winds are strong, the cloud-vector winds exhibit a marked easterly bias, i.e. the observed winds are lighter than those in the first-guess fields. This is an inherent problem with cloud-vector winds because, particularly in the jet-stream areas, the motion of the clouds does not necessarily correspond to the actual wind speed.

3.2.4 Drifting Buoys (DRIBU)

Reports of sea-level pressure are received at ECMWF from approximately 180 drifting buoys each month. The majority of the buoys are in otherwise data sparse areas so the quality of the first-guess fields should not be expected to be as good as in regions of dense data coverage. Nevertheless monthly statistics for each data platform are calculated and can be used to identify those buoys with problems.

An extract from the statistics for the month of October 1987 is presented in figure 11. All observations valid during the periods 2101 to 0300 UTC and 0901 to 1500 UTC were used to compile the values of mean (bias) and root mean square (RMS) observed minus first-guess differences (units are hPa). The total number of observations for each buoy valid during these periods is given in the fourth column, while the numbers accepted and rejected by the analysis are given in the fifth and sixth columns. Note that although some buoys report very frequently only one observation is used in each analysis so the total number of 'accepted' and 'rejected' reports is no greater than twice the number of days in the month.

It can be seen from the table that most of the buoys report on average within 2 hPa of the ECMWF first-guess. However, a typical DRIBU problem can be seen with number 32805. The RMS for this buoy was 6.3 hPa and the bias component was -6.1 hPa, a strong indication that the instruments were calibrated incorrectly. Moreover the reports were sufficiently close to the first-guess fields that they were rejected by the analysis on only 9 occasions out of 62 during the month.

3.2.5 Aircraft Reports (AIREP)

A relatively new activity at ECMWF has been the monitoring of the quality of wind reports from aircraft. One of the tools that has been developed for this purpose is the stratification of reports according to airline carrier. The results for September 1987 are presented in figure 12. It is assumed that each airline is uniquely identified by the first two characters of the aircraft report identifier, e.g. AY = Finnair, LH = Lufthansa. Only those airlines from which at least 150 reports were received have been included in the table, and only data valid during the periods 2101 to 0300 UTC and 0901 to 1500 UTC were used.

Although the RMS vector difference from the ECMWF first-guess fields may be used as a guide to the quality of the reports, it should be recognised that a direct comparison between airlines could be misleading when different geographical areas are involved.

4. ANNUAL MONITORING

When considering the quality of observations from individual reporting stations it is often useful to identify trends over a period of time. Currently ECMWF has software tools enabling visualisation of the monthly variability of the quality of geopotential and wind observations from radiosonde stations.

This facility is frequently used to identify sudden changes in the behaviour at a station. For example, Figure 13 shows the evolution of the mean monthly differences (in metres) between observations and first-guess over the 13-month period November 1986 to November 1987 for the station 32061. Separate curves for 500 hPa, 100 hPa and 50 hPa are shown; 00 and 12 UTC data are displayed in different graphs. The number of observations shown above each time graph box are the number of 500 hPa height reports received during each month. The most striking feature is the extremely high value of bias in August 1987, particularly at 100 and 50 hPa, contrasting markedly with the normal performance in July and September.

A second example demonstrates the monitoring of wind quality trends. Figure 14 shows the 13-month time series of mean observed minus first-guess differences of the westerly, or u-component of the wind for the Spanish station 08430 at 12 UTC between November 1986 and November 1987. Curves are plotted for two levels, 700 hPa and 250 hPa. Between December 1986 and May 1987 the negative bias of the westerly component of the wind increased steadily to become much larger than the previous normal level. It is assumed that on the evidence of the monitoring results (generally available to ECMWF Member States) action was then taken and the quality of the wind observations returned to normal within a month.

5. CONCLUDING REMARKS

ECMWF has at its disposal a wide range of tools facilitating comprehensive monitoring of data availability and quality, both in real-time and in delayed mode. On several occasions the daily monitoring procedure has led to the discovery of important data problems which have then been relayed to the data producers.

Monthly monitoring statistics are accumulated and made available to Member States. In addition many other international organisations have received statistics appropriate to their application. These include the World Meteorological Organisation (WMO), the European Space Agency (ESA), the National Data Buoy Centre, USA, and CLS Service ARGOS, Toulouse.

The introduction of the monitoring of long-term trends for radiosondes has introduced a different dimension when considering the quality of individual stations. In particular it has helped to identify those stations where sudden changes in behaviour have occurred during the preceding months.

REFERENCES

ECMWF (1987): Planning meeting on data for global models - final report

Hollingsworth, A., Shaw, D.B., Lönnberg, P., Illari, L., Arpe, K., and Simmons, A.J., (1986): Monitoring of observation and analysis quality by a data assimilation system. Mon. Wea. Rev. (114) 861-879

Lönnberg, P., Pailleux, J., and Hollingsworth, A., (1986): The new analysis system. ECMWF Technical Memorandum No. 125 (available from ECMWF)

Lorenc, A.C., (1981): A global three-dimensional multivariate statistical interpretation scheme. Mon. Wea. Rev. (109) 701-721

Shaw, D.B., Lönnberg, P., and Hollingsworth, A., (1984): The 1984 revision of the ECMWF analysis system. ECMWF Technical Memorandum No. 92 (available from ECMWF)

WMO (1985): Commission for Basic Systems, abridged final report of the extraordinary session, Hamburg 1985.

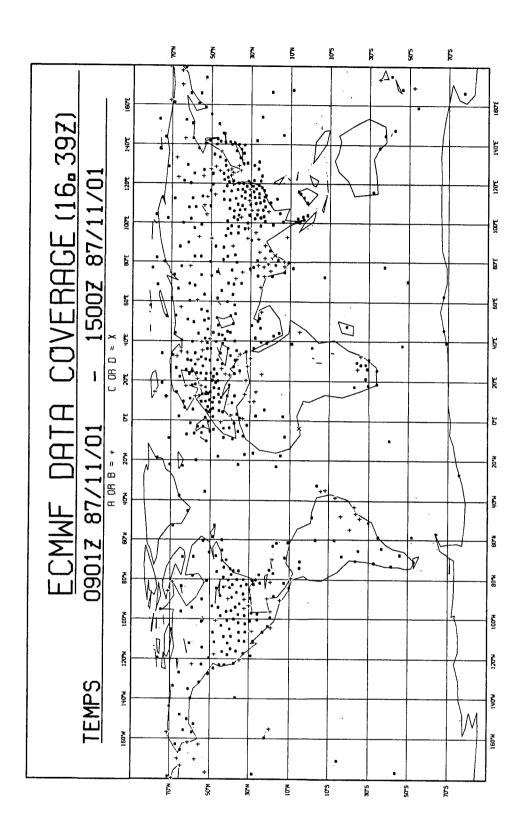


Fig. 1: 6-hourly data coverage map for radiosondes

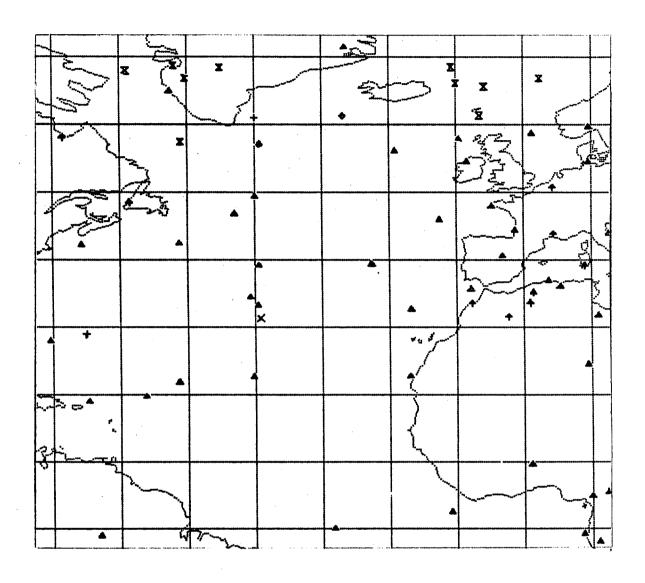


Fig. 2: Part of a data rejection map. Key to observation types:

- surface land or ship

+ - aircraft report
x - satellite cloud-vector wind

- drifting buoy

- radiosonde

\$ - satellite mean layer temperature

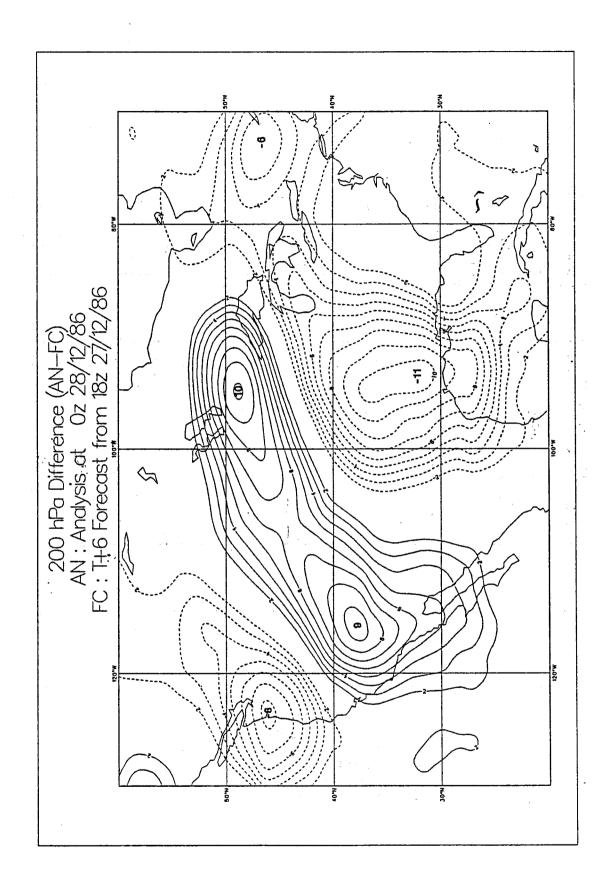


Fig. 3: Analysis minus first-guess 200 hPa geopotential increment map for 00 UTC 28 December 1986. Solid lines denote positive increments, dashed lines negative. Units are decametres.

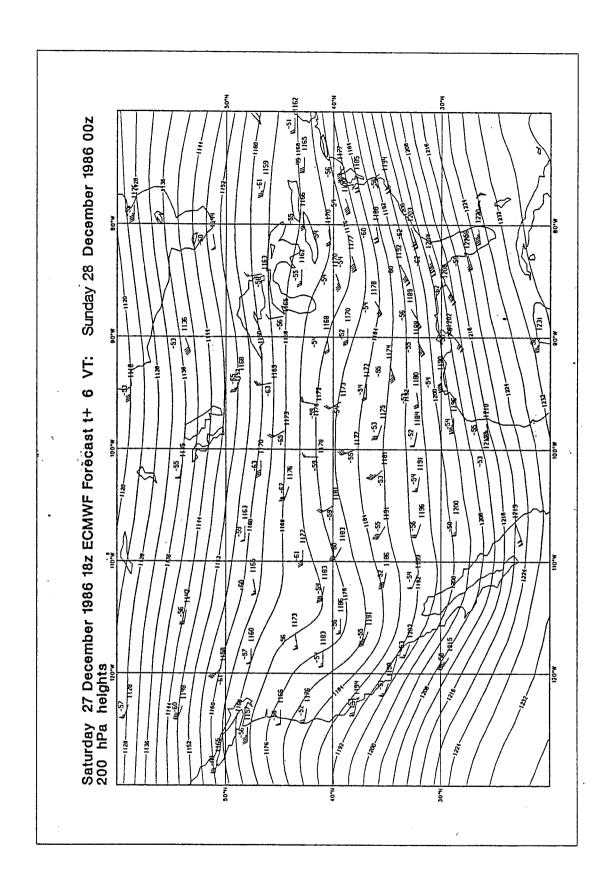


Fig. 4: 200 hPa geopotential first-guess field with reported observations from radiosondes over North America, valid at 00 UTC 28 December 1986.

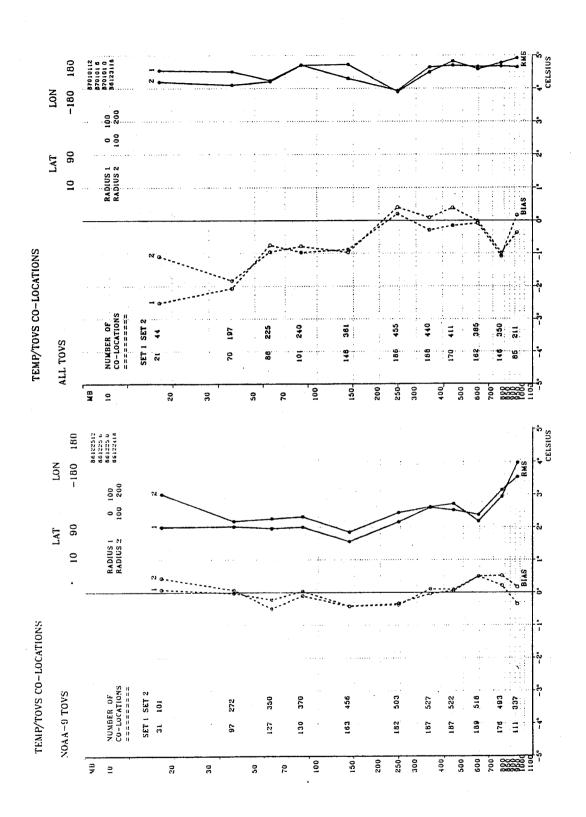


Fig. 5: TEMP/TOVS collocation statistics of mean layer temperatures for the four analysis cycles 18 UTC 24 December to 12 UTC 25 December 1986 (left) and 18 UTC 31 December 1986 to 12 UTC 1 January 1987 (right).

STATISTICS ON GLOBAL EXCHANGE DATA RECEIVED - TEMP/TEMP SHIP AND PILOT/PILOT SHIP.

EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS.

1 TO 30 NOVEMBER 1987.

NUMBER OF REPORTS RECEIVED WITHIN THE SPECIFIED PERIOD								
		AFTER OBSERV						
STATION		MP SHIP	PILOT/PI					
INDEX	HH+180MIN	HH+720MIN	HH+180MIN	HH+720MIN				
NUMBER	00 06 12 18	00 06 12 18	<i>00 06 12 18</i>	00 06 12 18				
12374	27 — 29 —		— 28 — 30					
12425	26 — — —		— 29 — —					
12843	27 25 28 30	3 3 2 —						
12982	26 — 26 —	1 — 2 —	— 10 — 15	_ 1				
13130	30 28	- 1 -	— 3 — 13 — 2					
13150			- 1 - 1 -					
13275	25 — — —	1						
13334			$\frac{-}{-}$ $\frac{-}{1}$ $\frac{-}{3}$					
15120	27 — 16 —	1 — 14 —						
15200			_ 1					
15310			_ 3 1 _					
15420	29 12 14	1 12 14						
15480	28 15	1 — 14 —	_ 3					
15614	28 24 26	1 - 2 -	27 20 25 —	_ 1 1 _				
15730	15 — 23 —		13 - 22 -	1 - 1 -				
16044	27 — 28 —	3 — 2 —	- 9 - 12					
16080	29 28	- - 1	- 9 - 15					
16144	26 26	1 — ż —						
16245	22 — 25 —	5 — 2 —						
16320	26 — 27 —	ī — ī —						
16429	27 26	3 — 3 —						
16560	29 — 29 —	1						
16597			— 5 — —					
16622								
16716	27 24	1 1	— 14 — 21					
16754		 1	2 14 15					
17030	25 — 25 —	5 — 2 —						
17062	27 — 28 —	3 1						
17130	26 — 27 —	3 2						
17220	24 28	4 2						
17240	25 — 28 —	4 — 2 —						
17280	23 — 29 —	5 — 1 —						
17300	1							
17352	28 — 28 —	2 2						
17601			25 26 21	 2				
17607			 28 1					
17609			19 12 18 14					
20046	29 — 27 29	 1 1						
20069	30 — 25 —			 1				
20107	27 — 20 —	1 5	- 4 - 1	— 1 — —				
20274	29 — 27 —							
20292	30 — 28 29	1 _						
20353	24 — 25 23 29 — 27 —							
20674								
20744	30 — 29 29							

Fig. 6: Extract from TEMP/PILOT reception statistics for November 1987.

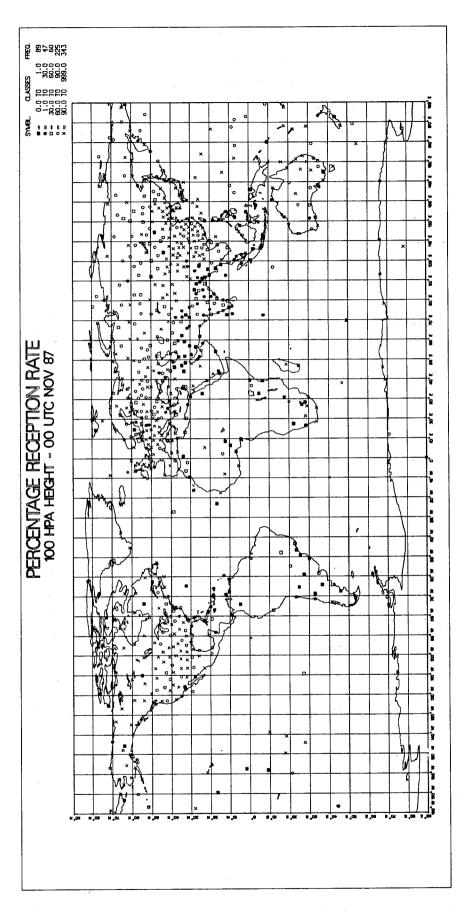


Fig. 7: Reception rate of 100 hPa geopotential height from individual radiosonde stations at 00 UTC for November 1987 (per cent).

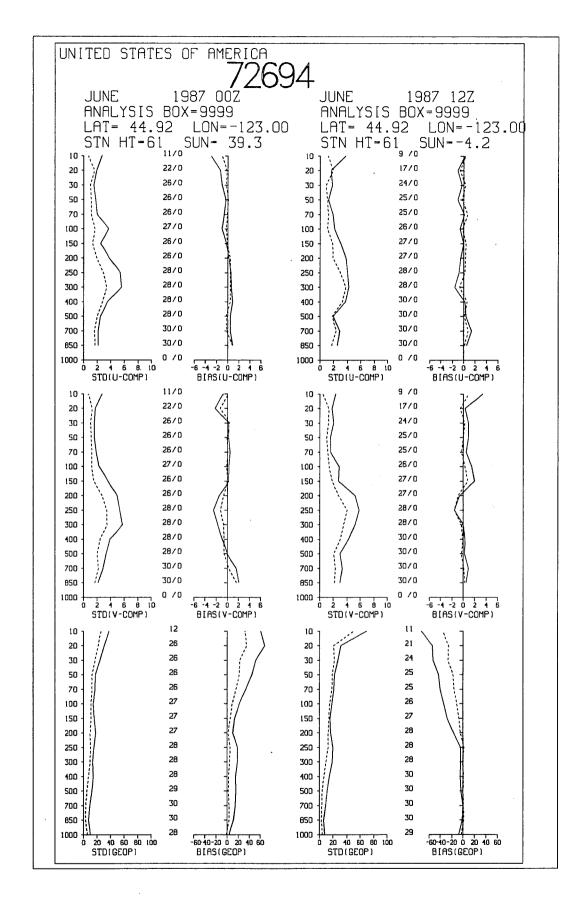


Fig. 8: Vertical profiles of observed minus first-guess (solid lines) and observed minus analysis (dashed lines) differences at radiosonde station 72694 for June 1987.

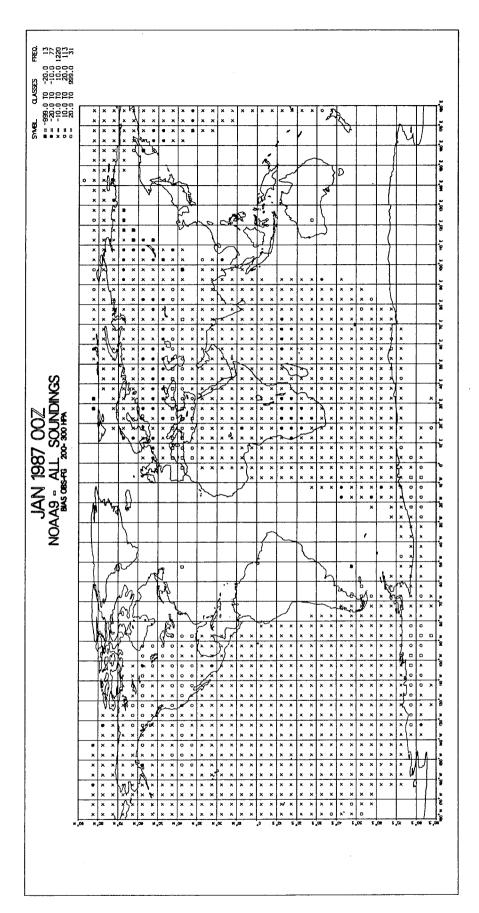


Fig. 9: Mean observed minus first-guess differences of mean layer temperatures from NOAA-9 at 00 UTC for January 1987. Statistics are averaged over latitude/longitude boxes of 5 x 5 degrees.

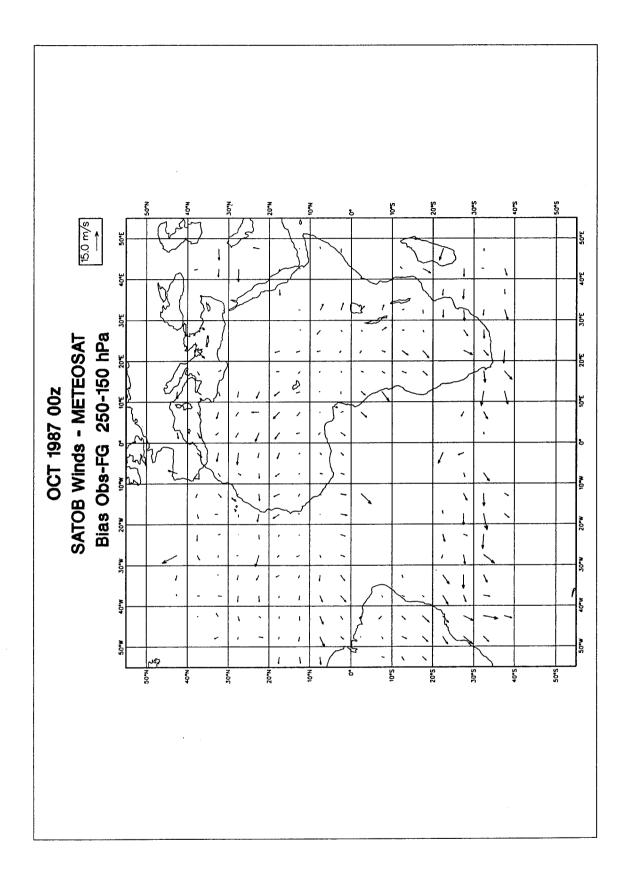


Fig. 10: Mean observed minus first-guess differences of wind reports from METEOSAT at 00 UTC for October 1987. Statistics are averaged over latitude/longitude boxes of 5×5 degrees.

DRIBU Ident	Bias (hPa)	RMS (hPa)	Total Obs	Number Accepted	Number Rejected
1dent 14801 14802 15501 15502 16802 17806 17807 17808 17809 17810 17811 17812 17813 17814 25521 25522 25523 25524 25525 25890 25892 25894 31804 31806 32801 32802 32803	(hPa) 0.3 0.6 -0.6 -0.5 -0.1 0.7 0.7 0.4 -0.3 0.6 0.5 -0.8 -0.3 0.2 1.7 0.8 1.2 -1.4 0.3 -3.5 -1.1 -66.7 -2.0 -1.0 1.9 1.7 1.5	(hPa) 1.2 2.1 1.3 3.0 1.8 1.7 1.7 1.2 1.7 2.5 2.8 1.8 2.4 5.6 4.1 3.1 2.9 3.7 12.7 3.0 67.3 2.1 1.8 2.1 2.0 1.8	Obs 161 147 69 68 198 178 197 201 59 195 165 105 218 100 53 64 76 37 88 109 9 2 11 159 162 173		
32805 32806 32807 33801 33802 33803 33804 33814 34809 34810 34811 34812	-6.1 1.1 0.2 -0.7 -0.2 -0.9 -1.7 1.2 -1.3 -0.9 1.5	6.3 1.5 0.9 1.2 3.2 1.6 3.5 2.2 2.3 1.8 1.9	172 182 139 94 167 67 142 145 170 156 192 163	53 62 62 61 62 50 60 45 62 60 62	9 0 0 0 0 0 0 0 0 0 0 0 0

Fig. 11: Extract from drifting buoy quality statistics for October 1987.

Airline Ident	Vector RMS (m/s)	Total Obs	Number Rejected	FG Diff 15-30 m/s	FG Diff > 30 m/s
AA	9.8	1665	12	117	15
AC	10.1	553	5	40	5
AF	9.2	543	5	23	5
AM	8.8	280	2	. 16	1
ΑY	8.8	160	1	7	3
AZ	10.7	272	3	24	3
BA	9.7	710	9	44	8
BR	10.0	154	1	21	. 0
CA	12.4	402	. 12	37	11
CI	9.6	860	12	53	14
CO	9.1	1821	13	108	14
CP	9.4	861	7	44	8
CX	9.9	242	4	20	3
\mathtt{DL}	10.1	2226	19	154	25
FT	9.1	455	3	37	0
GA	9.8	289	2	15	2
HA	14.9	823	10	84	11
IB	10.4	874	13	59	13
JL	9.6	3514	30	171	34
KE	9.0	1055	5	78	6
KL	9.6	828	10	58	11
KZ	8.6	218	2	5	3
LH	9.3	1466	16	98	11
M5	11.6	174	2	14	2
M6	8.2	213	0	14	0
MA	9.7	334	4	21	5
NH	9.4	425	5	30	6
NW 	10.3	4855	54	317	65
PA	10.3	1060	17	88	16
PR	9.2	312	3	16	3
QF	9.8	1122	18	31	18
RA	15.5	196	6	19	5
RR	11.1	254	6	30	3
SK	8.5	1050	8	42	8
SQ	8.6	895	4	51	6
SU	13.4	168	5	30	5
TE	11.7	734	4	31	7
TG TP	8.7	168	1	6	1
	11.1	227	3	30	3
TW	13.0	1061	. 15	109	17
UA	9.5	4673	36	230	44
UT	18.8	374	7	12	6
VV	9.3	276	3	13	. 3
. WD	13.5	281	3	17	3
XX	7.9	501	3	23	2

Fig. 12: Aircraft wind reports quality statistics stratified by airline carrier for September 1987.

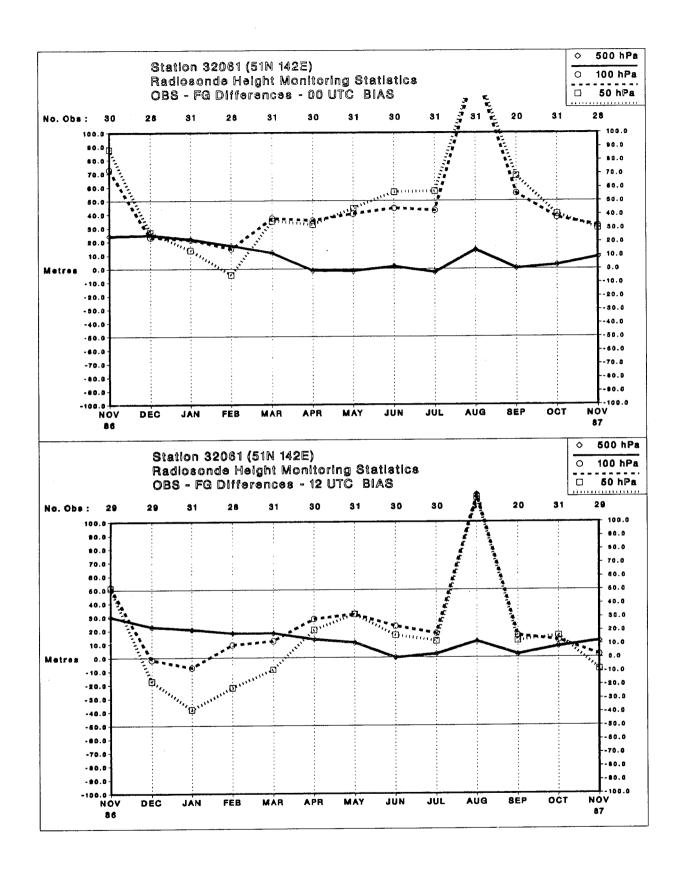


Fig. 13: Time graphs of mean monthly differences between observations and first-guess of geopotential height (metres) at station 32061, 00 UTC data (above) and 12 UTC (below).

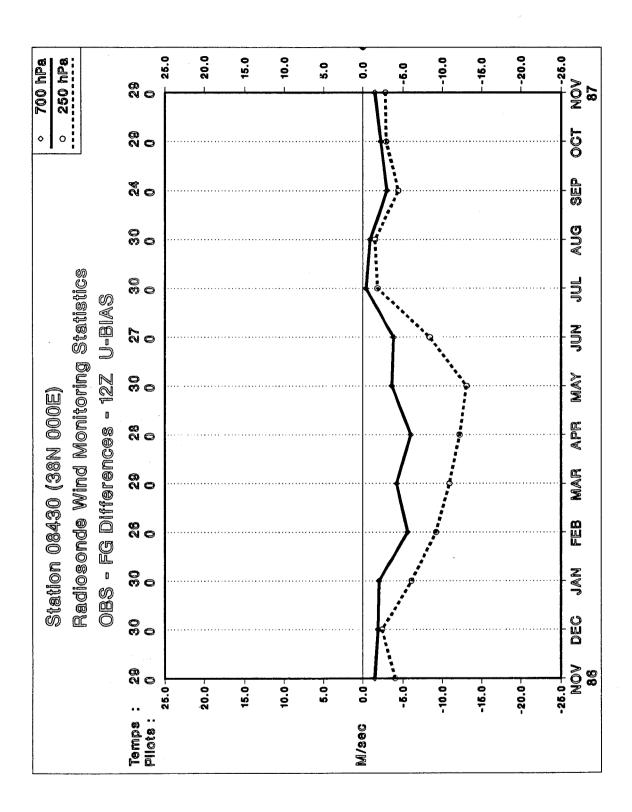


Fig. 14: Time graph of mean monthly differences between observations and first-guess of the zonal wind component (metres per second) at station 08430 for 12 UTC data.