

Forecasting Research

Short - Range Forecasting Research Division

Technical Report No 6

Tuning and Performance of the Atmospheric Quality Control

by

N.B. Ingleby

December 1991

**Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom**

Short Range Forecasting Research

Technical Report No.6

Tuning and Performance of the Atmospheric
Quality Control

by

N.B. Ingleby

December 1991

Short Range Forecasting Research
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SZ
ENGLAND

N.B This paper has not been published. Permission to quote from it must be obtained from the above Met. Office branch.

Tuning and Performance of the Atmospheric Quality Control

N B Ingleby

1. Introduction

This report documents the changes to the Unified Model quality control scheme from its incorporation into the trial suite in July 1990, through the operational implementation on 12 June 1991, up to the end of October 1991. It includes a discussion of the sensitivity of the background check to the input parameters which forms the basis for some of the changes (Appendix D). For some observation types time series of the percentages of rejections are presented for the period November 1990 to October 1991 which illustrate the effects of the changes. More detailed statistics for October 1991 are presented to give a summary of the current performance of the quality control. The emphasis throughout is on the automatic quality control as described by Ingleby and Parrett (UMDP 32), other components of the system such as block rejects of satellite data are only mentioned in passing. The essentially practical approach in this report is complemented by a more theoretical consideration of the properties of Bayesian quality control (in preparation).

The 'performance' is given mainly in terms of the percentages of observations rejected by different checks. Unfortunately no absolute standard of right or wrong decisions is available. Some rather limited subjective assessment is included as is some comparison with the previous quality control. A change to the quality control can be assessed objectively by conducting parallel quality control/analysis/forecast/verification cycles. This was not done because the time and expense for such parallel trials was not available and the scattered, usually small, impact of quality control changes can be very difficult to detect in objective scores. However all the changes were made in response to perceived deficiencies in the quality control and in general represent an improvement to the quality control.

2. Changes to quality control and timeseries of rejections

This section deals mainly with single level data. Information is presented as timeseries of the proportions of observations rejected. All the timeseries (and the statistics in section 3) have been calculated from data stored in the global model Observation Processing Database (OPD) using the SAS and SAS/GRAPH packages. Section 3 summarizes the performance of the quality control in October 1991 and treats single and multi level data.

2.0 Timeseries

Timeseries are presented starting from when the observation type was first stored in the parallel CRAY OPD (29th October 1990 for aircraft, surface and radiosonde observations - other types were added later) until the end of October 1991. Until 22nd May 1991 a comparison with the CYBER OPD is available (the main archive was changed to store the CRAY OPD a few weeks before the Unified Model became operational on 12th June). In general the sets of observations in the two OPDs do not compare exactly, partly due to slight differences in the time of extraction from the OPD, partly because some unusable observations (eg corrupt position) and duplicate reports are stored in the CRAY OPD for monitoring purposes.

The numbers of observations in the CRAY OPD is illustrated in Figure 1 (note that this does not include compressed code SATEMs). There are a few gaps in the OPD as shown by the reduced numbers during some weeks (the CYBER OPD also contained some gaps, generally not at the same time). The last point only represents three days rather than a week. The reduced numbers of radiosondes from 29th January to 26th March 1991 appear to be caused by an error of labelling within the OPD rather than any real loss of data.

The main purpose of the timeseries is to look at the effect of changes during the tuning period - usually made on a Tuesday, so the statistics are calculated by week beginning on a Tuesday. A list of changes can be found in Appendix A. Unfortunately it is difficult to compare the proportions rejected by the background and buddy checks: in the CRAY suite all observations undergo the background check but observations with CFO or permanent rejection flags are excluded from the buddy check (in the CYBER quality control such rejected observations were excluded from both the background and buddy check). However for some observation types the proportion of final rejections

from the CYBER suite provides a useful reference as the basic CYBER quality control did not change during the period. For satellite soundings (SATEMs and LASS) and satellite cloud track winds (SATOBS) the proportions of rejections are much higher than with the CYBER suite due to the incorporation of new types of check, and comparison of the two suites is of rather limited value.

2.1 Surface observations

a) Pressure

Figures 2 and 3 show the proportions of pressure rejections for SYNOPs and SHIPs respectively. Because P_{msl} is quality controlled and P_* is used in the assimilation there is a check on the consistency of the two - this applies to all surface observations but problems mainly arise for high level land stations. The $P_{msl} - P_*$ check was revised on 18th December 1990 to be more selective, particularly for high level stations, and this led to a small reduction in the numbers of SYNOPs flagged by this check (final flags minus background flags). However there was a long-standing inconsistency in the processing of P_{msl} from model output and from the observations. When this was corrected on 26th March 1991 there was a significant reduction in background, buddy and inconsistency flags for SYNOPs. From this point on the CRAY QC was rejecting less than half the 4% of SYNOP pressures rejected on the CYBER. All SYNOPs are used in the CRAY suite whereas on the CYBER Western European SYNOPs were thinned using a station list.

The introduction of the buddy check on 5th February 1991 had little effect on overall levels of flagging, perhaps causing a slight increase. However to avoid excessive flagging a 'fix' had been introduced: the total error correlation ρ was multiplied by 0.5. ρ was too large because the climatological background error estimate for P_{msl} was too large. It is of interest to compare the effects of the fix, and the damping (first method used), see Table 1. This shows that both modifications to the buddy check reduce the flagging to a more 'reasonable' level. But there is more sensitivity to the ratio between observation and background errors (through ρ) than to the damping.

	Bk only	0.5ρ	$\eta_1=0.5$	Buddy
SYNOP	2.6	3.2	4.4	6.0
SHIP	3.5	2.3	4.0	5.8

Table 1. Comparison of modifications to the buddy check of P_{msl} averaged

over three cases in March 1991 (12Z 4th, 00Z 7th, 00Z 22nd) - climatological background errors were used. The columns give the percentage rejected for a) background check only, b) $0.5p$ fix, c) damping with $\eta_1=0.5$, d) buddy check without either modification.

A subjective assessment of one case is reported in Appendix B, this shows that the buddy check improved on the background check alone for 49 decisions but made 13 decisions worse. On 9th April 1991 buddy check 'damping' was introduced, this was modified on 9th May 1991. Both changes reduced the flagging by the buddy check by a fraction of 1% for SYNOPs and SHIPs.

The 'damping' changes are complicated by the introduction of synoptic dependent background errors on 23rd April 1991 - this reduced the average background error from ~ 2 mb to ~ 1.5 mb. At the same time the 'fix' in the buddy check of P_{msl} was removed: this probably contributed significantly to the increase in buddy check flagging of P_{msl} (see Table 1). These changes had little effect on the background check of SYNOPs. The SHIP background check displays a jump at that point - this is partly the effect of the background errors being erroneously small for 3 days and partly coincidence as the longer term average has settled back to its previous value. The 23rd April changes do give a fairly clear increase in buddy check flagging: for SYNOPs this is roughly cancelled by the 'damping' changes, but for SHIPs the three April/May changes give a net increase in buddy check flagging relative to background check flagging.

The reason for the blip in SYNOP background flags on 31st July 1991 is unknown. The criterion for excluding data from the buddy check was relaxed on 13th August, see Appendix C for a discussion of this and rounding errors in the buddy check. This change is likely to reprieve a small number of correct observations with differences from background more than ~ 10 mb. The small increase in SYNOP buddy check flagging relative to the background check is due to bad observations which previously by-passed the buddy check and does not imply an increase in final flags.

There may be a seasonal effect with more background flags in the Northern Hemisphere winter relative to the summer as most of the observations are in the Northern Extratropics. This would explain the increase in flagging in October 1991, particularly for SHIPs. However the reduction in flagging in the first half of 1991 might also be partly due to improvements to the

Unified Model over this period.

The timeseries for drifting buoys is not shown because it is more noisy, tending to mask the effects of the changes. The thinning of drifting buoys (to a maximum of 5 in a six hour period) introduced on 19th February 1991 leads to the rejection of almost 20% of drifting buoy reports.

b) Wind

The comments for SHIP winds (figure 4) are broadly similar. A reduction in k on 18th December 1990 caused a small reduction in background flagging. The introduction of the buddy check on 5th February 1991 caused a noticeable increase in the number of final flags. Multiplying background wind speeds by 10% (19th February 1991) to reduce observation/background biases appears to have had little effect. The synoptic dependent background wind errors introduced on 23rd April were about 80% of the previous values, the observed errors were multiplied by 0.8 to keep the ratio roughly constant. This increased the background flagging by about 1% but had little effect on buddy check decisions relative to the background check. The 'damping' of the buddy check reduced flagging relative to the background check.

2.2 Aircraft observations

During the period considered there were three types of aircraft observation available: AIREPs - coded manually and transmitted verbally, ASDARs - an automated system, and PIREPs - manually coded observations over North America. AIREPs are the traditional and most plentiful source of data, however being manual they contain scope for human transcription errors etc. ASDAR offers more frequent observations, and the automation should ensure high quality data. ASDAR data is not included in the timeseries as it was only introduced in late November 1990 with some teething problems, but contingency tables for October 1991 are presented in Section 3.2. PIREP observations are automatically excluded from the analysis on quality grounds and are not considered further.

a) Wind

Figure 5 shows a timeseries of AIREP wind reports between 1 and 400 mb (nearly all AIREPs). Initially about 5% of observations were being rejected - about double the proportion in the CYBER suite. When the value of k was

reduced to a more realistic level (see Appendix D) on 18th December 1990 the rejection rate dropped to just over 3%.

On February 5th 1991 the buddy check was introduced, this increased the final flagging by about 1%. Two cases reported by forecasters showed the worst features of the buddy check: flagging of observations close to the background in data dense areas and sensitivity to the order of checks. These have been improved by the introduction of 'damping' into the buddy check. However the very large wind shears associated with jet streams may still cause occasional problems.

Buddy check

On 12Z 23rd February at 35.5°N, 150.0°E three AIREPs at 197 or 217 mb were flagged and two others were suspected by the buddy check despite being in good agreement with each other and the background. There were seven observations at 238 mb veered by 5-10° from the background and another seven observations lower down. The 10° differences at speeds of about 190 knots were sufficient to cause disagreement between the observations in the pairwise check. Changing the order of the buddy check caused the higher observations to pass but flagged two observations at 238 mb. Damping the buddy check caused all the observations to pass. Other issues raised were that the climatological vertical error correlations were too broad in this jet stream situation and that the reporting (by most aircraft) of wind directions to the nearest 5° causes significant rounding errors. On 9th April 1991 damping and extra sorting were introduced into the buddy check.

On 12Z 24th April there was another *faux pas* by the buddy check, at 41°N, 149.1°E (pointed out by D Ireland who noticed several useful cases). Three late AIREPs (after T+110 minutes) caught the leading edge of a jet as it moved forward and developed. They were reporting 110 or 120 knots compared to the background and other observations at about 60 knots. They were, not surprisingly, flagged by the background check but unfortunately not reinstated by the buddy check despite the three observations being consistent (within 5° and 10 knots of each other).

The solution implemented on 9th May 1991 was to change to an alternative form of damping (which gives more weight to checks between pairs of observations which agree than to pairs which disagree) and to narrow the vertical correlation scale somewhat. Narrowing the time correlation scale was

also considered but not found to be beneficial. Compared to average statistics the new vertical correlation is slightly on the narrow side, however in jet stream cases the model error can be very strongly sheared giving very narrow background error correlations. If the correlation is overestimated the pairwise buddy check seems to flag too many observations, whereas underestimating the correlation is less harmful. Earlier tests suggest that the AIREP buddy check is insensitive to the horizontal correlations used: because of current reporting patterns most close buddies are colocated in the horizontal and differ only in pressure or time.

There has been one other adverse comment on the AIREP buddy check (18Z 13th June 1991), again associated with a sharp change in wind towards the end of the time window. Because of the abruptness of the change (a veering of the wind and a sharp increase in speed) and some inconsistency between observations this was a difficult case for any buddy check. Despite the rejection of three 'good' winds the resulting six-hour forecast was found to be quite good.

All observations within the six hour window are compared with the background values in the middle of the window. In general interpolating the background in time would help. Using aircraft observations just outside the analysis time window within the buddy check might help in some cases. Since June the only complaints about the buddy check have been minor and have concerned surface BOGUS observations.

Calm winds

In April forecasters noticed that there were increasing numbers of AIREPs reporting calm winds (wind speeds equal to zero), apparently when the winds should have been coded as missing data. An SDB change was made resetting all calm AIREP winds to missing data - this affected both the CYBER and CRAY suites. On 27th August 1991 the SDB check was removed, all calm AIREPs now being rejected after the background check (QC change of 9th May), to enable the calm AIREPs to be monitored via the OPD. This is the reason for the jump in rejections of winds at the end of August. It seems likely that a few of the AIREP 'calms' are real - they are in areas where the background wind is very light - but the majority of them are wrong. In October 1991 there were 2655 AIREP calms, the percentage of these by airline (the first two letters of the callsign) reveals: 46.3% - DL, 20.6% - WA, 4.3% - UA, 3.8% - NW with smaller contributions from other airlines.

b) Temperature

Figure 6 shows a timeseries of AIREP temperature reports above 400 mb. Initially just over 2% of observations were being rejected, this was increased to about 3% when the intervention was introduced. The buddy check (5th February 1991) increased this by 1.5-2.0%. The buddy check damping (9th April 1991) removed most of this increase, the alternative damping and narrower vertical correlations (9th May 1991) appears to reduce the buddy check flagging relative to the background check, but any effect on the final rejections is less obvious. From mid-April to mid-May 1991 the rejection rate was about 3%, just slightly above that for the CYBER suite.

2.3 Satellite cloud track winds - SATOBs

It has been recognized for some years that SATOBs wind speeds tend to be biased low in jet stream regions. At ECMWF an 'asymmetric check' - stricter on observations with wind speeds more than 7.5 m/s weaker than the background - was tested and introduced (Lönnerberg, 1989, Eriksson, 1990). On 9th May 1991 a similar test was included in the Unified Model quality control (see Ingleby and Parrett, UMDP 32, for details). For winds above 400 mb (Figure 7) the asymmetric check caused a modest increase in flagging of METEOSAT winds and a sharp rise in GMS flagging. The impact on GOES is not obvious - partly because of the large fluctuations in GOES quality. From June to October 1991 the quality of high level GOES reports appears to have deteriorated. The asymmetric check is applied at all levels and also caused increased flagging of mid-level METEOSAT winds (Figure 8), it has little effect at low levels.

The flagging of SATOBs in the CRAY QC is much higher than in the previous QC. On the CYBER low level flagging was less than 0.5% for all satellites and at middle and high levels it was up to 2% with GOES showing the largest flagging rates.

Originally, for SATOBs the probability of gross error (PGE) in location after the background check, $P(G_1|O)$, was calculated based on temperature and wind with the initial PGE in location, $P(G_1)$, set to 0.02. The PGE in temperature was set to 0.0: the pressure is assumed to be based on the 'observed' temperature, thus a temperature error implies a (vertical)

location error. A large temperature difference from background therefore causes the whole observation to be flagged.

However it was found that low level SATOBs from the GOES satellite had pressure assigned climatologically and sometimes temperatures from a tropical profile inserted (eg 17° at 900 mb) - violating the assumption made above. These temperatures contain no useful information and were causing 'good' winds to be rejected in mid-latitudes. Thus on 9th May 1991 $P(G_1)$ was set to zero for SATOBs removing the influence of temperatures from the quality control of the winds (the results are then identical to quality controlling wind on its own). However on 16th July $P(G_1)$ was unwittingly reset to 0.02 when a change was made to SATOB permanent rejections. Figures 8 and 9 for medium and low level SATOBs respectively show that this temperature dependence substantially increases the rejection of GOES SATOBs. It has little effect on METEOSAT and GMS (GMS does not report temperatures).

K Wright (personal communication) looking at cases in August and September 1991 provided evidence that at high levels in the Southern Hemisphere some GOES winds were being allocated to the wrong pressure giving large temperature differences from background ($7-20^{\circ}$). The background winds at the level matching the observed temperature gave a much better fit to the observed winds. Some of these poor height allocations may be due to operator intervention at NESDIS (C Hall, personal communication), particularly as their numbers seem quite variable in time.

Thus there are two distinct problems with GOES SATOBs (affecting a small proportion of the observations): good pressure/poor temperature at low levels (particularly 900 mb) and poor pressure/good temperature at higher levels. To take account of this on November 19th 1991 $P(G_1)$ for GOES was set to 0.0 (no cross-check with temperature) below 800 mb and to 0.04 above. For GOES SATOBs at 300 mb this implies that temperature errors of about 7.5 (4.5) degrees would cause observations with vector wind differences of 0-10 m/s (15 m/s) to be rejected. Increasing $P(G_1)$ from 0.02 to 0.04 only changes the cut-off by a fraction of a degree so the effect of this change on high level GOES SATOBs will be small. The values for the other satellites remain at 0.02.

2.4 BOGUS observations

Figure 10 shows a timeseries of flagging of BOGUS pressure observations. (BOGUS observations are generated by forecasters either to support existing

observations or to insert features visible in satellite imagery.) The timeseries is quite noisy (even more so for other BOGUS variables - not shown). Initially in order that most BOGUS observations should pass the quality control k was set to very low values. However on 23rd April 1991 k was reset to the values used for radiosondes and the estimated background errors were increased (in the background check only). This was felt to be more realistic as in general BOGUS observations are inserted in areas of large background error.

It is obvious from Figure 10 that the new quality control is rejecting significantly more BOGUS observations than the old quality control - which hardly rejected any. However rejecting about 1.5% of the observations is probably realistic as they can be subject to typing errors. A case involving the rejection of several BOGUS pressure observations near Iceland (00Z 1st May 1991) was investigated. It was found that they were intended to support a ship but had been placed 5° too far South. Other position errors and value errors have occasionally been noted.

2.5 NESDIS Satellite soundings

Only SATEMs (500 km resolution) were included in the timeseries as the compressed code SATEMs (250 km resolution) were only stored in the CRAY OPD for a limited period. The flagging of these two sources is assumed to be similar except that in a particular area LASS soundings are selected in preference to NESDIS soundings and compressed code SATEMs in preference to SATEMs - with redundant data being given prior (permanent) reject flags. The usage of SATEMs increased significantly in September/October 1991 with problems in the reception/decoding of compressed code SATEMs.

Because much of the quality control of satellite soundings is now performed on blocks of layers the timeseries are presented on a whole sounding basis. The number of reports completely rejected (F) is the same as, or slightly more than the number permanently rejected (P) (the discrepancy in December 1990 is because only the top level was checked for a prior flag initially).

The vertical stability check and the multi-level checks were present throughout the period but were only included in the OPD in late March 1991. The vertical stability check was tightened on 21st May 1991, it can be seen that this caused an increase in flagging (of levels below 100 mb) from about

10% to just under 20%. The multi-level checks, marked troposphere and stratosphere in Figure 11 are much less significant affecting only one or two percent of the soundings. There is a suggestion of a step change in the flagging by the stability check on 16th July 1991, this is not due to any quality control change, but could be a result of minor changes to the unified model, or possibly to changes in the retrieval system at NESDIS.

The large percentage of SATEMs with a background flag on at least one level is largely due to temperature biases above 10 mb in the Unified Model.

2.6 Local Area Satellite Soundings (LASS)

Observations quality controlled against the limited area model are stored in the CRAY OPD between 22nd May 1991 and 30th June 1991, disrupting the global model timeseries. The other changes have only small effects on the timeseries which is therefore not shown. See Section 3.6 for current performance.

2.7 Radiosonde observations

The quality control of radiosondes was fairly stable over the period considered with only a reduction in k_{wind} on 5th February 1991 and a reduction in k_{RH} on 13th August 1991. The results before 13th August 1991 are somewhat unrepresentative because all reported levels which had a background PGE > 0.999 or a permanent flag were excluded from the vertical averaging and hence from the OPD. For these reasons time-series of radiosonde rejections are not shown. The processing change on 13th August 1991 had a rather minor effect on the observations presented to the analysis - in some cases a partial layer average was replaced by a full layer average flagged because it was corrupted by a suspect reported level (this affects only a small number of temperature and wind observations but rather more RH observations).

Radiosondes are assumed not to be susceptible to position errors, this is a reasonable assumption for land stations, but a study of four month's data suggests that at least four TEMPSHIPS a month have position errors and that action should be taken to detect them (see Appendix E).

3. Statistics for October 1991

This section presents a summary of quality control decisions in October 1991. For single level observations contingency tables are used enabling a comparison of the background check with the buddy and other checks. For satellite soundings and radiosondes the information is presented as bar charts by level (there is currently no buddy check for multi-level observations).

3.1 Surface

		Prior	CFO	Buddy	Other	Pass	
Background	Pass	0.54	0.00	0.34	0.00	98.30	99.19
	Fail	0.02	0.00	0.50	0.08	0.21	0.81
		0.56	0.00	0.84	0.08	98.51	100.00%

Table 2. Contingency table for **SYNOP pressure** decisions, October 1991, 241975 observations.

		Prior	CFO	Buddy	Other	Pass	
Background	Pass	0.83	0.67	1.42	0.00	93.51	96.42
	Fail	0.22	1.18	1.77	0.12	0.29	3.58
		1.04	1.85	3.19	0.12	93.79	100.00%

Table 3. Contingency table for **SHIP pressure** decisions, October 1991, 77816 observations.

		Prior	CFO	Buddy	Other	Pass	
Background	Pass	2.65	0.39	0.28	14.32	81.07	98.70
	Fail	0.07	0.43	0.42	0.18	0.20	1.30
		2.72	0.82	0.69	14.50	81.27	100.00%

Table 4. Contingency table for **Drifting Buoy pressure** decisions, October 1991, 33422 observations.

Table 2 shows that 99.19% of SYNOP pressures passed the background check whereas 98.51% passed overall (the final check). The flagging of SHIP

pressures (table 3) is larger in all categories with 6.2% rejected overall compared with 1.5% for SYNOPS. Table 3 also shows that two-thirds of the SHIP pressures rejected by CFO failed the background check anyway. When blacklisted and thinned observations (prior and 'other' flags) are excluded drifting buoys have much lower rejection rates than SHIPs (table 4).

All three tables show that the buddy check rejects more observations than it reinstates. This is only a small effect for SYNOPS (0.34% vs 0.21%) and drifting buoys (0.28% vs 0.20%) but is much larger for SHIPs (1.42% vs 0.29%).

		Prior	CFO	Buddy	Other	Pass	
Background	Pass	0.92	0.25	0.43	0.00	94.10	95.70
	Fail	0.13	0.48	3.29	0.00	0.40	4.30
		1.04	0.73	3.72	0.00	94.51	100.00%

Table 5. Contingency table for **SHIP** wind decisions, October 1991, 77803 observations.

The background check flags more SHIP winds than it does pressures (table 5 of table 3). The buddy check changes fewer decisions, rejecting only slightly more than it reinstates (0.43% vs 0.40%). As with pressures two-thirds of the CFO rejects failed the background check. SYNOP and drifting buoy winds are not used in the assimilation, but for comparison 1.12% of SYNOP winds and 5.72% of drifting buoy winds were rejected by the background check.

3.2 Aircraft

		CFO	Buddy	Calm	Other	Pass	
Background	Pass	1.12	0.30	0.93	0.00	95.37	97.72
	Fail	1.43	0.55	0.16	0.00	0.14	2.28
		2.55	0.85	1.09	0.00	95.51	100.00%

Table 6. Contingency table for **AIREP** wind decisions, October 1991, 130714 observations. Note that there are no permanent rejections.

		CFO	Buddy	Calm	Other	Pass	
Background	Pass	0.09	0.02	0.02	0.42	99.16	99.71
	Fail	0.00	0.23	0.00	0.00	0.06	0.29
		0.09	0.24	0.02	0.42	99.22	100.00%

Table 7. Contingency table for **ASDAR wind** decisions, October 1991, 13098 observations.

Tables 6 and 7 show the flagging for AIREP and ASDAR wind observations. They show quite clearly the benefits of the automated ASDAR system in reducing gross errors. 1.1% of the AIREPs are reporting calm winds which are automatically flagged (see section 2.3). Most of these were passed by the background check, but the proportion is far too high compared to the corresponding figure of 0.02% calms from ASDAR. 0.42% of the ASDAR winds were rejected for 'other' reasons - these were found to have missing pressures.

		CFO	Buddy	Other	Pass	
Background	Pass	1.06	0.28	0.00	96.88	98.21
	Fail	1.06	0.65	0.01	0.07	1.79
		2.12	0.93	0.01	96.94	100.00%

Table 8. Contingency table for **AIREP temperature** decisions, October 1991, 132386 observations. Note that there are no permanent rejections.

		CFO	Buddy	Other	Pass	
Background	Pass	0.21	0.09	0.41	98.75	99.47
	Fail	0.37	0.16	0.00	0.01	0.53
		0.58	0.25	0.41	98.76	100.00%

Table 9. Contingency table for **ASDAR temperature** decisions, October 1991, 13098 observations. Note that there are no permanent rejections.

Tables 8 and 9 for temperature again show the superiority of the ASDAR system. For aircraft temperatures the buddy check is particularly grudging only reinstating a very small percentage. For aircraft winds and temperatures over half of the reports rejected by CFO would have been rejected by the background check anyway (and some of the others may be correct - Ingleby and

Hammon (1990) found that in some cases CFO were rejecting the whole report when only the wind was wrong).

In October 1991 there were four ASDAR systems reporting, but one (BA010P) accounted for just over half of the observations. One of the other aircraft (CO006P) had 60 of its temperatures (2.3%) rejected by CFO.

3.3 Satellite cloud track winds - SATOBs

		Prior	CFO	Buddy	Pass	
Background	Pass	6.78	0.16	0.02	88.70	95.66
	Fail	1.50	0.40	2.22	0.21	4.34
		8.29	0.56	2.24	88.91	100.00%

Table 10. Contingency table for SATOB wind decisions (all satellites), October 1991, 187611 observations.

Table 10 shows that the buddy check reinstates ten times the number of observations that it rejects, the highest proportion reinstated is 0.28% for METEOSAT. The prior rejections include 1/3 of the 'bad' data as judged by the background check. About 70% of the observations rejected by CFO failed the background check.

3.4 BOGUS

		Prior	Buddy	Pass	
Background	Pass	0.14	0.32	98.22	98.67
	Fail	0.00	0.91	0.42	1.33
		0.14	1.23	98.63	100.00%

Table 11. Contingency table for BOGUS surface pressure decisions, October 1991, 5046 observations.

		Prior	Buddy	Pass	
Background	Pass	5.47	0.00	94.40	99.87
	Fail	0.10	0.00	0.03	0.13
		5.57	0.00	94.43	100.00%

Table 12. Contingency table for BOGUS surface wind decisions, October 1991, 3069 observations.

The automatic quality control of surface BOGUS observations flagged 1.23% of pressure observations (table 11) but no wind observations (table 12). Surface winds over land have a permanent rejection - hence the 5.5% with prior flags.

	Number	Prior	Bkgnd	Other	Pass
Wind	3030	0.43	1.12	0.13	98.32
Temperature	165	0.00	1.82	0.00	98.18
Humidity	718	0.00	0.00	2.23	97.77
Thickness	903	0.00	0.00	0.00	100.00

Table 13. Contingency table for **upper air BOGUS** observations, October 1991. Except for the number of observations all figures are percentages for that element. Bkgnd refers to the background check (no buddy check is applied to these data).

All the BOGUS thicknesses and about 98% of the upper air BOGUS winds, temperatures and humidities are used in the assimilation (table 13). The automatic quality control flags between 1 and 2% of the winds and temperatures. The only humidities rejected have temperatures below -40°C (listed as 'other' flags).

3.5 NESDIS satellite soundings

Figure 12 shows the numbers of SATEMs reported and rejected by layer. (Layer thicknesses are reported, these are converted to layer mean temperatures for assimilation.) Unlike the timeseries the rejections are taken to be exclusive - observations with prior rejections are not counted in any of the other categories etc. The SATEMs have a prior rejection over land and where they are colocated with LASS observations. The bottom layer is also rejected North of 20°N and the top layer (7-3 mb) is in effect also permanently rejected. CFO can specify a region within which to reject soundings, about 1% of soundings are removed in this way. Almost 20% of the soundings fail the vertical stability check (many of them in the Northern Hemisphere judging by the much smaller numbers in the bottom level).

The multilevel stratospheric check rejects about 2% of observations, the multilevel tropospheric check has very little effect. The background check

causes very few rejections (but note that the multilevel check is based on the background PGEs). The exception is the 10-7 mb layer where there is a known model bias. The layers near the top and bottom are flagged if they have PGEs > 0.3 leading to the 'other' flags.

3.6 Local Area Satellite Soundings (LASS)

Figure 13 shows the numbers of LASS observations flagged by various checks by level. (The LASS system provides temperatures at standard levels rather than thicknesses.) The top three levels are permanently rejected, partly because LASS does not use the Stratospheric Sounding Unit channels, partly because at one stage the retrievals were based on a CYBER first guess (this no longer applies). It is immediately apparent that the only checks which have a major effect are the LASS processing flags (~ 15%), the permanent rejections (over land) and, to a much smaller extent, the stability check. (Observations with both LASS and stability flags are included within the LASS bar only).

The LASS processing includes three checks:

- i) the Brightness Temperature Relative Probability (BTRP) check described by Barwell and Young (1991) (this flagged 2.84% of soundings in Oct. 1991)
- ii) a check for surface elevation over 500 m (12.66%)
- iii) retrievals which contain only climatological information are also flagged (3.31%).

In October 1991 15.39% of the soundings were rejected by the LASS processing (there is some overlap between the three flags).

There is a much lower flagging on the stability (~ 5%) and multi-level checks than for SATEMs, presumably because of the dependence on the background field. There was found to be considerable overlap between the elevation check and the tropospheric stability check. Further investigation showed that 80% of the stability rejections were over land (soundings over land are permanently rejected anyway). These poor retrievals may have been affected by poor model surface temperatures caused by excessive drying of the land surface. The NESDIS soundings did not show any excessive stability flagging over land.

Table 14 compares the flagging by the BTRP check with the stability and multi-level checks within the quality control. There is significant overlap between them given the small percentages flagged and the fact that the BTRP

check is less strict at present. This is of interest because it would be possible to replace the stability and multi-level checks on all satellite soundings by a more theoretically based check similar to the BTRP calculation (but using temperature rather than radiance differences).

		Stability	Multi-level	Neither	
BTRP	Pass	2.96	0.07	94.13	97.16
	Fail	1.65	0.09	1.10	2.84
		4.61	0.16	95.23	100.00

Table 14. Contingencies for **LASS soundings: flags on whole troposphere**, October 1991, 130824 observations.

3.7 Radiosondes

Incoming standard and significant level radiosonde data is averaged over model layers before being input to the assimilation, see Ingleby and Parrett (UMDP 32) for details. Figure 14 shows the numbers of reports from radiosondes by model level (the levels are terrain following at low levels so they do not correspond to fixed pressure levels, however nominal pressure levels have been added for guidance). There are reduced numbers of observations in the bottom few levels: the surface report is not used in the vertical averaging so the bottom level for which data is available depends on the lowest significant or standard level reported. Also the station height may be above the model height. The numbers of reports available peaks at about 800 mb then gradually declines due to balloon burst etc. There is a sudden decrease at 100 mb, due to some Part Cs and Ds not being reported or not reaching Bracknell. The numbers then reduce steadily with effectively no data above 10 mb.

Wind

Figure 15 shows the percentages of winds rejected. Permanent (prior) rejections only seem to be having effect above 500 mb - this is probably due to an error in the preprocessing of the observations (after subsequent changes to the permanent rejections there is a more even spread). A very small number of reports are rejected by CFO. The main flagging is by the background check - this is largest in the upper troposphere and lower stratosphere peaking at about 250 mb. The quality control is applied to the ascents both before and after averaging. If a reported level has a PGE >

0.999 then any average layer that uses that value is flagged, even if the average passes the quality control. This is the cause of the 'other' flags, which show their largest effect in the upper troposphere and stratosphere.

Temperature

Figure 16 shows the percentages of radiosonde temperatures rejected. There are many more prior rejections than for wind - many of them for Indian stations (note the different horizontal scale to figure 15). There are particularly large flagging rates for levels 1 and 2, presumably due to boundary layer effects. However the most noticeable feature is the large number of 'other' flags between 150 and 50 mb. As explained above these are due to checks on the original (unaveraged) data. One case was examined and it was found that these flags were mainly in the tropics and tended to have temperatures 6-8 degrees colder than the background - corresponding to a model bias (probably 3-4 degrees on average). If the model bias persists then the check on reported levels will be relaxed in January 1992 to about 8.5° (it is currently $5.5 - 6^{\circ}$). The same case showed some 10° errors in the reported temperatures - these had been correctly flagged and showed the usefulness of checking values before they are smeared out by the averaging. There were also a number of corrupt temperatures at 10 mb and some evidence of gravity waves in the stratosphere: these are probably partly responsible for the increased background flagging in the stratosphere.

Humidity

Figure 17 demonstrates the percentages of radiosonde humidities flagged (note another change of scale). If the temperature is flagged the humidity is automatically flagged as well. Flagging by the background check is 1% or less in the lower troposphere but reaches 20% at 200 mb (but humidity would not be used at this level anyway). The quality control at this time used $\sigma_o = 13\%$, $\sigma_b = 20\%$ at all levels. However it had become clear, by comparison with observation minus background statistics that these were significantly too large at low levels. On 19th November 1991 σ_o and σ_b were changed to 10% and 12% respectively below 850 mb (with 13% and 20% above 300 mb and a smooth transition) and the initial PGE for RH was reduced to 0.015 (from 0.05). The net effect of the changes was to slightly increase flagging at low levels and reduce it above about 550 mb.

Goldbeaters skin and other similar sensors are reckoned to be useless at

temperatures below -40°C , so all humidities above the -40° isotherm are flagged (shown in the 'other' flags). Because this is applied to the reported levels an averaged humidity with a corresponding temperature of -35° (or even -32°) can be flagged because one reported humidity contributing to the layer average was flagged. This is probably too severe and a change to check for averaged temperature below -40° may be made. The newer chemically-based humidity sensors are reliable at lower temperatures and so could be used at higher levels.

In general flagging of radiosonde winds and temperatures seems somewhat larger than with the previous (CYBER) quality control, partly because of the checks on the unaveraged data. However comparison is difficult because the permanently rejected observations were not subject to the automatic quality control in the previous system.

4. Summary

Actions that will be taken to improve the quality control are indicated by italics.

A. Examination of the basic equations and experience indicate that the background check is rather insensitive to the initial PGE and k - to have a significant effect k has to be changed by an order of magnitude. The quality control is rather more sensitive to the background and observation errors showing the importance of these estimates (this was particularly noticed for ship winds). The buddy check is quite sensitive to the ratio of observation and background error. *It is desirable to check values of k against histograms of rejected observations, however it is even more important that the estimates of background and observation error are approximately correct.*

B. The proportions of rejections of surface and aircraft data is broadly similar to that from the old quality control - partly due to tuning based on perceptions of the intervention forecasters. There is a slight increase in flagging of radiosondes. There are large increases in rejections of satellite soundings and satellite cloud track winds - largely due to the incorporation of extra checks pioneered at ECMWF. The previous system hardly rejected any BOGUS observations. This is still true of BOGUS thicknesses, humidities and surface winds but roughly 1.5% of pressures and upper air winds and temperatures are rejected now, some of these have incorrect positions. *The permanent rejection of high level radiosonde RHs should be removed for the newer sensor types.*

C. For all data types except satellite soundings the background check remains the main arbiter with other checks playing a lesser role. The buddy check is important because of its ability to detect large background errors, but it only affects the decisions on 0.5 to 2.0% of surface and aircraft observations. It tends to reject a slightly larger proportion of observations than it reinstates, except for SATOBs which it mainly reinstates. There is substantial overlap between the observations rejected by CFO and by the automatic quality control. *A more integrated system is needed in order to make better use of the manual effort.*

D. A significant difference from most other operational quality control schemes is that different observation types are quality controlled separately (this facilitates observation specific tests). Only surface, aircraft and SATOB data have a buddy check. These restrictions have not caused any noticeable problems. Because of the spatial distribution of observations, by far the most important buddies are those of the same type. At present at least, quality control by observation type works successfully.

E. The current (pairwise) form of the buddy check contains approximations and it was found necessary to introduce a method of 'damping' the buddy check to prevent excessive rejections. A theoretical study of methods of buddy checking is underway. However current performance is largely satisfactory.

F. Because of their distribution AIREPs are not sensitive to the function used to model the horizontal error correlation in the buddy check. They are more sensitive to the vertical and time correlations. (This may change with increasing numbers of automated systems.) It is desirable, but not straightforward, to improve the modelling of correlations in jet stream regions.

G. **Data problems.** There are specific problems with AIREPs erroneously reporting calm winds and with height assignment of GOES SATOBS. These have been addressed. The quality of SHIP data is relatively poor. The automated ASDAR systems perform better than voice reports from aircraft. *There is a worrying number of TEMPSHIP position errors each month (Appendix E), action to detect these is necessary.*

H. **Model problems.** Biases in the model temperature fields have caused problems to the quality control of satellite soundings and radiosondes.

Acknowledgement

The tuning was performed with C Parrett and under the overall supervision of S Bell. Others who have helped are mentioned in the text.

5. References

- Barwell, B.R. and Young, J.L.C. 1991 Quality control of satellite temperature soundings.
Short Range Forecasting Research Technical Note No.63. UK Meteorological Office.
- Eriksson, A. 1990 Use of cloud motion winds at ECMWF.
EUMETSAT, Proc. 8th Meteosat Scientific Users Meeting, pp 79-86.
- Graham, R. J. 1990 The impact of North Atlantic tempship observations on global model analyses and forecasts during a case of cyclogenesis.
Short Range Forecasting Research Technical Note No.46. UK Meteorological Office.
- Graham, R. J. 1991 A case study of the impact of North Atlantic tempship observations on a forecast of cyclogenesis.
Short Range Forecasting Research Technical Note No.60. UK Meteorological Office.
- Ingleby, N.B. and Hammon, O.M. 1990 A trial of the Bayesian Upper Air Quality Control - New Year 1990
Short Range Forecasting Research Technical Note No.41. UK Meteorological Office.
- Ingleby, N.B. and Parrett, C.A. 1991 Quality control of atmospheric data
Unified Model Documentation Paper 32
UK Meteorological Office.
- Lönnberg, P. 1989 Quality control and filtering of satellite data
ECMWF/EUMETSAT Workshop on the Use of Satellite Data in Operational NWP, pp 61-80.

Appendix A. List of changes

Only changes which affect quality control decisions have been included. Changes to the quality control of satellite soundings are listed separately at the end.

- 12/ 7/90 Quality Control introduced into CRAY assimilation.
- 17/ 7/90 SONDE a) temperature k reduced to 0.01 (from 0.05)
b) large gaps excluded from vertical averaging
c) levels with $PGE > 0.99$ excluded from vertical averaging
d) mid-level definition changed to $SQRT(P_1 P_2)$
- 18/ 7/90 SURFACE: wind k increased to 0.003 (from 0.0018)
- 1/ 8/90 SONDE a) error in partial layer temperatures corrected
b) partial layer values of u/v/RH changed from spot to mean
- 6/ 9/90 SONDE: background check repeated after vertical averaging
(replaced averaging of reported level PGEs)
- 4/10/90 SURFACE: $P_{msl} - P_*$ check limit increased to 1.5 mb (from 1.0 mb)
- 6/11/90 SURFACE: $P_{msl} - P_*$ check limit increased to 2.0 mb (from 1.5 mb)
SONDE: removed automatic rejection of all levels if ≥ 4 flagged
- 18/12/90 SURFACE a) $P_{msl} - P_*$ check: o-b P_{msl} scaled by P_*/P_{msl} and limit reduced to 1.0 mb
b) wind: k changed back to 0.0018 (see 18/7/90)
AIREP: wind k reduced to 0.0002 (from 0.0015)
- 8/ 1/91 SURFACE/AIREP/SATOB: intervention included
- 5/ 2/91 SURFACE/AIREP/SATOB: buddy check introduced (6/2/91 for SURFACE)
SONDE a) relax rejection limit to 0.999 in vertical averaging
b) wind k reduced to 0.0002 (from 0.001)
- 12/ 2/91 BOGUS: surface BOGUS introduced
- 19/ 2/91 SURFACE a) background 10 m wind increased by 10% for ships, Ocean Weather Ships and Buoys (only within QC)
b) thinning of drifting buoys introduced
BOGUS: upper-air BOGUS introduced
- 21/ 3/91 BOGUS: (surface) wind k reduced to 0.000005 (from 0.0001)
- 9/ 4/91 SURFACE/AIREP/SATOB:
a) damping introduced into buddy check ($\eta_1 = 0.5$)

b) extra sorting in buddy check (mainly for AIREPs)
 BOGUS: (upper-air) wind k reduced to 0.000005 (from 0.00002)
 23/ 4/91 BOGUS: a) k set equal to radiosonde values
 b) background error variance x4 in background check
 SURFACE: a) synoptic dependent background errors for P_{msl} and v_{10m}
 (upper air background errors displaced)
 b) 10m observed wind errors multiplied by 0.8
 c) P_{msl} buddy check: 0.5p fix removed
 SURFACE/AIREP/SATOB: minor error in observation pairing corrected
 (implied no buddies from next latitude band in data sparse areas)
 26/ 4/91 SURFACE: correction to synoptic dependent errors, 1σ added
 9/ 5/91 SURFACE/AIREP/SATOB:
 a) change to alternative form of buddy check damping ($\eta_2=0.5$)
 b) vertical correlation function narrowed to b=6 (from b=3)
 AIREP: 'calm' AIREP winds rejected after background check
 SATOB: a) PGE(location) changed to 0.0 (from 0.02)
 b) asymmetry (speed bias) check introduced
 12/ 6/91 AIREP/SHIP: CFO supported flag introduced
 16/ 7/91 BOGUS: BOGUS thicknesses introduced
 SATOB: PGE(location) reset to 0.02 (from 0.0)
 13/ 8/91 SONDE: a) relative humidity k reduced to 0.01 (from 0.05)
 b) flagged data included in vertical averaging
 BOGUS: initial PGE for upper-air BOGUS reduced to 0.005 from 0.015
 (station list change on 23/4/91 not effective due to code fix)
 SURFACE/AIREP/SATOB: change to exclusions from buddy check
 (from $PGE2 > 0.99999$ to $x^2/V > N_c^2$)
 10/ 9/91 upper-air background errors corrected
 19/11/91 SONDE/BOGUS: RH PGE reduced to 0.015 (from 0.05) and RH errors reset
 $\sigma_o=10\%$, $\sigma_b=12\%$ below 850 mb with transition to 13 and 20% at 300 mb
 changes to sonde wind and temperature errors and Hawson corrections
 GOES SATOBs: PGE(location)=0.0 below 800 mb, =0.04 above
 10/12/91 SURFACE: wind k reduced to 0.0010 (from 0.0018)

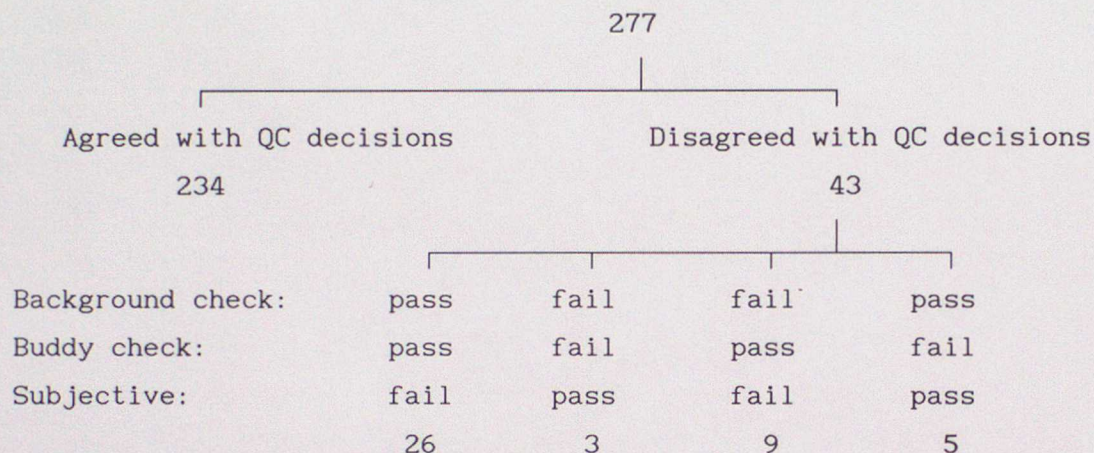
Satellite soundings

- 9/ 7/90 SATEM: PGE3 below 100 mb doubled when stability check not done
- 11/ 7/90 SATEM: double errors above 100 mb within QC only (FIX)
- 12/ 7/90 SATEM: triple errors of 50-30 mb layer within QC only (FIX)
- 19/ 7/90 SATEM: reduce -ve biases of 50-30 mb layer by 8° within QC (FIX)
- 22/ 8/90 SATEM: remove FIXES on stratospheric levels
- 14/ 9/90 SATEM: do not flag whole troposphere if only top level flagged and whole stratosphere flagged
- 6/11/90 LASS: quality control of LASS data introduced
- 18/12/90 SATEM: temperature k and PGE both increased to 0.12 (from 0.08)
- 21/ 5/91 SATEM: a) vertical stability check tightened by reducing
$$c_s \text{ to } 2.0 \text{ (from } 2.5) \text{ and } \varphi_{\min} \text{ to } 1.5^{\circ} \text{ (from } 2.5^{\circ})$$
b) PGE for lowest layer increased to 0.15 (from 0.12)
c) Part C SATEM check removed
- 17/ 9/91 LASS: Processing changed to NOAA11 from NOAA10.
- 19/11/91 SATEM: multiply PGE by 0.2 and error variance by 2 above 10 mb (FIX)

Appendix B. Subjective assessment of PMSL buddy check

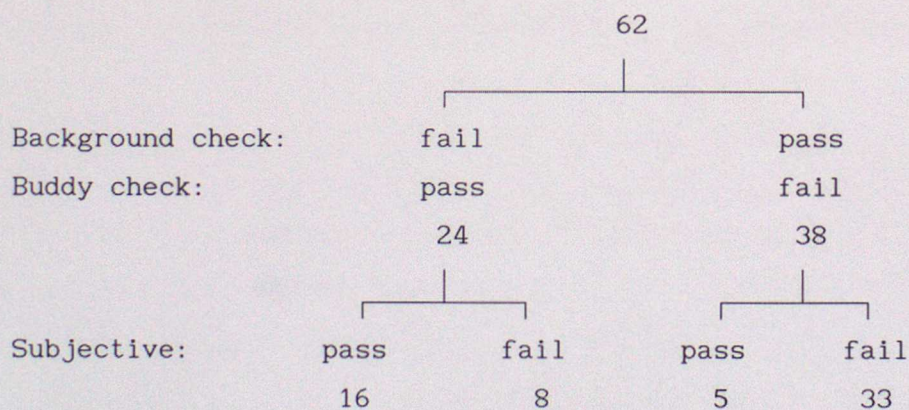
A subjective assessment of one case (12Z 27th January 1991) was made of the effect of the buddy check on quality control of pressure at mean sea level (PMSL) in the global model. Thanks are due to O Hammon for this assessment.

Total Number of Pmsl observations with PGE > 0.2



Total Number of Pmsl Observations with Critical Buddy Checks

(ie where the buddy check has reversed the background check decision)



So the subjective assessment agreed with the buddy check for 49 observations and disagreed for 13.

Appendix C. Buddy check exclusions and rounding error

When the buddy check was written it was initially considered that all observations could be included - the equations used will treat any deviation from background, but give much less 'weight' to an observation with PGE near 1. However the program failed on 29th January just as it was about to go 'live'. This was traced to two observations close together reporting P_{msl} of about 300 mb, the increments of about 700 mb caused the argument of the exponent function (eqn 4.15 UMDP 32) to exceed its maximum permissible value. The remedy used was to impose upper and lower limits of ± 170 on the argument and to exclude any observations with a PGE over 0.99999 from the buddy check. It was disturbing that when run with the argument limits but not the exclusions the two 300 mb observations passed the buddy check, presumably because of rounding errors.

The limits on PGE were used for several months until it was realized that it was excluding observations with differences from background greater than about 10 mb (the value depends on the background error estimate) - some of which were correct and which should be buddy checked. Adding further 9's to the critical PGE only increases the limit by less than 1 mb per digit (see Appendix 1 of UMDP 32) - leading to an unacceptable number of decimal places before the cut-off limit is increased significantly.

The solution adopted on 13th August was to change to a limit on the difference from background of $N_c = 25$ standard deviations (section 2.4 of UMDP 32). Two weeks later there was a failure in the operational surface buddy check and another in a 42-level assimilation caused by PGEs going slightly over 1 due to rounding error. The observations involved had differences from background of 22.1 and 17.2 mb, and the second agreed well with neighbouring observations. PGEs were chopped to a maximum of 1 from 10th September, but this did not stop a further failure in a 42-level assimilation caused by a PGE of exactly 1 (and some reordering of the calculations by the compiler?). To prevent recurrences $PGEs \geq 1$ were excluded from the buddy check and the factor raised to the power η_2 was checked for values ≤ 0 .

The limit on the argument of the exponent function was reexamined to see if the restriction on difference from background renders the limit unnecessary. The IBM has a restriction of $\arg \leq 174.673$ the CRAY a restriction of $|\arg| < 2^{13} \ln 2 = 5678.26$.

If $x_1/\sigma_1 = x_2/\sigma_2 = x/\sigma$ then equation 4.15 of Ingleby and Parrett (1991)

becomes

$$\frac{P(O_1 \cap O_2 | \bar{G}_1 \cap \bar{G}_2)}{P(O_1 | \bar{G}_1)P(O_2 | \bar{G}_2)} = \frac{1}{(1-\rho^2)^{0.5}} \exp\left(\frac{-\rho^2}{2(1-\rho^2)} - \frac{2x^2}{\sigma^2}(1-1/\rho)\right)$$

$$= \frac{1}{(1-\rho^2)^{0.5}} \exp\left(\frac{\rho}{(1+\rho)} - \frac{x^2}{\sigma^2}\right)$$

Similarly $x_1/\sigma_1 = -x_2/\sigma_2 = x/\sigma$ gives

$$\frac{P(O_1 \cap O_2 | \bar{G}_1 \cap \bar{G}_2)}{P(O_1 | \bar{G}_1)P(O_2 | \bar{G}_2)} = \frac{1}{(1-\rho^2)^{0.5}} \exp\left(\frac{-\rho}{(1-\rho)} - \frac{x^2}{\sigma^2}\right)$$

So the argument is bounded above by $0.5 (X/\sigma)^2$, but is not bounded below unless ρ is bounded above by $k < 1$ (the damped and vector forms of the equations would also have to be examined). For safety it was decided to leave a limit of $|\arg| < 170$ in place (any argument with larger modulus is almost certainly due to one or both of the observations being erroneous). It seems that when the observations are selected on $PGE < 0.9999$ (as in previous versions of the Bayesian quality control) the argument check is not necessary.

In summary the buddy check is more sensitive than desirable to rounding error, although the observation combinations causing problems are fairly rare and containing action has been taken. It is not clear whether the equations can be rewritten so as to make them more robust. It is interesting that all the specific problems were caused by P_{msl} and perhaps indicates that P_{msl} has a longer-tailed distribution than other variables.

Appendix D. Estimation of input parameters and background check limits

The input parameters V_0 and $P(G)$ used in the Met Office system were estimated from root-mean-square observation minus background statistics (for 'good' observations) and the proportion of 'bad' data respectively. This requires an approximate separation into observations with and without gross errors from a previous version of the quality control. k should be determined from histograms of the density of 'bad' data, but in practice it has been tuned more empirically.

For the model of observation errors to be meaningful the nominal range (usually $-1/2k$ to $1/2k$) of the 'tophat' function should cover the region where the density of normal observation errors is non-negligible, and in particular should include the background check rejection limits $\pm x_{0.5}$ (typically $x_{0.5}$ is several times the value of σ). These can be derived by setting $P_2 = 0.5$ in the equations given in Appendix 2 of Ingleby and Parrett (UMDP 32). For scalars we should have

$$x_{0.5} < 1/2k, \text{ where } x_{0.5}^2 = -2V \ln \left((2\pi V)^{0.5} k P_0 / (1-P_0) \right)$$

$1/2k$ just greater than $x_{0.5}$ should also be treated with suspicion.

For wind the distributions are two dimensional and so $k=1/\text{area}=1/4r_{\square}^2$, if it is considered that observations can lie in the square region $-r < x < r$ and $-r < y < r$. If the circular region $x^2+y^2 < r^2$ is used instead then $k=1/\pi r_{\circ}^2$.

$$x_{0.5} < r, \text{ where } x_{0.5}^2 = u^2+v^2 = -2V \ln \left(2\pi V k P_0 / (1-P_0) \right)$$

$$\text{and } r_{\square} = (1/4k)^{0.5} \text{ or } r_{\circ} = (1/\pi k)^{0.5}$$

For typical values of P_0 (less than 0.1) the $1-P_0$ in the denominator is approximately constant and x is effectively a function of kP_0 . The dependence is weak because $x_{0.5} \propto \sqrt{\ln(f)}$ where $f \propto kP_0/(1-P_0)$. In example 1 dividing k by 10 has less effect than increasing the background error from 1 mb to 1.5 mb. This insensitivity to parameters that are only known approximately is desirable. Because x is more sensitive to the estimates of background and observation error it is important to get these approximately correct. To a first approximation $x_{0.5} \propto \sqrt{V}$, but the factor multiplying V decreases slightly as V increases because of the presence of V inside the logarithm.

Examples

- 1) For ship pressures $P_0 = 0.06$ and $\sigma_0 = 1.0$ mb.
 If $\sigma_b = 1.0$ mb, $k = 0.0430 \text{ mb}^{-1}$ then $V = 2.00 \text{ mb}^2$ and $x_{0.5}^2 = 9.3V = (4.30)^2$.
 If $\sigma_b = 1.5$ mb, $k = 0.0430 \text{ mb}^{-1}$ then $V = 3.25 \text{ mb}^2$ and $x_{0.5}^2 = 8.8V = (5.34)^2$.
 If $\sigma_b = 1.0$ mb, $k = 0.0043 \text{ mb}^{-1}$ then $V = 2.00 \text{ mb}^2$ and $x_{0.5}^2 = 13.9V = (5.27)^2$.
- 2) For relative humidity, which has a range of 100%, k should equal $0.01\%^{-1}$, however it was erroneously set to $0.05\%^{-1}$ initially. The error standard deviations were $\sigma_0=13\%$, $\sigma_b=20\%$ and $P(G)=0.05$. This gives $\sigma=23.9\%$ and $x_{0.5}=45.9\%$ compared to $1/2k=10\%$. Also for zero difference from background $P(G|0)=0.14$ - the excessive value of k has caused the background check to increase the PGE despite the observation agreeing exactly with the background. (The above equations should be used with care for RH because of the finite range involved.)
- 3) For AIREP winds $k=0.0002 \text{ (m/s)}^{-2}$ which gives $r_{\square} \approx 35 \text{ m/s}$ or $r_{\circ} \approx 40 \text{ m/s}$. A typical value of $x_{0.5}$ is 23 m/s for $V \approx 41 \text{ (m/s)}^2$. The value used initially $k = 0.0015$ was certainly too large giving $r_{\square} \approx 13 \text{ m/s}$ and $x_{0.5} \approx 19 \text{ m/s}$. Note that r changes much more rapidly than $x_{0.5}$.
- 4) For surface ship winds $k=0.0018 \text{ (m/s)}^{-2}$ and $V \approx 13 \text{ (m/s)}^2$ from December 1990. This gave $r_{\square} \approx 11.8 \text{ m/s}$ and $x_{0.5} \approx 11.2 \text{ m/s}$. These two values seem too close together so k was reduced to $0.0010 \text{ (m/s)}^{-2}$ on 10th December 1991. This gives $r_{\square} \approx 15.8 \text{ m/s}$ and $x_{0.5} \approx 11.7 \text{ m/s}$. This brings k for ships more into line with the value for AIREPs and radiosondes, it should not be the same because the variance is less at the surface.

Appendix E. TEMPSHIP position errors

For the months January to April 1991 plots of TEMPSHIP tracks were examined, along with listings of the positions, for the suspect observations the reports were then examined (all information was taken from the OPD). Table E1 gives details of those observations that appear to be wrong. In February there were two cases of rather small position errors and one case where some doubt remained. The other observations can be identified with some confidence as having serious position errors. This gives an estimate of about four position errors a month if we take the 20 erroneous reports from UMFV as exceptional. Even this is probably an underestimate, there may be other errors not detected due to gaps in the OPD, insufficient data (only occasional reports from a ship), curious but possible tracks (sometimes from research vessels), rather small position errors or simply an oversight in the subjective checking. Figures E1 and E2 show the position errors in March and April 1991 respectively.

Possibly two thirds of the errors could be detected by considering the previous track, including about one third of the errors where there is an observation at the correct position. In some of these cases the error is obvious such as a sign change or tens digit wrong and the position could be corrected. At least two of the erroneous reports were over land, but no systematic checking of this was done.

The automatic quality control cannot detect TEMPSHIP position errors at present because it considers each element at each level individually. Only a minority of levels had been flagged - some of the differences from background were rather larger than usual but the quality control of radiosonde data has wide tolerances because the vast majority of the data is considered good. Some of the erroneous reports only contained a few levels, but some were full ascents. The old (CYBER) QC scheme was examined in this investigation. Results with the new QC scheme are likely to be similar, or worse since the check 'if 4 or more levels of one element flagged then flag all levels of that element' was removed.

In absolute terms the number of position errors is small, but TEMPSHIPS are an important source of data (eg Graham, 1990, 1991) and with a few island radiosonde stations provide the most complete profiles of information over the data sparse ocean areas. Because of the lack of other data one misplaced TEMPSHIP which passes the quality control can cause a significant bulls-eye

in the analysed fields through much of the depth of the atmosphere. One such bulls-eye from a TEMPSHIP position error was seen in the OWSE-NA experiment in 1987.

TEMPSHIPS with only one or two reports were also examined and in many cases could be identified as having incorrect callsigns, these are noted in table E2. This suggests that there are about 10 incorrect callsigns a month, with some ships giving several errors over the four month period. Four distinct cases where both callsign and position were in error were found.

Most of the errors were either one letter corrupt or missing or two letters transposed. It would be possible to provide an algorithm to choose the correct callsign from a list perhaps 75% of the time, subjective checking particularly with a movement history would be able to do even better. It is suggested (recommendation 3) that any unknown callsigns be rejected - this raises the question whether SHIP should be allowed or not (see table E2 for numbers). Incorrect callsigns degrade attempts at monitoring ships and checking their positions. The sample investigated here suggests that they are more likely than average to have a position error.

Thanks are due to A Burhan for his help with the checking.

Recommendations

1. That the SDB movement check applied to ships be extended to TEMPSHIPS. This should include cross checking between ships and TEMPSHIPS.
2. That the SDB land/sea check be extended to TEMPSHIPS. This checks for a navigable waterway within 1° of the reported position.
3. That a list of TEMPSHIP callsigns be kept up to date, any report with a callsign not on the list should be flagged and not assimilated, unless it is reinstated by CFO.
4. That the intervention forecasters examine any TEMPSHIPS flagged by the SDB/QC checks, a graphical display of TEMPSHIP movement should be available to help them. They should have a facility to correct the callsign or the position, or to set an 'ignore SDB movement check' flag.
5. That a check be made for more than one report at the current time (this is not included within the SDB movement check). This is best done within the quality control when the maximum amount of information is available to decide which of the two observations is correct.

6. That the flags from the SDB movement and land/sea checks (1 and 2) be incorporated into the automatic quality control. The land flag should provoke an automatic rejection, the movement flag (which can be due to an error in the previous report, rather than the current one) would be used as input to a position check (7).

7. That the automatic quality control include a position check involving all elements in the report.

Date/time	Callsign	Position	Correct	Report at correct posn?	Notes
20JAN 12Z	ERES	(34.4, -40.6)	(44.4, -40.6)	Yes	
25JAN 12Z	OFNRH	(45.9, -8.8)	(45.9, -18.8)	Yes - FNPH	
25JAN 12Z	FNRH	(45.9, -8.8)			
27JAN 00Z	FNOU	(11.1, -57.5)	(19.1, -57.5)	Yes	
27JAN 12Z	OZJP	(53.3, -45.3)	(59.0, -45.3)	No	
4FEB 12Z	DZHA	(46.2, 7.2)	(46.2, -7.2)	No	
8FEB 00Z	EREA	(28.0, -27.4)	(28.6, -27.4)	Yes	Minor.
9FEB 12Z	FNPH	(23.3, -53.9)	(21.2, 54.6)	No	Probable.
15FEB 00Z	FNOU	(30.0, -36.3)	(30.5, -36.3)	Yes	Minor.
4MAR 12Z	LBBH	(-19.4, -47.1)	(-19.4, -7.1)	Yes - DBBH.	
9MAR 12Z	UMFW	(-59.6, -21.8)	(-59.6, 21.8)	No	
.... 18 other sign errors from UMFW, 10 with report at correct position					
29MAR 00Z	UMFW	(-67.1, -73.6)	(-67.1, 73.6)	Yes	
11MAR 00Z	UPUI	(-5.2, -23.3)	(-4.3, -27.2)	No	
15MAR 00Z	EFEH	(20.7, 14.4)	(20.7, 142.1)	Yes - EREH.	Land.
20MAR 00Z	DBBH	(-13.7, -10.2)	(-13.7, 10.2)	No	
1APR 12Z	UUQR	(82.7, 43.4)	(72.7, 43.4)	Yes	
2APR 00Z	DBBH	(-25.6, -1.7)	(-5.6, -1.7)	No	
11APR 00Z	VSBW3	(50.2, -76.1)	(50.2, -36.1)	Yes - VSBV3.	Land.
13APR 00Z	DZHA	(59.3, 2.4)	(52.4, 9.3)	No	

Table E1. TEMPSHIPS with position errors, January to April 1991.

The positions are given as (latitude, longitude) with North and East as positive. The estimated positions (*italics*) are obtained from the track of the vessel, and if possible from permutations of the reported position. Where

there is another report that is consistent with the track then its position is given as correct.

Notes. Minor - only small difference in position. Probable - probably wrong but cannot be deduced with certainty. Land - the incorrect report is at a land position.

January: 1034 observations, 38 from SHIP. 7 or 8 callsign errors:

DWR (LDWR), FRE (UFRE), VSBV (VSBV3), ZC/K (ZCSK), ZCSK UUB (ZCSK), OFNRH ** (FNPH), FNRH ** (FNPH), 7JDU (? isolated report)

February: 879 observations, 64+2+1 from SHIP+SHIPA+SHIPB. 11 callsign errors:

DKQ (DKQP), F (FNOU), FNOUS (FNOU), FNOW (FNOR), FNRQ (FNRS), G/CA (GACA), SHIP (GACA), U (FNOU), UPHA (VPHA), USBV3 (VSBV3), ZHA (DZHA)

March: 960 observations, 17+23+1 from SHIP+SHIPA+SHIPB. 15 callsign errors:

DBB (DBBH), DRBH (DBBH), EFEH ** (EREH), FN S (FNRS), FN/U (FNOU), LBBH ** (DBBH), OVJY (OVYA), RSBV3 (twice - VSBV3), UM Y (UMAY), VYA (OVYA), UMF (three times - UMF), SUIPA (SHIPA)

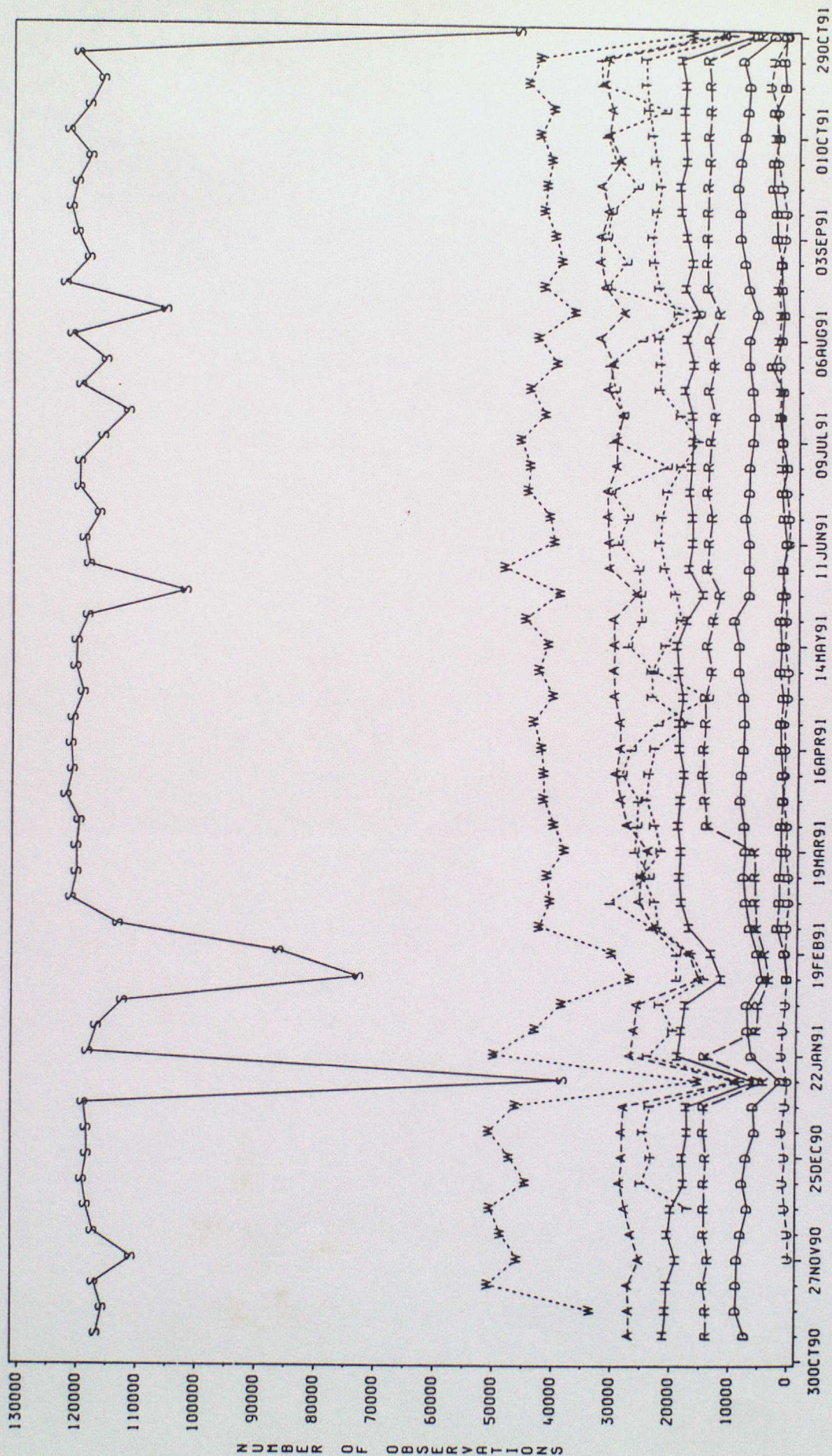
April: 962 observations, 7+20+1 from SHIP+SHIPA+SHIPE. 11 callsign errors:

DBFH (DBBH), DKQQ (DKQP), DZAH (DZHA), ERVC (EREC ?), FNMU (FNOU), FNX (?), IZGH (UZGH), VOBV3 (VSBV3), VSB3 (VSBV3), VSBW3 ** (VSBV3), SHZP (SHIP)

Table E2. Numbers of observations and callsign errors. Observations with identical callsign/latitude/longitude/time are not included in the numbers. The correct callsigns are given in brackets - indicated with a question mark if there is some doubt about the identification. ** indicates a position error, see Table E1.

Figure 1.

CRAY OPD: NUMBERS OF OBSERVATIONS



TYPE S-S-S SHIP S-S-S SHIP
 A-A-A AIREP U-U-U ASDAR
 R-R-R SONDE H-H-H SHIP
 B-B-B DRIBU U-U-U SHIP
 T-T-T SATM S-S-S SHIP
 W-W-W SATOB L-L-L LASS
 COMRESSED CODE SATEMS ARE NOT INCLUDED

Figure 2.

SYNOP STATISTICS: Pressure

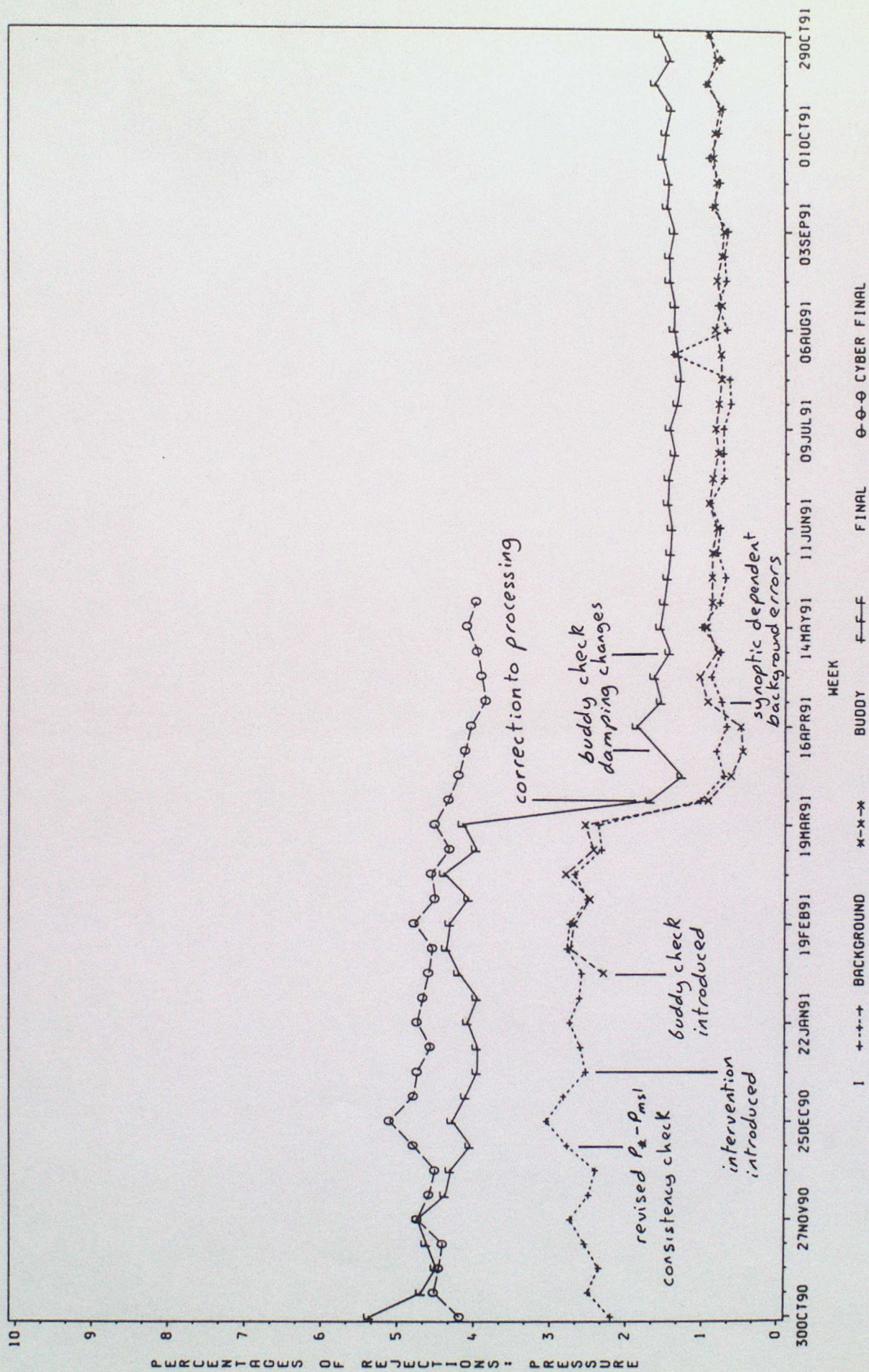


Figure 3.

SHIP STATISTICS : Pressure

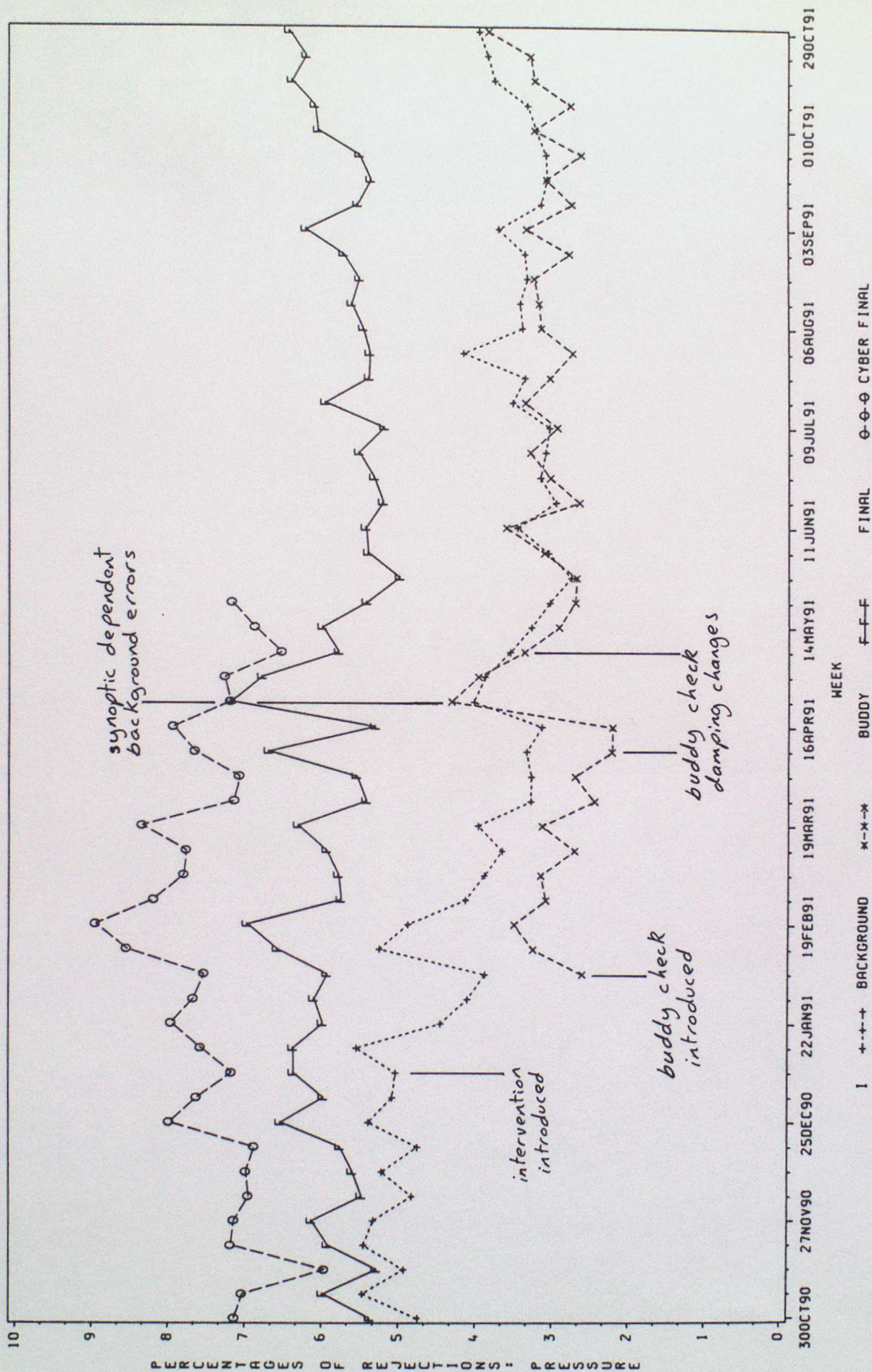


Figure 4.

SHIP STATISTICS : Wind

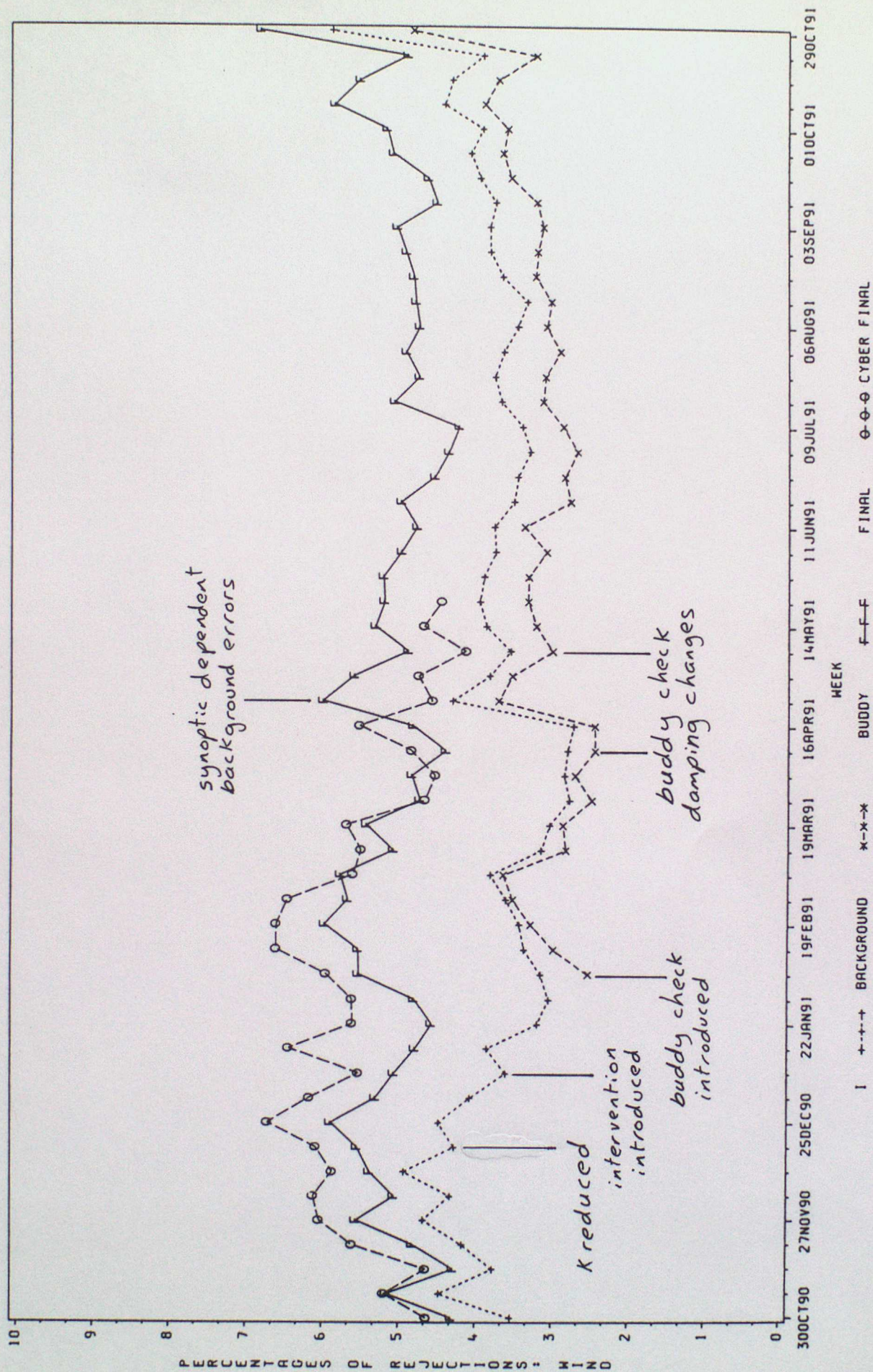


Figure 5.

AIRCRAFT STATISTICS : Wind

TYPE=AIREP PRESSURE=1-400 MB

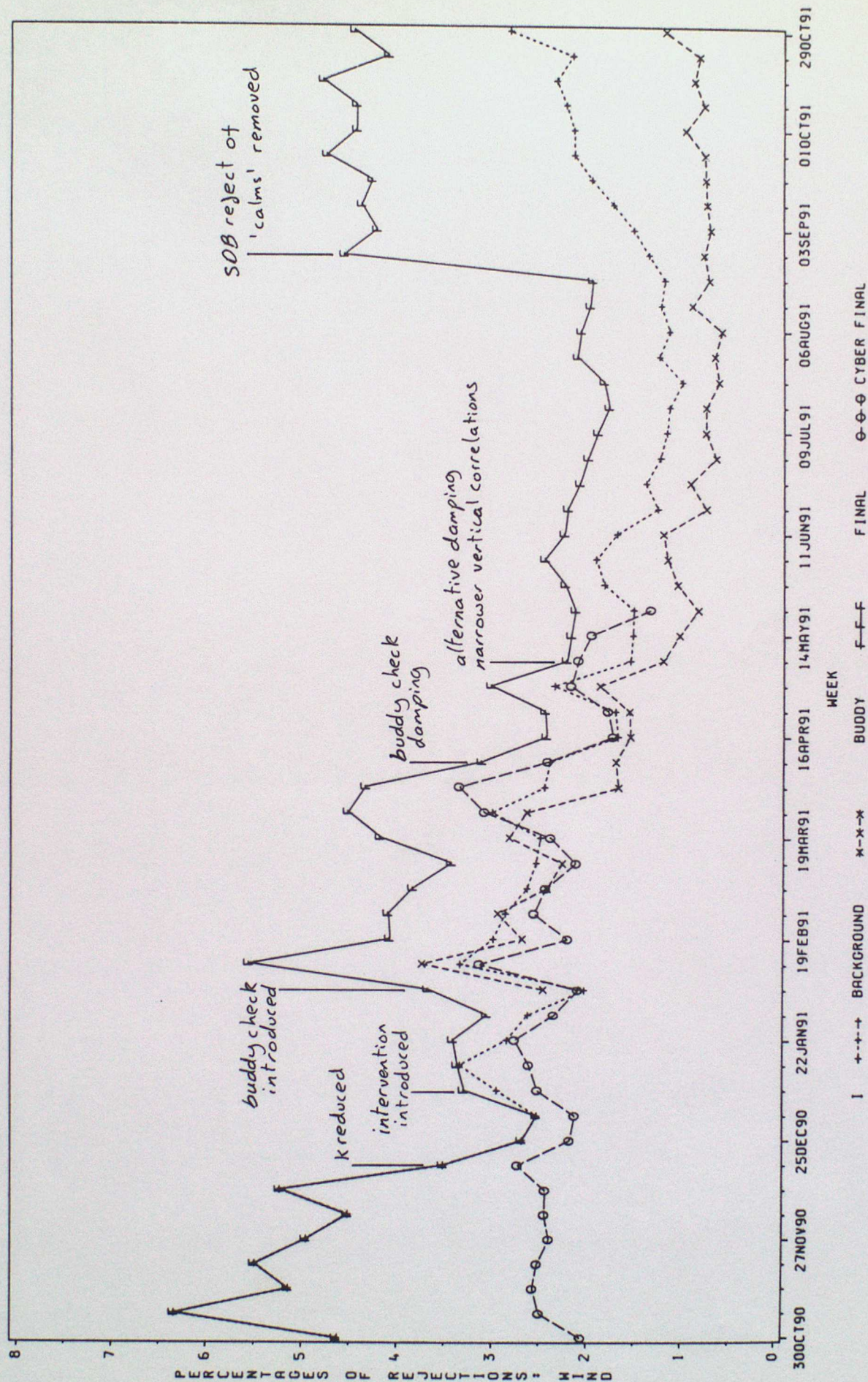


Figure 6.

AIRCRAFT STATISTICS: Temperature

TYPE=R1REP PRESSURE=1-4.00 MB

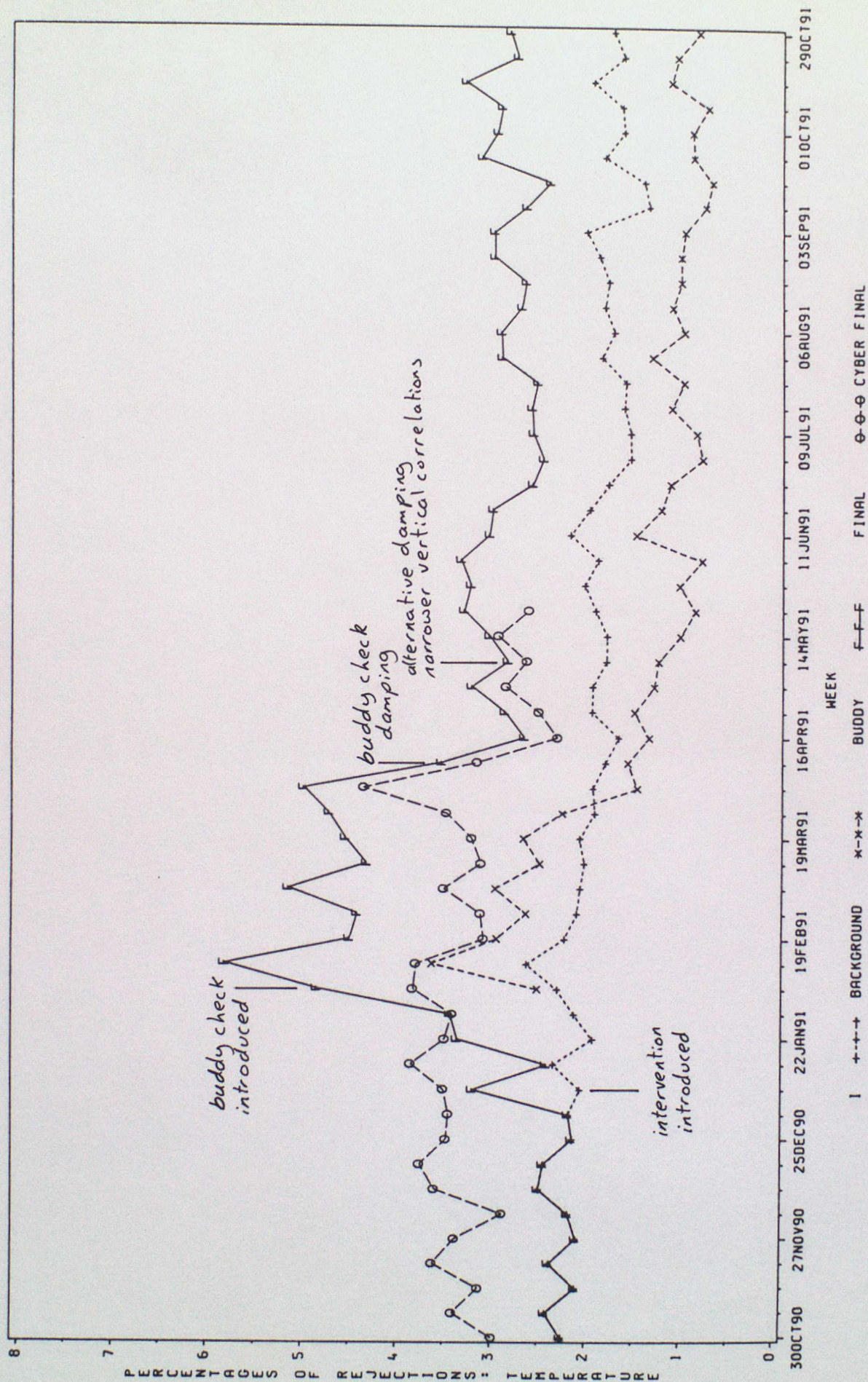
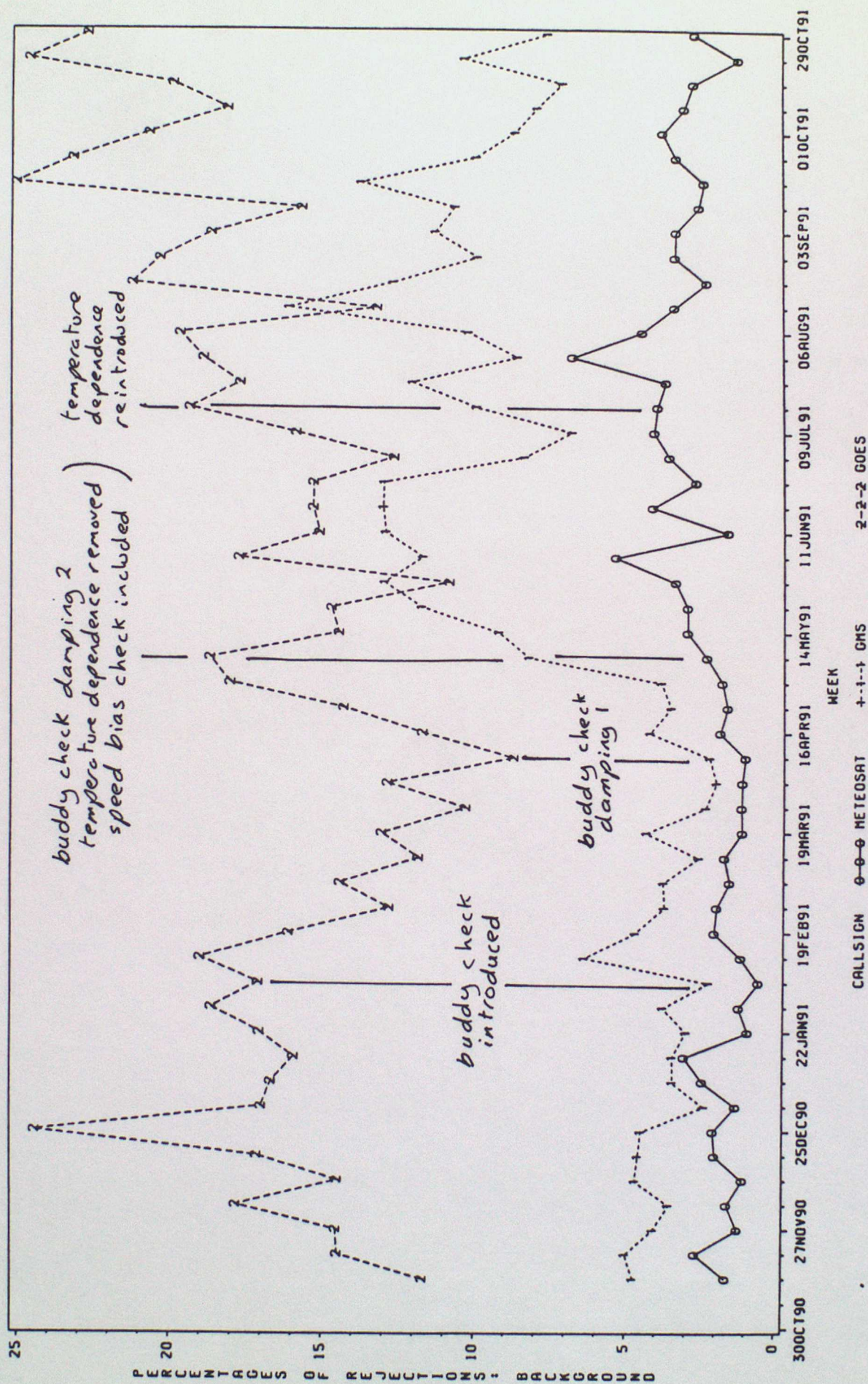


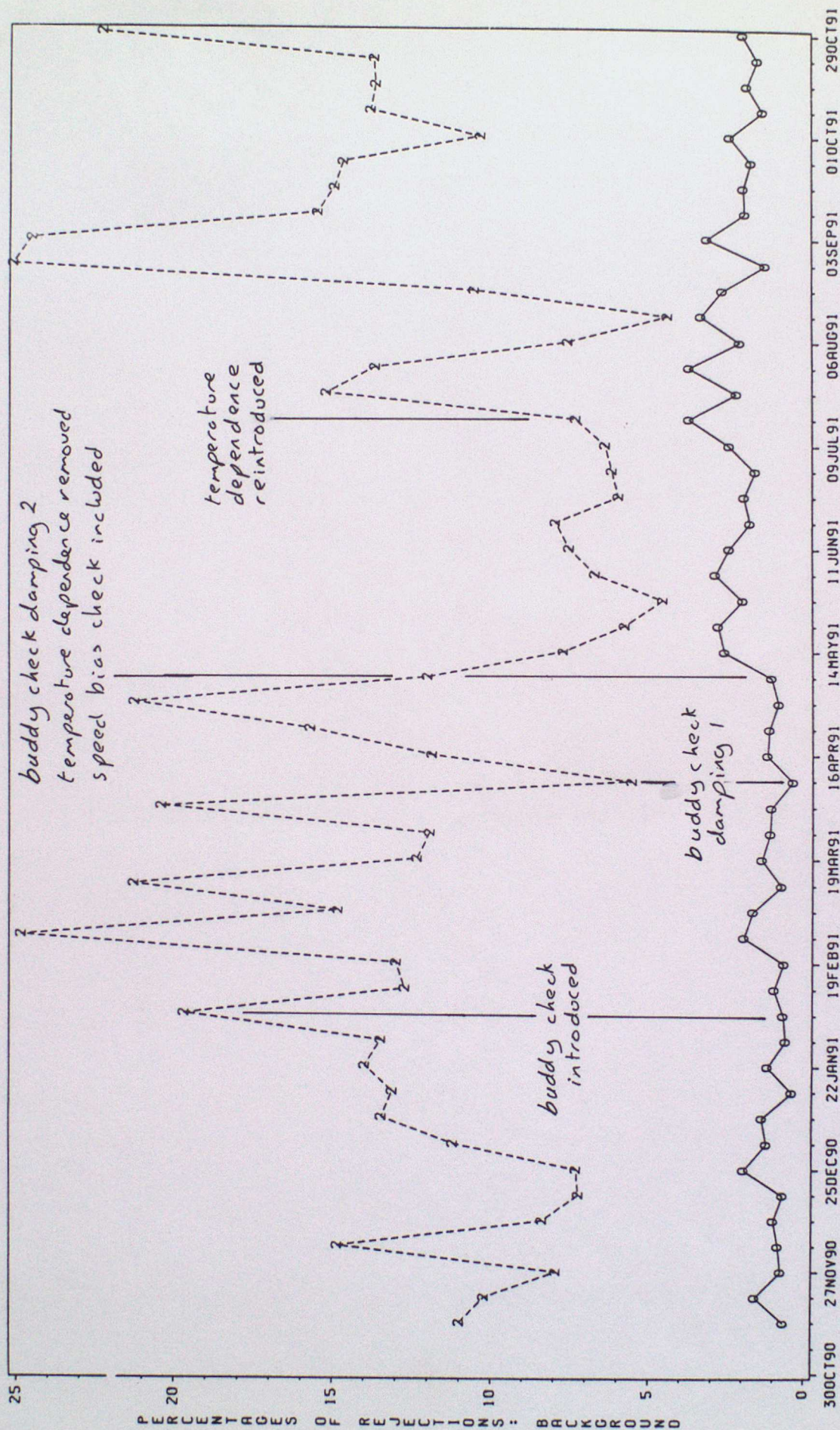
Figure 7.

SATOB STATISTICS: CRAY : High level
PRESSURE=1-400 MB



SATOB STATISTICS: CRAY : Mid level PRESSURE=400-700 MB

Figure 8.



SATOB STATISTICS: CRAY : Low level PRESSURE = 700-1000 MB

Figure 9.

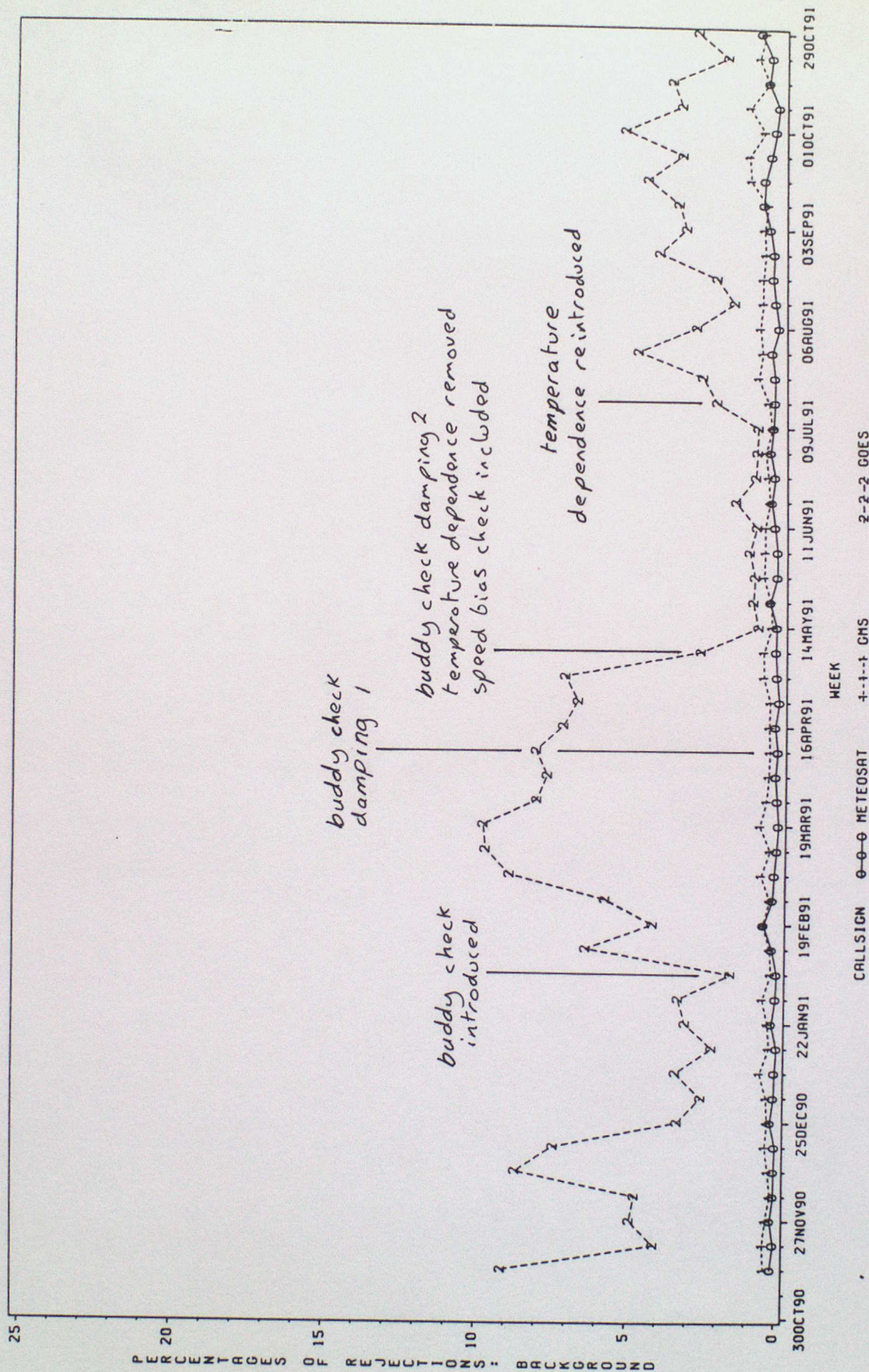


Figure 10.

SURFACE BOGUS STATISTICS: Pressure

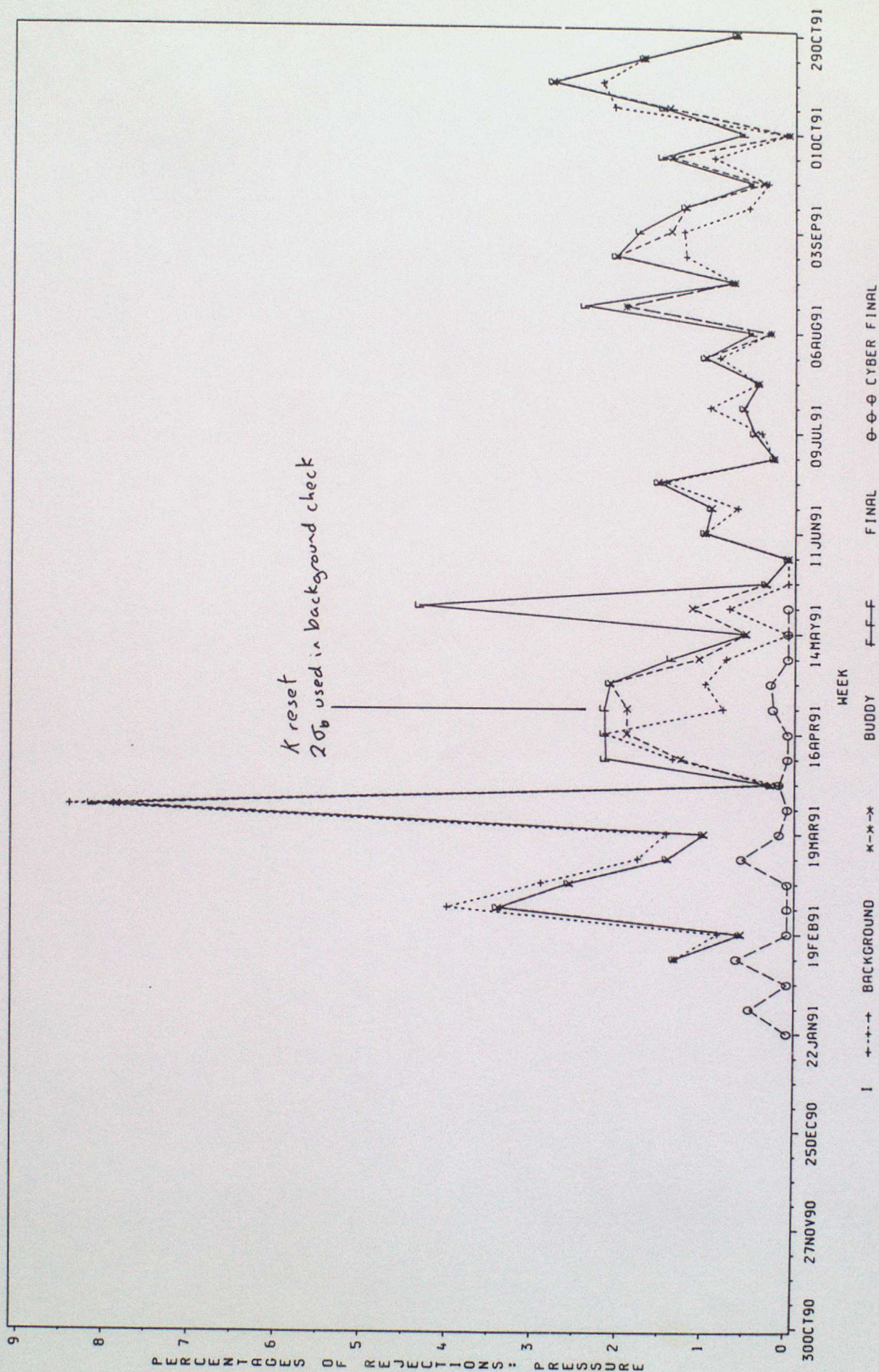


Figure 11.

SATEM STATISTICS: CRAY

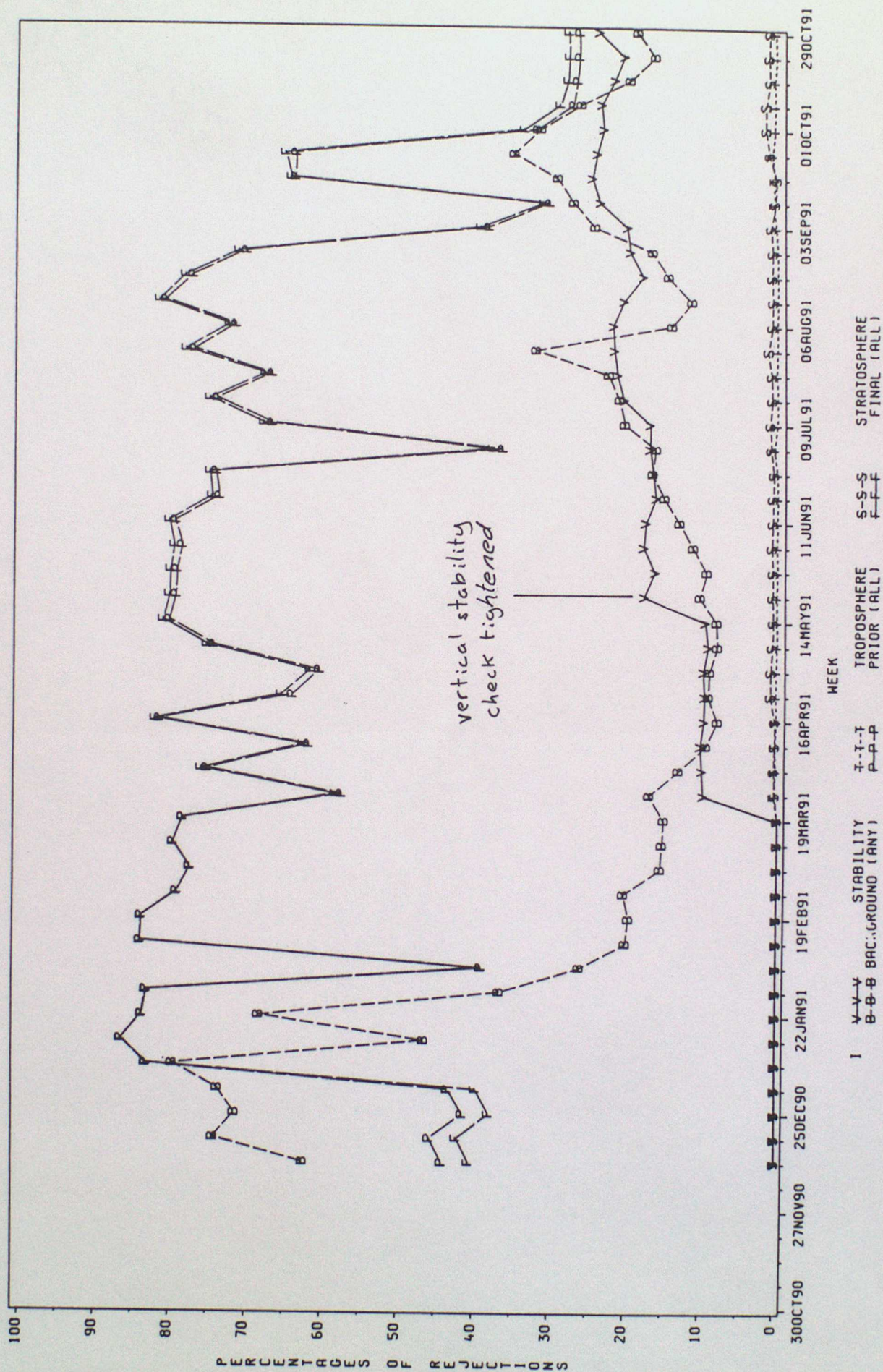


Figure 12.

SATEMS: OCTOBER 1991

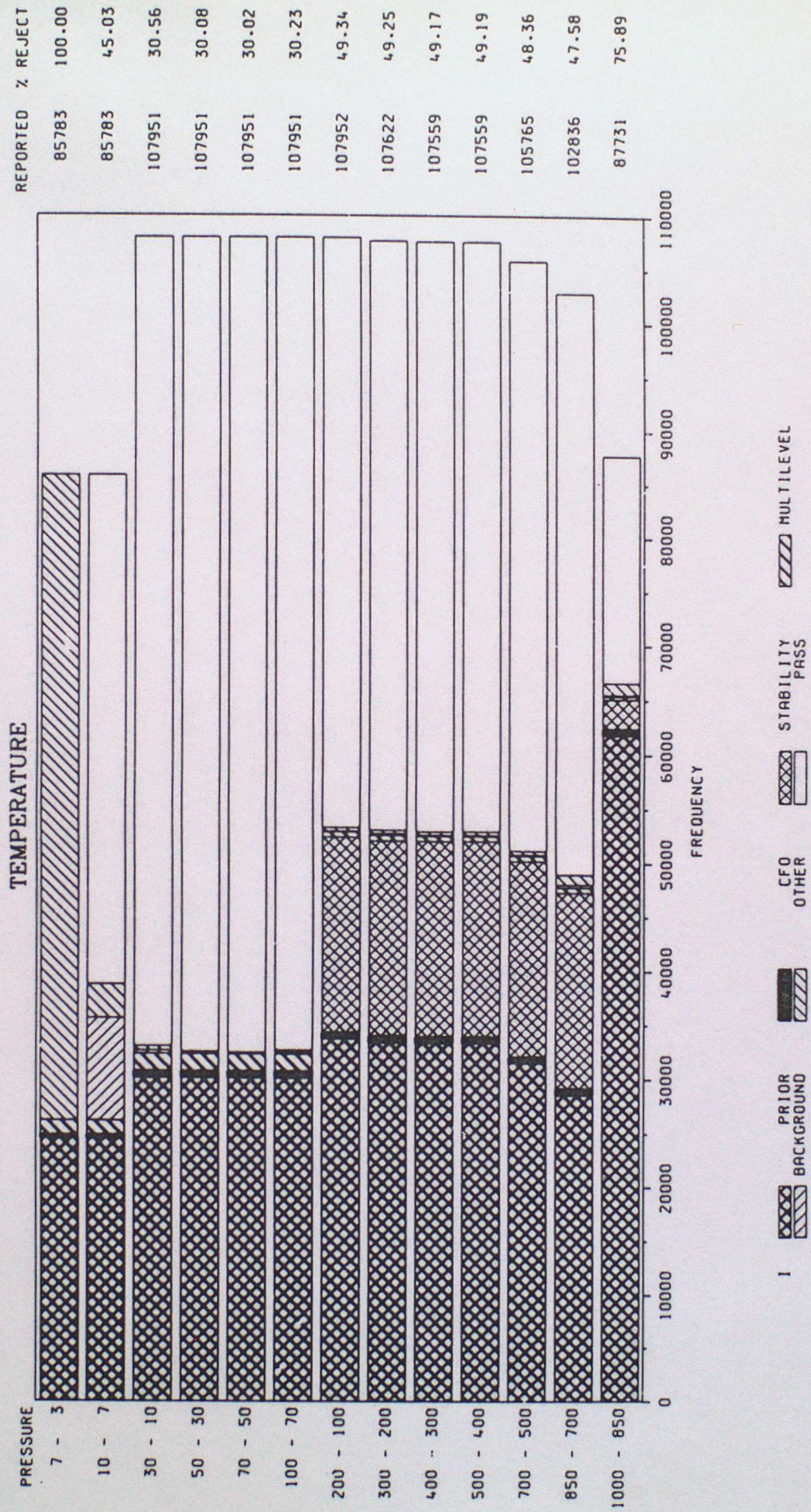


Figure 13.

LASS: OCTOBER 1991

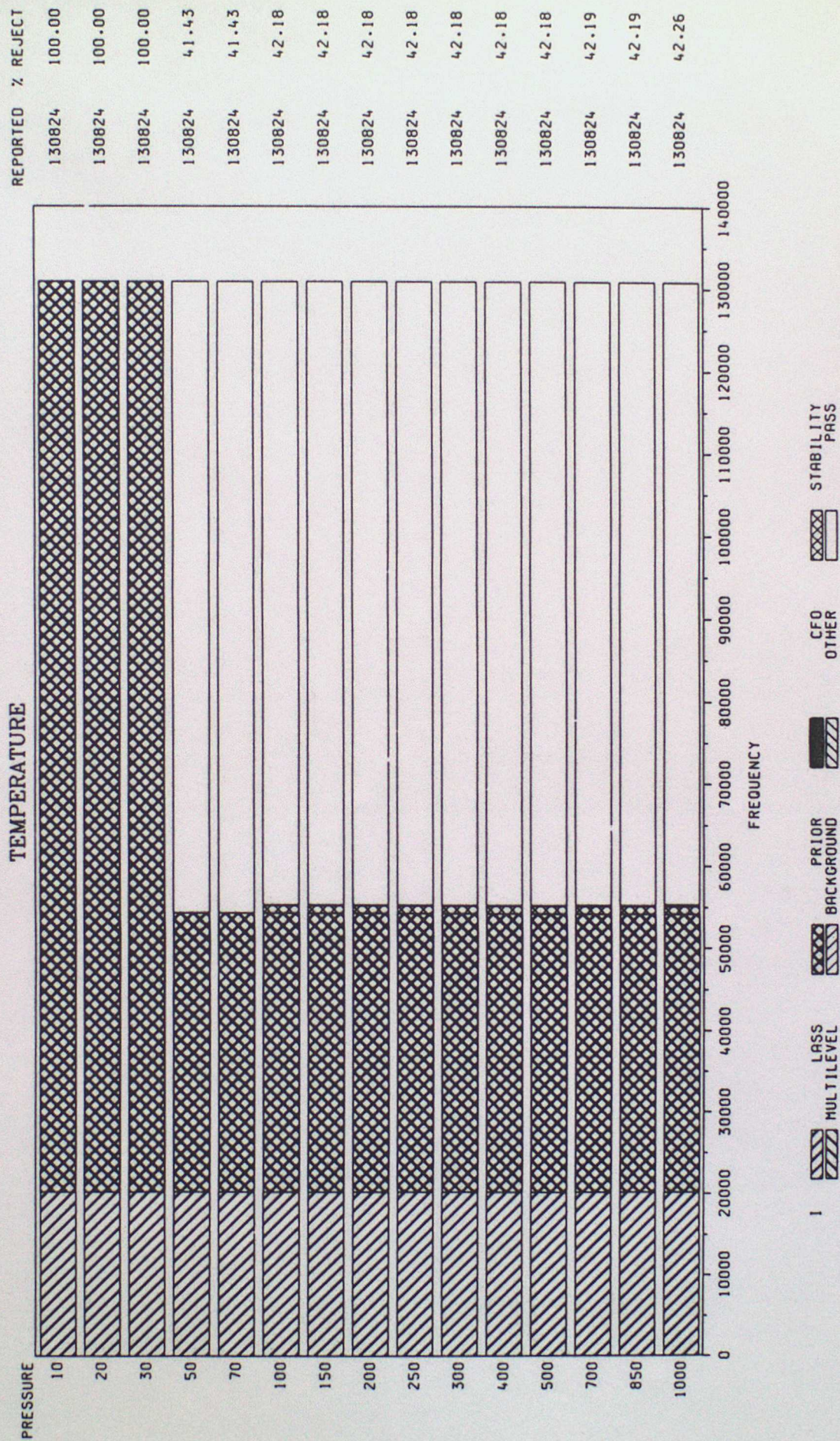


Figure 14.

RADIOSONDES: OCTOBER 1991
THE NOMINAL LAYER PRESSURES CORRESPOND TO $P_{\text{m}} = 1000 \text{ MB}$

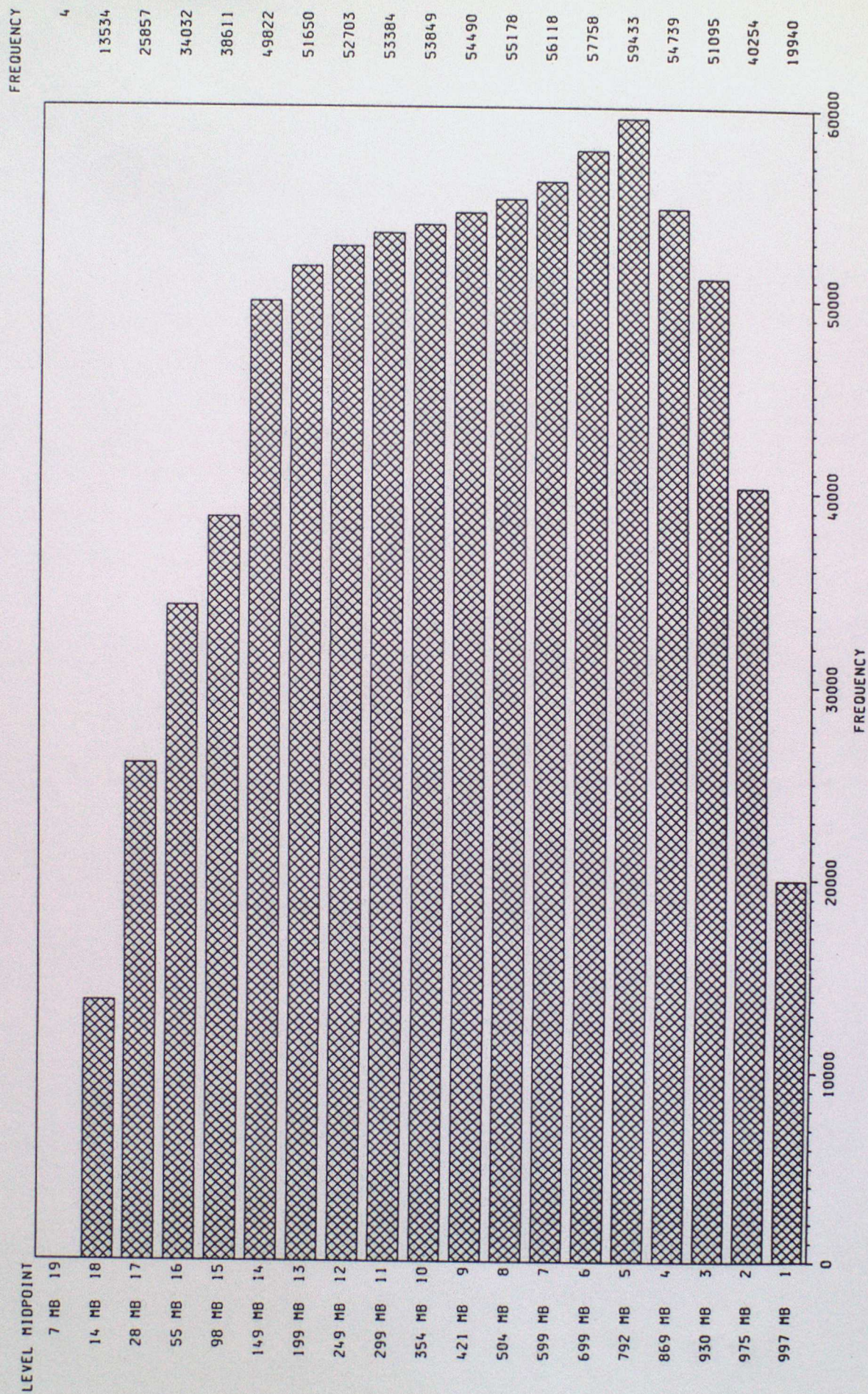


Figure 15.

RADIOSONDES: OCTOBER 1991 THE NOMINAL LAYER PRESSURES CORRESPOND TO P_m = 1000 MB

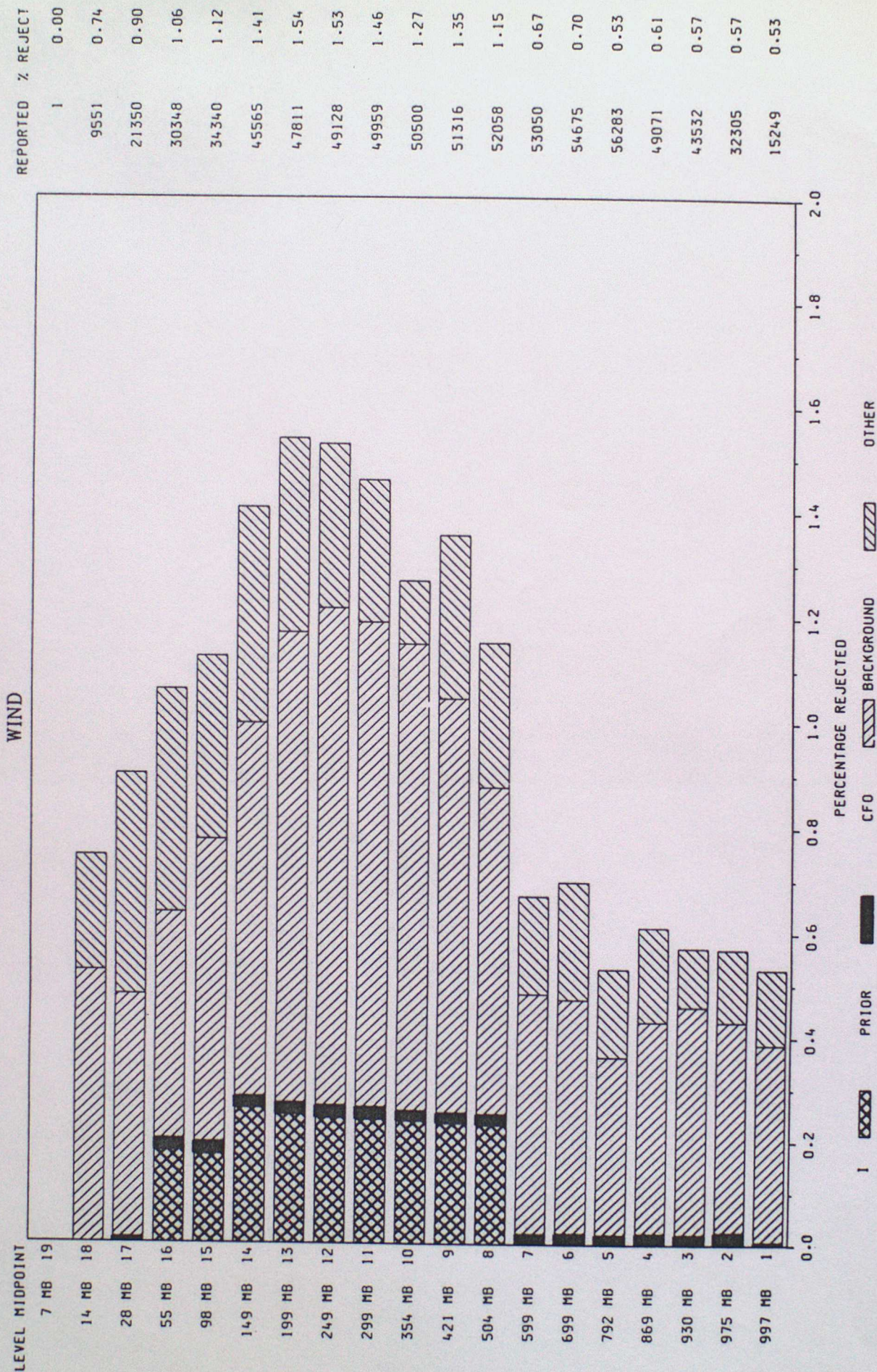


Figure 16.

RADIOSONDES: OCTOBER 1991 THE NOMINAL LAYER PRESSURES CORRESPOND TO $P_{*} = 1000$ MB

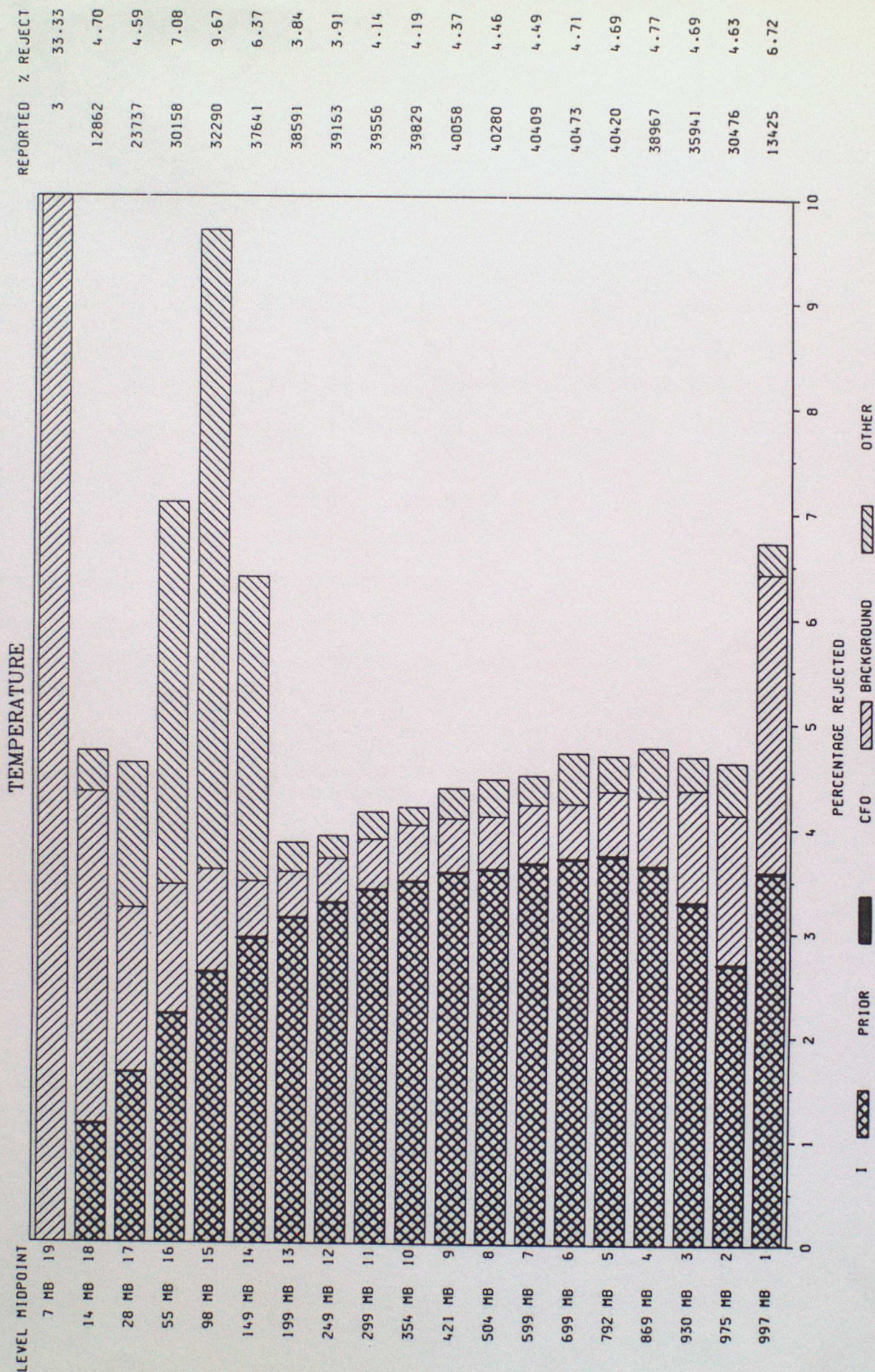


Figure 17.

RADIOSONDES: OCTOBER 1991 THE NOMINAL LAYER PRESSURES CORRESPOND TO P₀ = 1000 MB

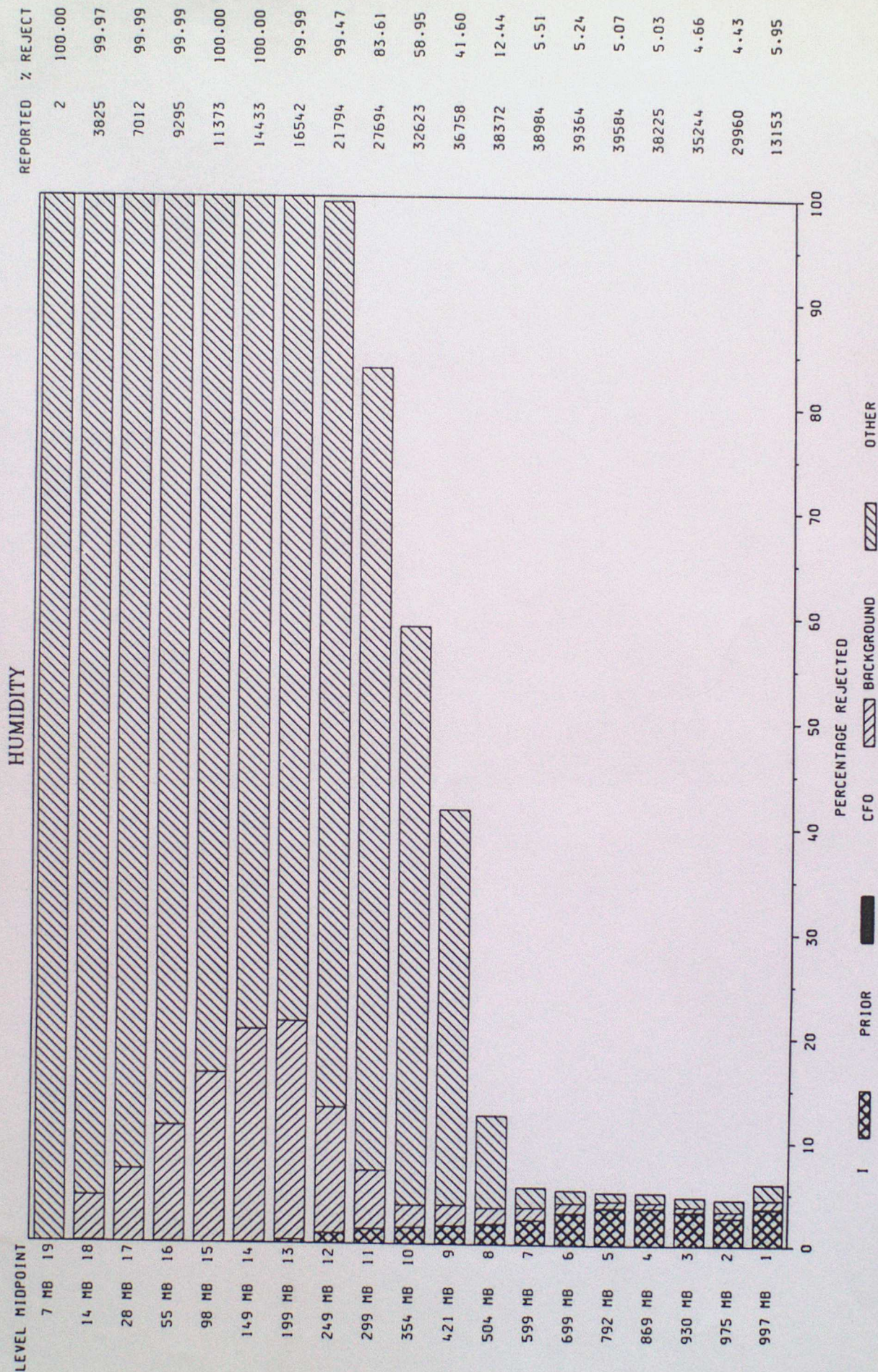


Figure E1.

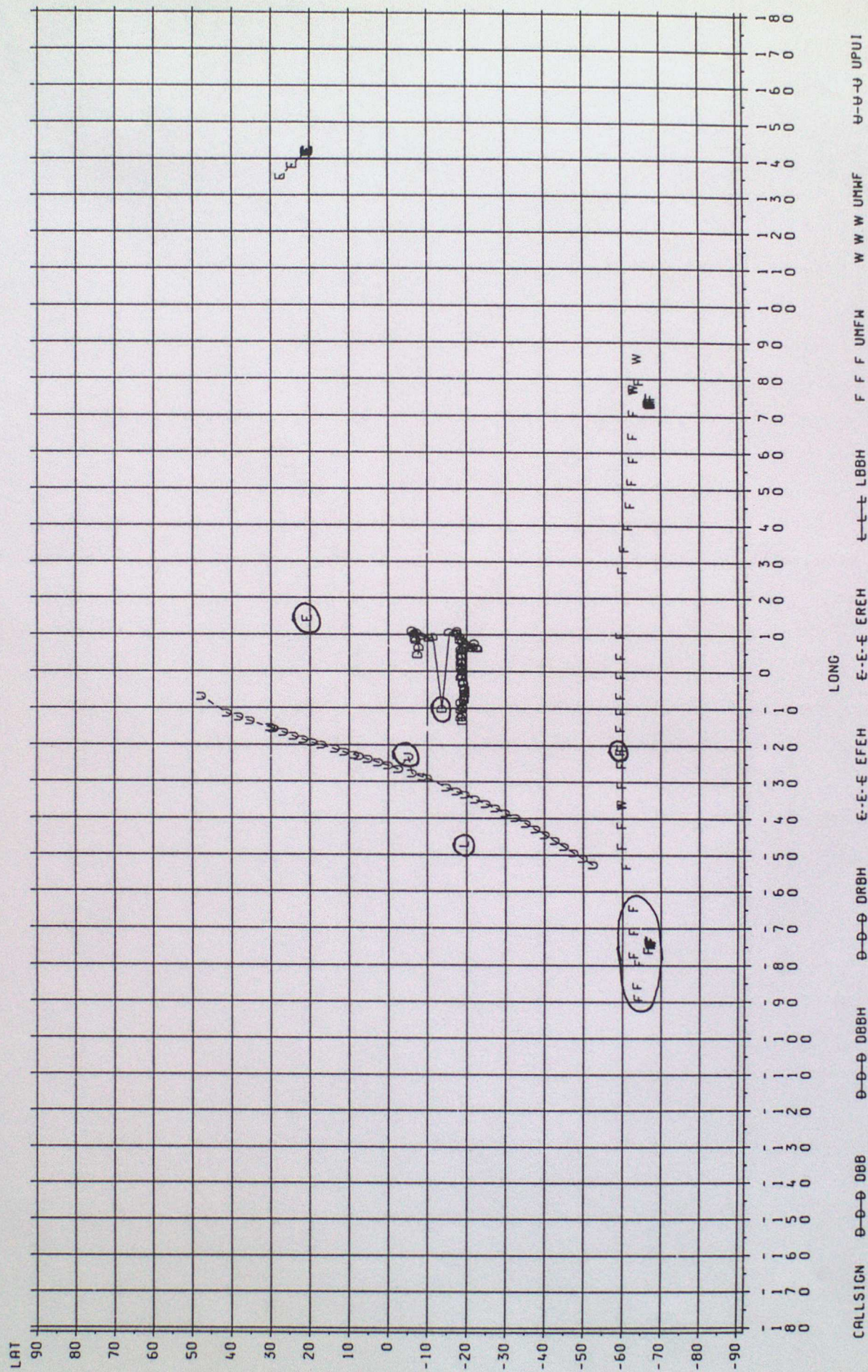


Figure E 2.

TEMPSHIPS WITH POSITION ERRORS: APRIL 1991
POSITION ERRORS ARE CIRCLED

