Forecasting Research

Met O 11 Technical Note No. 10

The Automatic Quality Control of Surface Observations from Ships. The Final Trial, Latest Statistics, Operational Implementation & Future Work

by

C.A. Parrett

May 1988

Meteorological Office (Met O 11)
London Road ... G12 2SZ, England

Met O 11 Technical Note no. 10

THE AUTOMATIC QUALITY CONTROL OF SURFACE OBSERVATIONS FROM SHIPS :

THE FINAL TRIAL, LATEST STATISTICS, OPERATIONAL IMPLEMENTATION AND FUTURE WORK

by

C. A. Parrett

May 1988

Met O 11 (Forecasting Research)
Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ

1. INTRODUCTION

A new scheme for the objective automatic quality control of surface marine data has been developed within the Meteorological Office. The scheme uses Bayesian probability theory to give a probability of gross error [P(G)] for each observed datum, based on (i) historical statistics, (ii) a comparison with forecast first-guess fields (the background check) and (iii) a check against neighbouring observations (the buddy check). If the final P(G) is above a threshold value (50%) the datum is considered to have a gross error and is flagged, otherwise it is accepted as being a good observation [see Lorenc and Hammon (1986) and Barwell and Parrett (1987) for details of the scheme and case study results]. The scheme has been run pseudo-operationally twice daily (on 00GMT and 12GMT global data) since March 1986, with statistics accumulating in an observation processing database (OPD). Some changes were made to the program in July 1987, as described in Barwell and Parrett (1987), and this version was run until 23rd March 1988. Recommendations were made in Barwell and Parrett (1987) for some further changes to be made to the program, and these were tested before being included in a version of the program that was used in an operational trial of the quality control scheme in December 1987. The results from this trial indicated that the new quality control scheme performed better than the existing operational scheme, and also showed some skill at matching the manual intervention performed by forecasters in the Central Forecasting Office (CFO). On this basis permission was given to implement the scheme operationally, to quality control all surface marine data used operationally in the Met. Office's coarse-mesh and fine-mesh models; it was subsequently implemented operationally on 23rd March 1988.

There were two main improvements to the quality control scheme that were included for the operational trial. Firstly, wind and temperature data from coastal SYNOPs were not used in buddy-checks with ships or buoys, and vice versa. Secondly, the scheme was extended to include the quality control of surface BOGUS data produced by CFO. The initial probabilities of gross errors for the BOGUS observations were estimated to be the same as those for ships (5%) for the trial. To obtain estimates for the background and observation errors for the BOGUS data, (observation - background) [(O-B)] differences were obtained from the operational OPD and analysed: the latitudinal dependence of the mean square errors was judged to be similar enough to the basic latitudinal dependency term used for the background error for other observation types, to simply scale the latter by an appropriate factor [3.75 for mean sea level pressure (Pmsl), 5.5 for vector wind and 4.0 for 1.5m temperature] to obtain the background error for each element. That part of the (O-B) difference assumed to be observation error was estimated to be about 1½ times that used for ships.

The results from the operational trial are presented in section 2. Then a noteworthy case is presented in section 3, which stood out when looking through ship reports that had large (O-B) differences in Pmsl. Section 4 contains information on statistics obtained from the twice-daily runs of the quality control program, from July 1987 to February 1988. Section 5 is concerned with the implementation of the scheme operationally, and includes tuning changes deemed necessary following results from the operational trial and analysis of the latest statistics. Finally, in section 6, future work is discussed, including the addition of an automatic blacklist of ships, extension of the scheme to include all land synops, and the use of fields of background errors.

2. THE OPERATIONAL TRIAL

2.1 Introduction

The operational trial of the new ship quality control scheme took place between 9th and 16th December 1987. The scheme was run on data for the 00GMT coarse-mesh update analyses, simulating operational use as closely as possible. The quality control information was incorporated in the data extraction datasets for transmission across the link to the Cyber, and analyses were produced. CFO and SDB flags were correctly passed to the quality control program, and output was successsfully sent across the link and correctly treated by the CONVERT/SORT program. Charts of Pmsl and surface wind for two areas (the data-dense North Atlantic and the data-sparse South Atlantic) were output and compared with corresponding charts for the operational analyses: there were only small differences between the two sets of charts.

In addition to the runs described above, further runs of the program were conducted for the 00GMT and 12GMT update analyses, and output was accumulated in an OPD. For these runs the program was modified so that it ignored CFO intervention flags (but recognised BOGUS data). This enabled an unbiased comparison with CFO's manual quality control to be made. After each run a program collating the quality control decisions with the corresponding final (mode 3) flags from the operational OPD was run, to show how the two schemes compared in the two areas referred to above. Output was sent in near real time to the intervention forecasters, who were invited to comment on observations treated differently by the two schemes and those flagged as 'suspect' by the new scheme. A number of useful comments were obtained.

2.2 Analysis of the comparison presented to CFO

Figures 1 and 2 give a breakdown of all the quality control decisions made during the trial, concentrating on differences between the operational decisions and those made by the new ship quality control scheme. The classification into "correct", "wrong" or "questionable", refers to the subjective remarks made by the intervention forecasters in CFO as they assessed the computer printouts. It does not refer to any objective criterion for deciding which observations should be passed to the analysis.

The results were also looked at in detail by a member of the Assessment Group in Met O 11 (with the help of additional plotted charts) to gain an independent opinion of the performance of the new quality control scheme.

7% of the total number of observations (both Pmsl and wind) were treated differently by the two schemes and 2.25% were treated 'wrongly' by the new scheme (an average of only 5 observations each data time) compared with 2.6% for the operational scheme. 'Suspect' observations (i.e. those with a probability of gross error of 30-50% in the new scheme) amounted to about 1½% of pressures and 3% of winds: the ability of the new scheme to bring these to the attention of CFO forecasters was seen as a valuable feature. The new scheme flagged over 2/3 of the observations flagged by CFO and showed some skill in detecting bad or borderline ones which they had missed.

For pressure observations, the new scheme performed better than the operational automatic scheme, and also compared well with CFO's flags in most cases. The problem of excessive data flagging due to large numbers of buddy checks (Barwell and Parrett, 1987) was alleviated near coasts with the use of station lists in data-dense areas. Although there were a few cases (in the N.E. Atlantic and North Sea) of poor quality control decisions, which seemed to be the result of the observations having too many buddy checks. The maximum number of buddy checks per ship was 20 for the trial, but since this limit applies to both buddies in each pairwise buddy check, there were often >30 buddies for observations in data dense areas.

About 8% of wind observations were flagged by the new scheme (about 4/5 on the buddy check), compared with about 4½% operationally (most of these by CFO intervention). Although about a quarter of the 8% were 'wrongly' flagged, CFO comments suggested that many of these were near sharp frontal troughs or were influenced by coastal effects. A forecaster can readily take these local effects into account but a numerical model would have difficulty in resolving them. Thus, although indicated as "wrongly flagged" in Figures 1 and 2, it may be that flagging of some of these observations was appropriate as far as the computer analysis was concerned. Most of the remainder were flagged on the buddy check and many of these seemed to have too many buddies.

Some pressure observations were also 'wrongly' flagged near small-scale features, particularly where there were positional errors in the background field (e.g. south of Ireland at 12GMT on 15/12/87 - see section 2.2(ii)).

A noticeable feature of the new scheme is its ability to assign an error probability for each element of an observation. Thus it can identify a good wind report while flagging a bad pressure or vice versa. In CFO, the whole observation is often flagged (this being easier than flagging one element), with the consequent loss of some good data. A number of instances were noticed where this had happened. Also, the new program detected the wrong treatment of drifting buoy 25892, which was on CFO's permanent rejection list but was found to be reporting good quality data (see Fig. 2(a)).

2.3 Examples from the trial

The following cases show a few of the more interesting examples of different quality control decisions made by the two schemes during the trial:

(i) Figure 3 shows the surface observations and 6 hour forecast (T+6) Pmsl field, used as the background for the quality control schemes, near to Northern Scotland at 00GMT on 13th December. Five Pmsl observations (those underlined) were flagged by the new scheme but passed operationally. With the close proximity of good observations over the land it appears that all 5 reports were too low by about 2 to 3mb. Thus, the new scheme did well to flag them, despite the fact that the 5 observations would have supported each other in buddy checks. In a re-run of the new scheme that included the tuning detailed in section 5.1, four of the five were still flagged, but one just passed the quality control (although it would have been classed as 'suspect', with a final P(G)=45%).

(ii) At 12GMT on 15th December, the new quality control scheme flagged 8 ship Pmsl reports near to a frontal trough approaching the British Isles from the south-west (see Figure 4). CFO commented that 4 ships were 'wrongly' flagged (FHDM, FPPF, SLCH and EOIE) and 3 were 'correctly' flagged (URYX, VRLE and YTEQ), with ship URAE "possibly 2mb too low, but OK to leave in". The only ship flagged operationally in this region was YTEQ, which was flagged by CFO. The excessive flagging by the new scheme here, seems to have been caused largely by too many buddy checks on the ships concerned and also by the nearness of the trough. A re-run of the new scheme for this case, incorporating the tuning detailed in section 5.1, which included a lower limit on the number of buddies allowed per ship, resulted in two of the 4 ships 'wrongly' flagged being passed (SLCH and FPPF) and in reduced final probabilities of gross error for the other two such that they were only just flagged; ship URAE was also passed by the quality control. The new scheme correctly flagged SLCH's wind report on both runs and also that of ship ESRT (also flagged operationally). (URYX's Pmsl report was wrongly passed by the quality control on the re-run, because of a duplicate report (with a different call-sign) from the ship to the north-west (reporting a Pmsl of 978.8mb), which had not been received for the trial run.)

The area near to Portugal in Figure 4 is also of interest, with a trailing cold front lying close to Iberia and a lot of ships reporting in the area. Operationally, 4 Pmsl reports were flagged: ships FNNC, V2DG and Y5LQ were flagged by CFO and ULDB was flagged mode 2 - all 4 were flagged by the new scheme. The new scheme also flagged GWAP's Pmsl 'wrongly', due to excessive buddy checking, but it was passed on the re-run mentioned above. There was one worse decision (according to CFO) in the re-run, which was the passing of V2DG's Pmsl; but this decision was not surprising considering the Pmsl reports close to it with similar (O-B) differences. Also, ships ESET and UYOT were flagged by the new scheme (on the re-run as well); there was no comment on ESET, but CFO considered UYOT's Pmsl to be "OK - near a front". Nevertheless, both observations' Pmsl reports were somewhat low compared with nearby observations and both were flagged on the buddy check. The new scheme also correctly flagged ESWB's wind report, which was flagged operationally.

Thus, this example shows that the new scheme tended to flag Pmsl reports rather too readily in data dense areas, particularly where small-scale synoptic features were present; but that this tendency is improved by including the tuning changes listed in section 5.1, particularly reducing the maximum number of buddy checks for each ship and also reducing the background error correlation length scale.

(iii) Figure 5 shows the background Pmsl field and surface observations for 12GMT on 15th December in the South Atlantic: a data-sparse area where the background field was in error. Such a situation is known to be a difficult one for the new scheme, which did slightly worse than the operational scheme here by flagging ship URPD (this ship was passed operationally). As can be seen from Figure 5, both ships URPD and UOLF were reporting Pmsl's more than 15mb lower than the background, and they were flagged by the Synoptic Data Bank (SDB). After also being flagged against the background in the new scheme, they were not re-instated by the 5 buddies (including the two BOGUS obs). This may be a good argument for reducing the initial probability of gross error for this SDB flag on Pmsl, since an observation is only flagged by the SDB if its Pmsl is more than 15mb different from the background, in which case it will certainly be flagged on the background check anyway. If its initial probability of gross error was lower, then supporting neighbouring observations might be able to reinstate it on the buddy check; and in a re-run of the new scheme with values reduced from 95% to 75% ship URPD was reinstated. Although, in the re-run with the tuning changes listed in section 5.1 (initial P(G)=90%), URPD was still flagged. It was decided not to reduce this initial P(G) below 90% operationally, without further tests to determine whether such a reduction would lead to the passing of a few bad observations (although this seems unlikely with a value of 75%). The BOGUS wind at 55S 39W was also flagged in the trial, but in the re-run with lower initial probabilities of gross error for BOGUS observations (see section 2.3), it was passed.

2.4 Treatment of CFO BOGUS observations during the trial

A few BOGUS observations were flagged by the new scheme (3 pressures and 6 winds, out of a total of 60), mostly in the South Atlantic. All of these cases were investigated in detail to see if some tuning of the new scheme was necessary to handle BOGUS data more accurately.

At 12GMT on 11th December a BOGUS observation inserted at 53N 28W was flagged by the new scheme after 17 buddy checks, of which 15 were with multiple reports from two drifting buoys (whose Pmsl reports differed by about 10mb). The new ship quality control scheme used all observations within the time-window of +/- 3 hours and all off-time reports from drifting buoys (DRIBUs) were treated as separate observations (except that they did not buddy check each other). This is obviously undesirable, since if a DRIBU has a bias in its Pmsl reports, these reports would support each other in buddy checking other ships, possibly causing wrong flagging of these ships. In the 11th December case both buoys' reports looked correct; they were simply about 450km apart in a strong pressure-gradient, and the buoy furthest from the BOGUS observation had too much influence on it, with 8 reports in the 6 hour window all buddy checking it (although after 5 buddy checks with this buoy the BOGUS observation was still acceptable). At the moment, only the DRIBU report nearest to the analysis time is used in the analysis, with the selection of these reports taking place within the Convert/Sort stage, which is after the ship quality control (see Barwell (1987) for details of where the ship quality control fits into the operational system). If this selection were to be brought forward to be before the quality control, it would remove the problem highlighted here; but all of the multiple reports should be quality controlled for verification and monitoring purposes, and in the future the Analysis Correction scheme will be able to use them all, in which case the selection procedure will need to be removed anyway. We have alleviated the problem by only updating the probabilities of gross errors of the drifting buoy itself in buddy checks if its report is more than 1 hour from the data time. However, this may result in a few single off-time reports not being used in buddy checking other observations. Also, in order to treat off-time drifting buoy reports correctly, their influence should be reduced as a function of their time difference from the analysis time, and their associated background error should be increased somewhat.

Another BOGUS observation was flagged by the new scheme at 00GMT on 13th December (see Figure 6), where an attempt was made by CFO to move and fill a depression in the South Atlantic (near Gough Island). The Pmsl and wind from the

BOGUS observation (labelled number 4 in Figure 6) were both flagged on the background check alone; the other BOGUS observations and the synop report (Gough Island) were just outside the buddy check range (350km). The BOGUS observation's Pmsl was 5.1mb higher than the background and the vector wind was 35kts different (including being in the opposite direction!), but it is still undesirable to have flagged this BOGUS observation.

The final BOGUS Pmsl flagged was again in the Southern Hemisphere and was flagged on the background check alone, being 7.1mb different from the background. Five out of the 6 BOGUS wind observations flagged were also due to the background check, having either none or just a few buddies (most of which were on the opposite sides of pressure troughs and therefore had conflicting winds).

To overcome the problem of flagging BOGUS observations too readily, their initial P(G)'s were reduced prior to operational implementation from those assumed for ordinary ships (about 5%) to those assumed for Ocean Weather Ships (about ½%). Re-runs with this change included improved these quality control decisions, with the BOGUS observation in Figure 6 (number 4) being passed (both Pmsl and wind), along with the BOGUS wind previously flagged in Figure 5.

3. A NOTEWORTHY CASE - 00 GMT 18th NOVEMBER 1987

This is another case where the T+6 forecast Pmsl field used as the background was in error (by almost 20mb here); this time close to a deep depression situated in the North Pacific, near the Kamchatka peninsula (see Figure 7). Sixteen ship reports of Pmsl on the northern side of the low were flagged operationally, because they were on average about 15mb higher than the background. The new scheme flagged them all on the background check, but because they supported each other and agreed with land synop reports on Kamchatka and islands to the south, 14 were reinstated on the buddy check. A subjective analysis has been superimposed on the background field in Figure 7, drawing to the land synops and majority of ships, and it shows that the new scheme was correct in almost every case to pass these 14 observations.

The two ship Pmsl reports flagged by both the new scheme and the operational system (3FJR2 and 3EOU) had been flagged by the SDB, and their wind reports were also flagged by the new scheme; whereas two other ships given SDB Pmsl flags (A8FW and DUFC) were reinstated, largely because their wind reports were acceptable (and maybe also because of the order of their buddy checks).

Several ships on the western flank of the low were flagged by the new scheme but passed operationally. JHJS's Pmsl agreed with the background, but was flagged after buddy checks with other observations disagreeing with the background (e.g. EWZC). UUQN and UFCS were marginal decisions, but the new scheme was probably correct to flag them; and it was definitely correct to flag EUEM. Ship ELET8 on the south-west side of the low was also flagged operationally, but was passed by the new scheme after 3 buddy checks; although its wind report was flagged by both schemes. The strong wind reported by ship UFCS was flagged only by the new scheme (P(G)=97%), and 3 more wind reports

(7JMF, ZCKA and D5EE) were only just flagged by the new scheme (P(G)=58%, 55% and 65%, respectively).

This is clearly a major success for the new quality control scheme over the operational system, and operational use of the new scheme would almost certainly have lead to a substantial improvement in the analysis in this region.

4. LATEST O.P.D. STATISTICS

Statistics from the Ship Quality Control OPD have been analysed for the period from October 1987 (when an error in the data extraction was corrected) to February 1988, in order to feed back information to the quality control scheme, and tune it where necessary. Care is needed when interpreting these long-term statistics: for example, to avoid unrealistic positive feedbacks between changes to the values chosen for use in the program and consequent changes in the statistics produced by the program (e.g. lowering the initial probability of gross error may lead to more observations having a lower final probability of gross error, and analysis of the statistics would then imply that the initial probability should be further reduced - or vice versa). Sometimes case studies can highlight necessary changes hidden by the statistics, e.g. the average final probability of gross error for all Pmsl observations flagged by the SDB was close to 100%, which has been shown by studying individual cases (see section 2.2(iii)) to be too high, since some of these observations are good and should not be flagged, which implies that the initial probability is set too high (not too low, as implied by the statistics alone).

Most of the initial probabilities of gross error in use until March 1988 seemed to be valid, although for observations that had failed the SDB pressure-tendency test the values seemed rather high, which confirmed a finding from the trial that these observations were too readily flagged (see Figure 2(b)). The error variances were also little changed from previous statistics, except those for drifting buoys which seemed to be too high (reports from drifting buoys seemed subjectively to have improved over the previous year or so). Thus a little tuning of these parameters should prove beneficial, and this was done before operational implementation (see Table 1 for values of these parameters in use operationally).

Histograms of (O-B) differences of Pmsl, 1000mb vector wind and 1.5m temperature have been plotted for 6 months data (July 1987 to January 1988). Figure 8(a) shows the data passed by the quality control scheme and confirms the approximately Gaussian distribution of the differences; Gaussian distribution curves with the same mean and standard deviation as the (O-B) differences have also been plotted (dashed lines). The Pmsl and temperature distributions are very close to being Gaussian, with means of 0.1mb and 0.6°C and standard deviations of 1.84mb and 2.6°C, respectively. The histogram of the modulus of vector wind difference shows a minimum near zero, as expected, indicating that the observed wind rarely agrees exactly with the background value, but above about 3 m/s the distribution is still not far from Gaussian. (Since all vector wind differences are positive, the comparable Gaussian curve shown is only the positive half of a Gaussian distribution calculated assuming symmetry about zero.) If we assume that the distribution for the vector wind differences is Gaussian in a two-dimensional frame, then when the modulus of this is taken, it has a shape similar to the actual distribution shown in Figure 8(a) (ii), with a dip near zero.

Histograms for data flagged by the quality control scheme are shown in Figure 8(b), along with the top-hat distributions (dashed lines) assumed for the representation of gross errors. There is some disagreement between the two sets, but, as with the other statistics, care needs to be taken before changing the distributions assumed in the program, so as to prevent too much positive feedback producing unrealistically high or low equilibrium values. The (O-B) distribution for Pmsl is somewhat above that assumed (for |O-B| < 5mb), but since the new scheme has no marked tendency towards flagging too few observations we are reluctant to increase the height of the top-hat (also, there may be a slight seasonal dependency). The (O-B) distribution for vector wind does seem to be well below the assumed value (0.086 s/m), which may be one reason why the ship quality control scheme tended to flag wind observations too readily. Consequently, the top-hat distribution was changed to that shown as a dotted line (0.060 s/m) prior to operational implementation. The dip near zero is accentuated for the vector wind because of the one-dimensionality of the diagram (as for the observations passed by the quality control scheme); it is planned to change the assumed top-hat distribution in the future to take account of this minimum (i.e. to a two-dimensional top-hat), and this will also be accompanied by a similar change to the distribution assumed for the observations passed by the quality control scheme, since the ratio of the two is important in calculating the final probability of gross error.

It was found that the distributions of (O-B) differences for Pmsl were little changed in the tropics compared with the extra-tropics, so an earlier assumption that the distribution should be narrower in the tropics has been removed. Distributions have also been obtained for wind components parallel and perpendicular to the observed wind direction, to see if there was any bias in either component of the difference - there was no noticeable bias in either direction.

5. OPERATIONAL IMPLEMENTATION

Before the program was implemented operationally (on 23rd March 1988), it was tested on data for the fine mesh model, following the availability of the necessary background fields on the fine-mesh printfiles. Fine-mesh quality control runs were compared with coarse-mesh runs for the same data times, and generally there seemed to be only small differences in the final probabilities of gross error; although, for a few (usually 2 or 3) observations at each data time different quality control decisions were reached. These differences were due to the different background fields (and assumed background errors) being used, or to different data being used in the buddy checks (due to the different data cut-off times and station lists, and observations near the fine-mesh boundaries having fewer buddies).

In operational implementation, the output from the new ship quality control program is written initially to a temporary dataset and then copied so as to overwrite the old data extraction datasets. If a program failure occurs, the latter datasets are therefore not overwritten and the new program is simply bypassed; quality control is then automatically passed over to the previous operational scheme. Although BOGUS data are used in the quality control scheme, the flags set (if any) are not passed to the analysis at present, so quality control is carried out by the old scheme. This omission from the new scheme should have little impact, since very few BOGUS observations should need to be flagged anyway.

The quality control program is run at the end of the data extraction jobs, before the individual type datasets are merged for passing to the Cyber/Eta. The latest T+6 forecast is used as the background for each coarse-mesh run. For the two main fine-mesh runs, T+3 forecast fields from the final assimilation cycle are used. For most other fine-mesh runs data is extracted for 2 or 3 data times together, so that the latest fine-mesh background fields available are those from either the previous main run or intermediate run. The 06GMT intermediate (1I) run uses background fields from the 00GMT main (OF) run: T+3 for 03GMT and T+6 for 06GMT data times. The 1I run provides T+3 background fields for data time 09GMT; while, for data times 15, 18 and 21 GMT, background fields from the 12GMT main (2F) run (valid at T+3, T+6 and T+9, respectively) are used. If the relevant printfile is unavailable (e.g. if a forecast has failed), then the quality control program will automatically use the latest printfile available, valid at the appropriate data time. If all fine-mesh printfiles out to T+24 are unavailable, the quality control for the main and intermediate hour fine-mesh runs will use the latest available coarse-mesh printfile. If all coarse-mesh printfiles out to T+60 are unavailable (very unlikely), the quality control for both the coarse-mesh and fine-mesh runs will use the latest available Washington printfile (out to T+60), or the latest coarse-mesh printfile out to T+144 for data times 00GMT and 12GMT.

Until we have accumulated some statistics for background errors for the fine-mesh quality control runs, the background errors are being calculated using the coefficients in use previously for the coarse-mesh T+6 background fields, but then a suitable scaling is applied. This scaling takes account of the differences in errors between the coarse-mesh and fine-mesh forecasts and also the tendency for errors to grow through the forecast period. Verification statistics were obtained from Met O 2b for the period from July to December 1987, from which the scaling factors were obtained for forecasts valid out to T+48 for fine-mesh runs, and to T+144 for coarse-mesh runs, for Pmsl, surface wind and 1.5m temperature; the RMS errors for all three variables were found to increase approximately linearly through the forecast period (see table 2). (The values used for the coarse-mesh forecasts are also to be used for the Washington forecasts, if they are needed.)

The probabilities of gross error will be written to the operational OPD when appropriate changes have been made to the OPD format, and in addition more detailed output (including buddies and error variances used, etc., which will be useful for any future case studies) is being written to one of two ship OPD's, which are then archived daily to Met O 11 tape datasets. The format of these ship OPD's is the same as that for the one which was maintained pseudo-operationally for 00GMT and 12GMT coarse-mesh main runs from July 1987 to March 1988. As with the operational OPD, information is written to a coarse-mesh ship OPD from the quality control runs for the update analyses at 00GMT and

12GMT and the secondary analyses at 06GMT and 18GMT, and to a fine-mesh ship OPD from the main runs at 00GMT and 12GMT and the 2 assimilation cycles (1I for 03GMT data, 1F for data times 06 and 09 GMT, and 3F for data times 15, 18 and 21 GMT). Four 6250 bpi tapes are being used to archive the ship OPD's (including 2 as backup copies); each tape should take about 5 months of data.

5.1 Tuning included prior to operational implementation

Following the results from the operational trial and analysis of the latest statistics from the ship OPD, the following tuning changes were incorporated into the trial version of the quality control scheme, prior to it being implemented operationally :

(a) The initial P(G)'s for BOGUS observations were reduced to ½% (to be the same as those for OWS's).

(b) There was a slight reduction in the initial P(G)'s for Pmsl observations with SDB flags (see Table 1 for the latest values).

(c) The maximum number of buddy checks allowed for both observations in each pairwise buddy check was reduced from 20 to 12.

(d) The number of range scans for buddy checking was increased from 3 to 4, to complement (c) and use nearer buddies first.

(e) The correlation length scales were reduced from 390/300km to 275/200km for Pmsl/wind, to be consistent with values used in the Analysis Correction scheme.

(f) For drifting buoys there was a slight reduction in observation error variances (for Pmsl and temperature), and in initial P(G) in location, to be closer to those for ships (see Table 1).

(g) The decrease in variances for Pmsl observations in the Tropics was removed.

(h) The top-hat distribution for gross errors in vector wind was made somewhat broader [as shown by the dotted line in Figure 8 (b) (ii)].

(i) Drifting buoy reports off-time by more than 1 hour are no longer allowed to buddy check other observations, although they are still buddy-checked themselves.

N.B. Re-runs of the quality control program with the above changes included showed an overall improvement in quality control decisions, with very few cases of worse decisions having been found.

6. FUTURE WORK

6.1 Extension to include all land synops

The extension of the ship quality control program to include the quality control of all land synops is a natural progression. The program has already been run with all synops, but using the error characteristics and initial P(G)'s assigned to coastal synops at the moment; there is obviously a need to obtain statistics from the OPD that apply to all land synops. There may also be a need to change the background error characteristics for land stations to incorporate a dependence on station height and possibly use a different latitudinal dependence, although the synoptic dependence may still be valid. Since land stations have fixed positions (stored in a look-up table dataset), there should be no problems with positional errors (unless there are errors in the look-up table), although there will probably be the occasional corrupt call-sign.

Many synops (about half of those extracted) report station level pressure as well as (or instead of) Pmsl, and at the moment the data extraction takes the station level pressure in preference to Pmsl. Consequently, values of Pmsl will need to be calculated within the quality control program, to compare with the background Pmsl fields, for which the various coefficients in use in the program have been calculated. Alternatively, a surface pressure field could be used as the background, but the pressure at the model surface height would need to be calculated from the station height pressure (or from Pmsl for ships near coastlines where the model surface is above sea level), and new coefficients would need to be calculated for the background errors. The latter alternative may be preferable in the long term to calculating Pmsl from surface pressure for land stations, since the height differences involved in the interpolation/extrapolation should be much less on the whole than those involved in the extrapolation to Pmsl over land.

High altitude synops that report a standard pressure level height, instead of Pmsl or station level pressure, will require these additional height fields (850mb, 700mb and possibly 500mb) as background fields. Statistics on background errors for these fields will need to be obtained from the OPD, and appropriate error variances assigned.

6.2 Inclusion of an automatic blacklist of ships

A blacklist of 'rogue' ships (i.e. those with consistently bad and/or biased reports) has been shown to be beneficial at ECMWF (Illari, 1983), and the need for such a list with the new quality control scheme was shown in Barwell and Parrett (1987). The data stored in the OPD can be used initially to obtain recent historical statistics on individual ships, and so identify those with consistently large root mean square (RMS) errors and/or large mean errors (biases). To be consistent with the new quality control scheme, these statistics should be based on (observation-background) differences, rather than (observation-analysis) differences. It may be necessary to take into account the geographical variation of model background error, so as not to exclude ships from the analysis just because they happen to be in an area where the background field commonly has large errors (which is where they are most needed!). Statistics on the final probabilities of gross errors may be useful in this respect.

Once an initial statistical dataset of ships has been created, it can be stored as a separate on-line dataset to be updated after each run (so that the OPD would not need to be used operationally). The list could be kept in alphanumeric order for ease of updating; for which purpose a new alphanumeric list could be created and merged with the pre-existing one (so that any new ships could be included). The old ship's statistics would need updating, even if the ship did not report at the time in question. Use of a weighting function (w_i) to give most weight to the most recent observation from each ship, would ensure that ships that had not reported for a long time (6 weeks, say?) could be removed from the dataset, so preventing it from becoming too large.

In order to calculate the necessary RMS and a weighted bias for each ship, it can be shown that we need to store just 3 statistics : $\sum w_i$, $\sum w_i x_i$, and $\sum w_i x_i^2$, where x_i = (O-B) for observation i. There would also be a need to store a parameter indicating whether or not to include each ship (and each element) on the blacklist - a 'confidence indicator'. This confidence indicator could also be used to produce a 'white-list' of reliable ships (possibly useful for verification purposes), or to put each ship on a graded black-white list. There is also the possibility of using the confidence indicator directly within the quality control program to amend the initial P(G)'s for each ship, i.e. for bad ships setting the initial P(G) close to 100%, for good ships setting it close to 0%, and for in-between ships either leaving their P(G)'s unchanged or grading them appropriately; but doing this would cause additional complications (e.g. in producing 'clean' statistics), and initially it may be best to keep the blacklist separate from the quality control program. There may also be a need to store a running mean of the final probabilities of gross errors.

Obviously, ships that have been blacklisted and excluded from the analyses, but then start to send in good quality reports, should be reinstated as soon as possible; although there should be no need to try and re-instate a blacklisted ship just because its latest report seems to be correct, since it would need to agree with several buddies (and probably also the background) before any confidence could be placed in it, and in this case the observation would almost certainly be redundant. Experiments will need to be carried out to determine how rapidly to reinstate such observations.

It should be possible to treat separately each element reported; so that at the moment, where only Pmsl and wind are used by the analysis, either could be blacklisted alone. This would entail making further changes to the quality control program, so that the correlation between elements in the background check could be removed in this case, as well as buddy-checking with the blacklisted element. It may be best not to include common-named reports (e.g. 'SHIP', 'RIGG', 'PLAT', 'BUOY', blank) on the blacklist, since some good quality reports could possibly be blacklisted amongst them; although, if we were to store positions for 'PLAT', 'RIGG' and moored buoy reports, then these could be included as individual stations.

In conjunction with the blacklist, CFO may wish to have an option to display (either graphically or by report) time-series of reports from any named observation (to include, for example, call-sign, time of report, position, observation values, (O-B) values, final P(G)'s and possibly (O-A) values). Such information could be obtained from the operational on-line OPD for the last 16 days (see Smith and Ashcroft (1988) for details about the OPD), or from a separate dataset containing the necessary information for just the blacklisted ships (for the last month, say).

Use of the OPD enables us to identify those ships with a consistent bias, and this information could be passed (automatically?) to Met O 1 (and possibly on to WMO) to notify those concerned and thus remove the bias. These ships with a consistent bias could also be corrected automatically (either as part of the blacklist program or separately), although some daily monitoring would be desirable to check that any bias had not been corrected at source. To decide on whether to apply a bias correction to an observation of Pmsl (or temperature, etc.), a test could be made on the value of $(RMS^2-bias^2)^{1/2}/RMS$, to see if it was below a critical value (to be determined by experiment). A further test could be carried out on the percentage of times that an observation fell below this critical value, to prevent an individual extreme report 'biasing' the decision. It would be more difficult to correct biases in wind reports, although it might be possible to apply corrections independently to the wind direction and wind strength, if statistics were obtained for these variables.

There are several aspects of the blacklist that are still to be decided, but the above ideas are a basis on which to carry out some preliminary experiments before finalising our plans. (The automatic correction of biases will not be attempted until after an automatic blacklist has been set up.)

6.3 Incorporating fields of background errors

At the moment, the ship quality control program calculates a value for the background error variance at each observation position (using a combination of a synoptically varying term and a latitudinally dependent term). In Barwell and Parrett (1987), it was suggested that it would be useful to have these error variances in the form of fields on the model grid. This would have several advantages :

(a) It would allow the calculation of an 'envelope' of maximum values over a reduced number of grid points, which would remove unrealistic minima in the fields near the centres of depressions (for example).

(b) A similar (but less potent) effect to (a) could be produced by just smoothing the fields.

(c) The fields of background errors could be displayed on a screen in CFO, and possibly modified interactively so as to allow for a very poor background forecast in certain areas (e.g. near Kamchatka on 18/11/87).

(d) There would be the possibility of including a ∇^2 term in the calculation of background errors (instead of (a) or (b)?), which is rather expensive to do for individual observations.

(e) It may be possible to include a value for the density of observations (since this is needed for the Analysis Correction scheme) in the calculation of the background error, to take account of the fact that the analyses and therefore the forecasts are more likely to be in error in data-sparse areas.

6.4 Other work on the Ship Quality Control program

(a) Improvements to the buddy checking procedure :

At the moment, the buddy checking is done in range scans, with all observations within each range being buddy-checked in the order in which they are extracted. It may be better to firstly find all the buddies (up to a maximum of 15, say) within a somewhat larger range, dropping-off the furthest away one each time a closer one is found. They could then be buddy-checked in order of increasing distance from the observation concerned (in range scans still), and each observation could have a limit on the number of buddies it should have had at a certain range, beyond which no more buddy-checking need be done; this would allow for a larger effective buddy check range in data-sparse areas compared with data-dense areas.

(b) Improvement of vector wind errors :

The correct error distributions for vector wind are a two-dimensional Gaussian for the observations passed by the quality control and a two-dimensional top-hat for flagged observations. These distributions should replace the ones used at present, which are based on wind components. The formula for the error correlation used for vector wind is also an approximation at present, being the same as that used for Pmsl (with a shorter length scale); to be correct, and consistent with the Analysis Correction scheme, the formula used for the correlation for streamfunction (rather than wind) should be the same as that used for Pmsl, so that the correlation can be different in directions parallel and perpendicular to the wind direction (i.e. greater in the along-stream direction and less in the cross-stream direction) (see Lorenc and Dumelow, 1985).

(c) Improvement of the treatment of off-time drifting buoy reports:

As mentioned in section 2.3, the influence of off-time drifting buoy reports should be reduced as a function of their time-difference from the analysis time; and the background error associated with off-time reports should be increased, to allow for changing synoptic patterns.

(d) Inclusion of preliminary quality control runs for CFO, to aid intervention. (To be set up by Met O 2.)

(e) Further tuning:

Continued monitoring of OPD statistics, to check the performance of the scheme and coefficients in use, will probably lead to further tuning being necessary in the future, particularly for the fine-mesh (e.g. the calculation of separate regression coefficients for background errors from fine-mesh OPD data).

References

Barwell, B R, 1987 : 'The incorporation of Bayesian quality control techniques into the operational suite.' *Met O 11 Working Paper no. 88.*

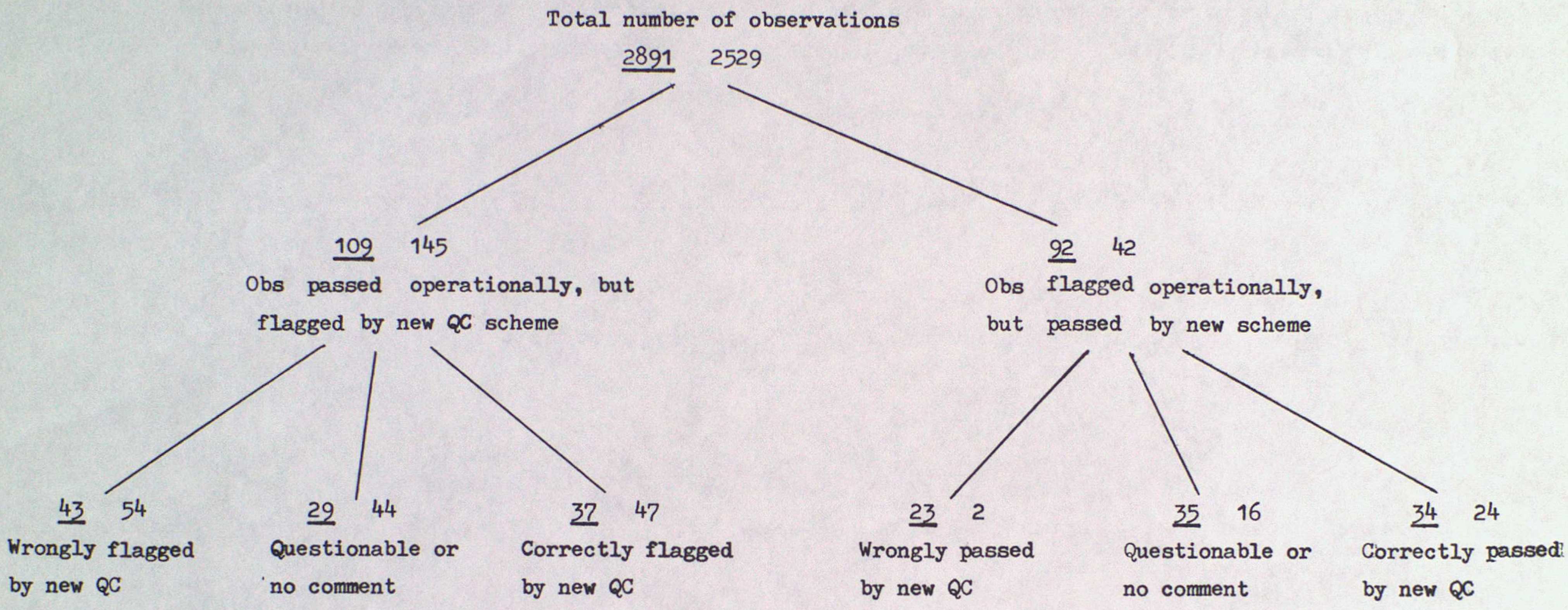
Barwell, B R and Parrett, C A, 1987: 'Modifications to the automatic quality control of ship data and an assessment using case studies.' *Met O 11 Tech. Note no. 261.*

Illari, L, 1983 : 'Automatic blacklist of ships.' *ECMWF Research Dept. Memorandum R2314.*

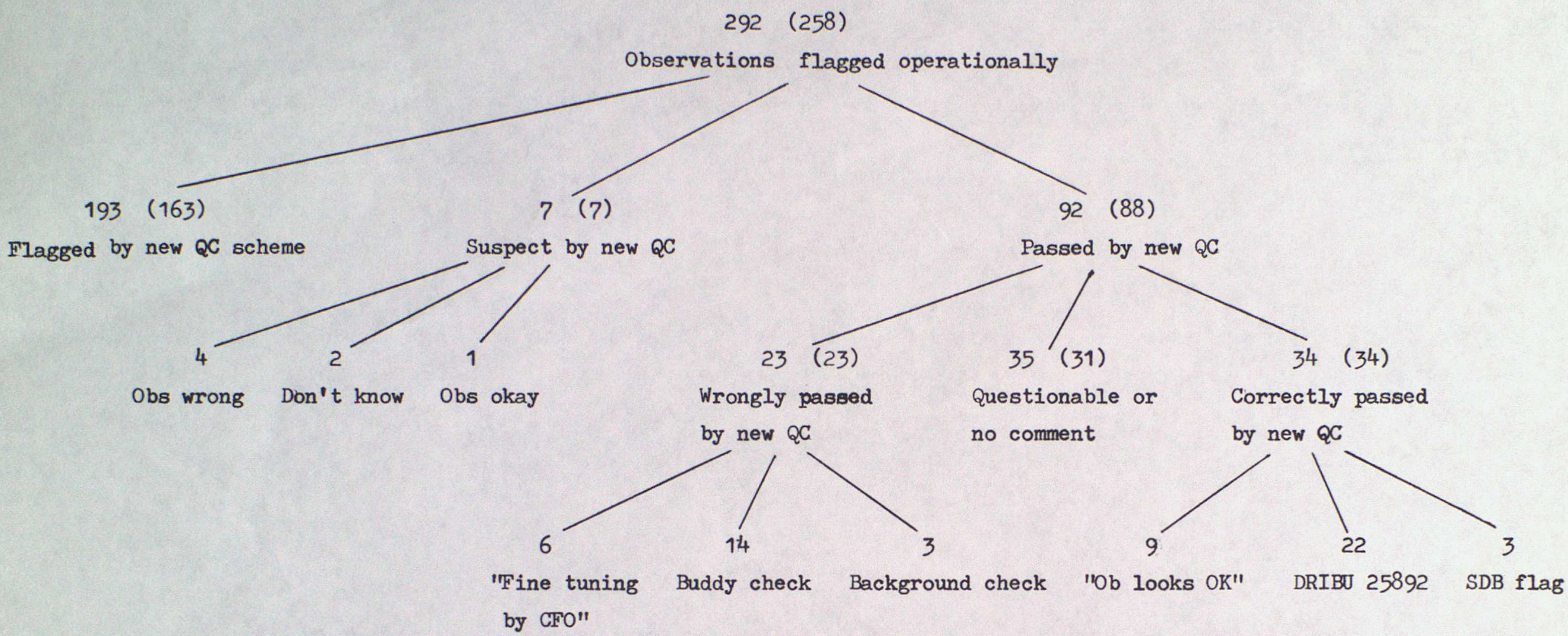
Lorenc, A C and Dumelow, R, 1985 : 'Four-dimensional analysis by repeated insertion of observations into a NWP model.' *Met O 11 Technical Note no. 224.*

Lorenc, A C and Hammon, O, 1986 : 'Objective quality control of observations using Bayesian methods - theory and a practical implementation.' *Met O 11 Technical Note no. 225.*

Smith, S G and Ashcroft, J, 1988: 'The Observation Processing Database - its contents and its derivation from the operational suite.' *Operational Numerical Weather Prediction System Documentation Paper 2.3.*



<u>Figure 1.</u> A summary of quality control decisions made during the Operational Trial of the New Ship Quality Control scheme (9th to 16th December 1987), for both Pmsl (underlined) and surface wind.



<u>Figure 2(a).</u> A breakdown of Pmsl observations flagged operationally during the Operational Trial of the New Quality Control scheme (9th to 16th December 1987). The numbers in brackets refer to observations flagged by CFO.

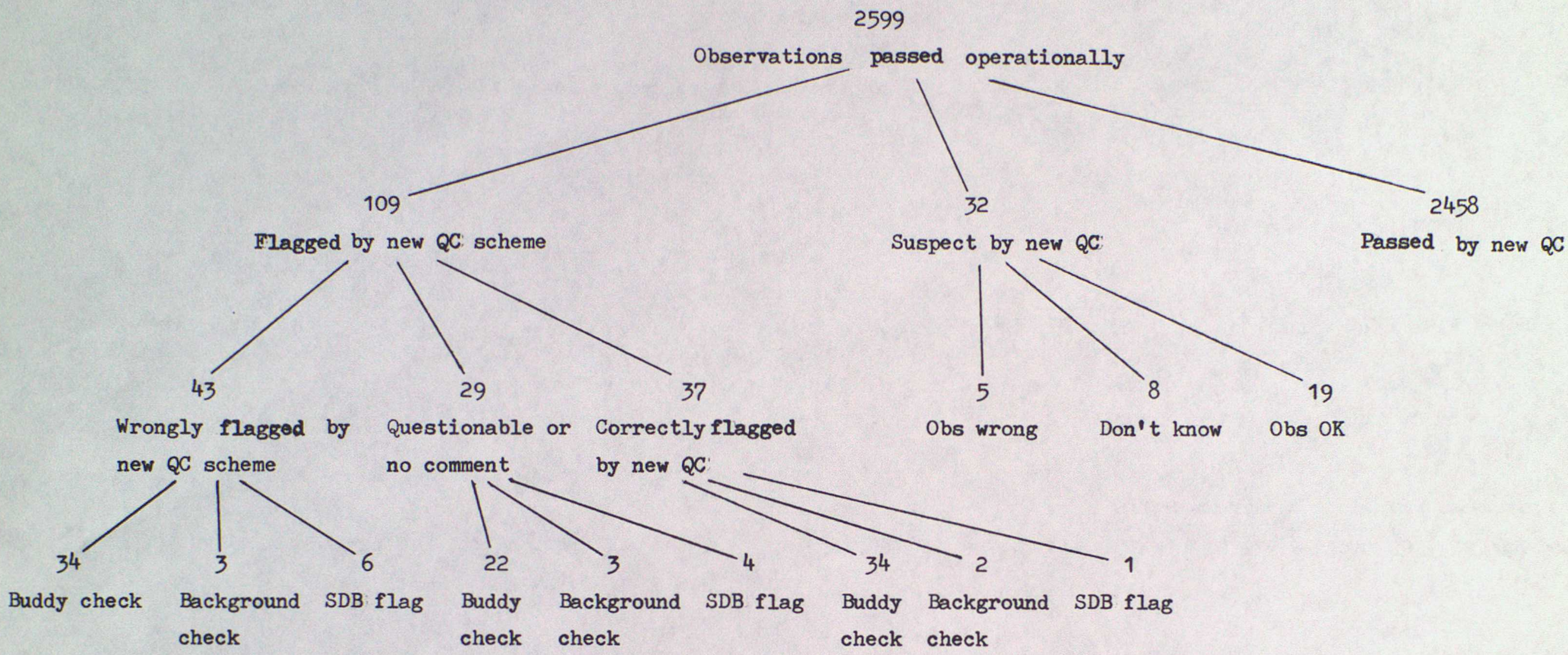


Figure 2(b). As figure 2(a), but for observations passed operationally.

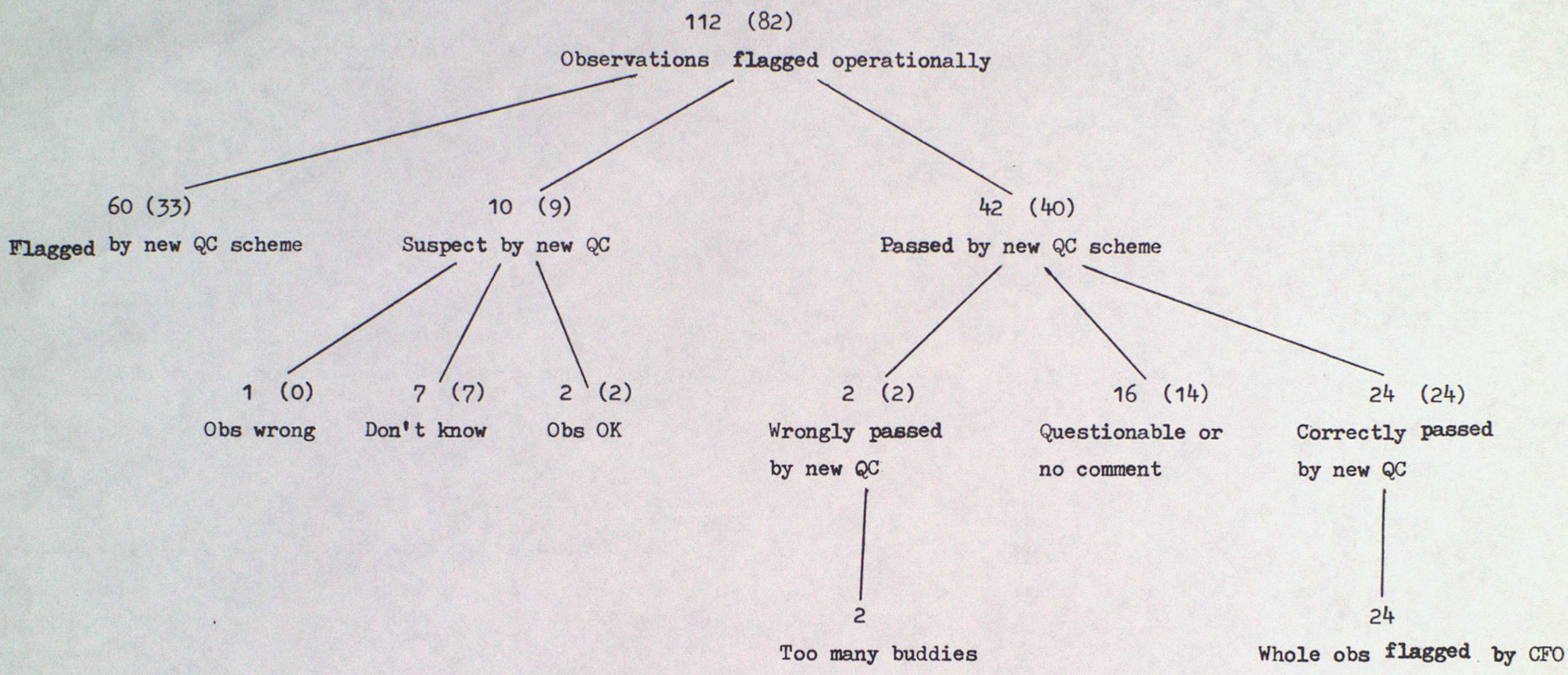


Figure 2(c). As figure 2(a), but for observations of surface wind.

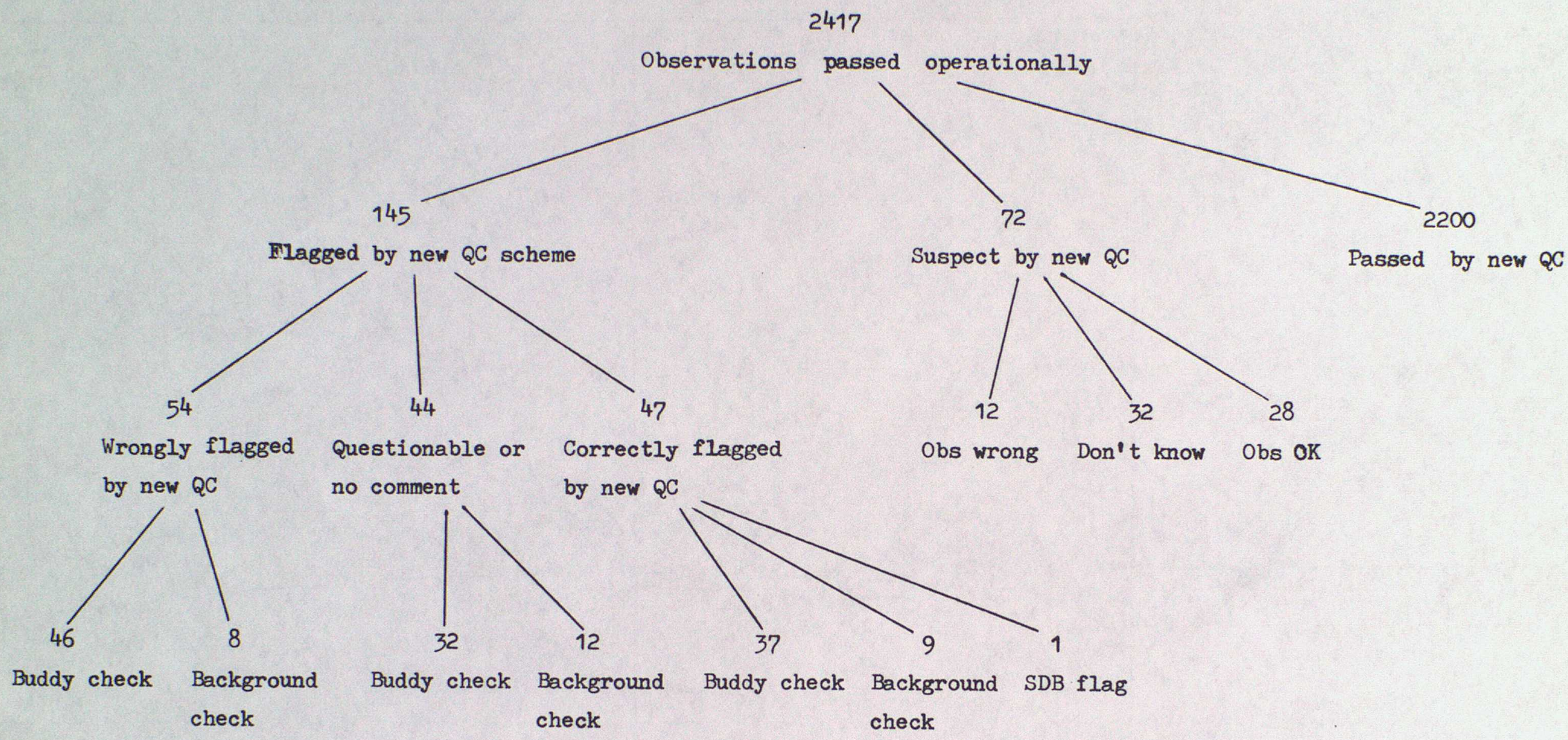
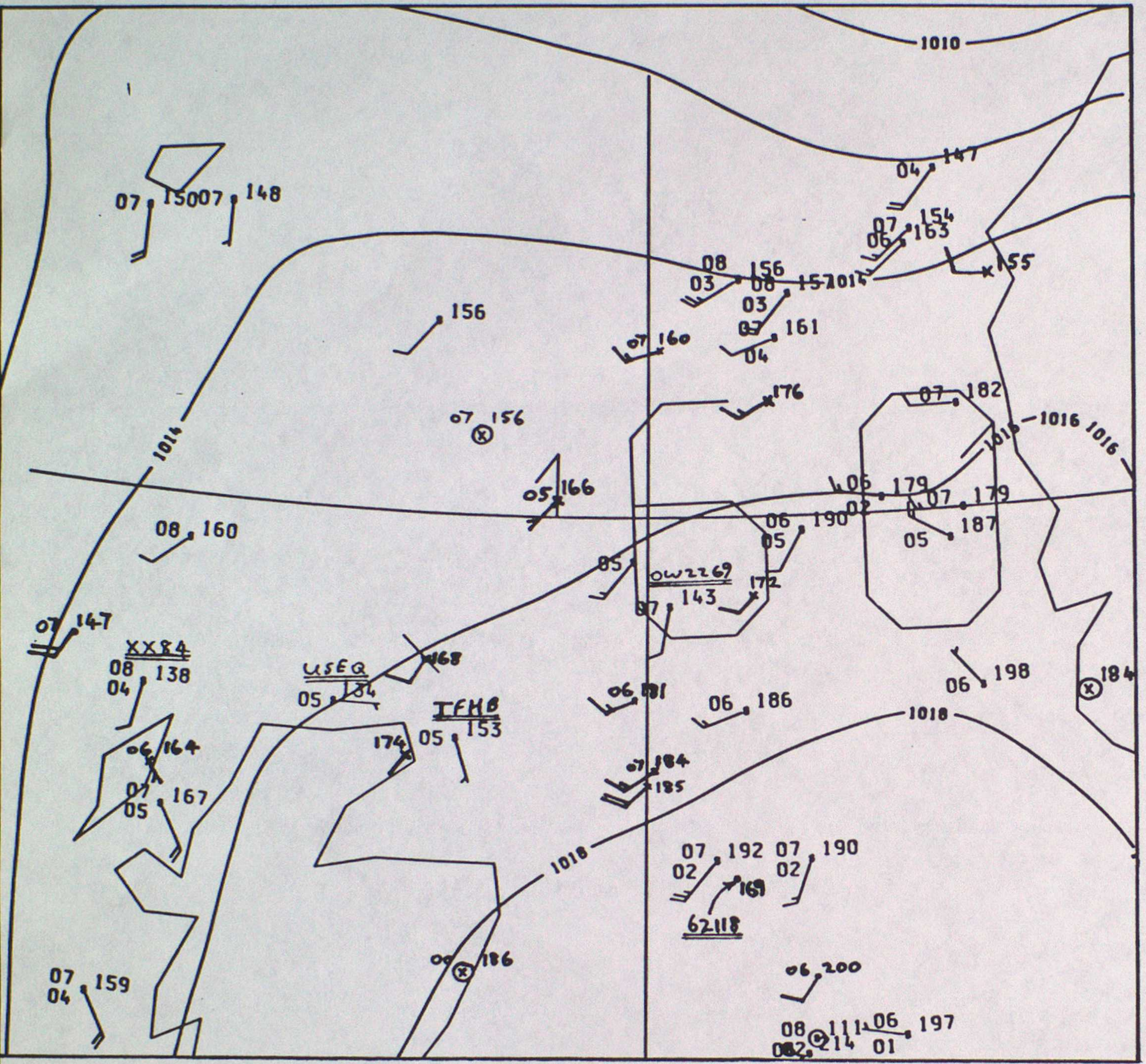


Figure 2(d). As figure 2(a), but for observations of surface wind passed operationally.

T + 6 BACKGROUND Pmsl FIELD & SURFACE OBS

FOR 00Z 13/12/87.

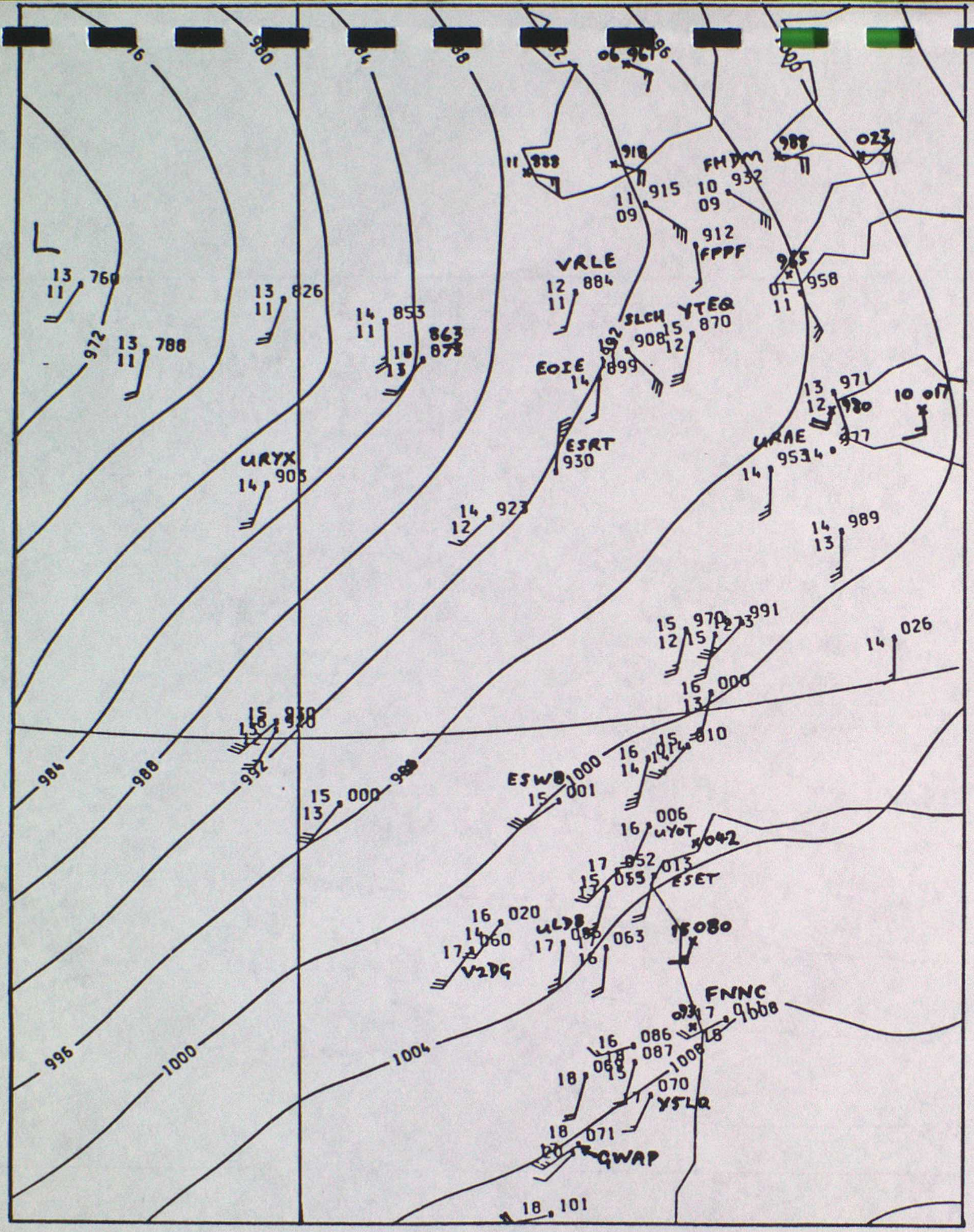


<u>Quality control **details** for ships Pmsl</u>

<u>obs **flagged** by the new scheme only</u> :

Call-sign	Prob(Gross error) (+ no. of buddies) 1. Trial run	2. Re-run	(O-B) (mb)
OW2269	100% (23)	100% (16)	-2.3
TFHB	100% (27)	81% (16)	-1.5
USEQ	100% (27)	100% (17)	-2.6
XX84	90% (21)	45% (14)	-1.2
62118	100% (21)	75% (12)	-1.6

Figure 3.



T + 6 BACKGROUND Pmsl FIELD AND SURFACE OBS FOR 12Z 15/12/87.

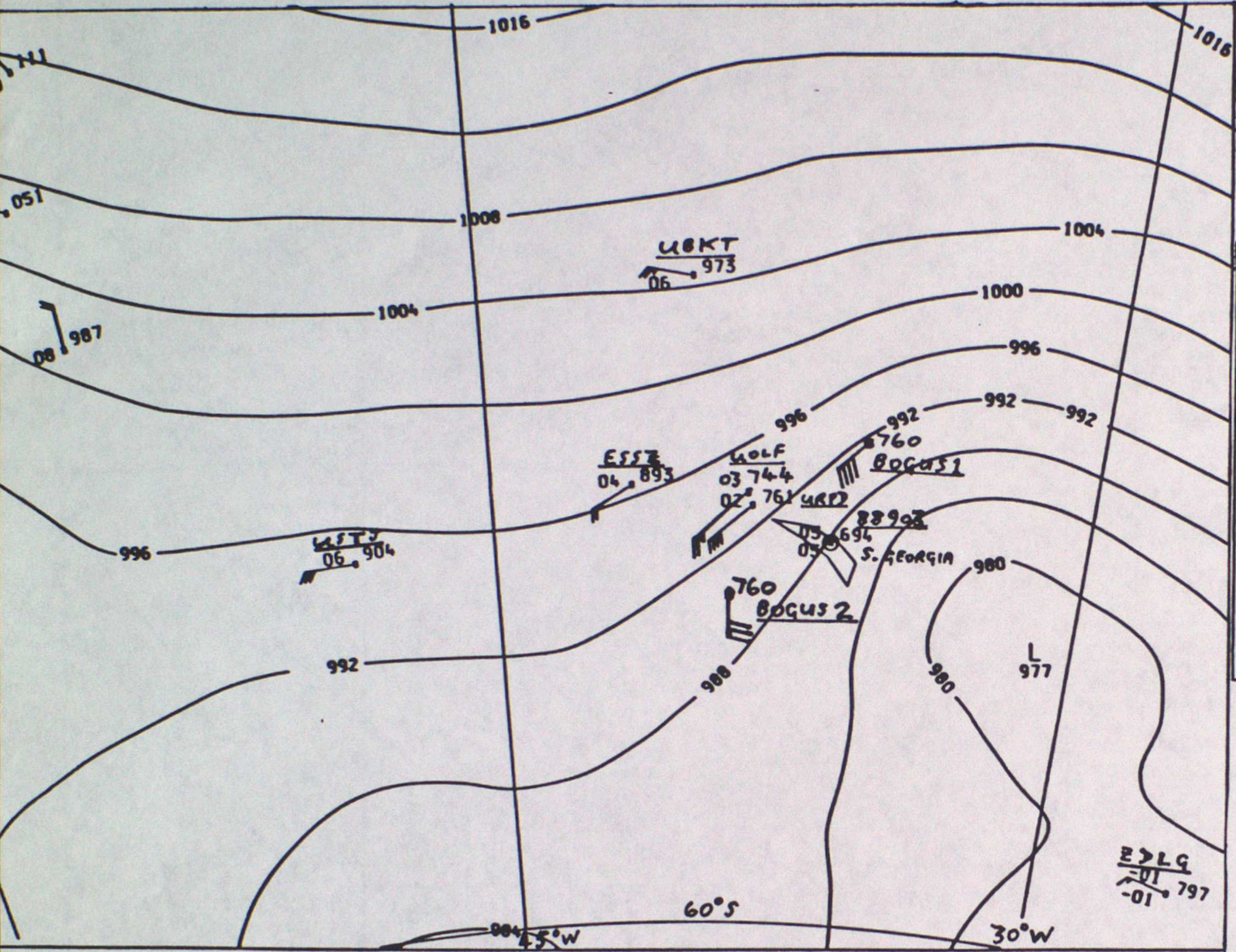
Quality control decisions for ships Pmsl obs flagged by one or both schemes :

Call-sign	Oper.	New Q.C. scheme (+ no. buddies) Trial run	Re-run
EOIE	P	100% (22)	79% (14)
ESET	P	100% (22)	100% (14)
FHDM	P	89% (28)	55% (16)
FNNC	F	100% (19)	100% (14)
FPPF	P	99% (26)	47% (16) - P
GWAP	P	71% (16)	37% (12) - P
SLCH	P	94% (22)	18% (15) - P
ULDB	F	100% (21)	100% (17)
URAE	P	94% (22)	22% (15) - P
URYX	P	86% (9)	10% (9) - P
UYOT	P	100% (23)	100% (16)
VRLE	P	100% (20)	99% (15)
V2DG	F	86% (19)	24 % (13) - P
YTEQ	F	100% (24)	100 % (14)
Y5LQ	F	95% (17)	73 % (13)

(F=flagged, P=passed)

Figure 4.

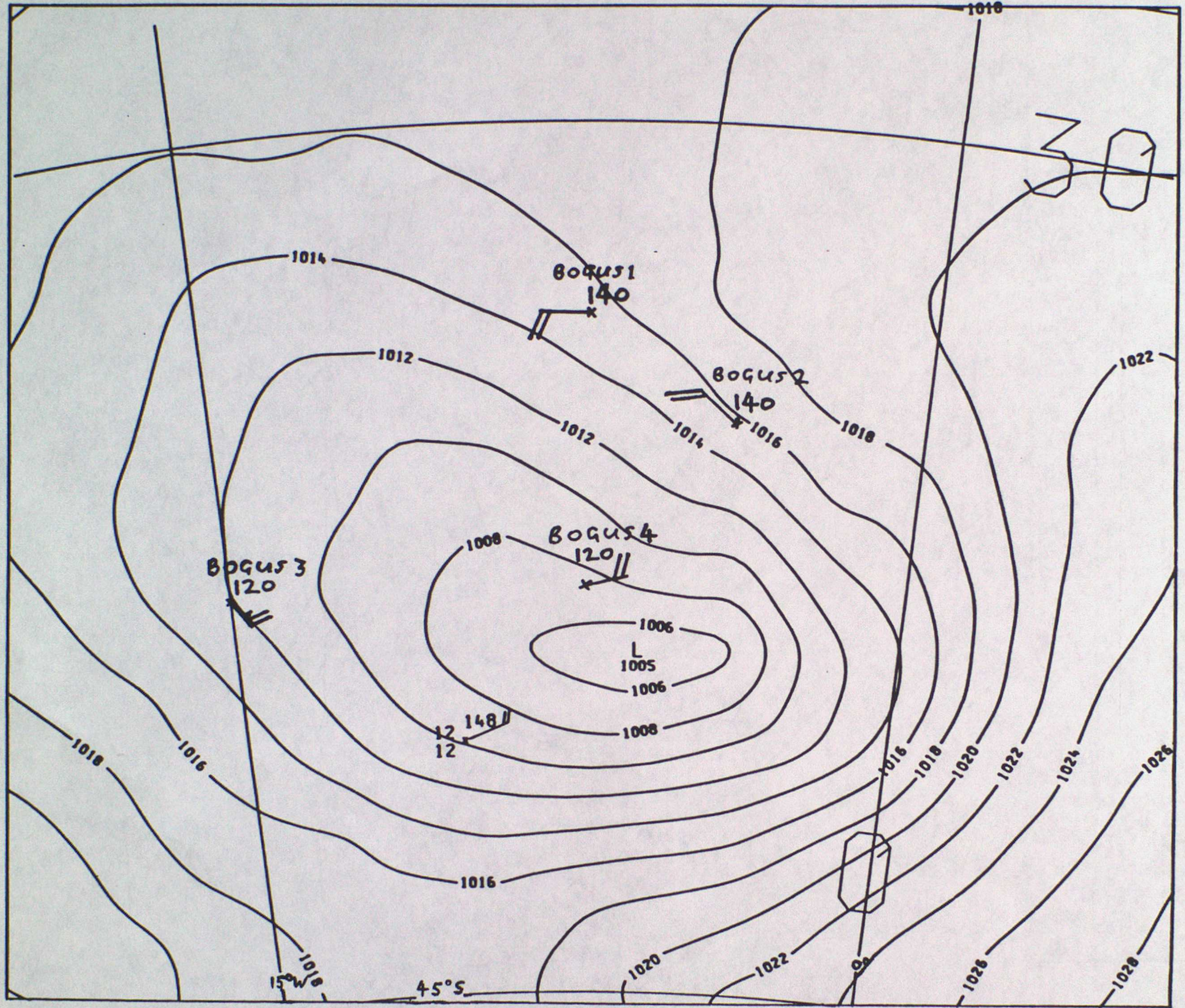
T + 6 BACKGROUND Pmsl FIELD AND SURFACE OBSERVATIONS FOR 12Z 15/12/87.



Quality Control decisions :

Call-sign	Oper. P	Trial run P.	Trial run wind	Re-run P.	Re-run wind	(O-B) P. (mb)
BOGUS1	P	3%	39%	0%	12%	-15.2
BOGUS2	P	1%	75%	0%	31%	-14.2
ESSZ	P	3%	4%	1%	3%	-7.2
UBKT	P	3%	8%	4%	5%	-7.3
UOLF	F	100%	4%	100%	2%	-19.2
URPD	P	100%	8%	100%	4%	-16.9
USTJ	P	13%	8%	8%	10%	-4.6
ZDLG	P	3%	5%	4%	3%	0.9
88903	P	36%	94%	0%	98%	-18.7

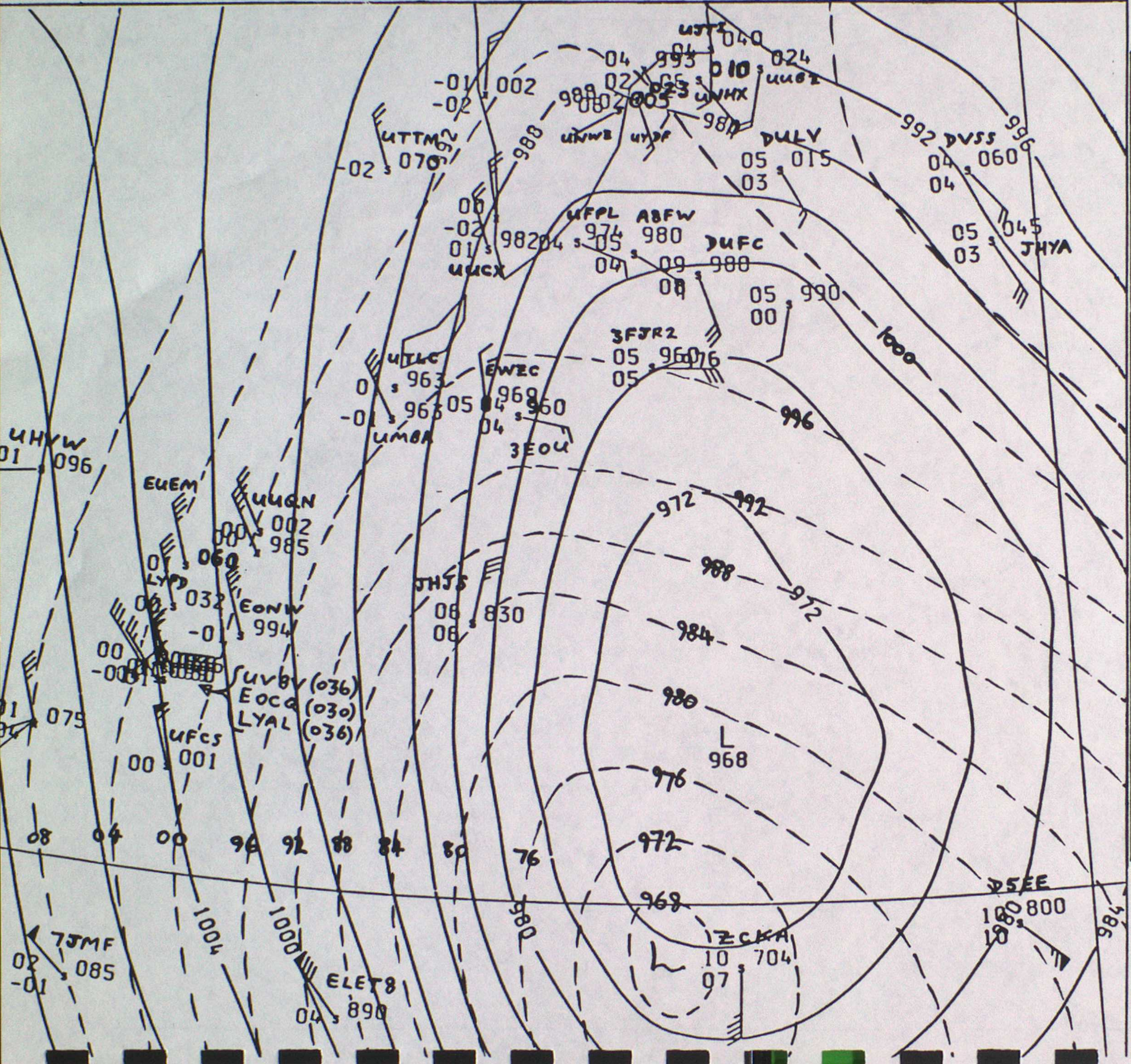
Figure 5.



T + 6 Pmsl FIELD & SURFACE OBS FOR 00Z 13/12/87.

Figure 6.

T + 6 BACKGROUND Pmsl FIELD AND SURFACE OBSERVATIONS FOR 00Z 18/11/87 + HAND ANALYSIS (dashed lines)

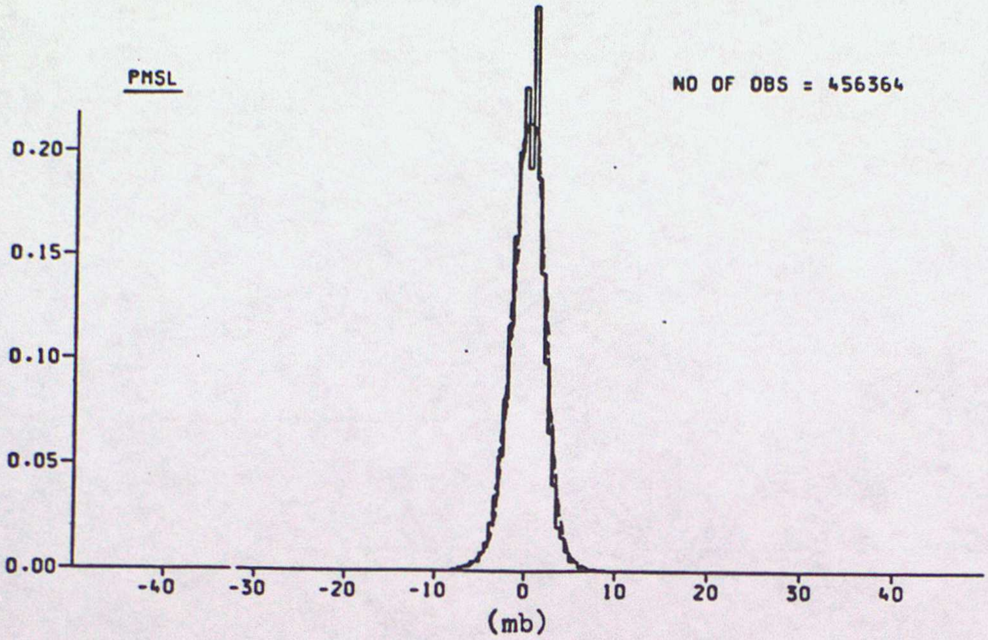


Call-sign	(O-B) Pmsl (mb)	Quality control decisions: Oper.	New scheme: no. bud.'s (+final P(G))	
A8FW	17.0	F	P (0%)	19
DUFC	19.0	F	P (0%)	18
DULV	15.4	F	P (0%)	16
DVSS	13.6	F	P (0%)	7
D5EE	-0.6	P	P (8%)	3
ELET8	-10.1	F	P (3%)	3
EOCQ	-0.7	P	P (0%)	13
EONW	0.4	P	P (7%)	14
EUEM	4.4	P	F (100%)	16
EWZC	12.9	F	P (0%)	15
JHJS	0.7	P	F (97%)	13
JHYA	13.0	F	P (0%)	4
LYAL	0.1	P	F (56%)	13
LYFD	0.6	P	P (0%)	16
UFCS	-4.0	P	F (97%)	14
UFPL	14.5	F	P (0%)	19
UHVW	0.9	P	P (1%)	14
UJTI	13.4	F	P (0%)	15
UMBA	5.5	P	P (1%)	15
UNHX	11.6	F	P (0%)	16
UNWZ	12.8	F	P (0%)	19
UTLC	5.5	P	P (0%)	16
UTTM	13.0	F	P (0%)	15
UUBZ	11.4	F	P (0%)	15
UUCX	11.0	F	P (0%)	17
UUQN	2.5	P	F (77%)	16
UVBV	0.5	P	P (0%)	13
UYDF	14.2	F	P (0%)	17
ZCKA	-2.7	P	P (7%)	1
3EOU	14.8	F	F (100%)	15
3FJR2	19.9	F	F (100%)	17

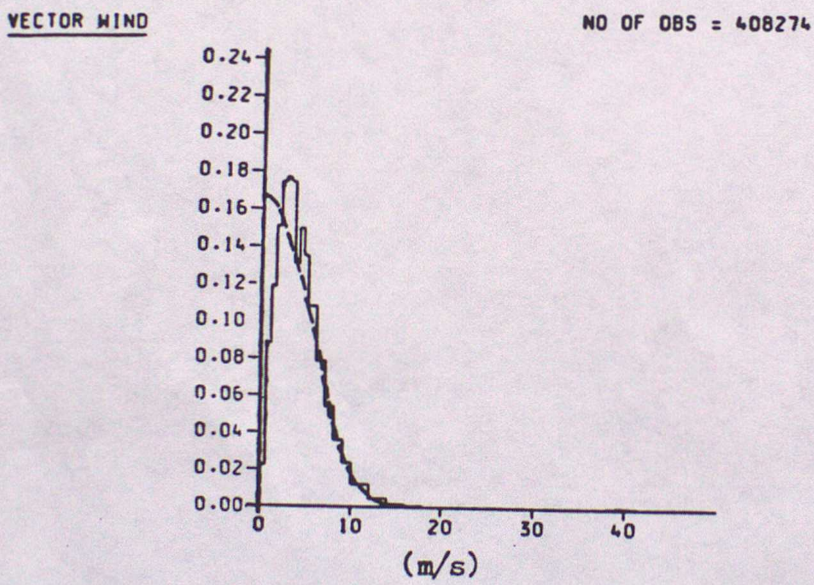
(F=flagged, P=passed q.c.)

Figure 7.

(i)



(ii)



(iii)

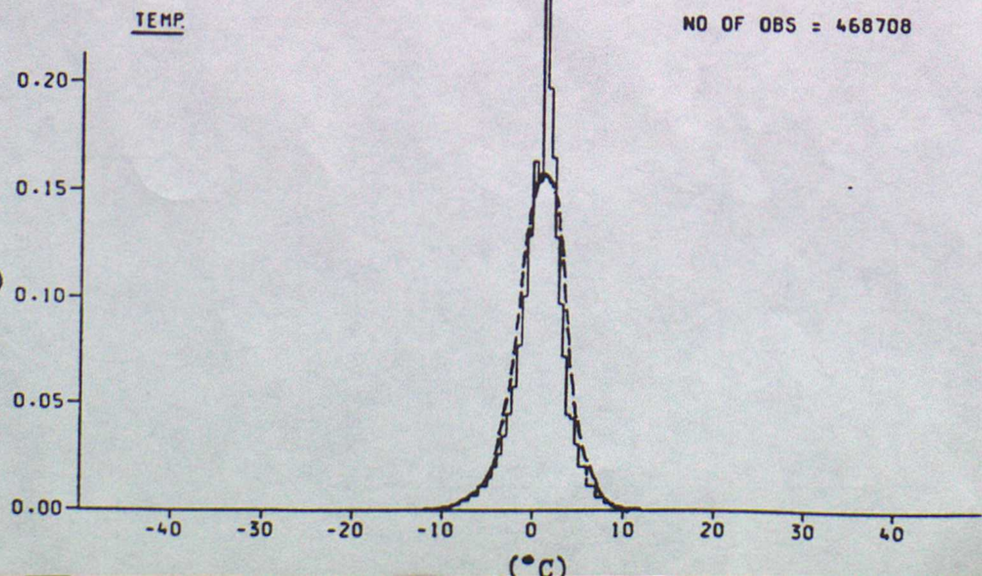
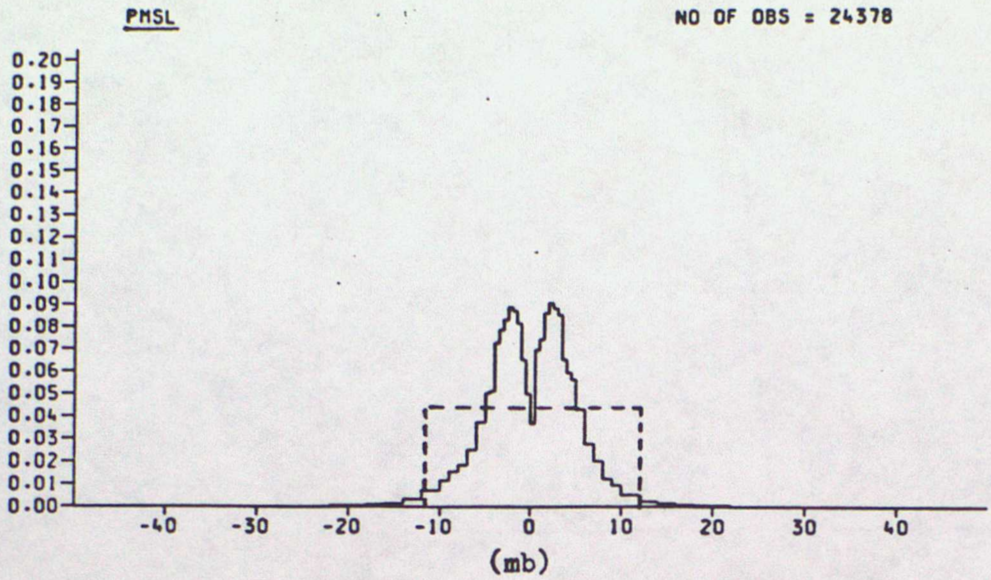


Figure 8(a).

Surface marine data passed by new quality control scheme: (O-B) differences.

(i)



(ii)

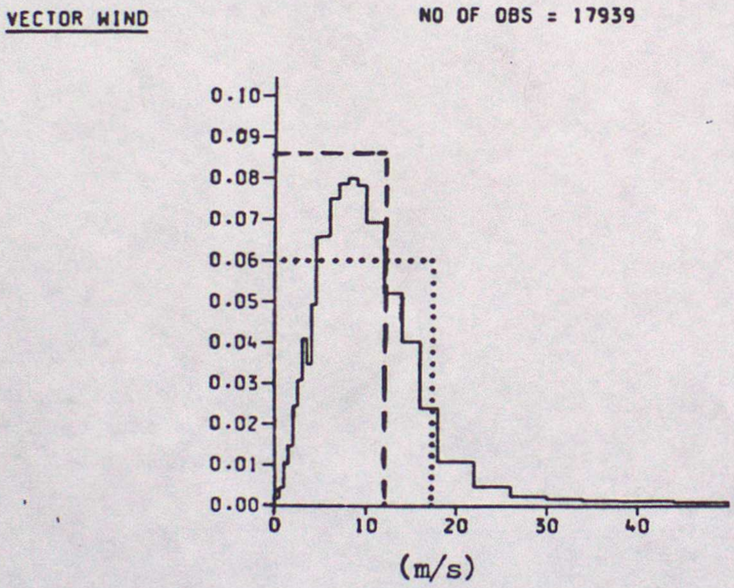
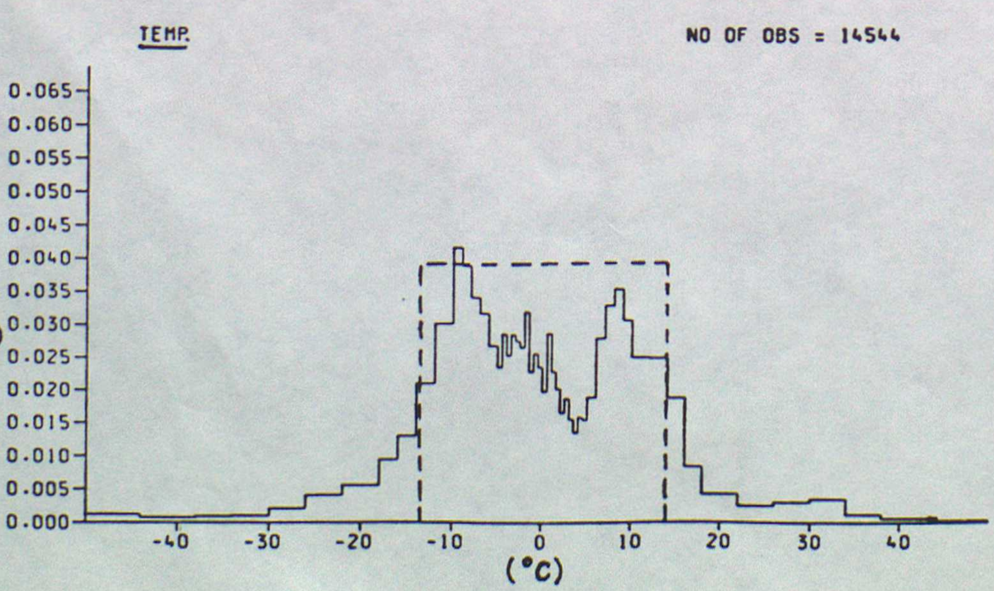


Figure 8(b).
Surface marine data flagged by new quality control scheme: (O-B) differences.

(iii)



	OWS	SHIP	SHIP (AUTO)	RIG/ PLAT	R/P (AUTO)	MOORED BUOY	DRFT. BUOY	SYNOP	SYNOP (AUTO)
Probability of gross error in location (%)									
Observation failed SDB sequence test	75	75	--	--	--	--	--	--	--
Observation failed other SDB test	90	90	90	90	90	90	90	90	90
Observation passed SDB sequence test	0.4	2	--	--	--	--	--	--	--
No SDB test done or flag set	1	4	4	2	1	1	4	0.8	1
Probability of gross error in pressure (%)									
Observation failed SDB tendency test	65	65	65	65	65	65	65	65	65
Observation failed other SDB test	90	90	90	90	90	90	90	90	90
Observation passed SDB tendency test	0.4	5	5	--	--	2	--	0.6	1.5
No SDB test done or flag set	0.8	6	6	4	3	3	5	2	2
Probability of gross error in temperature (%)									
Observation flagged by SDB	90	90	90	90	90	90	90	90	90
Observation not flagged by SDB	0.5	7	7	4	3	3	8	1.5	1.5
Probability of gross error in wind (%)									
Observation flagged by SDB	90	90	90	90	90	90	90	90	90
Observation not flagged by SDB	0.5	6	6	4	3	4	8	2	2
Observation error variances									
Mean sea level pressure (mb^2)	1.0	1.0	1.0	1.0	1.0	1.0	1.6	1.0	1.0
Temperature ($^oC^2$)	2.3	4.0	4.0	3.2	3.2	4.4	5.0	4.4	4.4
Vector wind ($(m/s)^2$)	9.7	12.5	10.5	10.5	7.2	10.5	18.0	11.5	8.8

Table 1. Initial values of probabilities of gross errors for different obs types and levels of SDB flags, used operationally. Also tabulated are observation errors used by the quality control program for each obs type.

	Coarse-mesh runs	Fine-mesh runs
Mean Sea Level Pressure (mb)	$(1.5+0.05\Delta t)/1.8$	$(1.6+0.062\Delta t)/1.8$
Surface Vector Wind (Knots)	$(7.8+0.05\Delta t)/8.1$	$(8.7+0.05\Delta t)/8.1$
Surface Temperature (°C)	$(2.8+0.01\Delta t)/2.86$	$(2.5+0.004\Delta t)/2.86$

Table 2. Formulae used for calculating the scaling factors for background errors, compared with coarse-mesh T+6 forecasts (for which coefficients had previously been calculated). The functions of the forecast time (Δt - in hours) were obtained from Met O 2b verification statistics for the 6 months from July to December 1987. (The denominator in each expression is the RMS error for the coarse-mesh T+6 run.)