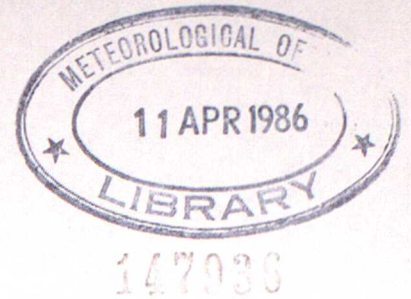


KW



MET O 11 TECHNICAL NOTE NO. 225

Objective quality control of observations using Bayesian methods  
- Theory, and a practical implementation

by

A C Lorenc and O Hammon

April 1986

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

Note: This paper has not been published. Permission to quote from it should be obtained from the Assistant Director of the above Meteorological Office branch.



## 1. INTRODUCTION

A major problem for operational numerical weather prediction is the quality control of the observational data used in the analysis of the initial conditions for the forecast. Unfortunately a significant proportion of the data are corrupted by instrument, operator or telecommunication errors. Using such data can cause incorrect forecasts; on the other hand rejecting extreme but correct data can cause the forecast to miss important events. Case studies of differences between analyses from different analysis systems, presented with the same data, often show that these are related to different quality control decisions (eg Hollingsworth et al 1985).

Many errors can be detected by checks using internal consistency and climatological extremes, before the central numerical weather prediction processing. However, plausible but incorrect data are best identified using all available information, such as forecasts from earlier data and observations from other sources, as part of the numerical weather prediction analysis process. In the past quality control algorithms have been developed in a rather pragmatic way, with methods and tolerances empirically tested and tuned. Automatic analysis systems are now processing such a wide variety of types of data, with different reliabilities and accuracies, that a more formal framework is desirable.

This work attempts to provide a framework by studying the quality control process in terms of Bayesian probability theory. In section 2 we apply the theory to the analysis problem of combining data and a background, where both are susceptible to errors. In section 3 we discuss models of the error distributions. We propose that the error in each datum is either a normal observational error, from a known Gaussian distribution, or a gross error, in which the datum gives no information. In section 4 we apply the Bayesian theory to this model, and show that the usual operational practice, which is to reject outlying data and to treat the rest as if their errors are Gaussian, is a reasonable approximation to the correct Bayesian analysis. Appropriate rejection criteria are derived in terms of the observational error, background error, and prior probability of gross error.

These ideas have been implemented in a computer program to check surface pressure, wind, temperature and position data from ships, weather ships, buoys, and coastal land stations. Data are checked against a forecast background and against neighbours, the results of these checks being quantified as probabilities of gross error and combined using Bayes' theorem. Section 5 gives details of this program.

Without an objective means of providing the necessary estimates of error statistics, the Bayesian approach is hardly more objective than other methods of setting rejection tolerances. The collection of these statistics thus needs to be an integral part of the checking scheme, and an observation processing database (OPD) has been set up to facilitate this. In section 6 we describe how the OPD can be used to check and update the estimates used in the program, and to monitor its behaviour. Results from an extended trial of the program are given.



Of course the practical benefit and efficiency of the program need to be assessed more concretely than by overall statistics. Some cases were selected for careful study, to see if the decisions made by the program were reasonable, and to compare them with results from the current operational system which uses both human and automatic quality control. Results of these case studies are presented in section 7.

Finally in section 8 we summarize our conclusions, and discuss further work.

An earlier version of some of this work was presented at a workshop on the use and quality control of meteorological observations. The report of that workshop (ECMWF 1985) provides a useful survey of quality control techniques currently used in many operational centres.

## 2. BAYESIAN THEORY

In this section we present a simple example of Bayesian probability theory, and show how this relates to the analysis of observations with errors. Further discussion of Bayesian Theory can be found, for example, in Winkler (1985). A more complete and rigorous derivation of the analysis equations is given by Lorenc (1986).

Bayes' Theorem enables us to calculate the likelihood of any event, based on prior estimates and new evidence. For events A, B it can be written as

$$P(A|B) = P(B|A) P(A)/P(B) \quad 1$$

In words, we say that the posterior probability of A, given that B has occurred, is proportional to the probability of B, given that A has occurred, multiplied by the prior probability of A. The constant of proportionality (which is often not needed to use the theorem) is the reciprocal of the prior probability of B.

Let us apply this to estimation of a single parameter from the events:-

T = the true value is t

B = the prior (background) estimate is b.

O = the observed estimate is o.

Without the observation or any other information all background and true values are assumed equally likely, so  $P(B) = P(T)$  and Bayes' Theorem gives us

$$P(T|B) = P(B|T) \quad 2$$

$P(B|T)$  is the error distribution of our background estimate b. For a parameter which can take any value and assuming that the distribution depends only on the deviation  $b-t$ , we say that the probability that the background is between b and  $b+db$  is given by  $P_b(b-t)db$ .

Thus we get

$$P(T|B) = P_b(b-t) \quad 3$$



From now on we shall assume knowledge of B always, and not represent it explicitly, e.g. we shall write  $P(T)$  for  $P(T|B)$ . Now consider the information added by the observation. Bayes theorem states that

$$P(T|O) = P(O|T) P(T)/P(O) \quad 4$$

The rule for addition of probabilities for mutually exclusive events gives us

$$P(O) = \int_{t=-\infty}^{\infty} P(O|T) P(T) dt \quad 5$$

$P(O|T)$  is the observational error distribution of  $o$ ; in a similar way to the background error we write it as  $P_O(o-t)$ . If these error distributions are both Normal (Gaussian), with variances  $V_b$  and  $V_o$  respectively, then

$$P_b(b-t) = N(b-t, V_b) \quad 6$$

$$P_o(o-t) = N(o-t, V_o) \quad 7$$

$$\text{where } N(x, V) = (2\pi V)^{-1/2} \exp(-x^2/2V) \quad 8$$

Substituting in (4) and (5), and using the algebraic properties of the Gaussian distribution, gives us

$$P(T|O) = N(a-t, V_a) \quad 9$$

$$1/V_a = 1/V_b + 1/V_o \quad 10$$

$$a = (o/V_o + b/V_b)V_a \quad 11$$

We call the posterior distribution given by (9) the analysis distribution.  $a$  is the analysed value and  $V_a$  is the analysis error variance. Examples are shown in figure 1 for four different values of  $(o-b)$ .

(11) and (10) can be derived directly as the minimum variance estimate of our parameter, and its variance, without using probability theory. For this example, with Gaussian distributions, the Bayesian approach adds little, since the mean, median and mode of a Gaussian are identical, and the mean and variance are sufficient to define the entire distribution.

### 3. ERROR MODELS

#### 3a. Observation errors

Let us suppose that there are two distinct types of error in observations: normal observational errors (instrumental), and gross errors (mistakes). The former are usually assumed to have a Gaussian distribution; if the observational error is due to an accumulation of small unrelated perturbations then the Central Limit Theorem justifies this assumption. Observations with the latter, gross, error we will assume to be useless, being unrelated to the true value, although they will still fall within some finite range of possible values related to the climatological variance. (We are assuming that implausible observations are eliminated at an early stage). Plausible values are thus equally likely, with



probability density  $k$ . If we write  $G$  for the event "the observation has a gross error", an overbar for "not" and  $\Omega$  for "and", the complete observation error distribution is given by

$$P(O|T) = P(O|\overline{T}\overline{G})P(\overline{G}) + P(O|\overline{T}G)P(G) \quad 12$$

$$P_O(o-t) = N(o-t, V_O) P(\overline{G}) + k P(G) \quad 13$$

Of course (13) is only an approximation, valid in the range of plausible values. It does not have the correct integral or variance. This has no practical significance; we can use (13) as an observation error distribution defined by the observed value and three physically interpretable parameters, each of which can depend on observation type, and the known reliability and accuracy of the observing and telecommunication systems.

#### b. Background errors

In practice our background information is dominated by that from a numerical forecast. Models of the errors in these are a crucial component of objective optimal interpolation analysis schemes, and have been discussed by (for instance) Lorenc (1981), Thiebaut (1985), Hollingsworth and Lonnberg (1986). We have in this work used a very simple model of background errors, ignoring important information such as geostrophy. For any parameter (eg sea level pressure) the expected error covariance between points  $i$  and  $j$  is assumed to be given by

$$\langle (b_i - t_i) (b_j - t_j) \rangle = F_i F_j \mu_{ij} \quad 14$$

where  $b-t$  is assumed to be from a Gaussian distribution with variance  $F^2$ , and  $\mu_{ij}$  is the correlation between points  $i$  and  $j$ .

#### c. Errors of representativeness

In objective analysis the desired resolution determines a scale, which may be described by saying a grid-point value represents an average in space and time. This scale is usually also that of the background, and it is usual to convert observations to the same scale. This means that errors of representativeness have to be added to the purely instrumental observational errors.

In contrast, we wish our quality-controlled observations to be available for use in analyses with various resolutions, so it is helpful to repartition errors of representativeness. They can be considered explicitly as the error in the generalized interpolation from background to observation (Lorenc 1986). In fact we compromise, including very small scales (eg gusts) in the observational error, and others as part of the background error. We assume these errors to be Gaussian.

In some situations there can be large errors in the background estimate of observed values because of discrepancies in the scales represented, especially when there are topographic features not resolved by the forecast model. For instance the observed wind from a ship in a fjord will not be well predicted by a forecast model which does not resolve the fjord. Ideally one should allow for this by having a sophisticated model of errors



of representativeness, which causes the background to be given low weight when checking the observation, and which in an objective analysis scheme causes the observation to be given low weight when analysing for the forecast model. If the model of errors of representativeness is inadequate, and does not do this, then a single unrepresentative observation is likely to be rejected as having a gross observational error. This will improve results from an objective analysis using a similarly inadequate model of errors of representativeness. However it will not provide reliable error detection for other users of the data. Moreover several correct but unrepresentative observations will support each other in the quality control, and (being given incorrect weights) will degrade the objective analysis. An objective quality control scheme cannot distinguish between such cases and others where the forecast is in error unless it "knows" that the relevant errors of representativeness can be large.

#### 4. BAYESIAN ANALYSIS WITH NON-GAUSSIAN ERRORS

We now repeat the analysis of section 2 using the observational error model (13). Bayes' theorem gives

$$P(T|O) = (N(o-t, V_o) P(\bar{G}) + k P(G)) N(b-t, V_b)/P(O) \quad 15$$

as the equivalent of (9). The normalization factor  $P(O)$  is given by

$$P(O) = N(o-b, V_o + V_b) P(\bar{G}) + k P(G) \quad 16$$

The direct use of Bayes' theorem as in (15) is equivalent to its use to get the posterior probability of gross error:-

$$P(G|O) = P(O|G) P(G)/P(O) \quad 17$$

then substituting in the identity

$$P(T|O) = P(T|O\bar{G}) P(\bar{G}|O) + P(T|OG) P(G|O) \quad 18$$

to give (15). Implementation of the method in section 5 is in terms of this two stage derivation; in this section we discuss (15) directly.

Figure 2 shows the analysis distribution for four values of  $o$ , for comparison with figure 1. (In the examples shown in figures 1-7 we have chosen values typical of those for surface pressure from ships:  $V_o = 1 \text{ mb}^2$ ,  $V_b = 3.6 \text{ mb}^2$ ,  $P(G) = .03$ ,  $k = 0.043 \text{ mb}^{-1}$ ).

Note that for  $(o-b)^2/(V_o + V_b)$  large the analysis distribution with the possibility of gross errors becomes distinctly bimodal (figure 2). Practically, since operationally we need a single 'best' analysis; we then have the problem of picking the 'best' value. One can argue for the mean, the median, or the mode; the variation of these with  $o-b$  is shown in Figure 3. If we take the mean then we can see from (18) that this is equivalent to reducing the weight given to the observation by the factor  $P(\bar{G}|O)$ , shown in figure 4. This procedure has been advocated by Gutowski and Hoffman (1985). However for bimodal distributions the mean value can be rather unlikely to actually occur. Frequently when using Bayes' theorem the maximum likelihood (mode) value is chosen. This however takes no account



of the width of the peaks; perhaps the local mode nearest the mean is best. The preceding discussion illustrates that there is no unique definition of 'best', while figure 3 shows that whatever is used, there is a region of rapid variation. For the mean this occurs near where  $P(G|0)$  equals  $1/2$ ; this is true also of the median and the mode nearest the mean when  $P(G)$  is small. Thus we can approximate the 'best' analysis by

$$\begin{aligned} a &= (o/V_o + b/V_b)V_a \text{ when } P(G|0) < 1/2 & 19 \\ a &= b & \text{when } P(G|0) > 1/2 \end{aligned}$$

For small  $P(G)$  this is a very good approximation to the behaviour of the mode nearest the mean, and a fair approximation to the behaviour of the median or mean. The errors in the approximation are only significant where there is already uncertainty as to what is 'best'. Substituting in (17), we find that the jump in (19) occurs when

$$(o-b)^2 = (V_o + V_b)T^2 \quad 20$$

$$T^2 = 2 \ln (P(\bar{G})/P(G)) - \ln ((V_o + V_b)k^2) - \ln (2\pi) \quad 21$$

We have thus derived an expression for the checking tolerance which is an empirically tuned parameter in many existing quality control schemes (eg Lorenc 1981).  $T$  is plotted in figure 5.

In the less common case of  $P(G)$  large, (19) is not as good an approximation. For instance we see from figure 4 that for  $P(G) = .9$ ,  $P(G|0) > 1/2$  for all values of  $0$ . Thus the analysis value given by (19) will be  $b$  always. All of the options for the "best" analysis plotted in figure 3 give significant weight to such an observation for small deviations from the background.

The approximation (19) has equivalents for the probability of gross error, and for the analysis variance:-

$$P(G|0) = 0 \text{ when } P(G|0) < 1/2 \quad 22$$

$$P(G|0) = 1 \text{ when } P(G|0) > 1/2$$

$$V_a = V_b V_o / (V_b + V_o) \text{ when } P(G|0) < 1/2 \quad 23$$

$$V_a = V_b \text{ when } P(G|0) > 1/2$$

Although (19) is a good approximation to the 'best' analysis, (22) and (23) are much less good measures of the reliability of the datum and the accuracy of the analysis. The actual variance of the posterior distribution (15) about its mean, and about the value given by (19), are illustrated in figures 6 and 7. The approximation (23) can be a gross underestimate. If the analysis is to be used to check other observations this is important, since the analysis variance appears in the equivalent of (20). Lorenc (1981) found that a statistical quality control scheme based on a succession of approximations (22), performed better with an empirically determined increase to the variance given by (23).



Figure 6 also illustrates a difference between the Bayesian analysis method and the minimum variance method usually used to derive the optimal interpolation equations. The addition of information from an observation with a non-Gaussian error distribution can actually increase our uncertainty about the true state, as measured by the variance  $V_a$ . A better measure of uncertainty is related to entropy, and is proportional to  $\langle \ln P \rangle$ . This is decreased by the observation.

## 5. A PROGRAM TO QUALITY CONTROL DATA FROM SHIPS

As a prototype for a comprehensive system checking all types of observation, using these ideas, a system for checking surface data from ship reports has been set up. These data were chosen because the frequency of gross error is rather high, and considerable manual effort is put into their quality control in the current operational system. With the ships it is natural also to check similar data from buoys and coastal land reports.

The objective of the prototype is to test both the practical utility of the Bayesian probability ideas, and methods of accumulating the necessary statistics. The physical ideas are therefore kept simple. Observations of pressure, wind and temperature are compared with a forecast background and with nearby observations (ie surface observations within 150 km). A sequence test comparing with the observation from the same ship 6 hours earlier, and other flags set in tests of position and internal consistency of the report, which are part of the current operational system, are also used. No attempt has been made yet to improve these tests as part of this work, nor to introduce new physical ideas like allowing for constant calibration errors, or comparing wind and pressure geostrophically.

None of the tests used can confidently detect all gross errors, or pass all good data, so the Bayesian approach, with a probability of gross error instead of simple pass/fail flags, should be of benefit in combining their results. The tests are described in more detail in the rest of this section.

### a) SDB tests

The synoptic data bank (SDB) programs perform tests on observations as they are stored, before they are used by other programs. Details are given by Met O 12c (1985) and Atkins (1984). Results are stored as flags on individual elements of the report, and further flags indicating whether an earlier report was available to perform the sequence test. For ships, the relevant checks for this work are that the position is in the sea, that the position is consistent with reported speed and direction and last position in the sequence test, that the pressure is not more than 15 mb from the background, that the pressure is consistent with reported tendency and last pressure in the sequence test, that the wind speed is less than 70 kts or a special group is reported, and that the temperature is consistent with reported snow etc. The tolerances in the sequence tests are 100 miles and 5 mb. For weather ships and land stations the observation 3 hours earlier is used, with a tolerance of 2 mb.

On the basis of the flags set in these tests, prior probabilities of gross error in position, surface pressure, wind and temperature are assigned. We denote these by  $P(G_p)$ ,  $P(G_s)$ ,  $P(G_u)$ ,  $P(G_T)$  respectively.  $G_p$  is also taken



to include other gross errors affecting the whole report. The values assigned are chosen on the basis of judgement, and the records of reliability of a particular flag. Some examples are given in section 6.

#### b. Comparison with background

The basic method of section 4 is readily extended to include consideration of  $G_p$ , which links checking of pressure, wind and temperature data.

If we extend the use of the suffixes p,s,u,T to other parameters, and allow for the two sources of gross error  $G_p$  and  $G_s$ , then for surface pressure (13) becomes

$$P(O_S | T_S) = N(O_S - t_S, E_S^2) P(\bar{G}_p) P(\bar{G}_s) + k_S (1 - P(\bar{G}_p) P(\bar{G}_s)) \quad 24$$

Similar equations are used for wind, in terms of the absolute vector error  $|u_O - u_t|$ , and temperature. Application of Bayes theorem to  $G_p$  gives

$$P(G_p | O_S \Omega O_U \Omega O_T) = P(O_S \Omega O_U \Omega O_T | G_p) P(G_p) / P(O_S \Omega O_U \Omega O_T) \quad 25$$

Assuming the various instrumental and gross errors are independent, we have

$$\begin{aligned} P(O_S \Omega O_U \Omega O_T) &= P(O_S | G_p) P(O_U | G_p) P(O_T | G_p) P(G_p) \\ &\quad + P(O_S | \bar{G}_p) P(O_U | \bar{G}_p) P(O_T | \bar{G}_p) P(\bar{G}_p) \end{aligned} \quad 26$$

where, for instance

$$P(O_S | G_p) = k_S \quad 27$$

$$P(O_S | \bar{G}_p) = k_S P(G_s) + N(O_S - b_S, E_S^2 + F_S^2) P(\bar{G}_s) \quad 28$$

Similar equations to (25) can be written for  $P(G_s | O_S \Omega O_U \Omega O_T)$  etc. Although it is assumed that  $G_p$  and  $G_s$  are a priori independent events,  $G_p | O$  and  $G_s | O$  are not necessarily independent, so it is important to calculate those posterior probabilities needed for future processing. As explained below, the buddy check against nearby observations is done independently for pressure, wind and temperature, without explicitly considering errors such as  $G_p$  which relate them, so we need to calculate the probability that for instance surface pressure is unaffected by gross error from either  $G_p$  or  $G_s$ .

$$P(\bar{G}_p \bar{G}_s | O_S \Omega O_U \Omega O_T) = P(O_S \Omega O_U \Omega O_T | \bar{G}_p \bar{G}_s) P(\bar{G}_p \bar{G}_s) / P(O_S \Omega O_U \Omega O_T) \quad 29$$

Evaluating expressions like (25) and (29) using identities like (26), (27) and (28) is a straightforward algebraic task, extendable to observations with more parameters and types of gross error, albeit with rapidly increasing complexity.

#### c. Buddy check

The above equations are also extendable in principle to several observations, as long as the joint probability distribution of their deviation from the background can be expressed. However to simplify the



algebra and search algorithms, we derive an expression for just two data, and assume that its iterative application is a reasonable approximation in cases where there are several close observations.

Let us consider observation events  $O_1$  and  $O_2$ . The buddy check is useful because, without gross errors, the events are not independent, ie.

$$P(O_1 \Omega O_2 | \bar{G}_1 \bar{\Omega} \bar{G}_2) \neq P(O_1 | \bar{G}_1) P(O_2 | \bar{G}_2) \quad 30$$

We shall still consider other events such as  $G_1$  and  $G_2$  to be independent, so we are not considering observations which might have a common source of error. To evaluate the joint probability we use

$$P(O_1 \Omega O_2 | \bar{G}_1 \bar{\Omega} \bar{G}_2) = P(O_1 | \bar{G}_1 \Omega O_2 \bar{\Omega} \bar{G}_2) P(O_2 | \bar{G}_2) \quad 31$$

Because we are excluding gross errors, all distributions in (31) are Gaussian, and we can use the well known technique of statistical (optimal) interpolation (OI) to evaluate a best estimate at position 1 using the background and observation 2, and to calculate the error variance of this estimate.

The OI weight  $w$  is given by

$$w = (E_2^2 + F_2^2)^{-1} F_1 \mu_{12} F_2 \quad 32$$

The interpolated value, and its variance are given by

$$a_1 = b_1 + w (o_2 - b_2) \quad 33$$

$$A_1^2 = F_1^2 - w F_1 \mu_{12} F_2 \quad 34$$

Thus

$$P(O_1 | \bar{G}_1 \Omega O_2 \bar{\Omega} \bar{G}_2) = N(o_1 - a_1, E_1^2 + A_1^2) \quad 35$$

Before the buddy check both observations' probabilities of error are known from (29); afterwards they become

$$P(G_1 | O_1 \Omega O_2) = P(G_1 | O_1) P(O_1) P(O_2) / P(O_1 \Omega O_2) \quad 36$$

$$P(G_2 | O_1 \Omega O_2) = P(G_2 | O_2) P(O_1) P(O_2) / P(O_1 \Omega O_2) \quad 37$$

These equations are derived in the appendix, where we also show that the multiplicative factor, which differs from unity because of (30), is given by the reciprocal of:-

$$\begin{aligned} & P(O_1 \Omega O_2) / P(O_1) P(O_2) \\ & = 1 - P(\bar{G}_1 | O_1) P(\bar{G}_2 | O_2) (1 - P(O_1 \Omega O_2 | \bar{G}_1 \bar{\Omega} \bar{G}_2) / P(O_1 | \bar{G}_1) P(O_2 | \bar{G}_2)) \end{aligned} \quad 38$$

Substituting (31) to (35) enables (38) to be evaluated.

This procedure is applied independently to each observed element from a pair of ships. The checking of several neighbouring ships is done sequentially, in pairs, at each step modifying the current estimate of the



probability of gross error for each using the factor (38). This is an approximation, since for example, if observations 1 and 2 have already both been compared with observation 3, then their estimated probabilities of error will no longer be strictly independent.

## 6. ACCUMULATION AND FEEDBACK OF STATISTICS

The methods described in section 5 require estimates of the coefficients for the error distributions and correlations. Without an objective means of determining these, the methods would need empirical tuning, and would have few advantages over older, ad hoc checking algorithms. The collection of the necessary statistics thus needs to be an integral part of the system.

All observed data, background values, and probabilities of error (prior, intermediate, and final) used in the checking program, are recorded in an observation processing database (OPD) in a form suitable for sorting and calculation of statistics. This OPD can be used to derive coefficients for the error models described in section 3, as well as testing the validity of new models. Each of the accumulated statistics depends on more than one of the coefficients, so an immediate determination of them all is not possible. For example the mean square deviation of data without detected errors from the background depends on the background error variance, the observation error variance, and the detection of gross errors. However by making the assumption that prior estimates of some of these are reasonable, new estimates can be made, and the repeated feedback from a series of such statistics can converge towards good estimates for all the coefficients.

Experience with other schemes, and a short initial trial, were used to make preliminary estimates of the required coefficients. These were then used in an extended trial of the scheme using the operational global 6 hour forecasts to check observations from the operational synoptic data banks. Sea surface data and coastal land reports valid at 00 and 12 GMT were checked from mid August to mid October 1985. In this section we present some statistics from this trial to illustrate how the prior estimates can be checked and modified.

Our prior estimates for the observational errors (E) were guided by those used in operational OI analysis schemes. Our prior estimates for the background errors (F) were obtained rather crudely, from the differences between operational 12 hour forecasts and verifying analyses from mid-June to mid-July 1985. Global fields of the error variance of surface pressure, vector wind and temperature were obtained. These were used directly to represent the 6 hour forecast error variance, augmented by a term dependent on the grid-scale roughness of the current background, meant to represent the error of representativeness of the model. Using an analysis as "truth" in the calculation of these statistics leads to underestimates, since in data voids the analysis will reflect the forecast background used. This is counteracted by the use of 12 hour forecasts instead of 6 hour, and also by arbitrarily adding  $9K^2$  to the temperature variance since surface temperatures were not analysed. Table 1 shows how these estimates matched the actual statistics. It shows, for data accepted by the program, the mean square deviation of observations from background, and the mean assumed observation and background error variances, classified by observation types and area. Coastal land reports are those from SYNOPs with a station



elevation less than 100m which are within 150 km of a ship or buoy observation (this is the limit used in the buddy check routine). If the estimates are good, and the checking was good, and the assumption of no correlation between observation and background error is valid, then they should approximately satisfy

$$(o-b)^2 = E^2 + F^2 \quad 39$$

It is clear from table 1 that many of the estimates were too large. For temperature and pressure most of the mean background errors alone were greater than the mean squared deviations. There is similar evidence that the assumed observational errors for ocean weather ship winds and temperatures were too large.

As discussed more thoroughly by Lorenc (1986), the partition of errors between  $E^2$  and  $F^2$  is somewhat arbitrary, depending on where one assigns errors of representativeness. The error correlation models used must however be consistent with the partition chosen. We assumed that observational errors were uncorrelated, while the background error correlation used was taken from the 12 hour forecast error fields of surface pressure used for the variances. These fitted quite well the simple second order autoregressive model suggested by (among others) Thiebaut (1985).

$$\mu_{ij} = (1 - r_{ij}/c) \exp (- r_{ij}/c) \quad 40$$

where  $\mu_{ij}$  is the correlation,  $r_{ij}$  is the distance between points, and  $c = 400$  km.

There is less experience in quantifying estimates of the prior probability of gross error. This has to be done for  $P(G_p)$ ,  $P(G_s)$ ,  $P(G_u)$  and  $P(G_T)$ , for each observation type, for sub-classifications corresponding to different settings of the SDB flags (see section 5a). Here we shall only present results for errors affecting surface pressure

$$P(G_p \text{ or } G_s) = 1 - P(G_p) P(G_s) \quad 41$$

Table 2 shows prior and mean final estimates of this, and the percentage finally rejected (ie. with final probability greater than  $1/2$ ), for various types and sub-classifications. The final probabilities calculated by the program depend on the assumed background and observational errors, shown in table 1, and on the assumed probability distribution of gross errors, discussed below. Where these are believed to be accurate, or where there is a very clear change in probabilities in table 2, the final probabilities can be used to provide new estimates. For instance it seems clear from table 2 that too much importance has been attached to the sequence test for pressure; for ships that passed it too low a probability was assumed, for ships and coastal land reports that failed it too high a probability was assumed. Physical arguments can be advanced to support this. Ships passing the test might still have a large calibration error (as seen in some case studies). A failure of the test might be due to an error in the pressure tendency, or an undetected error in the earlier observation, rather than the current pressure datum. Other rows in table 2 confirm the general reliability of land reports and of moored buoys, and also the unreliability of drifting buoys. As one might hope, ocean weather ships



were too reliable for the percentages rejected to be significant. Out of 473 reports, 1 failed the position sequence, 1 failed the pressure sequence, 2 were rejected in the comparison with the background, and 1 was rejected in the comparison with neighbours.

There is even less experience in quantifying prior estimates of the probable distribution of gross errors, yet this is necessary before we can attempt to decide whether a datum comes from such a distribution rather than from the normal distribution of observational errors. In the simple model proposed in section 3, the gross error distribution is specified by a single parameter  $k$ , the probability density within the range of plausible values. We used an ad hoc relationship between  $k$  and an estimated climatological variance, which for surface pressure gave  $k_s = 0.043 \text{ mb}^{-1}$ . Figure 8 shows a histogram of the deviation from the background of all rejected ship surface pressures. Also plotted are a Gaussian distribution with the same mean and variance, and the assumed  $k_s$ . The distribution is clearly far from Gaussian. The similar histogram for accepted surface pressure observations (not shown) was nearly Gaussian, with standard deviation 1.85 mb. Within the range of plausible values (ie about  $\pm 3$  standard deviations) of the latter, it can be seen from figure 8 that the assumed  $k_s$  is indeed a reasonable approximation. Note that getting an appropriate estimate for  $k$  is important; its direct impact can be seen in (20) and figure 5. To test the effect of reducing  $k$ , a run was made with all values multiplied by 0.6. This resulted in a mean final probability of gross error for all variables and a total number of data rejected which were 0.8 of those from the standard run.

As well as providing statistics for error estimates, the OPD is also useful for monitoring the behaviour of the checking algorithms, and the relative importance of the various checks. As an example table 3 shows the relationships of the estimates of the probability of gross error affecting surface pressure (ie either  $G_s$  or  $G_p$ ) after interpretation of SDB flags (prior), after comparison with the background, and finally after comparison with neighbours within 150 km (buddies). Table 3a summarizes the effect of the background check. The range of probabilities has been divided into eight equal intervals; only the extreme intervals are occupied by the prior probabilities. The first row shows the mean prior probability for each interval, and the second the mean probability for the same data after the background check. The bottom row shows the total number in each prior category, and the intermediate rows how these are distributed amongst the posterior probability categories. We see that agreeing with the background is not sufficient to reduce the probability of gross error below 50% for data initially assigned high probabilities, although some have it considerably reduced. Disagreement on the other hand can change the probability from small to large.

Table 3b in a similar way summarises the effect of the buddy check. The columns correspond to the rows of table 3a. The rows correspond to the final probabilities after the buddy check. The larger values down the diagonal show that many ships had their probabilities unchanged, usually because they had no near neighbours. The buddy check was rather more symmetrical in its effect; agreement with neighbours was enough to restore confidence in a datum. Table 3c shows the combined effect of the two checks.



It is clear from table 3b that in the interaction of the two checking algorithms the full range of available probabilities is used. Any attempt to summarize the results of the first check as a simple reject/accept flag would change and presumably degrade the final results. However in the final results most data are in the extreme categories; the fraction of borderline decisions is small.

Table 4 repeats the same information in a more summarised form, with only two probability categories, corresponding to those used finally for acceptance and rejection. We see that 3248 data were rejected by the SDB checks, 2531 by the background check, and 1637 by the buddy check. These figures add up to greater than 6969, the number finally rejected, because out of the first two 238 and 209 respectively were reinstated by the buddy check. Tables 5 and 6 give similar statistics for wind and temperature, and tables 7, 8 and 9 for coastal land reports.

## 7. CASE STUDIES

Case studies were chosen to test the effectiveness of the statistical quality control scheme described in the preceding sections. When discussing the results, we will refer to this scheme as the "test scheme". The test scheme was run in parallel with the current operational system. This relies heavily on a human "intervention" forecaster in the Central Forecast Office (CFO), with simple interactive computer aids such as comparison with the background, to flag or correct erroneous reports. Remaining data are subjected to a two stage automatic check using (20) with  $T^2 = 12$ . Firstly, in mode 1, observations are compared with the forecast background. Secondly, in mode 2, they are compared with an analysis using up to 7 nearby observations not flagged in mode 1.

Detailed results from a run of the test scheme for 12 GMT 3rd October 1985 were compared with those made operationally. The results are discussed in section 7(a) below. An extended trial was held in the Central Forecast Office during the week 20th-28th November 1985, with ten cases being compared in detail. Forecasters were asked to assess all ship observations given a probability of error  $\geq 30\%$  by the test scheme for mean sea level pressure and wind, and to state whether they considered the observation to be correct or incorrect. These results will be discussed in section 7(b).

To make the assessment easier, ship observations of mean sea level pressure and wind will be regarded as being rejected by the test scheme if the probability of error is greater than or equal to 50%, and as borderline if the probability of error lies between 30 and 49%. Temperature flags were not assessed.

### a) Case Study from 12 GMT, 3rd October 1985

The observations listed as rejected or borderline by the test scheme were compared with operational rejections. The number of observations used in the test quality control program differed slightly from that used operationally. We will be concerned only with those ship observations used in both schemes. The results are shown in Table 10. Both schemes listed a similar number of observations, but there were 49 different important decisions made during the analysis. The 23 observations rejected by the test scheme but passed operationally are analysed in Figure 9.



This diagram shows that there are one or two problems regarding rejection of observations due to synoptic data bank flagging. If a ship observation is flagged by the synoptic data bank, it is given a high probability of error, usually  $\geq 90\%$ , which is only reduced to about 70% after the background check even if the observation agrees almost exactly with the background pressure. Hence the synoptic data bank flagging will be fatal unless the observation is backed by buddies. The lone ship flagged for position was situated very close to the African coast and probably was not checked by CFO. Of the 8 ship observations flagged for pressure, 3 had obvious errors; e.g. 10.0 mb, which were corrected by the forecasters. In two other cases, the mean sea level pressure was correct but the pressure tendency was wrong. An example is the ship GPAJ which at 06 GMT reported 1025.6 mb and at 12 GMT reported 1027.6 mb with a pressure tendency of 10.1 mb in 3 hours. These obvious errors can be more easily detected and corrected by a human than by an automatic system. The remaining three observations in the S.D.B category are more difficult to judge. The mean sea level pressure observations were close to the background values and looked correct, whereas the 06 GMT observations looked doubtful. An example is the ship URHH, whose reported pressure at 12 GMT differed from the background by 0.9 mb. This, and the two previous reports were:-

00 GMT	POSITION 38.6N 149.6E	PMSL	1017.4 mb	Tendency 3.6 mb
06 GMT	POSITION 37.2N 150.0E	PMSL	1028.7 mb	Tendency 0.0 mb
12 GMT	POSITION 36.5N 152.4E	PMSL	1020.2 mb	Tendency 2.2 mb

The 06 GMT pressure observation seems to be 10 mb in error. We believe the operational judgement to be correct in these nine cases. It shows that the probability of error, given after the synoptic data bank flags an observation, may be set a little too high.

The seven observations rejected by the test scheme for pressure following the background check, had typically a 70% probability of error and a deviation from the background of 10 mb. These are examples of the background itself being incorrect. The human analyst is better at recognizing synoptic situations where the background is likely to be incorrect. Two examples will be discussed in section 7(b).

The rejections for wind were mainly due to the ship's wind speed being stronger than that from the background and neighbouring observations. Two out of the six cases looked obviously wrong.

The 26 observations, rejected by the operational quality control but passed by the test scheme, may be divided into three main sections:-

(i) those pressures considered to be incorrect.  
e.g. Three ships, DNFV, DHNE, ZCKE situated on the western side of a mid Atlantic depression, were rejected operationally because they were considered to be 4 to 7 mb too high. However, the ship's pressures fitted the background field closely and were passed by the test scheme. The history of these three ships between August and October 1985 showed that they were reliable during this period. Also the pressures could be analysed as shown in figure 10. In these three cases, the test scheme decision was probably correct. Both schemes rejected the pressure from the ship VRJH to the east of the depression. Other observations to the south



are obviously erroneous, given the density of data in this case. They illustrate the errors that can occur and the difficulties there would be in detecting them with a less dense distribution.

(ii) Ships rejected by the intervention team which had a history of unreliability or of bias in their observations.

(iii) some ships were rejected by the intervention team even though their deviations from the background were small, in slack situations. For instance a ship observation 1-2 mb low could be rejected near the centre of a high.

The buddy check is a powerful tool in the automatic quality control of observations. Figure 11 shows a case where two quality control judgements were reversed following the buddy check. Ship UIJQ was rejected after the background check because its pressure was 6 mb higher than the background value, whereas ship 3EWH was accepted because its pressure was only 2.5 mb lower than the background. However, the buddy check showed that the background pressure field was 3-4 mb too low in this region. The quality control decision was reversed following the buddy check, with UIJQ being reinstated and 3EWH being rejected. This decision agreed with that made operationally. Ship 3EWH is a good example of the utility of having available an extended history of a ship's performance, in order to detect systematic errors. Table 11 shows every occurrence of the ship in the OPD generated during the extended trial from August to October 1985. (Note that only 0 and 12 GMT data were checked, and that a few times were missed because of operational difficulties). Clearly 3EWH was reporting pressures consistently several millibars too low. On several days it reported at 06 as well as 12 GMT, passing the SDB sequence test with the pressure tendency at 12 GMT. On these occasions the test program assigned only a 3% prior probability of gross error affecting pressure, and the background comparison could not with confidence flag it as erroneous. On most occasions the buddy check could do so, but on the 4th October, the day after that shown in figure 11, the ship was in the Atlantic and had no buddies. A human seeing its record would recognise it as erroneous; the test system had no algorithms programmed to do this.

Figure 12 shows another example of a reversal in the quality control judgement following the buddy check. Ship UVUR's pressure fitted the background pressure field closely and at this stage had only a 2% probability of error. However the buddy check showed the background pressure was 2 to 3 mb too low in the vicinity of this ship. Following the buddy check, the probability of error was raised to 99% and UVUI was rejected. Again, this decision agreed with the operational one.

Figure 13 shows a difference in quality control judgement between the operational and test schemes. Operationally, the oil rig platform at 60.9N 0.9E, reporting 994.5 mb was rejected. The test scheme rejected the oil rig at 61.3N 1.5E, reporting 991.8 mb, after the buddy check showed its pressure to be 2 to 3 mb too low. The platform at 61.2N 1.1E, reporting 992.2 mb, was not rejected, but was assigned a 22% probability of error.

On occasions, in data-rich areas, the buddy check can be stringent, with the rejection of an observation which is less than 2 mb in error. A single buddy can cause the reversal of a quality control decision.



b) CFO Trial 20-28th November 1985

Ten analyses were selected during this period and a detailed comparison made between the test automatic quality control decisions and those made by human quality control. The human judgement of the test scheme's results is shown in table 12. For rejected data there was agreement between the automatic test quality control system and forecasters on 81% of occasions for mean sea level pressure and 63% of occasions for winds. For the data found to be borderline, but not rejected, the forecasters agreed with 50% of the pressure decisions and 67% of the wind decisions. Considering the pressure quality control decisions, the forecasters thought that 43 observations, given a probability of error  $\geq 50\%$  by the test scheme, were, in fact correct. Of these 43, 13 gained a high probability of error due to synoptic data bank flagging, 15 gained a high probability of error following the background checks and 15 following the buddy checks. Similarly of the 47 disagreements with wind decisions, 18 gained a high probability of error due to synoptic data bank flagging, 25 due to background checks and 4 due to buddy checks.

The results show that the forecasters were twice as likely to disagree with automatic quality control wind decisions than pressure decisions. The problems with the synoptic data bank checks have already been described and these results reinforce the opinion that the initial probability of error following a synoptic data bank flagging is set a little too high. Problems with the background error checks generally occur when the background itself is incorrect and the observation correct. The 19 disagreements with buddy check decisions may indicate that the check is too stringent on occasions.

Table 13 shows the Test Scheme's assessment of the ship observations rejected by human intervention during the CFO trial. 77 observations, assigned a probability of error  $< 10\%$  are of most concern, since these would have passed the automatic quality control scheme easily. Of these 77, 22 were rejected operationally because they were clearly 2 to 3 millibars incorrect near the centre of an anticyclone or in a col; 28 were rejected due to the ship's pressure being considered biased or unreliable (these ships may have been on an unofficial black-list); the remaining 27 were difficult to assess since the observations of pressure fitted the operational analysis and some may have been rejected in error.

It has already been stated that the forecasters are better at recognizing synoptic situations in which the background pressure is likely to be incorrect. In such cases, observations which differ greatly from the background may be correct, but would be rejected by the automatic scheme. This kind of situation is most likely to occur near a deepening depression. The following examples illustrate two interesting cases when the test scheme rejected observations near the centres of depressions.

(i) Figure 14 shows the 12 GMT analysis of a mid-Atlantic depression on 28th November. Two ships, situated near the centre of the depression were rejected by the test scheme because of a large background error. Table 14 gives the details. There were no buddies to assist the quality control judgement and both ships were rejected. However figure 14 shows that both observations were correct and used in the operational analysis.



(ii) Another interesting case concerned the analyses at 00 GMT and 12 GMT of tropical storm Kate centred near the coast of Carolina on November 23rd. Table 15 shows the details of the quality control decisions made by the test scheme on three ships in the vicinity of the storm. Buddy checks saved four of these important pressure observations from final rejection.

## 8. DISCUSSION AND CONCLUSIONS

This work has set out an objective method for deriving rejection tolerances in quality control checks which do not give completely certain results. Similar variable tolerances (empirically derived) are already in use in the operational quality control scheme at the Meteorological Office. They avoid an unfortunate effect which occurs when (20) is used with a constant T: specifying a lower observational error variance for reliable reports such as ocean weather ships makes them more likely to be rejected. In the objective method (21) relates the tolerance to statistics on the percentage of bad reports.

The work also shows how to combine results from independent checks, by estimating and modifying probabilities of gross error. The prototype scheme for quality control of ship data demonstrated the utility of this approach for two checks, against the background and against neighbours.

In section 6 we demonstrated that the necessary statistical estimates for the objective quality control could be checked and updated using an observation processing database (OPD) containing records of data, background, and quality control results. Models of error characteristics and distributions are required, and these are best devised using physical knowledge of the observing or forecasting process, but the error model parameters can be updated from the practical results, and careful monitoring of results can help suggest areas for improvement to the models.

Comparison of the automatic prototype with careful human quality control revealed two important categories where the human could do better:-

(i) The human could recognize synoptic situations where the background was likely to be significantly in error, for instance near developing depressions. Observations with large deviations from the background in such areas could well be correct, whereas similar deviations in other quieter areas would be an almost certain indication of gross error.

(ii) Many ships had a history of unreliability or frequent bias in their data. A knowledge of this and a study of a sequence of reports could help in detecting gross errors.

Code for the prototype program was relatively simple to design and write, and well within the capabilities of a modern computer system to run routinely. However during the trial many areas for enhancement have been suggested. These include:-

(a) Calculation of background error estimates dependent on the local synoptic situation, as described in (i) above. Allow similar effects in the observational error model, for instance letting the wind error increase with speed.



(b) Use of a complete record of recent reliability of individual ships to maintain a blacklist of unreliable ones, to compensate for calibration errors, and to do a more thorough check on movements as in (ii) above.

(c) Integration with an interactive scheme for human monitoring.

(d) Extension to other observed variables, eg humidity and sea surface temperature.

(e) Extension to other types of observation.

(f) Use of physical knowledge about the way gross errors actually occur (eg by transposition of digits in transmission) to improve the model of the distribution of gross errors, and to check for specific types of error.

(g) Rapid automatic notification of errors to observers, so they can be corrected at source.

To achieve a steady enhancement in these areas, while maintaining a system simple enough for easy monitoring and efficient regular running, is a challenge for the future.



APPENDIX Derivation of equations for buddy check.

Bayes theorem gives

$$P(G_1 | O_1 \Omega O_2) = P(O_1 \Omega O_2 | G_1) P(G_1) / P(O_1 \Omega O_2) \quad A1$$

$$P(G_1 | O_1) = P(O_1 | G_1) P(G_1) / P(O_1) \quad A2$$

If there is a gross error in observation 1, then we are assuming that the reported value is random, and independent of the value from observation 2, so

$$P(O_1 \Omega O_2 | G_1) = P(O_1 | G_1) P(O_2 | G_1) \quad A3$$

The value from observation 2 is independent of gross errors in observation 1, so

$$P(O_2 | G_1) = P(O_2) \quad A4$$

Substituting A3 and A4 in A1, and using A2 to eliminate P(G), gives equation (36):-

$$P(G_1 | O_1 \Omega O_2) = P(G_1 | O_1) P(O_1) P(O_2) / P(O_1 \Omega O_2) \quad A5$$

To evaluate the factor in (A5), we need to split  $P(O_1 \Omega O_2)$ , remembering that  $G_1$  and  $G_2$  are assumed to be independent:-

$$\begin{aligned} P(O_1 \Omega O_2) &= P(O_1 \Omega O_2 | G_1 \Omega G_2) P(G_1) P(G_2) + P(O_1 \Omega O_2 | G_1 \Omega \bar{G}_2) P(G_1) P(\bar{G}_2) \\ &\quad + P(O_1 \Omega O_2 | \bar{G}_1 \Omega G_2) P(\bar{G}_1) P(G_2) + P(O_1 \Omega O_2 | \bar{G}_1 \Omega \bar{G}_2) P(\bar{G}_1) P(\bar{G}_2) \end{aligned} \quad A6$$

$O_1$  and  $O_2$  are assumed to be independent if either of them has a gross error, so all except the last term in (A6) can be factorized:-

$$\begin{aligned} P(O_1 \Omega O_2) &= P(O_1 | G_1) P(O_2 | G_2) P(G_1) P(G_2) + P(O_1 | G_1) P(O_2 | \bar{G}_2) P(G_1) P(\bar{G}_2) \\ &\quad + P(O_1 | \bar{G}_1) P(O_2 | G_2) P(\bar{G}_1) P(G_2) + P(O_1 \Omega O_2 | \bar{G}_1 \Omega \bar{G}_2) P(\bar{G}_1) P(\bar{G}_2) \end{aligned} \quad A7$$

Note that we have also assumed that  $O_1$  is independent of  $G_2$  etc.

We now eliminate  $P(G_1)$  from (A7) using (A2), similarly eliminate  $P(G_2)$ ,  $P(\bar{G}_1)$ ,  $P(G_2)$ , and divide by  $P(O_1) P(O_2)$ , giving:-

$$\begin{aligned} \frac{P(O_1 \Omega O_2)}{P(O_1) P(O_2)} &= P(G_1 | O_1) P(G_2 | O_2) + P(G_1 | O_1) P(\bar{G}_2 | O_2) \\ &\quad + P(\bar{G}_1 | O_1) P(G_2 | O_2) + P(\bar{G}_1 | O_1) P(\bar{G}_2 | O_2) \frac{P(O_1 \Omega O_2 | \bar{G}_1 \Omega \bar{G}_2)}{P(O_1 | \bar{G}_1) P(O_2 | \bar{G}_2)} \end{aligned} \quad A8$$

Finally using  $P(G_1 | O_1) = 1 - P(\bar{G}_1 | O_1)$  etc gives (38).



# REFERENCES

- Atkins, M.A. 1984 'Quality control, selection and processing of observations in the Meteorological Office's operational forecast system'.  
in ECMWF Workshop proceedings "The use and quality control of meteorological observations".
- ECMWF 1985 Workshop on the use and quality control of meteorological observations. 6-9 Nov 1984.  
ECMWF, Shinfield Park, Reading RG2 9AX, UK. 427 pp.
- Gutowski, W.J. 1985 Robust data quality control in a statistical interpolation analysis scheme.  
and Hoffman, R.N. Preprints, 9th Conf on Probability and Statistics in Atmospheric Sciences.  
Am. Met. Soc. 428-433.
- Hollingsworth, A., 1986 The statistical structure of short range forecast errors as determined from radiosonde data. Part I: The wind field. Part II: the covariance of height and wind error.  
and Lonnberg, P. Tellus 38A 111-161.
- Hollingsworth, A., 1985 'The response of Numerical Weather Prediction Systems to FGGE IIb Data, Part I: Analysis'.  
Lorenc, A.C., Tracton, M.S., Quart. J.R. Met. Soc. 111 1-66.  
Arpe, K., Cats, G.,  
Uppala, S., Kallberg, P.
- Lorenc, A.C. 1981 A global three-dimensional multivariate statistical analysis scheme.  
Mon. Wea. Rev. 109 701-721.
- Lorenc, A.C. 1986 Analysis methods for numerical weather prediction.  
(Met O 11 Tech Note No. 214).  
To appear in Quart J. Roy. Met. Soc.
- Met O 12c 1985 Synoptic Data Bank: The quality control applied to surface land and ship reports.  
Met O 12c Tech Note No. 3.
- Thiebaux, H.J. 1985 On approximations to geopotential and wind-field correlation structures.  
Tellus 37A 126-131.



Winkler, R.L.

1985

'Bayesian Inference'.  
in 'Probability, Statistics and  
Decision Making in the Atmospheric  
Sciences eds Murphy and Katz. pp  
461-492 Westview Press.



## List of Figures

Fig 1. Probability density functions for background, observation, and Bayesian analysis, for four different observed values and a Gaussian observational error distribution.

Fig 2. As fig 1 for an observational error distribution equal to a Gaussian plus a small constant.

Fig 3. Mean, median, largest mode, and mode nearest the mean, of the analysis distribution, plotted against normalized observed minus background value, for various prior probabilities of gross error.

Fig 4. Posterior probability of an observation not having a gross error, plotted against normalized observed minus background value, for various prior probabilities of gross error.

Fig 5. Rejection tolerance  $T$ , plotted against prior probability of gross error, for various  $k^2(V_o + V_b)$ .

Fig 6. Variance of posterior analysis distribution, plotted against normalized observed minus background value, for various prior probabilities of gross error.

Fig 7. As fig 6 for the mean square deviation from the analysis value given by (19).

Fig 8. Histogram of observed minus background deviations, for pressure data rejected in the quality control. Also shown are the Gaussian with the same mean and standard deviation, and the assumed gross error probability  $k$ .

Fig 9. Breakdown of ships' data rejected by the test system but not by the operational system, for 12 GMT 3rd October 1985.

Fig 10. Manual analysis of an Atlantic depression at 12 GMT on 3rd October 1985, drawn to fit three ships' observations which were rejected in the operational scheme and accepted in the test scheme. The last three digits of the observed pressures in tenths of a millibar are plotted. Winds are plotted in the conventional manner, with each full fleche representing 10 kts (5m/s). Call signs referred to in the test are shown.

Fig 11. Observations in the English Channel at 12 GMT 3rd October 1985, plotted as in figure 10. Contours show the background PMSL field, from a 6 hour forecast, at 2 mb intervals. The scale is such that the figure covers 670 km square, centred at 49.5N 4.0W.

Fig. 12. As figure 11, centred at 52.5N 3.9E.

Fig. 13. As figure 11, centred at 61.0N 1.2E.

Fig. 14. Manual analysis of a mid-Atlantic depression on 28th November 1985, fitting two ships which were rejected by the test scheme. Plotting conventions as in figure 10.



## List of Tables

Table 1. Mean squared observation minus background differences, mean assumed observational error variances  $E^2$  and background error variances  $F^2$ , for data accepted in the quality control. The errors are averaged for each observation type, for the regions 90°N to 22°N, 22°N to 22°S and 22°S to 90°S.

Table 2. Mean prior and final probabilities of gross error (%), and the percentage actually rejected, of surface pressure data from various observation types classified according to the flags from the SDB checks.

Table 3. Two-dimensional histograms showing the change due to each check in the probability of gross error affecting surface pressure in ship observations. a) Background check. b) Buddy check. c) Background and buddy check. See text for further explanation.

Table 4. As table 3, summarized into 2 categories corresponding to final acceptance and rejection.

Table 5. As table 4 for wind from ships.

Table 6. As table 4 for temperature from ships.

Table 7. As table 4 for surface pressure from coastal land reports.

Table 8. As table 4 for wind from coastal land reports.

Table 9. As table 4 for temperature from coastal land reports.

Table 10. (Top) Number of data from ship observations rejected in the test and operational schemes for 12 GMT 3rd October 1985.

(Bottom) Cross-analysis of rejected and borderline data.

Table 11. History of pressure observations from ship 3EWH, from the OPD of the test scheme.

Table 12. Human assessment of ships' data given probabilities of error greater than 30% by the test scheme during the CFO trial 20-28th November 1985.

Table 13. Test scheme's assessment of ship's data rejected by human intervention during the CFO trial 20-28th November 1985.

Table 14. Ship observations rejected or nearly rejected in error by the test scheme near the developing depression shown in figure 14.

Table 15. Ship observations rejected or nearly rejected in error by the test scheme near tropical storm "Kate".



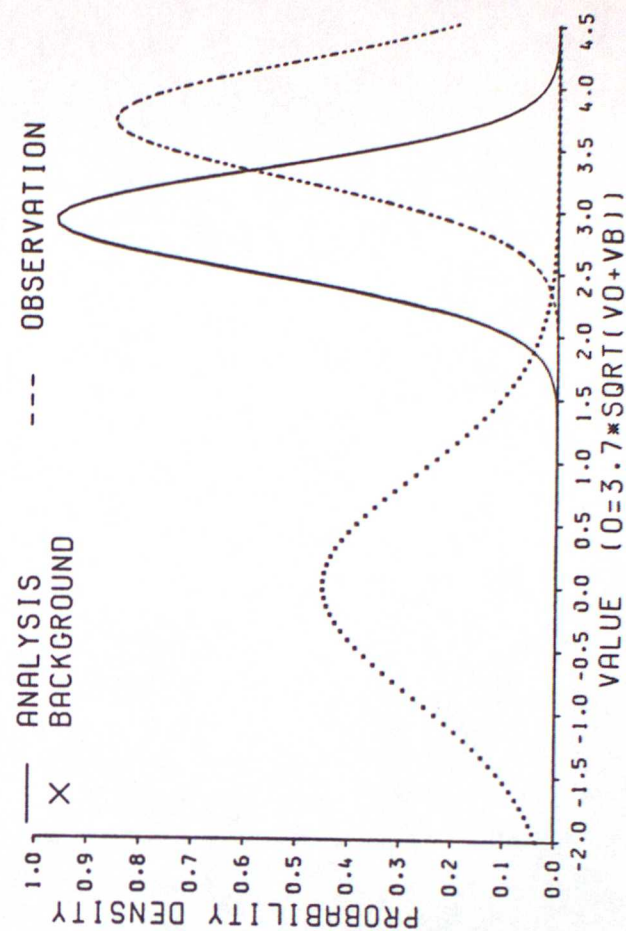
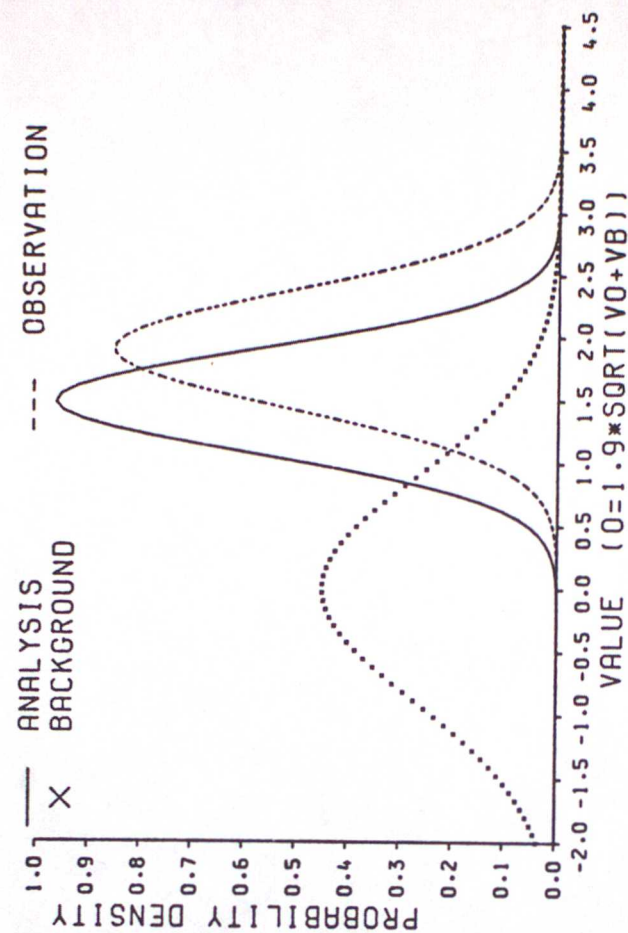
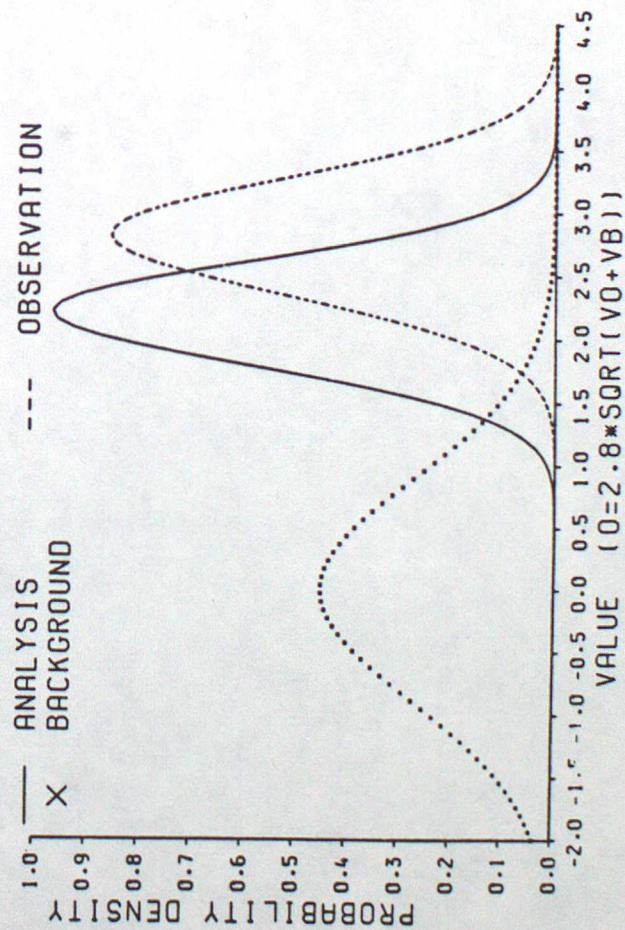
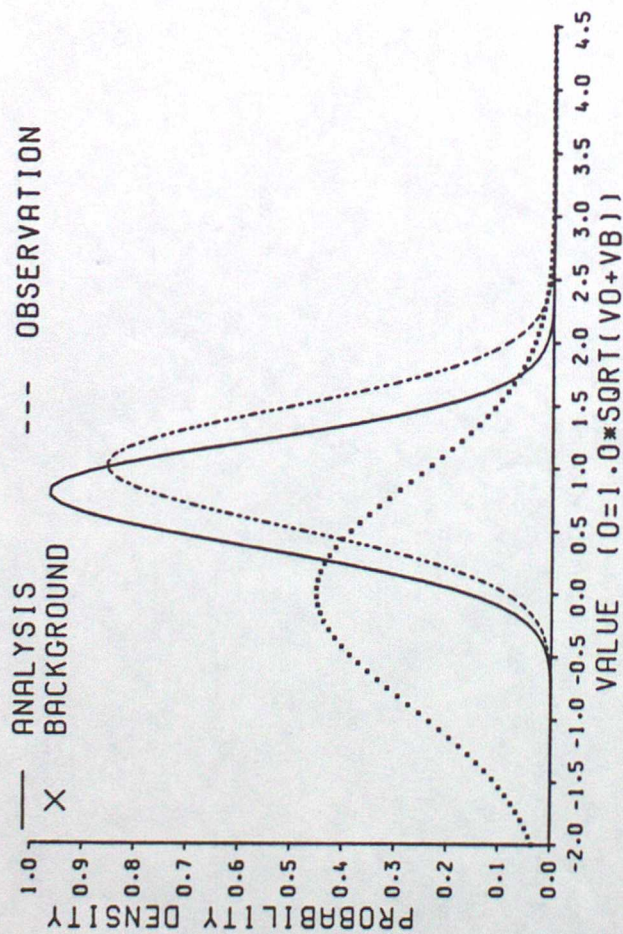


Fig 1. Probability density functions for background, observation, and Bayesian analysis, for four different observed values and a Gaussian observational error distribution.



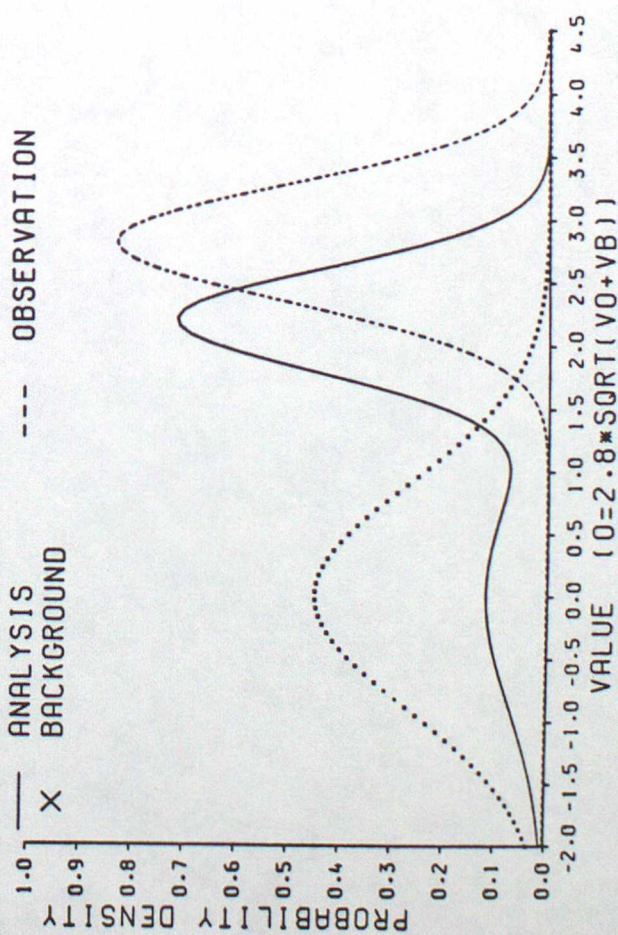
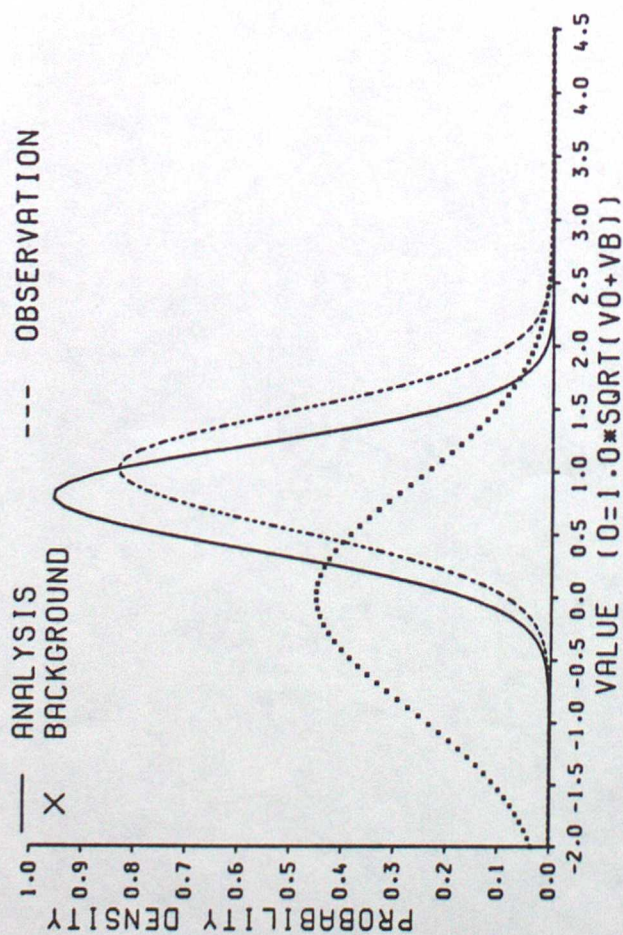
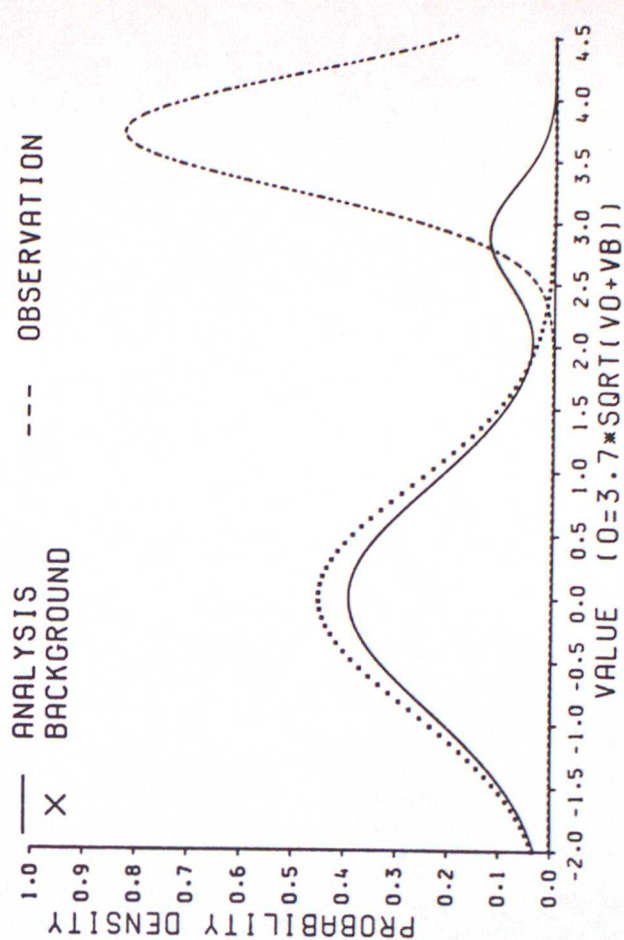
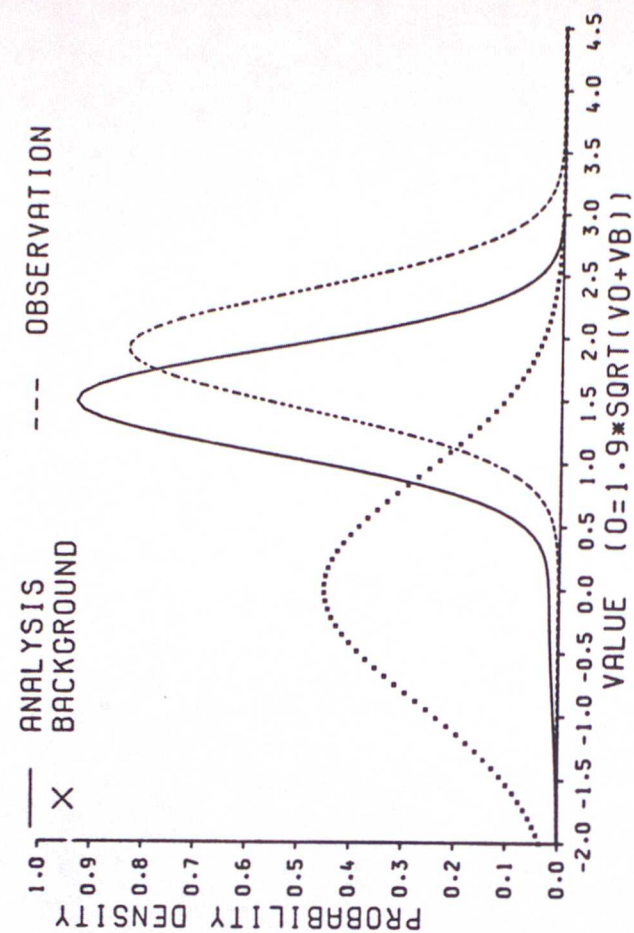


Fig 2. As fig 1 for an observational error distribution equal to a Gaussian plus a small constant.



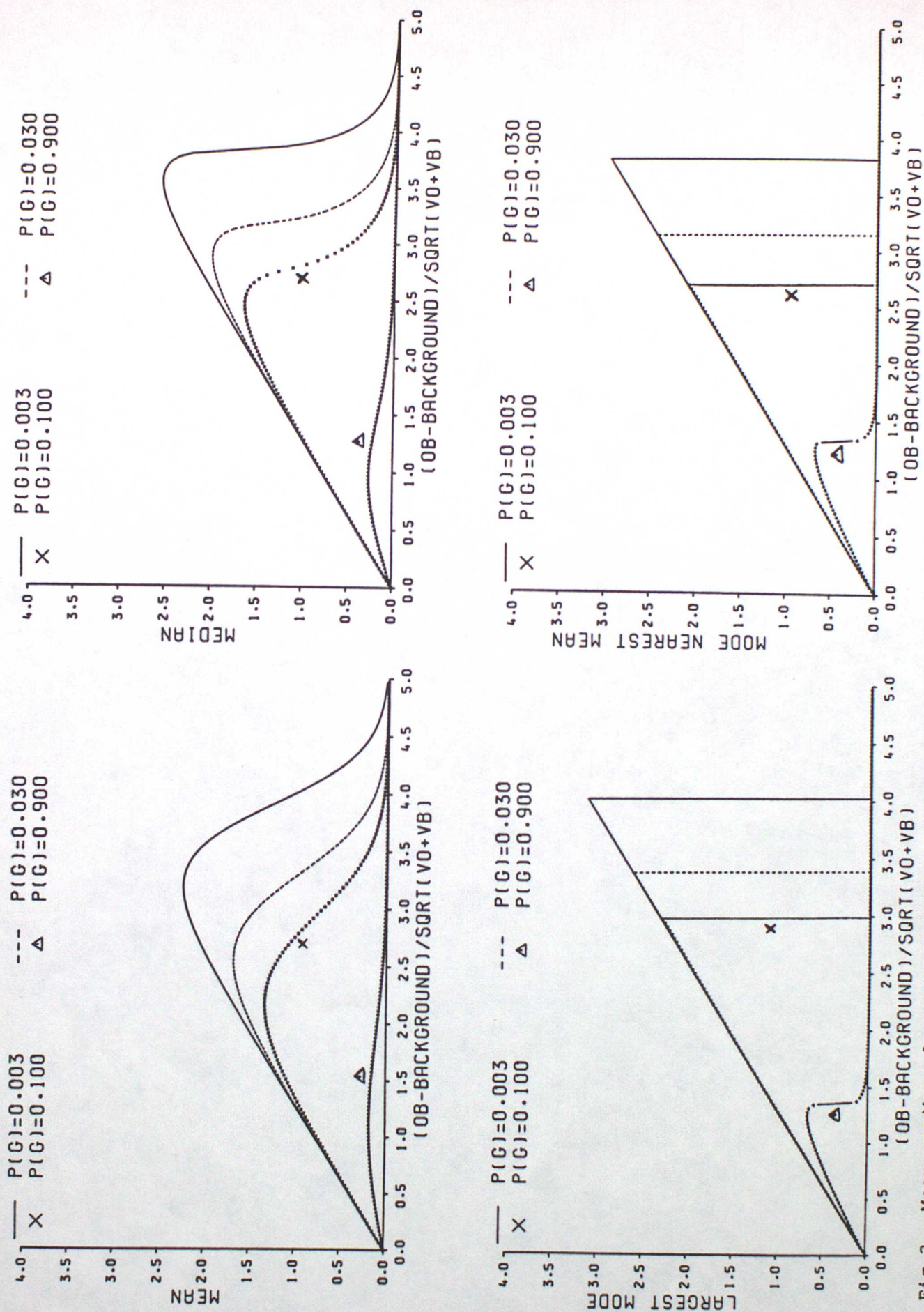


Fig 3. Mean, median, largest mode, and mode nearest the mean, of the analysis distribution, plotted against normalized observed background value, for various prior probabilities of gross error.



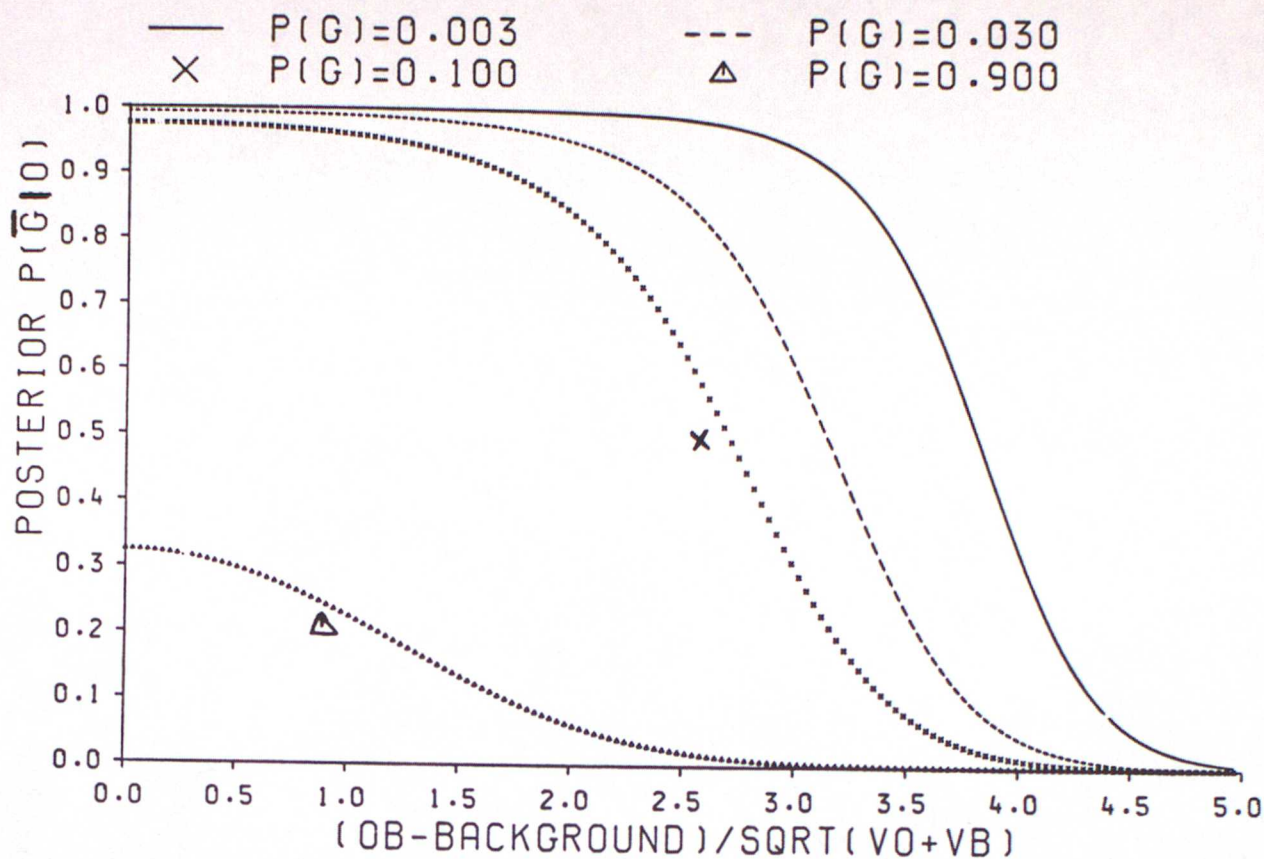


Fig 4. Posterior probability of an observation not having a gross error, plotted against normalized observed minus background value, for various prior probabilities of gross error.

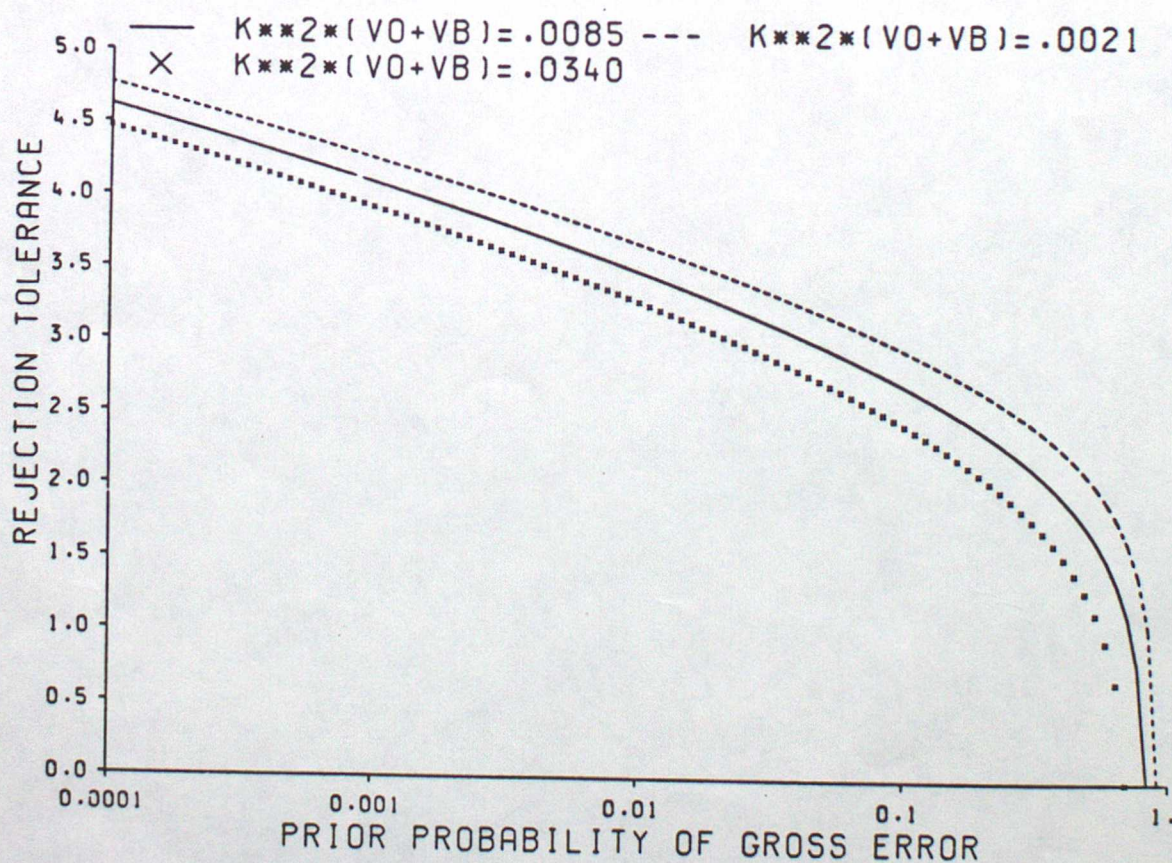


Fig 5. Rejection tolerance  $T$ , plotted against prior probability of gross error, for various  $k^2(V_0 + V_b)$ .



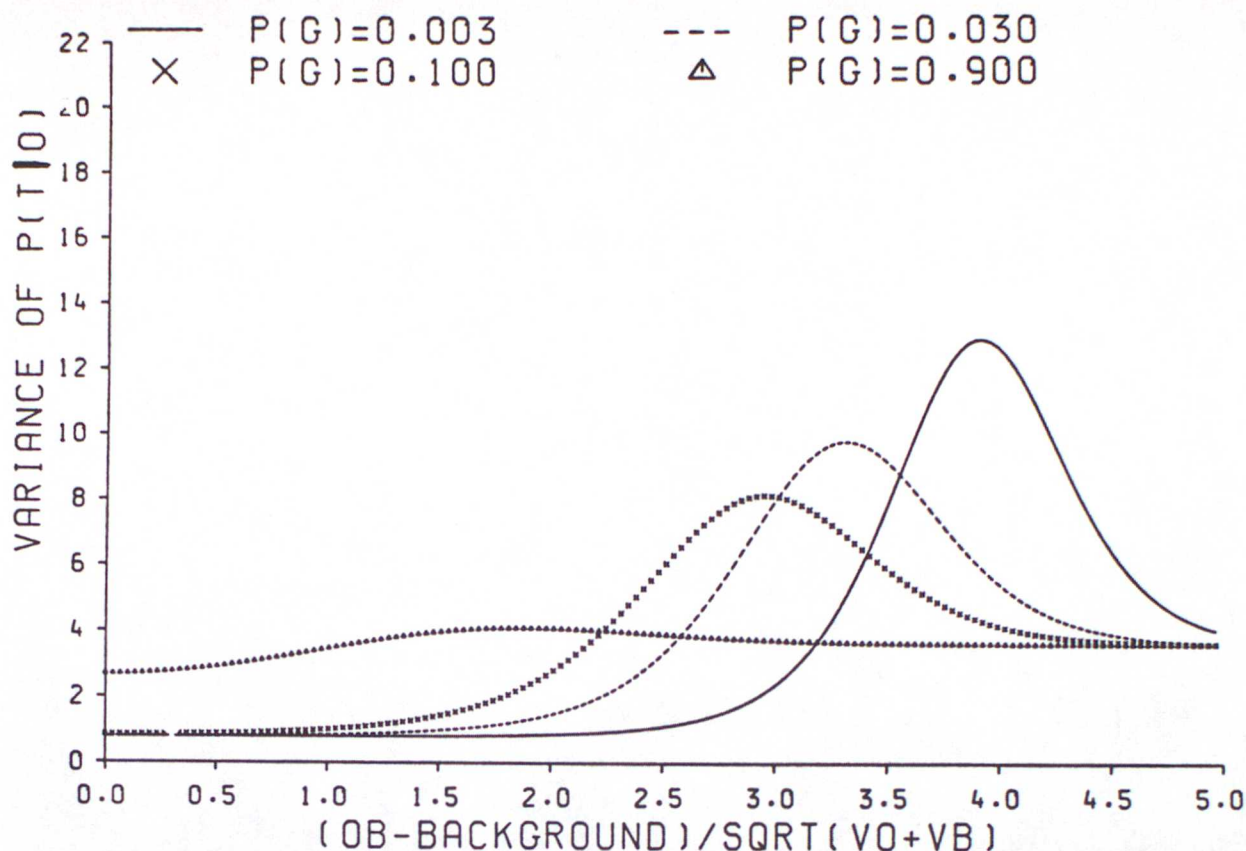


Fig 6. Variance of posterior analysis distribution, plotted against normalized observed minus background value, for various prior probabilities of gross error.

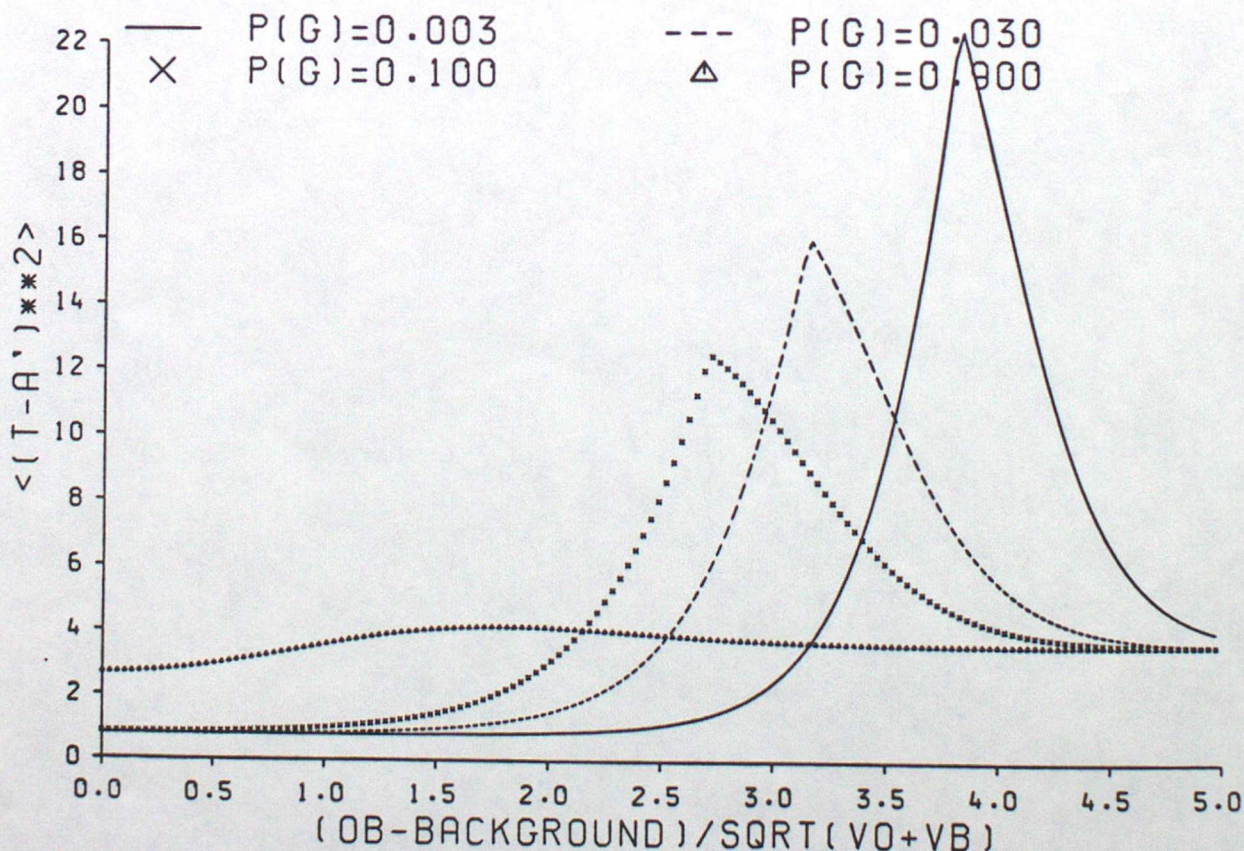


Fig 7. As fig 6 for the mean square deviation from the analysis value given by (19).



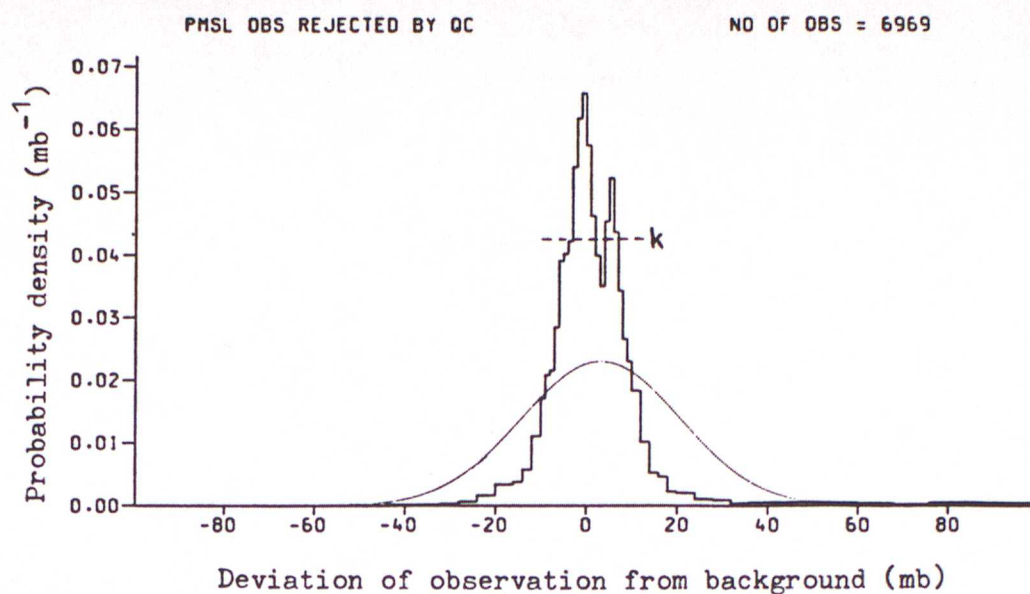


Fig 8. Histogram of observed minus background deviations, for pressure data rejected in the quality control. Also shown are the Gaussian with the same mean and standard deviation, and the assumed gross error probability  $k$ .

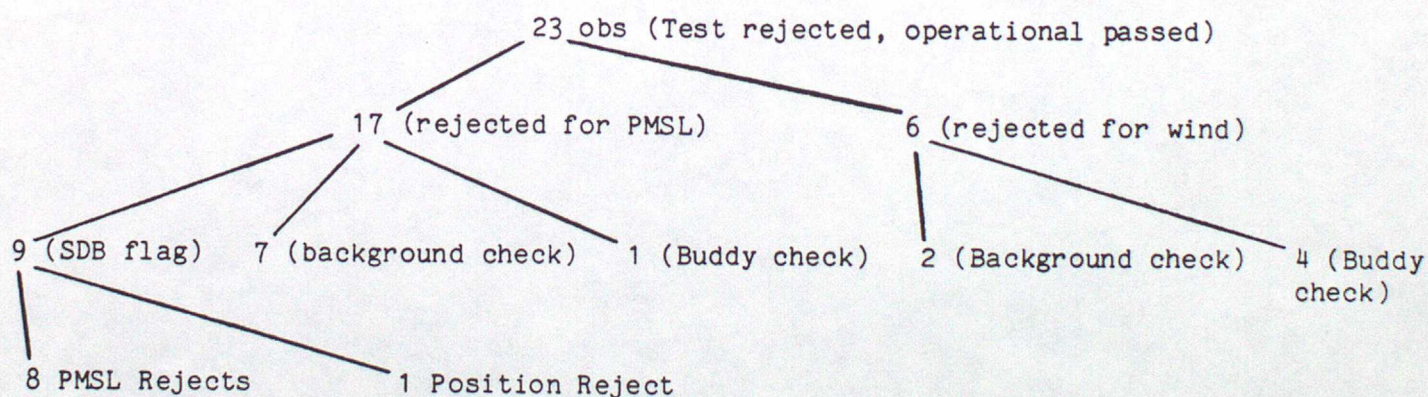


Figure 9. Breakdown of ships' data rejected by the test system but not by the operational system, for 12 GMT 3rd October 1985.



## OBSERVATIONS FOR 12Z ON 03/10/1985

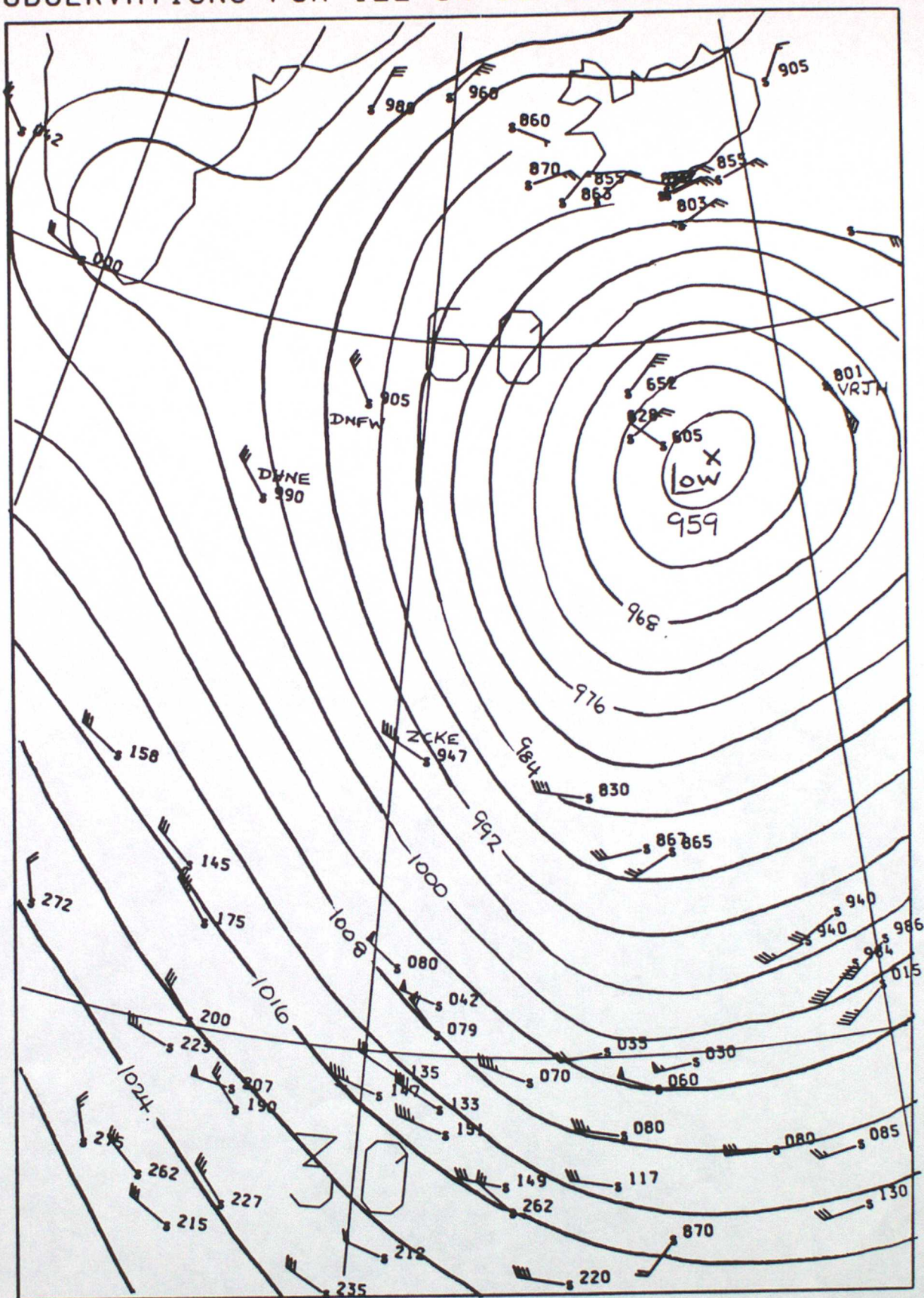


Fig 10. Manual analysis of an Atlantic depression at 12 GMT on 3rd October 1985, drawn to fit three ships' observations which were rejected in the operational scheme and accepted in the test scheme. The last three digits of the observed pressures in tenths of a millibar are plotted. Winds are plotted in the conventional manner, with each full fleche representing 10 kts (5m/s). Call signs referred to in the test are shown.



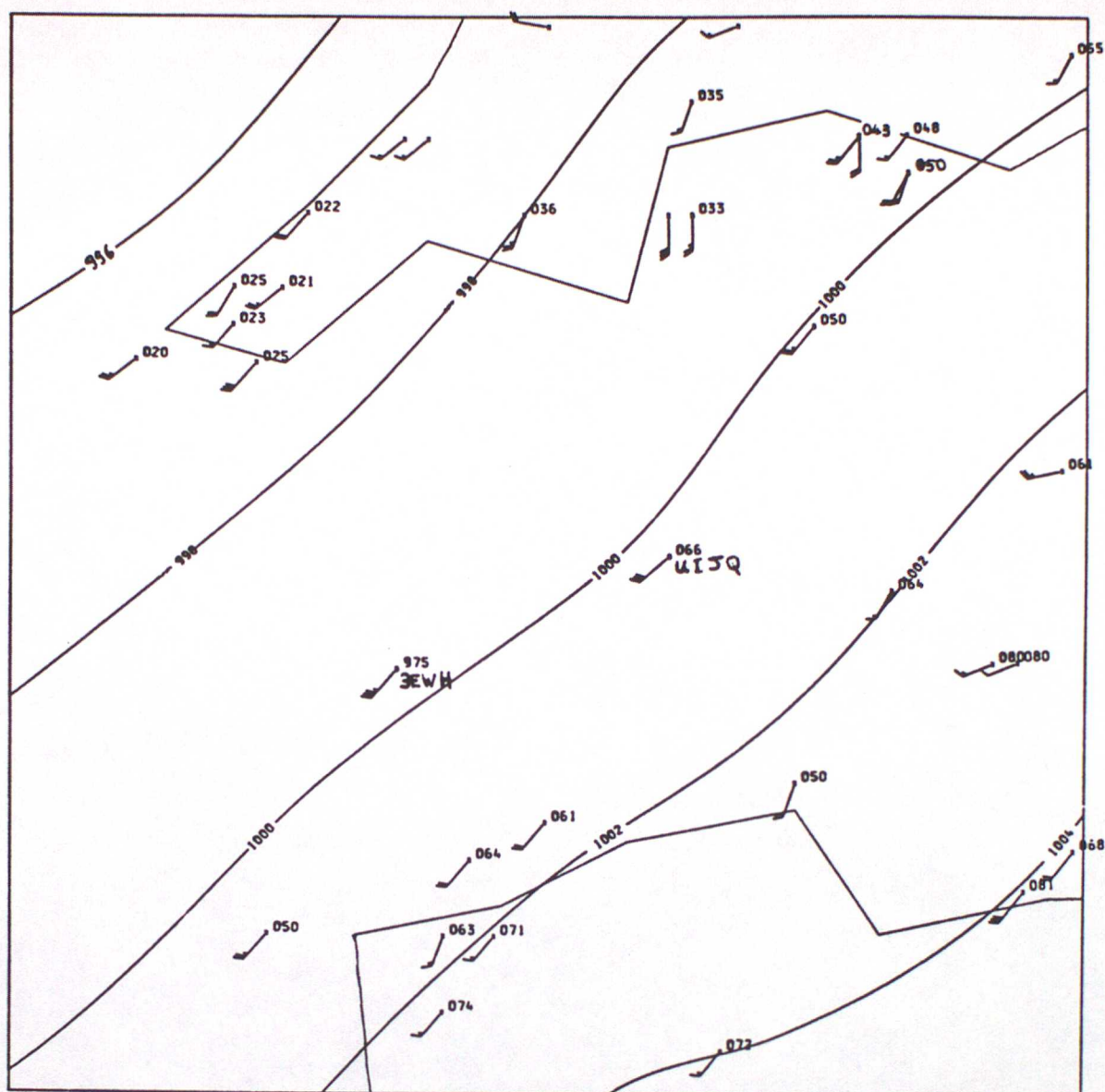


Fig 11. Observations in the English Channel at 12 GMT 3rd October 1985, plotted as in figure 10. Contours show the background PMSL field, from a 6 hour forecast, at 2 mb intervals. The scale is such that the figure covers 670 km square, centred at 49.5N 4.0W.







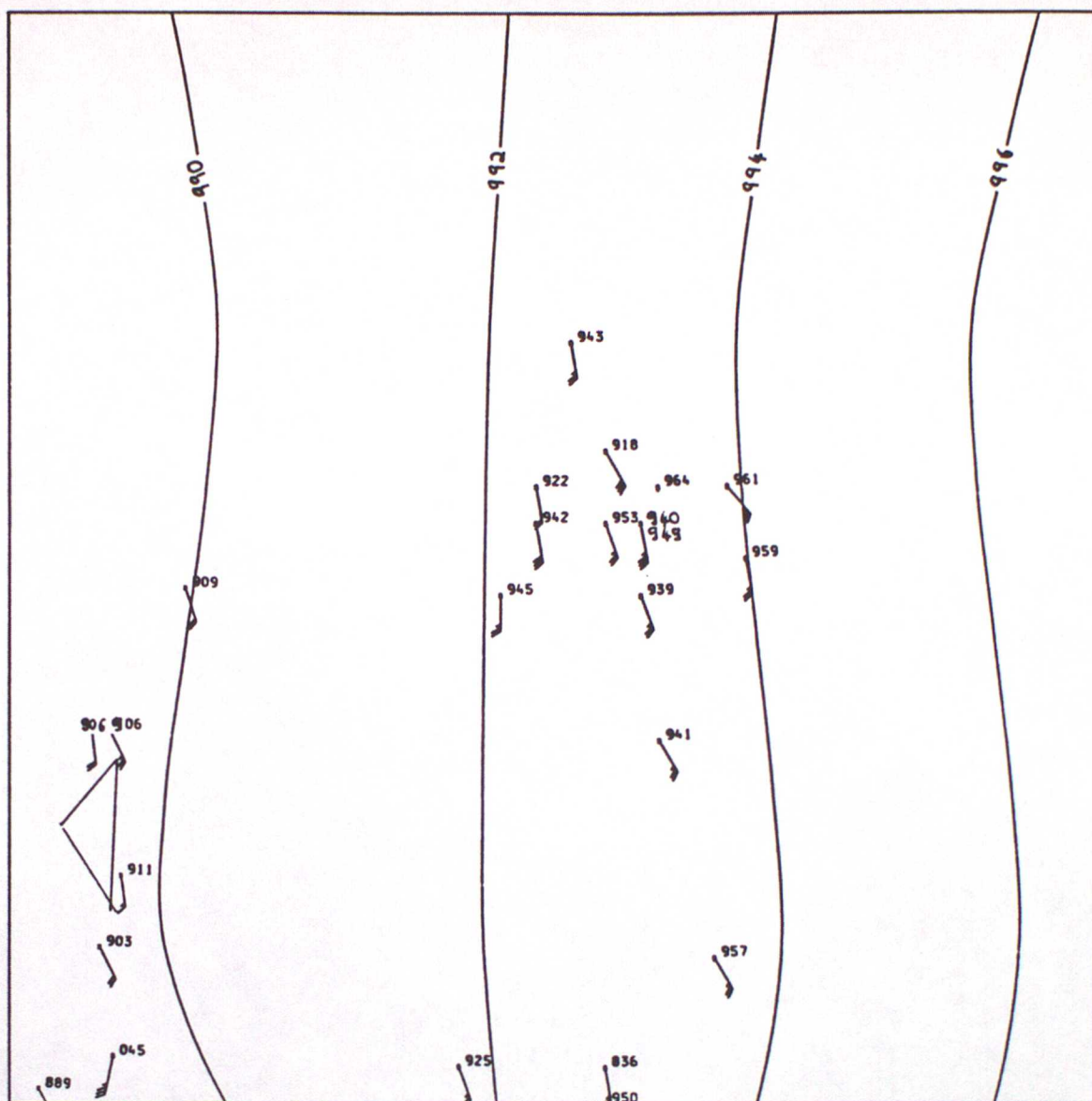


Fig. 13. As figure 11, centred at 61.0N 1.2E.



35



TABLE 1

		No.	PRESSURE (mb) <sup>2</sup>		No.	WIND (m/s) <sup>2</sup>		F <sup>2</sup>	TEMPERATURE (K) <sup>2</sup>		F <sup>2</sup>
			(o-b) <sup>2</sup>	E <sup>2</sup>		(o-b) <sup>2</sup>	E <sup>2</sup>		(o-b) <sup>2</sup>	E <sup>2</sup>	
SHIPS	N.Hem	44228	3.5	1.0	44259	26.3	20.3	9.8	47239	5.9	10.5
	Trop	14352	3.4	1.0	14431	19.9	20.3	9.5	15256	5.0	10.2
	S.Hem	5541	5.3	1.1	5623	31.3	20.3	13.3	6284	6.8	10.7
DRIBU	N.Hem	922	4.2	2.6	94	45.3	25.0	10.1	380	13.0	10.7
	Trop	99	2.5	2.6	20	44.0	25.0	6.3	89	1.8	9.2
	S.Hem	454	10.0	2.6	2	101.5	25.0	17.1	307	5.3	10.2
OWS	N.Hem	503	2.8	1.0	499	15.4	20.3	10.3	505	1.7	9.6
COAST LAND	N.Hem	40809	2.5	1.0	43144	19.7	20.3	6.7	44803	8.3	11.3
	Trop	2608	3.3	1.0	2085	23.2	20.3	10.3	2657	10.4	12.6
	S. Hem	1433	3.7	1.0	1306	40.5	20.3	13.5	1470	13.7	12.9
MOORED BUOY	N.Hem	3656	2.6	1.0	3564	17.7	20.3	9.2	3712	16.4	13.0

Table 1. Mean squared observation minus background differences, mean assumed observational error variances E<sup>2</sup> and background error variances F<sup>2</sup>, for data accepted in the quality control. The errors are averaged for each observation type, for the regions 90°N to 22°N, 22°N to 22°S and 22°S to 90°S.



TABLE 2.					
TYPE	SUB-CLASIFICATION	NUMBER	PRIOR	FINAL	REJECTED
	FROM SDB FLAGS		P(G)	P(G)	%
SHIPS	FAILED POSITION SEQ	1122	92	90	89
SHIPS	FAILED PRESSURE SEQ	872	90	79	88
SHIPS	PASSED SEQ TESTS	23471	3	6	4
SHIPS	NO SEQ. FAILED OTHERS	1154	95	94	99
SHIPS	NO SEQ. PASSED OTHERS	41259	8	10	7
LAND	FAILED PRESSURE SEQ	253	80	54	46
LAND	PASSED PRESSURE SEQ	38803	2	1	0.3
LAND	NO SEQ. FAILED OTHERS	29	95	85	93
LAND	NO SEQ. PASSED OTHERS	3714	2	2	1
BUOY	FAILED PRESSURE SEQ	52	90	96	94
BUOY	PASSED PRESSURE SEQ	3424	3	1	1
DRIBU		1627	10	14	9

Table 2. Mean prior and final probabilities of gross error (%), and the percentage actually rejected, of surface pressure data from various observation types classified according to the flags from the SDB checks.



TABLE 3. SHIPS CHECKED FROM MID-AUGUST TO MID-OCTOBER 1985

(A)											
PRIOR P(G) IN PRESSURE.											
10%	6%	19%	31%	44%	56%	69%	81%	93%			
MEAN	P(G,GIVEN BACKGROUND) IN PRESSURE.										
11%	7%	19%	31%	44%	56%	69%	81%	91%			
TOTAL											
3742=	1526	0	0	0	0	0	0	2216	58%	98%	
1005=	330	0	0	0	0	0	0	675	64%	82%	
642=	301	0	0	0	0	0	0	341	51%	69%	
390=	374	0	0	0	0	0	0	16	10%	56%	
502=	502	0	0	0	0	0	0	0	6%	43%	
974=	974	0	0	0	0	0	0	0	7%	31%	
2679=	2679	0	0	0	0	0	0	0	6%	17%	
62330=	62330	0	0	0	0	0	0	0	6%	3%	
TOTAL	---	---	---	---	---	---	---	---	MEAN	MEAN	
72264=	69016	0	0	0	0	0	0	3248	10%	11%	
(B)											
P(G,GIVEN BACKGROUND) IN PRESSURE.											
11%	3%	17%	31%	43%	56%	69%	82%	98%			
MEAN	P(G,GIVEN BACKG'D & BUDDIES) IN PRESSURE.										
12%	4%	21%	37%	46%	58%	63%	76%	95%			
TOTAL											
4873=	621	177	134	84	85	118	191	3463	78%	98%	
979=	154	43	22	17	17	16	581	129	66%	82%	
596=	106	26	21	8	11	302	95	27	57%	69%	
521=	141	43	26	14	193	58	27	19	42%	56%	
586=	188	52	29	234	17	36	19	11	32%	43%	
949=	288	79	473	23	14	35	20	17	26%	31%	
2293=	730	1355	91	41	12	32	20	12	16%	18%	
61467=	60102	904	178	81	41	45	52	64	3%	2%	
TOTAL	---	---	---	---	---	---	---	---	MEAN	MEAN	
72264=	62330	2679	974	502	390	642	1005	3742	11%	12%	
(C)											
PRIOR P(G) IN PRESSURE.											
10%	6%	19%	31%	44%	56%	69%	81%	93%			
MEAN	P(G,GIVEN BACKG'D & BUDDIES) IN PRESSURE.										
12%	8%	19%	31%	44%	56%	69%	81%	86%			
TOTAL											
4873=	2782	0	0	0	0	0	0	2091	44%	98%	
979=	419	0	0	0	0	0	0	560	55%	82%	
596=	328	0	0	0	0	0	0	268	44%	69%	
521=	430	0	0	0	0	0	0	91	21%	56%	
586=	533	0	0	0	0	0	0	53	14%	43%	
949=	904	0	0	0	0	0	0	45	11%	31%	
2293=	2256	0	0	0	0	0	0	37	8%	18%	
61467=	61364	0	0	0	0	0	0	103	6%	2%	
TOTAL	---	---	---	---	---	---	---	---	MEAN	MEAN	
72264=	69016	0	0	0	0	0	0	3248	10%	12%	

Table 3. Two-dimensional histograms showing the change due to each check in the probability of gross error affecting surface pressure in ship observations. a) Background check. b) Buddy check. c) Background and buddy check. See text for further explanation.



TABLE 4. SHIPS

PRIOR PIG) IN PRESSURE.					
10%	6%	93%			
PIG.GIVEN BACKGROUND)					
11%	7%	91%			
5779=	2531	3248	55%	89%	
66485=	66485	0	6%	4%	
TOTAL	---	---	MEAN	MEAN	
72264=	69016	3248	10%	11%	

PIG.GIVEN BACKGROUND)					
11%	4%	89%			
PIG.GIVEN BACKG'D & BUDDIES)					
12%	5%	86%			
6969=	1637	5332	72%	90%	
65295=	64848	447	4%	3%	
TOTAL	---	---	MEAN	MEAN	
72264=	66485	5779	11%	12%	

PRIOR PIG) IN PRESSURE.					
10%	6%	93%			
PIG.GIVEN BACKG'D & BUDDIES)					
12%	8%	86%			
6969=	3959	3010	44%	90%	
65295=	65057	238	6%	3%	
TOTAL	---	---	MEAN	MEAN	
72264=	69016	3248	10%	12%	

TABLE 7. COASTAL LAND

PRIOR PIG) IN PRESSURE.					
2%	2%	82%			
PIG.GIVEN BACKGROUND)					
2%	2%	73%			
301=	79	222	61%	81%	
42498=	42438	60	2%	1%	
TOTAL	---	---	MEAN	MEAN	
42799=	42517	282	2%	2%	

PIG.GIVEN BACKGROUND)					
2%	1%	81%			
PIG.GIVEN BACKG'D & BUDDIES)					
1%	1%	68%			
303=	102	201	63%	89%	
42496=	42396	100	2%	1%	
TOTAL	---	---	MEAN	MEAN	
42799=	42498	301	2%	1%	

PRIOR PIG) IN PRESSURE.					
2%	2%	82%			
PIG.GIVEN BACKG'D & BUDDIES)					
1%	1%	57%			
303=	159	144	40%	89%	
42496=	42358	138	2%	1%	
TOTAL	---	---	MEAN	MEAN	
42799=	42517	282	2%	1%	

TABLE 5. SHIPS

PRIOR PIG) IN WIND.					
9%	7%	92%			
PIG.GIVEN BACKGROUND)					
11%	9%	92%			
3936=	2041	1895	48%	84%	
67194=	67194	0	7%	7%	
TOTAL	---	---	MEAN	MEAN	
71130=	69235	1895	9%	11%	

PIG.GIVEN BACKGROUND)					
11%	7%	84%			
PIG.GIVEN BACKG'D & BUDDIES)					
10%	6%	82%			
3930=	271	3659	82%	85%	
67200=	66923	277	7%	6%	
TOTAL	---	---	MEAN	MEAN	
71130=	67194	3936	11%	10%	

PRIOR PIG) IN WIND.					
9%	7%	92%			
PIG.GIVEN BACKG'D & BUDDIES)					
10%	8%	91%			
3930=	2050	1880	48%	85%	
67200=	67185	15	7%	6%	
TOTAL	---	---	MEAN	MEAN	
71130=	69235	1895	9%	10%	

TABLE 8. COASTAL LAND

PRIOR PIG) IN WIND.					
3%	3%	95%			
PIG.GIVEN BACKGROUND)					
3%	3%	98%			
194=	162	32	18%	81%	
44189=	44189	0	3%	3%	
TOTAL	---	---	MEAN	MEAN	
44383=	44351	32	3%	3%	

PIG.GIVEN BACKGROUND)					
3%	3%	81%			
PIG.GIVEN BACKG'D & BUDDIES)					
3%	2%	74%			
275=	120	155	56%	83%	
44108=	44069	39	3%	2%	
TOTAL	---	---	MEAN	MEAN	
44383=	44189	194	3%	3%	

PRIOR PIG) IN WIND.					
3%	3%	95%			
PIG.GIVEN BACKG'D & BUDDIES)					
3%	2%	95%			
275=	244	31	13%	83%	
44108=	44107	1	3%	2%	
TOTAL	---	---	MEAN	MEAN	
44383=	44351	32	3%	3%	

TABLE 6. SHIPS

PRIOR PIG) IN TEMPERATURE.					
9%	7%	92%			
PIG.GIVEN BACKGROUND)					
11%	9%	91%			
3613=	1949	1664	46%	84%	
70666=	70666	0	7%	7%	
TOTAL	---	---	MEAN	MEAN	
74279=	72615	1664	9%	11%	

PIG.GIVEN BACKGROUND)					
11%	7%	84%			
PIG.GIVEN BACKG'D & BUDDIES)					
9%	6%	78%			
3523=	395	3128	79%	85%	
70756=	70271	485	7%	5%	
TOTAL	---	---	MEAN	MEAN	
74279=	70666	3613	11%	9%	

PRIOR PIG) IN TEMPERATURE.					
9%	7%	92%			
PIG.GIVEN BACKG'D & BUDDIES)					
9%	7%	83%			
3523=	2013	1510	43%	85%	
70756=	70602	154	7%	5%	
TOTAL	---	---	MEAN	MEAN	
74279=	72615	1664	9%	9%	

TABLE 9. COASTAL LAND

PRIOR PIG) IN TEMPERATURE.					
2%	2%	75%			
PIG.GIVEN BACKGROUND)					
3%	3%	75%			
103=	103	0	2%	76%	
46428=	46428	0	2%	2%	
TOTAL	---	---	MEAN	MEAN	
46531=	46531	0	2%	3%	

PIG.GIVEN BACKGROUND)					
3%	2%	76%			
PIG.GIVEN BACKG'D & BUDDIES)					
1%	1%	64%			
136=	62	74	51%	79%	
46395=	46366	29	2%	1%	
TOTAL	---	---	MEAN	MEAN	
46531=	46428	103	3%	1%	

PRIOR PIG) IN TEMPERATURE.					
2%	2%	75%			
PIG.GIVEN BACKG'D & BUDDIES)					
1%	1%	75%			
136=	136	0	2%	79%	
46395=	46395	0	2%	1%	
TOTAL	---	---	MEAN	MEAN	
46531=	46531	0	2%	1%	

Table 4. As table 3, summarized into 2 categories corresponding to final acceptance and rejection.

Table 5. As table 4 for wind from ships.

Table 6. As table 4 for temperature from ships.

Table 7. As table 4 for surface pressure from coastal land reports.

Table 8. As table 4 for wind from coastal land reports.

Table 9. As table 4 for temperature from coastal land reports.



Test Objective Quality Control Scheme				Operational Scheme		
Probability of error	Judgement	Number of obs			Judgement	Number of obs
≥ 50%	Reject	85	(74 PMSL) (11 wind)	Mode 2 Flag	Reject	19
30%-49%	Borderline	18	(9 PMSL) (9 wind)	Intervention	Reject	81
TOTAL 103				TOTAL 100		

Test Scheme - Rejected, Borderline Obs.				Operational Rejections			
Rejected by both schemes		56		rejected by both schemes		56	
Not used in operational scheme		12		not used in test scheme		18	
Test reject/operational pass		23		Operational reject/test		26	
Test borderline/operational pass		12		pass			
TOTAL		103		TOTAL		100	

Table 10. (Top) Number of data from ship observations rejected in the test and operational schemes for 12 GMT 3rd October 1985.

(Bottom) Cross-analysis of rejected and borderline data.

Time	Date	Position		SDB Seq Test	Prior Prob. Error	Obs-Bkgd (mb)	Bkgd Prob Error	No. of Buds	Final Prob Error
12Z	13/09	16N	18W	Pass	3%	-4.7	41%	1	97%
12Z	20/09	13N	18W		8%	-6.2	94%	1	100%
12Z	21/09	19N	18W	Pass	3%	-3.7	11%	2	98%
12Z	27/09	47N	7W		8%	-6.9	93%	4	100%
12Z	28/09	50N	2W	Pass	3%	-5.8	52%	25	100%
12Z	3/10	49N	5W	Pass	3%	-2.5	5%	16	100%
12Z	4/10	47N	8W	Pass	3%	-4.7	27%	0	27%
12Z	18/10	13N	18W	Pass	3%	-3.1	12%	2	57%
12Z	19/10	18N	18W		8%	-6.1	98%	2	100%

Table 11. History of pressure observations from ship 3EWH, from the OPD of the test scheme.



Mean Sea level Pressure				Winds			
Test Probability		Forecaster Decisions		Test Probability		Forecaster Decisions	
of error	OB is correct	OB is incorrect	Total	of error	OB Correct	OB Incorrect	TOTAL
30-39%	8	7	15	30-39%	19	9	28
40-49%	11	12	23	40-49%	12	6	18
50-59%	5	8	13	50-59%	10	13	23
60-69%	10	13	23	60-69%	8	5	13
70-79%	12	13	25	70-79%	4	7	11
80-89%	5	9	14	80-89%	7	10	17
90-99%	11	145	156	90-99%	18	46	64

Table 12. Human assessment of ship's data given probabilities of error greater than 30% by the test scheme during the CFO trial 20-28th November 1985.

Test Scheme Decisions on obs. rejected by Intervention						
Prob of Error				Number of obs		Decision
0	-	9	%	77	)	Pass
10	-	19	%	19	)	
20	-	29	%	6	)	
30	-	39	%	3	)	Borderline
40	-	49	%	4	)	
50	-	59	%	3	)	Reject
60	-	69	%	7	)	
70	-	79	%	5	)	
80	-	89	%	3	)	
90	-	99	%	73	)	

Table 13. Test scheme's assessment of ships' data rejected by human intervention during the CFO trial 20-28th November 1985.



Call Sign	Position	Time & Date	PMSL OB-BKG	PMSL Final Prob of error	Obs Wind deg kts	Bkg Wind deg kts	Wind Final Prob of error
A8JB	48.6N 35.5W	12Z 28/11/85	-13.4 mb	99	050 37	281 16	99
SDQT	50.1N 35.8W	12Z 28/11/85	-4.4 mb	43	070 24	229 06	65

Table 14. Ship observations rejected or nearly rejected in error by the test scheme near the developing depression shown in figure 14.

Call Sign	Position	Time & Date	PMSL OB-BKG	Bkg Prob of error	Final Prob of error	Bkg Wind deg kts	Obs Wind deg kts	Bkg Prob of error	Final Prob of error
SHIP	34.1N 75.9W	00Z 23/11/85	-5.7mb	98	41	192 24	260 40	99	99
WXWL	33.3N 76.5W	00Z 23/11/85	-6.0mb	53	41	205 26	230 30	53	68
NZSK	37.3N 74.5W	12Z 23/11/85	-7.0mb	98	17	350 25	001 29	39	40
SHIP	33.9N 76.6W	12Z 23/11/85	-6.2mb	85	85	002 27	360 32	34	34
WXWL	35.3N 74.6W	12Z 23/11/85	-7.3mb	99	8	-	-	-	34

Table 15. Ship observations rejected or nearly rejected in error by the test scheme near tropical storm "Kate".