Short-range Forecasting Research

Short Range Forecasting Division

Technical Report No. 23

Estimation of the Mean and Standard Deviation of the Random Component of Data also containing Non-random Errors

by

B R Barwell

July 1992

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

Short Range Forecasting Research

Technical Report No.23

Estimation of the Mean and Standard Deviation of the Random Component of Data also Containing Non-random Errors

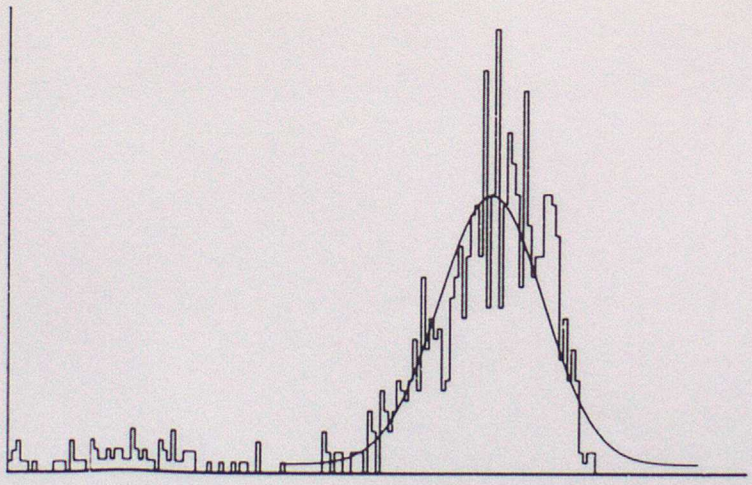
by

B R Barwell

S-Division, Meteorological Office,

London Road, Bracknell, Berkshire RG12 2SZ, England.

July 1992



1. Introduction

1.1 Random errors and gross errors

The mean μ and standard deviation σ are frequently used as measures of the statistical properties of a sample of data containing errors. These quantities are often regarded as parameters of a random (Gaussian) distribution

$$P(x) = \left(\frac{1}{\sigma(2\pi)^{1/2}}\right) \exp\left(-[x-\mu]^2/2\sigma^2\right). \qquad (1.1)$$

However, experimental and observational data are also prone to contamination by errors which we refer to as *gross errors* having an uncertain non-random distribution. Among the causes of gross errors are human observational errors (*e.g.* a barometer observation 10 mb too high or too low), transmission or coding errors (*e.g.* a pressure of 987 mb coded as 998.7 mb, or with transposed digits as 978 mb), or errors of representativeness (*e.g.* a strong observed wind due to a local squall at the time of observation).

For observations such as those from satellites, data processing can also be a source of gross errors. Satellite instruments can only measure radiation since no other information-carrying quantity can reach the instrument from the earth or atmosphere, but most computer models and other applications still require radiances to be converted into more familiar variables like temperature and humidity. The necessary data processing can involve assumptions which may be inappropriate for some observations. Examples are the assumption of a single cloud layer when removing the effect of clouds from infra-red observations for temperature soundings, and the selection of a vector wind from among several possibilities all consistent with scatterometer observations (de-aliasing).

As the work described in this paper was carried out in connection with research into monitoring the quality of satellite sounding data, the examples will use this data but the theory and techniques are quite general and can find application in verification or quality monitoring of data from a wide variety of sources.

1.2 Comtamination by gross errors

Use of the term 'gross error' does not *necessarily* imply errors which are of large magnitude although as their distribution is usually broader than that for random errors, points affected by gross errors will often appear as 'outliers' in a sample of data. Detection and flagging of these outliers is important for data assimilation into numerical weather prediction models and for the retrieval of temperature soundings from satellite data. These applications generally involve statistical techniques which assume that observations and numerical model fields have normally distributed errors. An observation with a large gross error can therefore be a serious problem if it is not detected beforehand. Gross errors comparable with or smaller than typical random errors are harder to detect but are much less damaging if undetected.

An outlier is relatively easy to detect automatically or manually if *a priori* values of the mean and standard deviation of the random errors are known. For example, suppose ten sea level pressure observations are compared with a model analysis and found to consist of nine 'good' ones with a zero mean error (μ) and a standard deviation (σ) of 1mb, and a single observation in error by 10mb. Applying the frequently used criterion of flagging data more than three standard deviations from the mean (which accounts for less than 0·3% of data from a normal distribution), the knowledge that σ is 1mb and μ is zero immediately causes flagging of the tenth observation as it deviates from the mean by ten standard deviations.

Suppose, however, that the mean and standard deviation of the errors are not known in advance and the available data provides the only way to estimate their values. Then we estimate

$$\mu = (9\times0 + 1\times10)/10 = 1\cdot0 \ ,$$

$$\sigma^2 = (9\times1^2 + 1\times10^2)/10 - \mu^2 = 9\cdot9.$$

The estimated standard deviation is then $(9.9)^{1/2} = 3\cdot146$. The 3σ criterion only flags errors outside the range -8·44 to +10·44 so fails to identify the outlying observation.

Figure 1 shows a histogram of the difference between the radiances measured by a satellite instrument and those computed froma numerical model. (Some cloud contaminated data which would normally be ignored is included in the histogram.)

Also shown is a normal distribution with the same mean and standard deviation as the data: clearly the long non-Gaussian tail of the distribution has caused the computed standard deviation to be too large and the mean to be displaced.

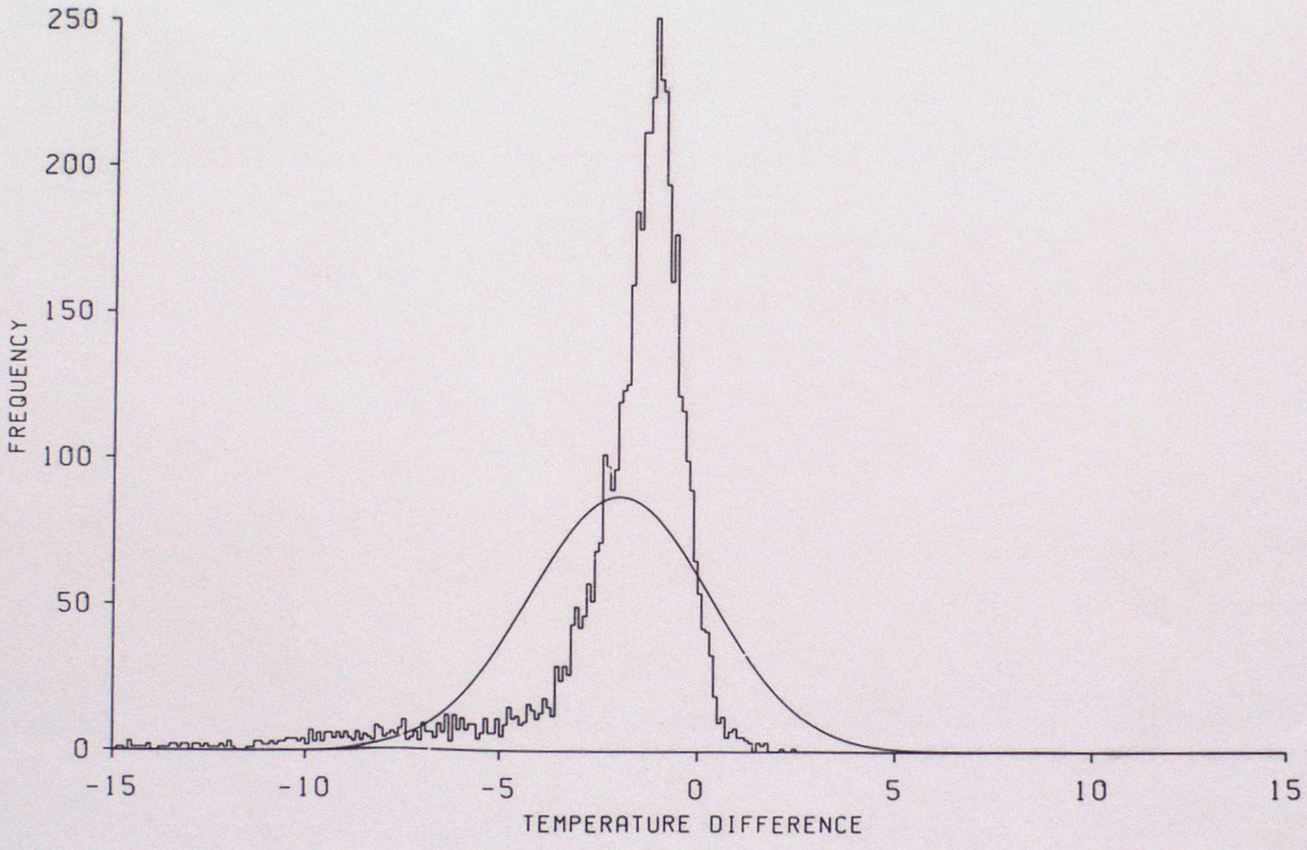


Figure 1. Histogram of observed-minus-computed radiances and a normal distribution having the same mean and standard deviation. (Data is for channel 5 of the HIRS instrument on NOAA-11 north of 50°N and includes cloud contaminated data.)

These examples illustrate the need for a more flexible error model such as a normal distribution combined with a simple model of gross errors to soak up the outliers. A technique for fitting the model to a given data distribution is also required. This report describes a method for meeting these needs. Naturally there will always be awkward data distributions where the method fails but if gross errors are not too frequent and the underlying distribution not too different from normal, better estimates of the mean and standard deviation will result than would otherwise be available.

An error model combining random and gross errors is derived in the next section, and section 3 describes the iterative technique used to fit the model to a given data distribution. Section 4 outlines a second level of iteration used to determine a suitable data range for the fitting and a method of choosing a good first guess to start things off. Section 5 gives examples of the performance of the method, and conclusions are presented in section 6.

The data to be fitted is assumed to be available as a frequency distribution having been sorted into 'bins' of uniform width δ. For generality, the computer programs work entirely in 'bin' units returning final values of mean and standard deviation in bin widths.

2. Error models

2.1 Combined normal and gross error distribution

Our model of the probability density function (PDF) for the errors consists of a weighted sum of two components; a normally distributed random component and a non-random 'gross error' distribution. Denoting the observed variable by x, the mean of the normal distribution by μ and its standard deviation by σ, we have

$$\text{Normal component} = \left(\frac{1}{\sigma(2\pi)^{1/2}}\right) \exp\left(-[x-\mu]^2/2\sigma^2\right). \qquad (2.1)$$

In the absence of any information about the shape of the distribution of gross errors, we assume a constant value over a finite range, namely

$$\text{Gross error component} = \begin{cases} h & \left(\frac{1}{2h} < x < \frac{1}{2h}\right), \\ 0 & \text{elsewhere.} \end{cases} \qquad (2.2)$$

The sudden cut-off of this function at $\pm 1/(2h)$ required for normalisation is rather unrealistic but is of little more than academic interest because we shall assume that this cut-off occurs outside the range of the data being processed.

If the probability of an observation having a gross error is p, the combined PDF is

$$P(x) = (1-p)(\text{Normal component}) + p(\text{Gross error component})$$

(Strictly speaking, we should also expect observations with gross errors to contain random errors as well, but the approximation involved in ignoring this complication is negligible and falls to zero if the gross error distribution is truly flat). Substituting from Equations (2.1) and (2.2), we find

$$P(x) = \left(\frac{1-p}{\sigma(2\pi)^{1/2}}\right) \exp\left(-[x-\mu]^2/2\sigma^2\right) + ph. \qquad (2.3)$$

At first sight, this appears to contain four unknown parameters (μ, σ, p and h) but the next section shows that with proper normalisation p and h can be combined into a single parameter.

2.2 Expected frequency distribution

The problem now is to fit a probability density function of the form (2.3) to a given data distribution. We assume that the data has been sorted into 'bins' of width δ covering a total range Δ. If x_o is the minimum x-value and x_{i-} and x_{i+} are the limits of the i^{th} bin, we have

$$x_{i-} = x_o + (i-1)\delta,$$

$$x_{i+} = x_o + i\delta,$$

and the probability of a data item in the range x_o to $x_o+\Delta$ falling within the i^{th} bin (P_i) is given by

$$P_i = \int_{x_{i+}}^{x_{i-}} P(x)dx \Bigg/ \int_{x_o}^{x_o+\Delta} P(x)dx. \qquad (2.4)$$

We can evaluate this by substituting from Eq. (2.3). If we assume that the range of integration in the denoninator includes essentially all of the normal component, this simplifies to

$$\int_{x_o}^{x_o+\Delta} P(x)dx = (1-p) + ph\Delta. \qquad (2.5)$$

Eq. (2.4) then becomes

$$P_i = \frac{[(1-p)I_i + ph\delta]}{[(1-p) + ph\Delta]} \qquad (2.6)$$

where

$$I_i = \int_{x_{i-}}^{x_{i+}} \left(\frac{1}{\sigma(2\pi)^{1/2}}\right) \exp\left(-[x-\mu]^2/2\sigma^2\right) dx. \qquad (2.7)$$

I_i can also be written in the form

$$I_i = \tfrac{1}{2}\left[\mathrm{erf}\left(\frac{x_{i+}-\mu}{\sqrt{2}}\right) - \mathrm{erf}\left(\frac{x_{i-}-\mu}{\sqrt{2}}\right)\right] \qquad (2.8)$$

which is the most suitable form if the error function is available as a standard function on the computer.

If we define a non-dimensional bin width γ (= δ/Δ, the reciprocal of the number of bins) and combine p and h into a single gross error parameter ρ defined by

$$\rho = \frac{ph\Delta}{(1-p)}\ , \qquad (2.9)$$

Eq. (2.4) then reduces to

$$P_i = \frac{(I_i + \rho\gamma)}{(1 + \rho)}. \qquad (2.10)$$

This is a function of only three unknown pararmeters, μ and σ (via I_i) and ρ. The parameter ρ can be interpreted as the ratio of the probability that a data item falling in the range x_o to $x_o + \Delta$ has a gross error to the probability that it has no gross error.

3. Fitting to actual frequency distributions

3.1 Fitting criterion

Initially, attempts were made to fit a distribution of the form (2.10) to data distributions using a least squares approach. This involves minimising a cost function S given by

$$S = \sum_i \left(\frac{n_i}{N} - P_i\right)^2 \qquad (3.1)$$

where n_i is the actual frequency in the i^{th} bin and N is the total number of data in all bins. P_i is the theoretical probability of an observation falling in the i^{th} bin and has the form of Eq. (2.10). Introducing a non-dimensional frequency m_i equal to n_i/N, this becomes,

$$S = \sum_i (m_i - P_i)^2. \qquad (3.2)$$

The approach eventually adopted, however, is a 'maximum probability' fit. The probability of getting the observed distribution if the actual PDF gives an expected probability of P_i for the i^{th} bin is

$$\text{Probability} \propto \prod_i \binom{N}{n_i} P_i^{n_i} (1-P_i)^{N-n_i}. \qquad (3.3)$$

Taking logarithms, we find

$$\text{Log(probability)} = \text{constant} + \sum_i \left(m_i \log P_i + (1-m_i)\log(1-P_i)\right)$$

which leads to a different cost function

$$S = \sum_i \left(m_i \log P_i + (1-m_i)\log(1-P_i)\right). \qquad (3.4)$$

With P_i given by Eq. (2.10) and I_i by Eq. (2.8), the problem is now reduced to finding values of μ, σ and ρ which *maximise* this value of S. (In the least squares formulation we *minimise* the cost function (3.1).)

3.2 Maximising the cost function, S

Newton's rule was used iteratively to find the maximum value of the cost function. This involves setting the three partial derivatives of S to zero, *i.e.*,

$$\frac{\partial S}{\partial \mu} = \frac{\partial S}{\partial \sigma} = \frac{\partial S}{\partial \rho} = 0. \tag{3.5}$$

This leads to the set of three equations

$$\sum_i \left(\frac{m_i - P_i}{P_i(1-P_i)}\right)\frac{\partial P_i}{\partial \varphi_j} = 0, \quad j=1,2,3 \tag{3.6}$$

where $(\varphi_1, \varphi_2, \varphi_3)$ is the vector (μ, σ, ρ) which will be denoted by $\mathbf{\Phi}$. [The least squares technique leads to the equation

$$\sum_i (m_i - P_i)\frac{\partial P_i}{\partial \varphi_j} = 0, \quad j=1,2,3.]$$

Equations (3.6) are solved iteratively using Newton's Rule

$$\mathbf{\Phi}_{i+1} = \mathbf{\Phi}_i - (\mathbf{S}_i'')^{-1}\,\mathbf{S}_i' \tag{3.7}$$

where $\mathbf{S}_i'$ is a column vector of the derivatives of S in Eq. (3.5) and $\mathbf{S}_i''$ is the corresponding Hessian matrix given by

$$\begin{pmatrix} \frac{\partial^2 S}{\partial \mu^2} & \frac{\partial^2 S}{\partial \mu \partial \sigma} & \frac{\partial^2 S}{\partial \mu \partial \rho} \\ \frac{\partial^2 S}{\partial \sigma \partial \mu} & \frac{\partial^2 S}{\partial \sigma^2} & \frac{\partial^2 S}{\partial \sigma \partial \rho} \\ \frac{\partial^2 S}{\partial \rho \partial \mu} & \frac{\partial^2 S}{\partial \rho \partial \sigma} & \frac{\partial^2 S}{\partial \rho^2} \end{pmatrix}. \tag{3.8}$$

To alleviate problems with poorly conditioned matrices, the increments applied at

any iteration are restricted in size to maximum values which are currently set as follows:

$$\begin{aligned} |\mu_{i+1} - \mu_i| &\leq 0.2\sigma_i \\ |\sigma_{i+1} - \sigma_i| &\leq 0.3\sigma_i \\ |\rho_{i+1} - \rho_i| &\leq 0.5(\rho_i - \rho_{min}) \end{aligned} \qquad (3.9)$$

where ρ_{min} is the minimum value of ρ which prevents P_i from becoming negative for any bin in which there is at least one observation ($m_i > 0$). This is the point at which singularities appear in the computation of terms in Eq. (3.6). In practice, ρ_{min} is a small negative quantity.

Newton's Rule only converges to an absolute maximum of S if all the eigenvalues of the Hessian (3.8) are negative (*i.e.* if the negative of the Hessian is positive definite). If this is not the case, the procedure may converge to a minimum or a saddle point. If negative eigenvalues are detected in any iteration, the values for the next iteration are found by taking a step up the local gradient of S, the size of the step being the maximum within the constraints (3.9). Similarly, if any of the diagonal terms of the Hessian (3.8) are positive, a step is made in the corresponding variable in the direction of increasing S and within the same constraints.

A further constraint during the iteration is to prevent ρ from becoming negative. If any iteration would tend to make ρ negative, the increments to all three variables are reduced by the same factor so as to make ρ zero for the next iteration. If ρ is already zero and the increment for the next iteration turns out to be negative, ρ is fixed at zero and increments for μ and σ are recomputed using Newton's Rule on these two variables alone.

3.4 Convergence criterion

After each iteration of Newton's Rule, several checks are done to see if satisfactory convergence has been achieved and, if not, to ensure that the procedure is still converging. The currently used criterion for convergence are as follows (where μ and σ are in units of bin width):

- Test on change in μ: $|\mu_i - \mu_{i-1}| \leq 0.005$.
- Test on change in σ: $|\sigma_i - \sigma_{i-1}| \leq 0.005$.
- Test on change in cost function, S:

$$|S_i - S_{i-1}| / \tfrac{1}{2}(S_i + S_{i-1}) \leq 10^{-4}$$

If all these conditions are satisfied, the iteration process is terminated and the values from the last iteration are taken as the solution.

The checks for divergence of the iteration are:

- Excessive number of iterations (currently limited to ten).
- Mean within 3 standard deviations of range of data (either $\mu_i - 3\sigma_i < 1$ or $\mu_i + 3\sigma_i >$ number of bins).
- Decreasing cost function ($S_i - S_{i-1} < -10^{-3}$). This test is not applied before the third iteration.
- Iteration could not be performed because the system of equations (3.7) was singular or too unstable. (This corresponds to an error code in the NAG routine used to solve these equations.)

If any of these conditions are satisfied the iteration process is terminated and an error code is set.

3.5 Appropriateness of the solution

The solution produced by the procedure described in this section may be mathematically correct but may not be the optimum for fitting the actual data distribution because we have not considered the possibility that the range of values over which the fit is done may be inappropriate. The test on $\mu \pm 3\sigma$ referred to above ensures that the data range is not too small but no check has been made for excessively large ranges which could mean large amounts of bad data being used in the fitting. Furthermore, experience showed that successful convergence of the fitting technique is aided by a good first guess, especially if data volumes are small. The next section addresses these problems.

4. Data selection and first guess

4.1 Selection of data range

In order for the fitting described in section 3 to perform well, the range of bins into which the data are classified must include essentially all of the normal component of the distribution. However, inclusion of a significant range outside this region may degrade the fitting as our assumption that the height of the gross error distribution is constant over the fitted range is less likely to be valid. Moreover, in most cases interest centres on the mean and standard deviation of the normal component rather than details of the gross errors.

If a prior estimate of the range of the normal component of errors is not available, the range of bins can be overestimated initially and the fitting routine allowed to reduce the range as it learns more about the characteristics of the data. This implies a second iteration to converge on a suitable bin range executed outside the Newton's Rule iteration described in the last section.

Initially, the fitting is done for the full range of bins, a first guess solution is chosen (see section 4.2) and a 'maximum probability' fit is carried out using the method of section 3 to find best values for μ, σ and ρ. From the results, the range of bins covering the region $\mu-4\sigma$ to $\mu+4\sigma$ is computed. If this contains less bins than that previously used, the fitting is repeated for the new range using the parameters from the last fit as the new first guess. The whole process is iterated until the number of bins being used no longer decreases. The outer iteration is continued regardless of whether the inner one converges, except that, if it does not converge, the value of σ used for the new bin range is the maximum of the values from the last two inner iterations. This is done to minimise occasions when a much reduced bin range is generated from a spuriously low σ taken from an iteration with convergence difficulties; an overestimate of the bin range is clearly safer than an unedrestimate in these circumstances.

4.2 The first guess

The example in section 1 indicated that the use of a mean and standard deviation computed from an entire data distribution may be unrepresentative of that for the normal component of errors since a relatively small amount of data with gross errors has a disproportionate effect on the calculation. A first guess which is more representative of the parameters of the normal component can

reduce the likelihood of convergence problems with the Newtonian iteration. A scheme was therefore devised for generating a reasonable first guess from the data distribution alone without any prior information.

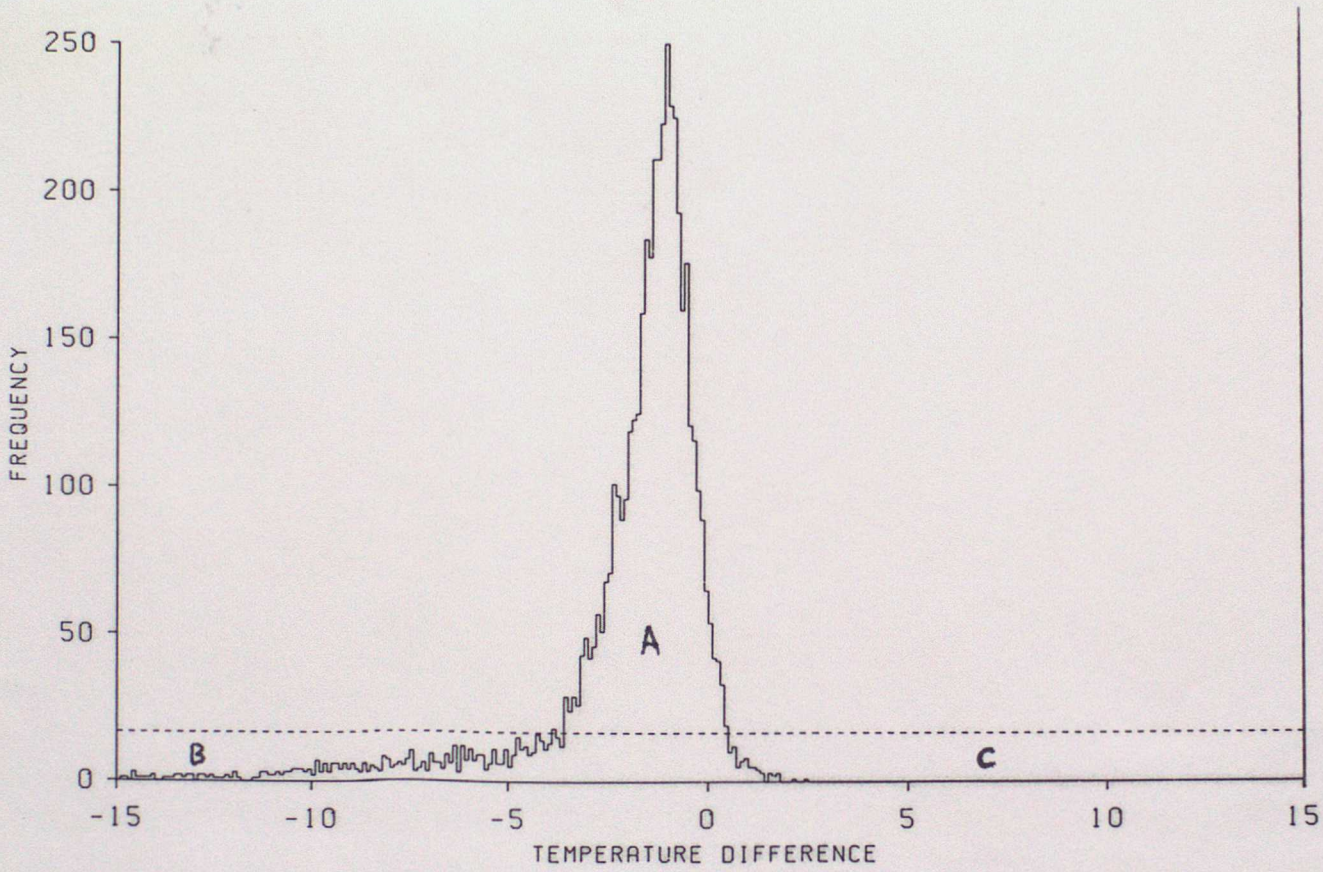


Figure 2. The histogram of Fig. 1 with the mean height marked (dashed line). Area A equals the sum of B and C.

If, as we have assumed, the gross error distribution has a constant value throughout the range of bins used, this will certainly be less than the average height of the distribution in that range. Therefore, if a horizontal line is drawn at average height in an ideal data distribution (Fig. 2, where area A equals the sum of B and C), the only variation above this line is due to the normal error component. The mean and standard deviation for area A (μ_a and σ_a) will then be unaffected by gross errors. Since the normal curve is symmetrical, μ_a is a suitable first guess for the mean of the normal component of the distribution but σ_a will obviously be an underestimate of the true standard deviation so further refinement is necessary. A method for extracting an improved first guess σ from σ_a and a correrssponding first guess ρ is described in Appendix A.

5. Examples

Figure 3 shows the effect of using the fitting procedure on a distribution which is fairly symmetrical but which falls off more slowly than a normal distribution in the region away from the peak. (The distribution represents differences between cloud cleared brightness temperatures from two independent cloud detection and removal schemes [Met. Office and NESDIS] for channel 11 of the HIRS instrument on the NOAA-11 satellite.) The data distribution shown by the histogram is the same for each part of the figure. Each diagram also shows a fitted curve plotted over the range of data used for the fitting. Diagram (*a*) shows that corresponding to the first guess (which used data in the range $-20° < x < 20°$ or four times the range shown); (*b*), (*c*), (*d*) and (*e*) show how the fit to the central part of the distribution is improved after 2, 3, 5 and 6 iterations of the data range (section 4.1) respectively, and (*f*) shows the best fit after the final (eighth) iteration. The following table shows details of the convergence of the iteration procedure.

Iteration	μ	σ	ρ	fitted range (°C)
1st guess	0.141	1.452	0.186	——
1	0.193	1.769	0.097	-20.0 to 20.0
2	0.077	1.152	0.320	-6.9 to 7.3
3	0.036	0.852	0.530	-4.6 to 4.7
4	0.036	0.683	0.719	-3.4 to 3.5
5	0.037	0.599	0.808	-2.7 to 2.8
6	0.037	0.567	0.831	-2.4 to 2.5
7	0.036	0.553	0.854	-2.3 to 2.4
8	0.036	0.542	0.865	-2.2 to 2.3

The diagrams show that the final result is a good fit the central peak of the distribution and ignores data well away from the peak.

The quantity ρ should be interpreted with care. It is not representative of the level of gross errors in the entire distribution but is only affected by the region near the peak. Even then, the apparent gross error level may contain a contribution from the non-Gaussian shape of the distribution. In Fig.3 for example, ρ is inflated because the distribution falls off less rapidly than a Gaussian curve, the "wings" of the distribution contributing to the apparent gross error component.

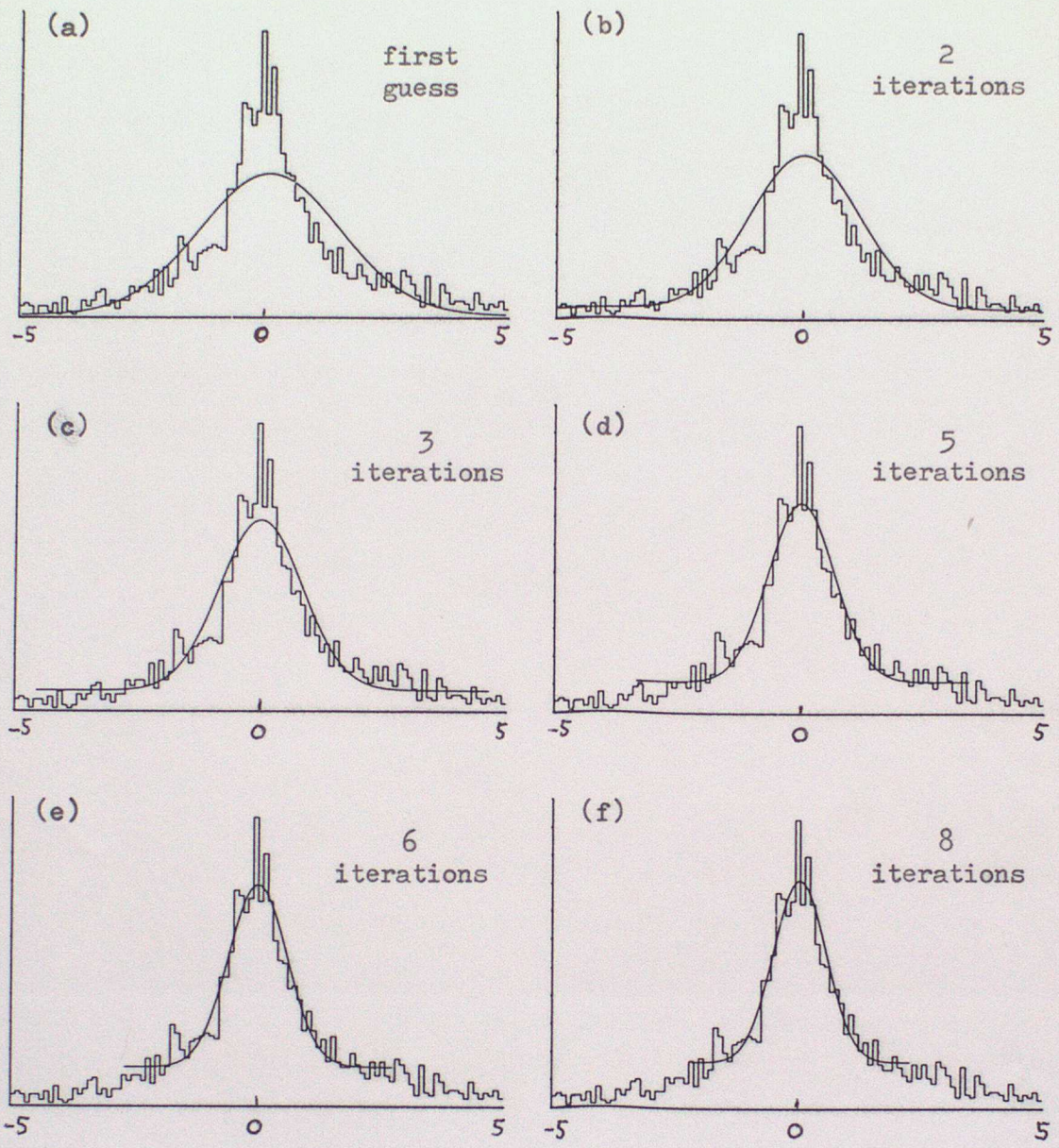


Figure 3. Stages in fitting a given data distribution (histogram). Curves represent the first guess (diagram (*a*)) and the fitted curves after 2, 3, 5, 6 and 8 iterations (diagrams (*b*) to (*f*)).

Figure 4 shows the fit for an asymmetrical data distribution. Clearly, with the symmetrical assumed distribution, the fitting must make a compromise in this case. The curve shows that the fitted mean is displaced but quite close to the peak of the histogram and the standard deviation is representative of the spread while the 'noise' at low x-values has been successfully ignored.

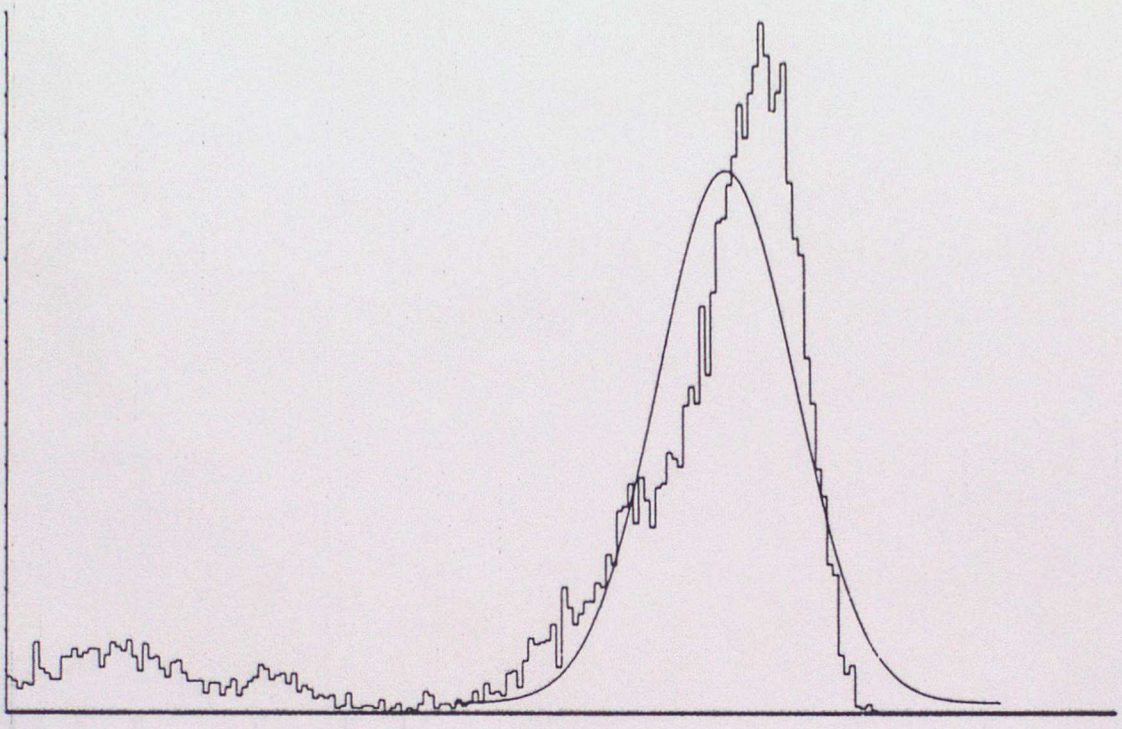


Figure 4. Gaussian curve fitted to an asymmetrical distribution.

6. Conclusions

An iterative method of fitting a combined normal and gross error probability curve to given data distributions has been developed. This has enabled objective computation of estimates of the means and standard deviations of the normal components of the distributions ignoring the effect of outliers and data with gross errors. Results have proved satisfactory, especially for distributions with a single peak and a degree of symmetry, though convergence difficulties can occur with distributions having broad or multiple peaks.

The method has applications in the monitoring of observational data where means and standard deviations are required without contamination by the occasional outlier. These may be needed for quality control purposes, for estimating model biases and variances, for input data for analysis or satellite sounding retrieval schemes *etc.* Another potential application is the detection and monitoring of changes in the quality of reports from individual stations.

Appendix A

Computation of first guess σ

In section (4.2) an initial estimate of the mean μ and standard deviation σ of the Gaussian part of a distribution was obtained by computing these quantities for the portion of the distribution above the mean (area A in Fig. (A.1)). The mean is indicated in the diagram by the horizontal line PQ whose height is N/Δ, N being the number of data points and Δ the total range of data used. Denoting the mean and standard deviation of area A by μ_A and σ_A, we have $\mu = \mu_A$ in an ideal distribution, so that μ_A can be used as an estimate of μ.

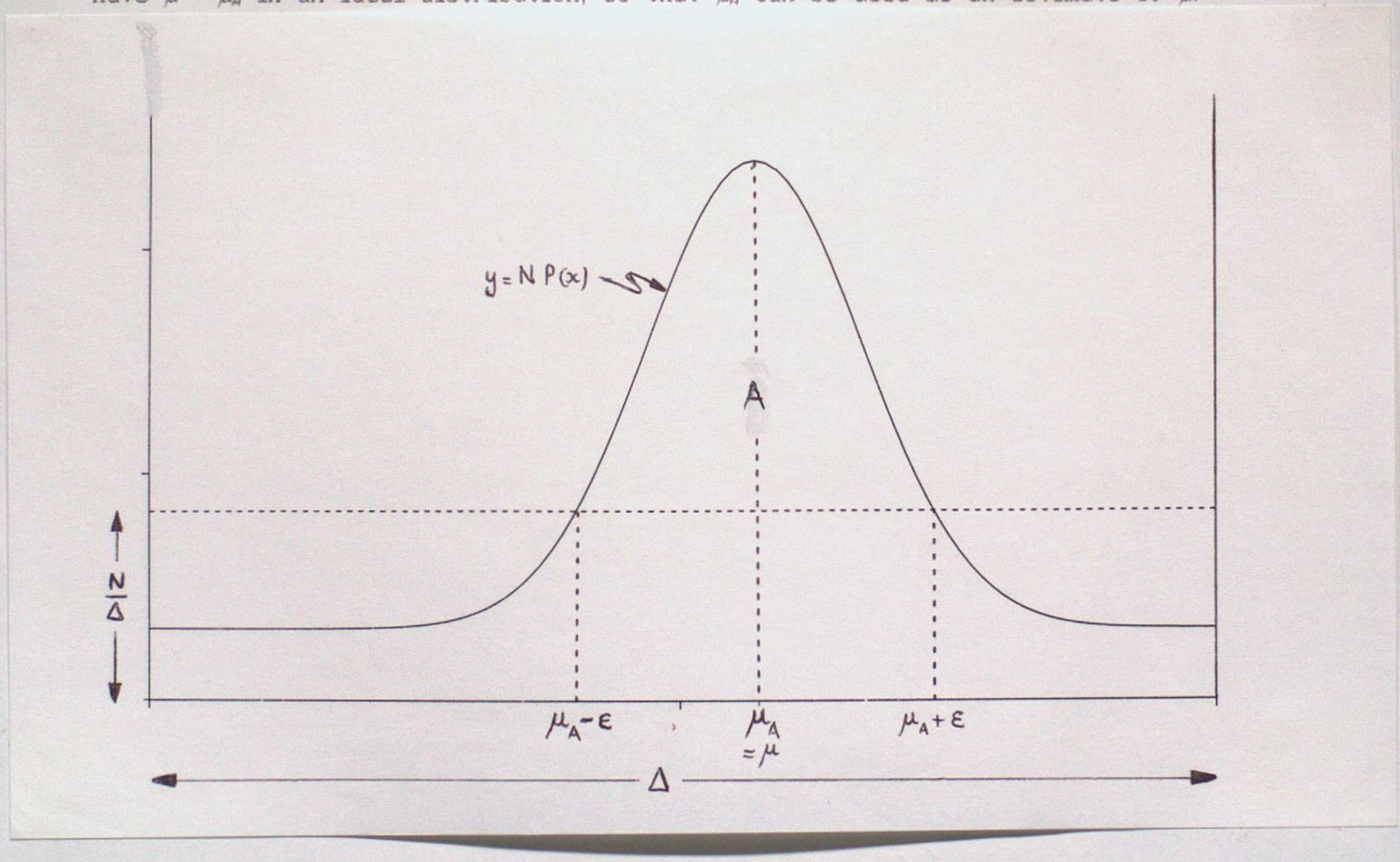


Figure A1. Geometry of combined Gaussian and gross error distribution.

According to our error model, the distribution normalised over the range Δ is given by Eq. (2.10) which for small bins of width dx becomes (using Eq. (2.7))

$$P(x).dx = \left(\frac{1}{1+\rho}\right)\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] + \frac{\rho}{\Delta}\right).dx \qquad (A.1)$$

If $\mu_A-\varepsilon$ and $\mu_A+\varepsilon$ are the x-values of the points where the mean line PQ cuts the curve, we have

$$P(\mu_A \pm \varepsilon) = 1/\Delta$$

which becomes on substituting from Eq. (A. 1),

$$\varepsilon^2 = 2\sigma^2 \log_e \left(\frac{\Delta}{\sigma\sqrt{2\pi}}\right) \qquad \text{(A. 2)}$$

If the number of data points corresponding to the area A is N_A, we have

$$\frac{N_A}{N} = \int_{\mu-\varepsilon}^{\mu+\varepsilon} \left(P(x) - \frac{1}{\Delta}\right) dx. \qquad \text{(A. 3)}$$

Similarly,

$$\frac{N_A \sigma_A^2}{N} = \int_{\mu-\varepsilon}^{\mu+\varepsilon} \left(P(x) - \frac{1}{\Delta}\right)(x-\mu)^2 \, dx. \qquad \text{(A. 4)}$$

Using Eq. (A. 1) these become

$$N_A = \left(\frac{N}{1+\rho}\right)\left(\operatorname{erf}\left[\frac{\varepsilon}{\sigma\sqrt{2}}\right] - \frac{2\varepsilon}{\Delta}\right), \qquad \text{(A. 5)}$$

and

$$\sigma_A^2 N_A = \left(\frac{\sigma^2 N}{1+\rho}\right)\left(\operatorname{erf}\left[\frac{\varepsilon}{\sigma\sqrt{2}}\right] - \frac{2\varepsilon}{\Delta} - \frac{2\varepsilon^3}{3\sigma^2\Delta}\right). \qquad \text{(A. 6)}$$

Dividing (A. 6) by (A. 5) to eliminate N_A and rearranging, we find

$$\sigma^2 = \sigma_A^2 \left[1 - \frac{2\varepsilon^3}{3\sigma^2\left(\Delta \operatorname{erf}\left[\frac{\varepsilon}{\sigma\sqrt{2}}\right] - 2\varepsilon\right)}\right]^{-1} \qquad \text{(A. 7)}$$

Equations (A.7) and (A.2) enable a suitable first guess for σ to be computed. Taking σ_A as an approximate value of σ, the latter equation can be used to find an estimate of ε which can be used in (A.7) to derive a better estimate of σ. This procedure can be iterated until convergence to provide even more refined estimates of σ, but as the result is required only for a first guess, three iterations are considered sufficient and no convergence test is used.

Having found σ, the first guess ρ can be computed from Eqs. (A.5) and (A.6). The simplest way is to eliminate the error function which leads to

$$\rho = \frac{2\varepsilon^3 N}{3\Delta N_A (\sigma^2 - \sigma_A^2)} - 1 \qquad \text{(A.8)}$$

If this value of ρ turns out to be negative, it is replaced by zero and σ is relaxed towards σ_A to some extent.

Short Range Forecasting Division Technical Reports

This is a new series to be known as Short Range Forecasting Division Technical Reports . These will be reports from all three sections of the Short Range Forecasting Research Division i.e. Data Assimilation Research (DA),Numerical Modelling Research (NM), and Observations (OB) . This series succeeds the series known as Short Range Forecasting Research / Met O 11 Technical Notes.

1.	ON THE TIME SAVING THAT CAN BE ACHIEVED BY THE USE OF AN OPTIMISED COURSE IN AN AREA OF VARIABLE FLOW	R.W. Lunnon A.D. Marklow September 1991
2.	Treatment of bias in satellite sea surface temperature observations	R.S.Bell August 1991
3.	FINITE DIFFERENCE METHODS	M.J.P. Cullen August 1991
4.	Representation and recognition of convective cells using an object-orientated approach	W.H. Hand 30th September 1991
5.	Sea-ice data for the operational global model.	C.P.Jones November 1991.
6.	Tuning and Performance of the Atmospheric Quality Control.	N.B. Ingleby. December 1991.
7.	More satellite sounding data - can we make good use of it?	R.S.Bell January 1992.
8.	WAM/UKMO Wind Wave model Intercomparison Summary Report	Heinz Gunther ECMWF Martin Holt UK Met Office January 1992
9.	Spin up problems of the UKMO Mesoscale Model and moisture nudging experiments	Akihide Segami JMA February 1992
10.	A comparison of 2nd generation and 3rd generation wave model physics	M.W. Holt B.J. Hall February 1992
11.	RETRIEVAL AND ASSIMILATION: SYSTEM CONSIDERATIONS	Andrew C Lorenc March 1992
12.	Detection of Precipitation by Radars in the UK Weather Radar Network	M. Kitchen P.M. Brown April 1992
13.	THE VALUE OF WIND OBSERVATIONS FOR WEATHER FORECASTING AND CLIMATE STUDIES	Andrew C Lorenc April 1992

Short Range Forecasting Division Technical Reports

14.	An investigation into the parameters used in the analysis scheme of the Mesoscale Model	G. Veitch B.J. Wright S.P Ballard May 1992
15.	THE VERIFICATION OF MESOSCALE MODEL FORECASTS OF LIQUID WATER CONTENT USING HELICOPTER REPORTS OVER THE NORTH SEA DURING WINTER 1991	M. Ahmed R.W Lunnon R.J. Graham May 1992
16.	Simulations of the Diurnal Evolution of Marine Stratocumulus Part I: The sensitivity of the Single Column Version of the Mesoscale Model to Changes in the . Turbulence Scheme.	S.D.Jackson S.P. Ballard May 1992
17.	Simulations of the Diurnal Evolution of Marine Stratocumulus Part II: A Comparison of Radiation Schemes Using the Single Column Version of the Mesoscale Model.	S.D.Jackson S.P. Ballard May 1992
18.	Quantifying the low level windshear aviation hazard for the UK: some research proposals	R.J. Graham R.W. Lunnon May 1992
19.	WAM/UKMO Wind Wave model Intercomparison Part 2 Running the UKMO wave model at higher resolution	M.W. Holt April 1992
20.	Sensitivity of Mesoscale Model forecasts of anticyclonic Statocumulus to the specifications of initial conditions and Boundary Layer mixing scheme.	B.J. Wright S.P. Ballard July 1992
21.	Evaluation of diffusion and gravity wave changes in the Global Forecast Model.	F. Rawlins O. Hammon 16 June 1992
22.	Background errors for the quality control and assimilation of (atmospheric) observations in the Unified Model - the situation in 1992.	C.A. Parrett July 1992
23.	Estimation of the Mean and Standard Deviation of the Random Component of Data also Containing Non- random Errors.	B.R. Barwell July 1992