Comments on 'Objective quality control using Bayesian methods' by A. C. Lorenc and O. Hammon (January B 1988, 114, 515–543)

By H. JEAN THIÉBAUX

UCAR Visiting Scientist Research Program, National Meteorological Center, Washington, D.C., U.S.A.

(Received 24 November 1987)

SUMMARY

The proposals for quality control (QC) with automatic updating of the QC system, which Lorenc and Hammon (LH) have presented are important and forward-looking contributions to the future of operational numerical weather prediction. With expanding computer capabilities for data management, including the maintenance of archives which keep up-to-date records of data quality by observation type and location, it is possible to make use of highly relevant information whose assemblage previously was unpractical. Lorenc and Hammon have suggested mechanisms for doing this; and their proposals offer promise for significant improvements in objective quality control. My comments are designed to enhance practical development of the authors' ideas, by focusing on some implications of the underlying theory.

The objective of the work under discussion is to establish an algorithm whose application will result in the retention of observations which are within credible range of a background (or forecast) field, together with numerical assignment of their 'credibilities', and the rejection of observations which are not judged to be credible. Furthermore, part of its exercise will be renewal of the criterion of 'within credible range of the background', using information stored in quality control performance files.

1. DERIVATION OF A POSTERIOR DISTRIBUTION FROM AN UNREALIZABLE PRIOR: OPTIONS

In dealing with the stochastic dependence of the atmosphere and its forecast fields, Lorenc and Hammon have taken a Bayesian statistical approach to deriving rejection tolerances for data quality control checks. This approach to deriving a QC mechanism requires specification of a prior distribution for 'truth', or 'realizations of meteorological variables', and a conditional distribution for 'a background', or 'forecast values of the variables'. The conditioning of the latter is on corresponding truth. The specifications of these distributions permit LH to derive first, the posterior distribution of the background

\[ P(B) = \int_{-\infty}^{+\infty} N(b - t, V_b) \, db \, P(t) \, dt \]

and then, from Bayes theorem

\[ P(T|B) = P(B|T) \times P(T) / P(B) \]

the conditional distribution of truth given a background (or forecast), which is the key to their quality control technique.

The authors' choice of a conditional probability distribution for \(B|T\) is a Gaussian distribution with mean equal to present truth and finite variance. The prior distribution for truth which LH have postulated to represent 'no prior knowledge' is an infinite variance Gaussian distribution; and this choice leads, through a mathematical derivation, to an infinite variance Gaussian distribution for the posterior distribution of the background. Both of these—the postulate and the corollary—are unrealizable; and the direct application of Bayes theorem to them is inadmissible, because their densities are everywhere zero. The mathematical device of writing the theorem in terms of large (finite) variance distributions and reducing the ratio to its \(V \rightarrow \infty\) form, to reach a result, seems irrelevant beside the fact that neither the chosen prior distribution of truth nor the derived posterior distribution of the background is a suitable representation of physical reality. The authors' Bayesian result (which is technically a correct consequence of their postulates) depends on the behaviour of the two distributions in the regions where they fail to be physically realizable. With an infinite variance distribution, the probability is more than \(\frac{1}{2}\) that a 'truth' will be realized that exceeds the range of anything that has ever been observed on earth. This is not a satisfactory prior distribution.

* Another way of saying this is that the prior distribution 'concentrates probability' well beyond the range of observed values which QC is anticipated to reject! (see Fig. 8 in Lorenc and Hammon).
A viable option open to potential developers/users of ‘adaptive quality control’ is to postulate a climatologically reasonable distribution for ‘truth’, which acknowledges a finite range of realizable truths and that some intervals of possible realizations occur more frequently than others, and apply Bayes theorem with this in place. Otherwise, I think that one must simply postulate the result claimed; because the assumptions from which Lorenc and Hammon have derived it are not all defensible. In fact, the latter option is not troublesome, although it removes the Bayesian label from the procedure. It amounts to postulating that the conditionings of truth on background and background on truth are reciprocal, except for inequality of variances. As a first approach to adaptive quality control, which ignores the complications of systematic forecast bias, this is a straightforward, defensible postulate; and one which can be readily modified as the QC system gains the experience of its monitoring archive.

2. The distribution of observations rejected by QC

Lorenc and Hammon presented the figure reproduced here as Fig. 1; this warrants discussion on two counts. The first relates to representing the histogram with probability density functions and the second, more importantly, to their shapes.

(1) Of the two densities shown by the authors in their figure, their rectangular gross-error distribution indicated by the horizontal dashed bar at 0-043 on the vertical axis, provides the closer representation for the histogram. However, the curves are not variance-matched, as the caption suggests. My Fig. 2 is the same, with two additional distributions overlaid: the dashed, bell-shaped curve is variance-matched with the $k = 0.043$ rectangular distribution, and the solid bar rectangular distribution with $k = 0.019$ is matched with the authors’ Gaussian curve. Clearly the one (variance) parameter for the zero-centred Gaussian can be tuned to provide a reasonable fit to the histogram, while neither the large variance Gaussian nor the matching rectangular distribution is acceptable.

The histogram indicates that the distribution of observed minus background deviations for rejected observations has more ‘mass’ in the tails than either of the dashed curves. In keeping with the LH proposition that errors are a mixture of instrumental observation errors, assumed to have Gaussian distributions, and gross errors (mistakes), with rectangular distribution over some finite range of possible values, it is possible to construct a mixed distribution which will accommodate

![PMSL obs rejected by QC](image)

**Figure 1.** Histogram of observed minus background deviations, for pressure data ‘rejected in quality control’. Also shown are the densities for a Gaussian error distribution and an assumed gross error distribution. (This is Fig. 8 of LH.)
the heavier tails. This is rather neatly handled with a small variance Gaussian and a broad (low)
rectangular distribution, with mixing ratio $q$:

$$q\left[1/\sqrt{2\pi V_0}\right] e^{-x^2/2V_0} + (1-q) \left[1/(2a)\right] \begin{cases} 1/(2a) & -a<x<a \\ 0 & \text{otherwise.} \end{cases}$$

![Figure 2. As Fig. 1 with a rectangular distribution variance-matched to the Gaussian and a Gaussian variance-matched to the assumed gross error distribution.](image)

![Figure 3. Histogram of observed minus background deviations as in Figs. 1 and 2, with a mixed Gaussian/rectangular distribution variance-matched with the histogram.](image)
For this distribution, the variance is \( qV_o + (1-q)a^2/3 \) and parameters \( V_o, a, \) and the mixing ratio \( q, \) are chosen to match variance with the histogram. For example, the mixed distribution with \( V_o = 45, a = 100, \) and \( q = 0.9 \) has been evaluated and Fig. 3 shows it overdrawn on the histogram. Thus, if it is important to obtain an accurate density representation, this provides a practical mechanism for its derivation.

(2) If the histogram of LH Fig. 8 has truly been constructed from the observed minus background deviations of observations rejected by the authors' QC algorithm, then it appears to have the wrong shape. This distribution for rejected observations has deviations from background concentrated at and near zero, where the observations agree best with the background. If the QC algorithm is accomplishing its design objective, this distribution should be bimodal, with virtually no mass in the immediate vicinity of zero deviation from background. Thus, the figure given would be more believable if were the histogram for observations not rejected by QC. Perhaps it has been mis-labelled.

Reply by A. C. LORENC

We thank Dr Thiébaut for her interest in this work. A sound theoretical basis will aid the design and development of the practical quality control schemes which are our main aim.

The argument in our appendix A was included to give the reader some understanding and justification of our use of the terminology 'background error'. The background is our prior information; before having the background we were in a hypothetical state of having no information. Looked at in this way, the background becomes like another observation with its own error distribution. But as Dr Thiébaut points out, there are some conceptual and mathematical difficulties with the hypothetical state of having no information. To avoid these we could use the concept (and terminology) 'prior uncertainty of truth' instead of 'background error', throughout the paper. The argument to which Dr Thiébaut objects would then become unnecessary. Alternatively, as she suggests, we could postulate their equivalence, instead of trying to justify it.

Dr Thiébaut's second remarks are about our Fig. 8, which she reproduces as her Fig. 1. Can we first of all reassure her that it has not been mislabelled, it is a histogram of all the data rejected by our scheme, which consisted of some simple consistency checks, and of buddy checks with neighbouring observations, as well as the background comparison. These former checks can lead to the rejection of data which agree with the background. The background check alone does give, as she suggests, a bimodal histogram with zero mass near zero deviation from background.

There is no physical reason why the gross error mechanisms we are hypothesizing should not by chance generate observations which agree with the background, but of course such errors will be undetected by the background check, and many will also be undetected by the other checks. Thus there is no reason to believe that the actual distribution has a bimodal shape at small deviations like that of the plotted histogram.

Dr Thiébaut has not realized the very large range over which the histogram is plotted; we stated in our paper that data accepted by the scheme have a much narrower, near Gaussian, distribution with standard deviation 1-85 mb.

Our practical interest in the distribution of gross errors is because assumptions about it affect quality control decisions. Thus we are not concerned with accurate modelling of the tails, as long as single observations with a large deviation from the background are clearly rejected, as is the case for our simple model. Most borderline quality control decisions are for observations which deviate by up to about 8 mb from the background; it is the behaviour of the assumed distribution of gross errors in this region which is critical. Our initial assumption was the simplest possible; a constant. The only deduction that can be made from the histogram is that our assumption was not too bad in the critical region; as discussed above, the histogram itself is not a good picture of the true distribution for small deviations from the background. Certainly, as Dr Thiébaut points out, there are many better ways of representing the histogram than our simple constant value. But given the unimportance of the tails and the uncertainty about the small deviations, the extra complexity they entail is not justified. It remains our opinion that refinements to the assumed distribution of gross errors should come through study of their physical causes, such as constant miscalibrations or simple mistakes in coding reports.