PARALLEL QUALITY CONTROL OF EUMETSAT WIND PRODUCTS, WITH AND WITHOUT THE USE OF FORECAST WIND FIELDS.

Simon S. Elliott

EUMETSAT
Am Kavalleriesand 31
D-64295 Darmstadt
Germany

ABSTRACT

It is widely felt among the numerical weather prediction community that the use of forecast wind fields in the quality control of atmospheric motion vectors should, in principle, be avoided. This is because of the concern that a feedback could be introduced into the system, whereby only those winds which are in any case similar to the forecast, would be disseminated for assimilation into the forecast models. Where the satellite winds could have the most useful impact on a forecast is in the very areas where the forecast is incorrect, and the satellite winds could help to get it back. However, in these cases a consistency check against forecast data would give the satellite wind a bad score and it would probably not be distributed.

Statistical analysis of co-location data shows that among the various tests used by EUMETSAT, the forecast test gives, on average, the most reliable measure of quality, measured in terms of the vector difference from radiosonde data. Removing it from the suite of tests, which form the automatic quality control scheme can, therefore, only worsen the overall performance this scheme. We will show, however, that by carefully performing the optimisation of the weights of the remaining checks, this effect can be mitigated. e use the flexibility offered to us by BUFR encoding to include quality control information calculated with and without the use of forecast data, separately, with each wind. The use of BUFR Table B entries defining the quality control process allow the two sets of information to be distinguished, and used independently if required.

1. Introduction

The mechanism by which we select the relative weights given to the various automatic quality control tests is as discussed in Elliott, 1998. The fundamental idea is to assume that the tests themselves are (a) fixed in terms of the values they give for a particular wind vector, and (b) mutually independent.

The first of these assumptions is certainly a simplification, but does not invalidate the basic process. In fact, the precise performance of the individual tests is continually analysed and the exact parameterisation of their functional forms is subject to constant review. As we will discuss below, the calculation of the weights does not appear to be unduly sensitive to the fine detail of the input data, and so unless an individual test has been significantly modified, the weights would not need to be re-calculated. If the tests were truly an independent set of measurements of the quality of the wind vectors, the processing of optimising the various weights assigned to them would be relatively straightforward. There is, however, a clear overlap between some of the tests. This is most obvious when we consider the impinging triptych of the speed, direction and vector consistency tests. The speed consistency test gives a score relating the difference in speed between the two components of a combined vector, a small difference giving a higher score. The direction consistency test works in a similar way but for the difference in direction. The vector difference between the constituent component vectors is used directly to calculate the third test score. Since the wind vectors themselves are only two dimensional, there cannot be three independent pieces of information to use for three separate tests. Any two of the three tests are sufficient to contain all the information, which is present.

A further dependency between the results of the tests derives from the vector extraction method itself. The tracking of a target across consecutive images is, in the vast majority of cases, either successful, or very bad. There are very few cases where the tracking is “not bad” or “quite good”. Since the three tests mentioned above are all measuring the quality of the target tracking, the nearly always give similar results.

Since the analysis presented in Elliott, 1998 was performed, an additional quality control test has been activated in the wind extraction process. The correlation consistency test gives a value relating to the mean of the height of the selected correlation peak, for each of the forward and reverse correlation surfaces. The strength of this test is that it actually is independent of the others used. Its usefulness is, however, limited by the fact that in the large majority of cases it has a very high value. In fact its behaviour can be summarised by stating that any wind (good or bad) can get a high score, but a low score is a strong indication of a bad wind. We use this test naïvely in a linear combination with the other tests for the following analysis. Investigation into the potential use of the correlation consistency test as a multiplicative filter on the combined score of the other tests are currently in progress (*priv. comm.*, J. Gustafsson, 2000).

2. Core tuning results

Fig. 1 shows the distribution of results from a standard set of experiments using various combinations of automatic quality control tests. All the co-location data were used. The x-axis is the standard deviation of the best-fit straight line drawn through the scatter plot of normalised vector difference (NVD)[§] as a function of final quality control score. The y-axis is the correlation between the NVD and the final quality control score. The correlation coefficient is always negative, because a wind vector with a large difference from the radiosonde should get a small quality control score, and vice versa. In an ideal case, there would be an exactly linear relationship between the two sets of data, and this being the case (i) the standard deviation would be zero, (ii) the correlation coefficient would be –1.0, and (iii) we would be able to write an equation of the form

$$NVD = (\mathrm{a} \times QC) + \beta, \qquad [1]$$

where QC is the final quality score of a wind vector, and a and β are arbitrary constants. The better a particular set of weighted tests perform in achieving a relationship of the form in Eq. 1, the further into the lower left corner of the plot our result will be.

The nine configurations of the quality control tests used are shown in Table 1. Test case (TC) 1 was a first guess at what a potential choice of tests could be for the quality control process without forecast data being used. TC 2 represents the best possible choice because all the tests are used and their weights are then optimised. The third and forth test cases represent the current operational configuration, and a similar configuration without forecast data respectively. In both cases, the weights used for the tests are prescribed based on experience, and are not calculated automatically. TCs 5 and 6 are almost the identical; neither includes the correlation test and both have the weights of the tests optimised, but TC 5 additionally excludes forecast data. We consider two possible configurations using only three tests in TCs 7 and 8. Both use the spatial and correlation tests, but TC 7 uses the direction test, where as TC 8 uses the vector test. Finally, TC 9 can be considered together with TCs 2, 5 and 6. Together these 4 can be juxtaposed to show the four binary combinations of with/without the forecast test, and with/without the correlation test.

The four results clustered together in the top left-hand corner are from TCs 1, 7, 8 and 9. These all use correlation test data, but exclude the forecast test. The two results in the top right corner are from TCs 4 and 5, both of which exclude the forecast test and the correlation test. These two groups differ primarily in their standard deviation, which can infer is reduced by the inclusion of the correlation test.

[§] The normalised vector difference, or NVD, is the vector difference between an extracted wind vector and the co-located radiosonde wind, normalised by the speed of the radiosonde wind.

The bottom left and right results are from TCs 2 and 6 respectively. Both these cases use the forecast test, but TC 6 excludes the correlation test (and so has a higher standard deviation, as we might expect). Comparing these two results with the first two clusters shows that the main impact of including the results of the forecast test is to improve the correlation coefficient, and therefore to do a better job of achieving a relationship like that given in Eq. 1.

Table 1. Details of the various combinations of quality control tests used. Where 'Opt.' appears, the weight assigned to the test was derived using the optimisation method. The highlighted cells containing 'None' indicate that that quality control test was not used for that case. The bold numbers are prescribed weights determined by experience of real data.

Test case	Description	Direction test weight	Speed test weight	Correlation test weight	Forecast test weight	Spatial test weight	Vector test weight
1	Intuitive without forecast	Opt.	Opt.	Opt.	None	Opt.	None
2	Optimal with forecast	Opt.	Opt.	Opt.	Opt.	Opt.	Opt.
3	Current operational	**0.167**	**0.167**	None	**0.167**	**0.333**	**0.167**
4	Operational without forecast	**0.200**	**0.200**	None	None	**0.400**	**0.200**
5	Optimal without forecast or correlation.	Opt.	Opt.	None	None	Opt.	Opt.
6	Optimal without correlation.	Opt.	Opt.	None	Opt.	Opt.	Opt.
7	Simple-I without forecast	Opt.	None	Opt.	None	Opt.	None
8	Simple-II without forecast	None	None	Opt.	None	Opt.	Opt.
9	Optimal without forecast	Opt.	Opt.	Opt.	None	Opt.	Opt.

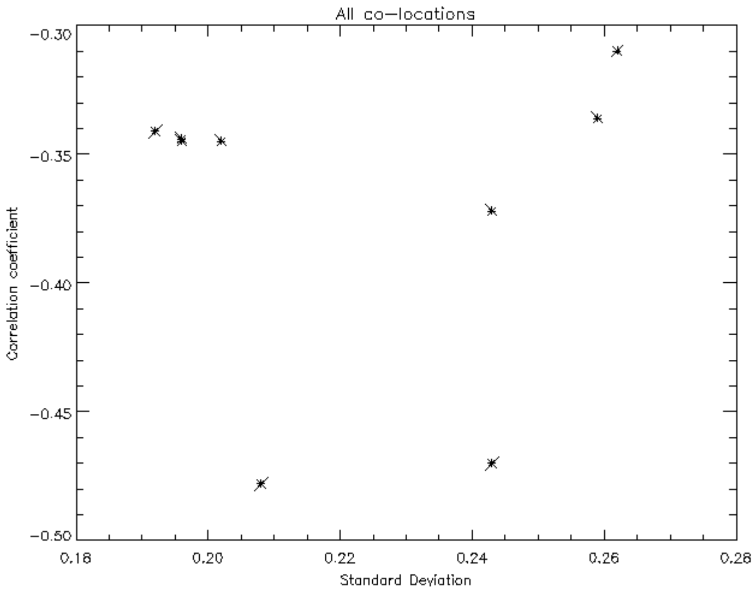

Figure 1. An inter-comparison of the performance of the various combinations of quality control tests used to derive an overall quality score. Those combinations that perform well fall towards the lower left part of the figure. These cases always include the forecast test.

TC 3 falls a little over half way up the plot on the left-hand side. This represents the operational configuration, which does use the forecast test. The plot suggests that simply by changing the weights given to the tests used operationally today, we could move toward TC 6 and by adding in the correlation test, we could move toward TC 2. It is also clear that in terms of standard deviation, the current operational setting is only marginally better than TCs 1, 7, 8 and 9, none of which use forecast data. This reinforces the idea that careful use of independent non forecast based tests can give a final quality control score which does almost as well as the current operational configuration.

3. Sensitivity to filtering by co-location vector difference

When one examines the co-location data set used for the analyses presented in this paper, it is clear that there is a small number of co-locations with a very large NVD. This can be the result of a failure in the semi-transparency correction mechanism, meaning that a fast moving thin cirrus tracer is mistakenly assigned to a low level. This vector is then co-located with a radiosonde wind at low level, which would have a much lower speed, and the resulting NVD can approach 100.0.

In order to ensure that the optimisation of the weights assigned to the quality control tests is not perturbed by these few rogue co-locations, it is possible to filter out co-locations where the NVD exceeds some threshold value from the tuning process. This step helps to ensure that the optimisation method is statistically robust. It is important, however, to consider the effect of these weights not only on only the selected set of co-location data used for the tuning process itself, but also on the entire co-location data set. Fig. 2 shows a similar set of results to that presented in Fig.1, but for a filtered set of co-locations with NVD < 5.0. The left hand plot shows the results when only the filtered-in set of co-locations used to determine the weights which are assigned to the tests are used for the evaluation. The right hand plot shows the result of applying the weights derived from the filtered set of co-locations to all the original co-locations.

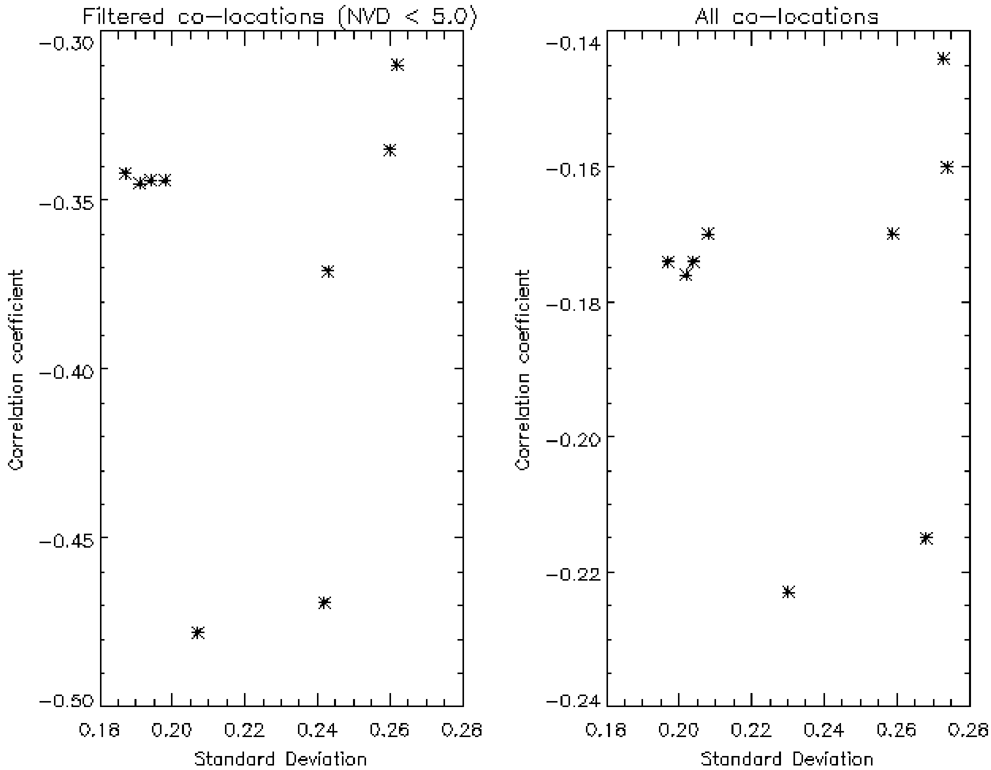

Figure 2. An inter-comparison of the performance of the various combinations of quality control tests, where the weights are selected using a set of co-locations filtered such that those with an NVD > 5.0 are excluded.

The results shown in Fig. 2 lead to two observations:

(i) In relative terms (comparing the left and right plots), there is very little difference between the results of applying the weights derived from the filtered set of co-locations to only those co-locations, or to the whole set. This means that the weights derived from the filtered set of co-locations can be used for all the co-locations with equanimity.

(ii) Comparison with Fig. 1 suggests that the exclusion of the outlying co-locations from the optimisation process actually made only a small difference. The general distribution of the various test cases remains essentially the same, and the values of the weights themselves were changed by typically less than 10 %.

The choice of a threshold value of 5.0 for the NVD was somewhat arbitrary. In an attempt to make a more calculated choice for the threshold value, the distribution of NVD within the co-location data set was examined. Fig. 3 shows this distribution.

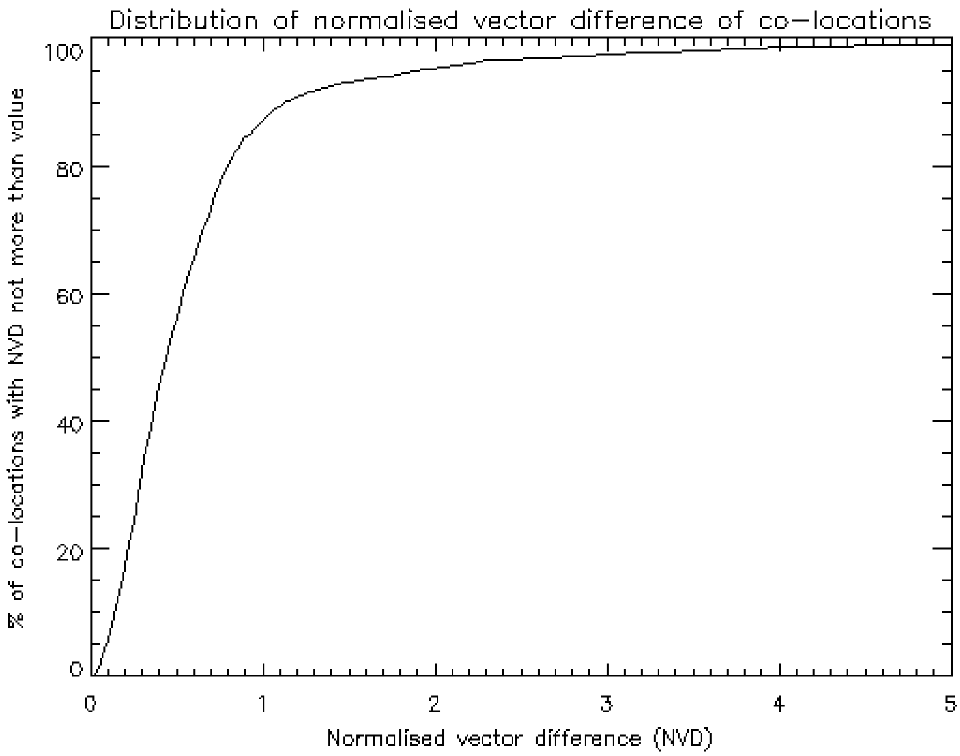

Figure 3. The distribution of NVD across the complete set of co-locations. There is a clear bend in the curve when the NVD is near 1.0, hence the choice of this as a preferred threshold value.

The graph in Fig. 3 shows that ~85% of co-locations have a NVD less than 1.0. The form of the curve is similar to that of a cumulative Poisson distribution function. For threshold values lower than 1.0, the number of co-locations remaining reduces rapidly. For this reason, the NVD threshold value of 1.0 was used, and the same set of test cases, as defined in Table 1, were re-run. The results are shown in Fig. 4.

By comparing Figs. 2 and 4, we can see that the only significant effect of reducing the NVD threshold value is to scale up uniformly the correlation coefficient values assigned to the test cases for the filtered set of co-locations. In fact, the only significant effect of filtering the co-locations used in the tuning process by NVD, is in the performance of the TC 1, 7, 8 and 9 cluster of results. When the weights are calculated using these filtered co-locations and applied to the whole data set, the aforementioned TCs have an improved correlation coefficient, relative to the other TCs in Fig. 1. This result alone does, however, indicate that there is value in applying an NVD threshold filter to the co-location data before the optimisation of the weights is performed.

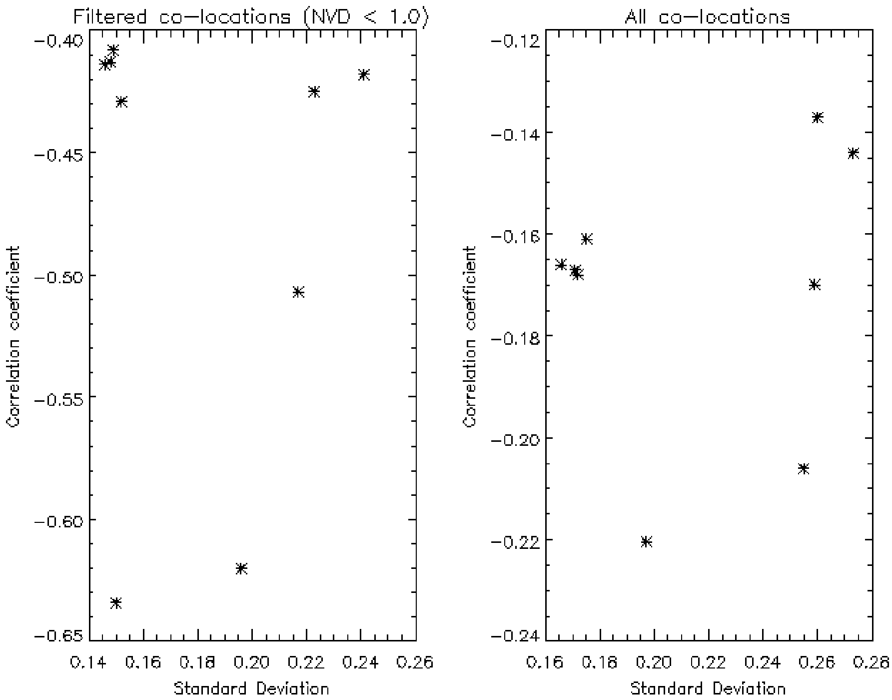

Figure 4. An inter-comparison of the performance of the various combinations of quality control tests, where the weights are selected using a set of co-locations filtered such that those with an NVD > 1.0 are excluded.

4. Limitations of optimisation method

As we have mentioned in Sect. 1, the entire optimisation process is critically dependent on the independence of the quality control tests, which we know is not the case. The performance of the optimisation when the tests are dependent is, at best, unpredictable, and can be in fact be improved upon by an intelligent choice of weights based on experience of the performance of the various tests.

The limitations of the optimisation method have been illustrated by several examples (*priv. comm.*, J. Gustafsson, 2000). For instance, when considering one set of low level wind co-locations, the following results were obtained,

- Manually weighted with forecast test: NRMS = 0.35
- Optimised with forecast test: NRMS = 0.36
- Forecast test only: NRMS = 0.37
- Operational setting: NRMS = 0.41
- Manually weighted without forecast test: NRMS = 0.58
- Optimised without forecast test: NRMS = 0.62,

where NRMS is the root mean square of the vector difference between the extracted wind vector and the co-located radiosonde wind, normalised by the mean radiosonde wind speed across the sample. In both the cases where test weights were derived using the optimisation method, the equivalent cases using manually assigned weights performed slightly better.

The reason for this limitation can be described by an analogy between optimising the results of *n* of test results and finding the co-ordinates of a point in an *n*-dimensional space. In the ideal case, each of out tests would be like an orthogonal basis vector in the space. The co-ordinates of the point are not very sensitive to small changes in the direction of a basis vector. If, however, the basis vectors spanning the space are all nearly parallel (so the tests all give similar results), the co-ordinates of the point change significantly if the direction of one of the vectors changes a little bit. This means that the weights assigned to tests that perform similarly to each other are very sensitive to the exact performance of the test on a particular set of co-location data.

For this reason, the results from the optimisation process are best considered as an indication of the approximate weight to be assigned to each test. We have also shown that changes in the weights of the order of 10% make only marginal difference to the relative performance of each test case (as described in Table 1).

5. Use of BUFR for representing parallel quality control data

The flexibility offered by the BUFR encoding of disseminated wind products can be exploited, to allow the inclusion of quality control information calculated with and without the use of forecast data, separately, with each wind. The use of BUFR Table B entries defining the quality control process allows the two sets of information to be distinguished, and used independently if required.

A hypothetical example applied to a wind consisting of only four parameters (speed, direction, temperature and pressure) is shown in Table 2. A real wind has 103 parameters, but the same principle applies.

Table 2. An illustration of the way in which BUFR can be used to include parallel quality control information within a message. The table should be read from left to right and then downwards. Fields highlighted in grey can be read down the columns to see what information is given for each parameter (Speed, Direction, *et c.*). Fields without a highlight do not correspond to any particular parameter.

Speed	Direction	Temperature	Pressure
QC follows	Define bitmap		
QC Included	QC Included	QC Included	QC Included
		EUMETSAT QC	QC1, with forecast
% confidence	% confidence	% confidence	% confidence
QC follows	Re-use bitmap	EUMETSAT QC	QC1, with forecast
Confidence threshold	Confidence threshold	Confidence threshold	Confidence threshold
QC follows	Re-use bitmap	EUMETSAT QC	QC2, without forecast
% confidence	% confidence	% confidence	% confidence
QC follows	Re-use bitmap	EUMETSAT QC	QC2, without forecast
Confidence threshold	Confidence threshold	Confidence threshold	Confidence threshold

Table 2 should be read left to right and from top to bottom. By reading down the columns and looking at the highlighted cells it is possible to see which quality control values are given for which parameter value.

Immediately after the measured parameters, a descriptor (222000) is used to indicate that quality control information follows. Next follows a descriptor (236000) to indicate the presence of a ‘data present’ bit map, followed by the bit map itself (four 031031 descriptors). Each entry in the bit map is set to indicate whether or not there will be quality control information following for the associated parameter. Then follows a descriptor (001031) indicating the originating/generating centre (EUMETSAT, 254), and one for the generating application (001032). It is by setting the value of this element that we can indicate which type of quality control has been performed, 1 being with forecast data being used, and 2 being without. Next a set of values relating to the parameters as per the bit map is given (033007 for % confidence, for example). Then follows the same pattern of descriptors for each successive set of quality control values, namely (i) quality control information follow (222000), (ii) re-use previously defined bit map (237000), (iii) originating/generating centre (001031), (iv) generating application (001032), and (v) the quality control values relating to the original parameters.

During the summer of 2000, EUMETSAT will use this method of BUFR encoding to send test bulletins on the GTS. The wind data in these bulletins will be the same as the operational wind data, but the quality control section will contain the two parallel sets of indicators, calculated both with and without the use of the forecast test.

6. Summary and conclusions

The principle conclusion to be drawn from this study is that a set of quality control values for the wind products can be calculated without using a forecast consistency test. The values calculated in this way will have a less well defined relationship to the actual quality of the winds (compared to co-located radiosonde data), but there will be less error correlation. The EUMETSAT wind products will be disseminated with both sets of quality control information attached.

The method used to optimise the weights assigned to each test is dependent on the tests being independent, which they are not. For this reason, the results from the optimisation process are best considered as an indication of the approximate weight to be assigned to each test. We have also shown that changes in the weights of the order of 10% make only marginal difference to the relative performance of each test case (as described in Table 1).

In an effort to make the optimisation more statistically robust, we have assessed the impact of filtering the set of co-location data used for the tuning process, by limiting the maximum NVD to be considered. The distribution of NVD within the sample of co-location data led to the choice of an NVD threshold value of 1.0, excluding ~15% of the data. This filter led to a slight improvement in the performance of the optimisation, particularly for the test cases using the correlation test, and not the forecast test.

During the summer of 2000, EUMETSAT will distribute test bulletins on the GTS, which will contain the two parallel sets of quality information, calculated both with and without the use of the forecast test. The users will then be able to decide which set of information they should use.

REFERENCES

Elliott, S. (1998). The application and implications of the use of a unified BUFR template for the exchange of satellite derived wind data. In *Proceedings of the fourth international winds workshop*. EUMETSAT, EUM P24.