TECHNICAL MEMORANDUM

250

Variational quality control

E. Andersson and H. Järvinen

Research Department

March 1998

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Variational Quality Control

By Erik Andersson and Heikki Järvinen

European Centre for Medium-Range Weather Forecasts.

Summary

We describe the implementation of variational quality control in the operational four dimensional variational data assimilation scheme at ECMWF. The quality control takes place during the iterative solution of the variational analysis problem itself, rather than in a separate step prior to the main analysis. This ensures that the quality control is consistent with the analysis in terms of error statistics, background and model constraints, in particular. All observational data are used and quality-controlled simultaneously. The probability of gross error is computed for each observation, given the preliminary analysis at each iteration. The weight of each observation is smoothly decreased with increased likelihood for gross error. The formulation, the implementation details and results from data assimilation and forecast experiments are presented.

1. Introduction

Many operational numerical weather prediction (NWP) centres have developed, or are developing, variational schemes for data assimilation (Parrish and Derber 1992; Lorenc 1995; Gauthier *et al.* 1996; Cohn *et al.* 1997; Gustafsson *et al.* 1997). At the European Centre for Medium-Range Weather Forecasts (ECMWF) a four-dimensional variational scheme (4D-Var) became operational in November 1997 (Rabier *et al.* 1997), replacing the three-dimensional version (3D-Var) implemented in January 1996 (Courtier *et al.* 1998; Rabier *et al.* 1998; Andersson *et al.* 1998). These variational schemes can be very flexible with respect to data usage, by allowing the relationship between model variables and observed quantities (the so called observation operators) to be non-linear and multi-variate (Lorenc 1986). The variational technique also offers new possibilities for the quality control of observational data (Ingleby and Lorenc 1993). In this paper we describe the quality control method employed in 3D and 4D-Var and present results on its performance. The scheme accounts for the possibility of gross errors in the data, by the use of non-Gaussian observation error statistics. The pre-analysis checks against the background are also described.

The variational quality control mechanism is incorporated within the variational analysis itself. It operates during the iterative minimisation, as part of the solution of the variational problem. The quality control is therefore fully consistent with the analysis use of the data, as the actual observation and background errors of the analysis also apply for quality control. All used data of all observed quantities influence the quality control through their observation operators, through the multivariate background constraint and, in the case of 4D-Var, through the model constraint. In 4D-Var there is the potential for an efficient time continuity check since observations are used at their correct time with respect to the model's evolution over the assimilation period (currently six hours).

The method is named variational quality control (VarQC), and is based on Bayesian probability theory (Gelman *et al.* 1995, Lorenc and Hammon 1988). In VarQC a modification of the observation objective function (or cost function) to take into account the non-Gaussian nature of gross errors, has the effect of reducing the analysis weight given to data

with large departures from the current iterand (or preliminary analysis). Data are not irrevocably rejected, but can regain influence on the analysis during later iterations if supported by surrounding data. This is in contrast to traditional quality control procedures which detect and discard questionable data prior to the main analysis (e.g. Lorenc 1981; DiMego *et al.* 1984; Woolen 1991). There is also no need for a separate, often complicated, decision making algorithm.

The approach of treating non-Gaussian errors directly within the assimilation was suggested by Purser (1984) and Lorenc (1984). The idea was first applied by Dharssi *et al.* (1992) to a set of simulated LIDAR data in a two-dimensional, iterative and nonlinear OI analysis. The paper by Ingleby and Lorenc (1993) developed VarQC theoretically and explained how VarQC, Optimum Interpolation quality control (OIQC, Lorenc 1981) and other quality control procedures are inter-related. In VarQC the weights of observations are a function of the magnitude of the departure. With purely Gaussian statistics, on the other hand, it is implied that every observation improves the analysis regardless of the distance between observation and analysis.

3D/4D-Var allows the observation operators to be (weakly) non-linear, as is the case for TOVS radiance data (Andersson *et al.* 1994), ERS scatterometer data (Stoffelen and Anderson 1997) and surface data (Cardinali *et al.* 1994) for example. The incorporation of quality control inevitably adds non-linearity to the variational problem. Non-Gaussian observation error statistics lead to a non-quadratic cost function. There may also be multiple minima in the VarQC cost function, each representing the rejection/non-rejection of individual data (see examples given in Dharssi *et al.* 1992). The technique used by Dharssi *et al.* was to set the observation errors to very large values initially and then gradually reduce them during the iterations to their 'true' values. Our approach to limit this problem has, instead, been to start from as good as possible an initial state. In this paper we discuss how this is achieved by partially minimising the cost function without quality control, before VarQC is switched on.

VarQC is a check on the departure between data and analysis. It effectively rejects data which are in strong disagreement with surrounding data. There is nevertheless a need for a variety of quality control checks prior to the analysis (such as those detailed by Sarukhanian, 1989; Gandin, 1988; Collins and Gandin, 1990; Atkins 1984; Järvinen and Undén, 1997). The checks against the background field are particularly important. The background check (BgQC) will be described in this paper as it had an influence on how we chose to implement VarQC. The pre-analysis quality control checks and data selection criteria of the ECMWF operational data assimilation system have been documented by Järvinen and Undén (1997) and are also summarized in Table 1 of Courtier *et al.* (1998).

For each observation the results of VarQC and BgQC are archived for the purposes of data monitoring (Hollingsworth *et al.* 1986; Hollingsworth 1989), scientific validation and refinement of the quality control and the data assimilation scheme as a whole. Some results of the QC validation will also be presented in this paper. Sections 2. and 3. present the VarQC method for uncorrelated and correlated data errors, respectively. Implementation details and the calibration of the system are described in Section 4. Results from assimilations with and without VarQC are presented in Section 5., followed in Section 6. by a summary and conclusions.

2. Algorithm for uncorrelated data

The variational method for data assimilation, as described by e.g. Lorenc (1986), comprises minimizing a cost function made up of two terms, J_o and J_b, measuring the distance to the observations and to the background, respectively. Both

cost functions are given a quadratic form, which assumes that the errors in both observations and background are Gaussian in nature. The expression for the observation cost function is thus

$$J_o = \frac{1}{2}(\mathbf{y} - H\mathbf{x})^{\mathrm{T}} \mathbf{O}^{-1}(\mathbf{y} - H\mathbf{x}), \tag{1}$$

where $\mathbf{y}$ is the array of observations, with error covariance matrix $\mathbf{O}$, $\mathbf{x}$ is the model state and H the observation operator (Lorenc 1986). Eq. (1) degenerates in the case of uncorrelated data to a sum of individual J_o-contributions, i.e.

$$J_o = \sum \frac{1}{2}\left(\frac{y - H\mathbf{x}}{\sigma_o}\right)^2 \tag{2}$$

with σ_o the observation error standard deviation.

As in Dharssi *et al.* (1992) and Ingleby and Lorenc (1993) we shall generalize Eqs. (1) and (2) by assuming that an observation error belongs to either of two populations: one which follows the normal Gaussian distribution, representing random errors, and one which is modelled by a flat distribution, representing the population of data affected by gross errors. Other models to represent the probability density of incorrect data could also be used (e.g. Huber 1977; Gelman *et al.* 1995). The choice of a flat distribution is convenient since it corresponds to the assumption that those data provide no useful information to the analysis.

2.1 Formulation

With a prior probability of gross error A and a probability of not having a gross error $1 - A$ we write the probability density function (pdf) p^{QC} for a single observation as a sum of two terms:

$$p^{\mathrm{QC}} = (1 - A)N + AF \tag{3}$$

N and F are the Gaussian and the flat distributions, respectively:

$$N = \frac{1}{\sigma_o\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y - \hat{y}}{\sigma_o}\right)^2\right] \tag{4}$$

$$F = \frac{1}{D} = \frac{1}{2d\sigma_o}, \text{ if } |y - \hat{y}| < D/2, \text{ zero otherwise} \tag{5}$$

where the observed value is y and the model equivalent, $\hat{y}$, of the observed quantity is the observation operator H applied to the model state $\mathbf{x}$, i.e. $\hat{y} = H\mathbf{x}$. The flat distribution is defined over an interval D (centred at zero) which in Eq. (5) for convenience has been written as a multiple d of the assumed observation error standard deviation σ_o. It is assumed in the following that absolute departures $|y - \hat{y}|$ greater than $D/2$ never occur or have been removed in a pre-analysis check against (for example) the background (see Section 4.1).

The observational cost function of the variational analysis (Lorenc 1986) is given by:

$$J_o = -\ln p + c \tag{6}$$

We obtain the normal quadratic cost function J_o^N by inserting in Eq. (6) the Gaussian pdf $p^N = N$, and arbitrarily choosing $c = -\ln(\sigma_o\sqrt{2\pi})$:

$$J_o^N = \frac{1}{2}\left(\frac{y-\hat{y}}{\sigma_o}\right)^2 \tag{7}$$

Its gradient ∇ with respect to the observed quantity $\hat{y}$ is:

$$\nabla_{\hat{y}} J_o^N = -\frac{1}{\sigma_o}\left(\frac{y-\hat{y}}{\sigma_o}\right) \tag{8}$$

Similarly substituting the modified pdf, Eqs. (3) to (5), into Eq. (6), we obtain after re–arranging the terms, an expression for the QC–modified cost function J_o^{QC} and its gradient $\nabla_{\hat{y}} J_o^{QC}$:

$$J_o^{QC} = -\ln\left[\frac{\gamma + \exp(-J_o^N)}{\gamma + 1}\right] \tag{9}$$

$$\nabla_{\hat{y}} J_o^{QC} = \nabla_{\hat{y}} J_o^N \left[1 - \frac{\gamma}{\gamma + \exp(-J_o^N)}\right] \tag{10}$$

(if $|y-\hat{y}| < D/2$, $\nabla_{\hat{y}} J_o^{QC} = \nabla_{\hat{y}} J_o^N$ otherwise), with γ defined as:

$$\gamma = \frac{A\sqrt{2\pi}}{(1-A)2d}. \tag{11}$$

The term $\gamma/(\gamma + \exp(-J_o^N))$ modifying the gradient in Eq. (10) can be shown to be equal to the *a-posteriori* probability of gross error P, given $\mathbf{x}$, and assuming that $H\mathbf{x}$ is correct (see Ingleby and Lorenc, 1993, *section 2g*):

$$P = \frac{\gamma}{\gamma + \exp(-J_o^N)} \tag{12}$$

Note also that the VarQC modification of the cost function is a function of A, d and the normalized departure $(y-\hat{y})/\sigma_o$, only.

Following Eq. (10) we define a VarQC weight W^{QC} such that

$$\nabla_{\hat{y}} J_o^{QC} = \nabla_{\hat{y}} J_o^N \; W^{QC} \text{ , i.e.} \tag{13}$$

$$W^{QC} = 1 - \frac{\gamma_i}{\gamma + \exp(-J_o^N)} = 1 - P \tag{14}$$

Eq. (14) shows that data which are found likely to be incorrect ($P \approx 1$) are given reduced weight in the analysis. Conversely, data which are found likely to be correct ($P \approx 0$) are given the weight they would have had using purely Gaussian observation error pdf.

2.2 Application

An *a-priori* estimate A of the probability of gross error and the range of possible values d are assigned to each datum, based on study of historical data, such as differences from the background and from the analysis (Section 4.3). Then, at each iteration of the variational scheme, an *a-posteriori* estimate of the probability of gross error P and the VarQC weight W^{QC} are calculated (using Eq. (14) etc.), given the current value of the iterand (the preliminary analysis). The calculated *a-posteriori* probability P depends on $\mathbf{x}$ and assumes that $H\mathbf{x}$ is correct, which means that it is important to have a preliminary analysis that is as good as possible at the start of the variational quality control. This is achieved in practise by performing the minimisation without quality control for a number of iterations (currently 40) before VarQC is switched on. VarQC then remains switched on until the end of the minimisation, i.e. during the remaining 30 iterations (Rabier *et al.* 1997). The most obviously wrong data do not influence the minimisation at all, as they have been removed in a pre-analysis check against the background (BgQC), to be described in Section 4.1.

The VarQC-modified cost function and its gradient are computed thereafter. The modified gradient becomes the input to the adjoint observation operators and provides the forcing for the adjoint integration in 4D-Var. This is a quality control algorithm without conditionals or threshold values.

In the case of many observations, all with uncorrelated errors, J_o^{QC} is computed as a sum (over the observations i) of independent cost function contributions:

$$J_o^{QC} = -\ln \prod_i p_i + C = -\sum_i \ln p_i + C = \sum_i J_{oi}^{QC} \tag{15}$$

The global set of observational data includes a variety of observed quantities as used by the variational scheme through their respective observation operators. The application of VarQC is always in terms of the observed quantity.

2.3 Illustration

Eqs. (9) and (10) are illustrated for one single observation in Fig. 1a and 1b. The figures show the normal and the VarQC cost functions (top), the VarQC gradient (middle) and the QC weight, W^{QC}, i.e. $1 - P$ (lower panel), plotted against observation departure $y - \hat{y}$. The parameters are $A = 0.01$, $d = 5$ and $\sigma_o = 2$. Figure 1a shows VarQC with a flat distribution for gross errors (as used in the ECMWF 3D/4D-Var). The cost function flattens out for large values of the departure, and its gradient goes towards zero, as the probability of the observation being correct also drops towards zero. The interval within which the observation is partly used/partly rejected is relatively narrow for this set of parameters.

The second example (Fig. 1b) shows a similar diagram resulting from using a different statistical model for F (the probability density for incorrect data). Here a Gaussian with three times the standard deviation ($3\sigma_o$) has been used rather than a flat distribution. We can see that the resulting analysis does not reject the data with a high P but uses them with the larger observation error - thereby remaining robust to outlier data (Huber, 1977).

2.4 Rejection limits

VarQC does not require the specification of threshold values at which rejections occur - so called rejection limits. Rejections occur gradually and are proportional to the obtained estimate of P. We can, however, derive from Eq. (14) an expression for the normalized departure $(y-\hat{y})/\sigma_o$ corresponding to any given value of P:

$$\left(\frac{y-\hat{y}}{\sigma_o}\right)^2 > 2\ln\frac{P}{(1-P)\gamma} \tag{16}$$

If we define a datum as rejected if its gradient has been reduced by a factor 0.25, i.e. $P = 0.75$ we obtain a relationship for the effective rejection limit ρ which is a function of A and d only:

$$\rho_i = \sqrt{2\ln\left(\frac{6d(1-A)}{A\sqrt{2\pi}}\right)} = \sqrt{2\ln(3/\gamma)} \tag{17}$$

with γ defined in Eq. (11). At first sight it might appear that the effective rejection limit ρ is independent of the background error σ_b. It should, however, be noted that the relationship has been expressed in terms of the normalized departure from $\mathbf{x}$, of the current iteration, which itself depends on specified background error statistics of the J_b-term as well as on observational data in the vicinity. If VarQC were applied from the start of the minimisation, however, data may be rejected initially where the background is poor, without reference to the estimated quality of the background, σ_b. We therefore reiterate the importance of starting VarQC from a preliminary analysis rather than directly from the background.

2.5 Vector wind data

Vector wind observations are used in 3D/4D-Var as u- and v-component data. The quality control of wind data should be designed in such a way that the whole wind observation is rejected if either of the two (or both) components are found incorrect.

With VarQC, the combined pdf for the two wind components p_{uv}^{QC} is the product of two pdfs (one for u and one for v) each with two terms, as in Eq. (3):

$$p_{uv}^{QC} = p_u^{QC}p_v^{QC} = [(1-A_u)N_u + A_uF_u]\,[(1-A_v)N_v + A_vF_v] \tag{18}$$

The expanded product includes four terms each representing a different QC outcome: both u and v correct, u but not v affected by gross error, v but not u affected by gross error and finally both u and v affected by gross error.

The desired behaviour can be achieved by neglecting the cross-terms including the products $N_u F_v$ and $N_v F_u$, and assigning their probabilities to the event representing rejection of both components, i.e.:

$$p_{uv}^{QC} = (1-A_u)(1-A_v)N_u N_v + [1-(1-A_u)(1-A_v)]F_u F_v \tag{19}$$

This allows the probability of error in u - and v -components of wind observations to be computed jointly. It corresponds to the assumption that gross errors in the wind data always affect both u - and v -components. The expressions for the cost function and its gradient remain as in Eqs. (9) and (10) with J_o^N replaced by

$$J_{uv}^N = J_u^N + J_v^N \tag{20}$$

and γ replaced by

$$\gamma_{uv} = \frac{[1-(1-A_u)(1-A_v)]/(2d_u 2d_v)}{(1-A_u)(1-A_v)/(2\pi)} \tag{21}$$

The two wind-components will then either be both rejected or both accepted (or something in between), with an *a-posteriori* probability of gross error:

$$P_u = P_v = \frac{\gamma_{uv}}{\gamma_{uv} + \exp[-J_{uv}^N]} \tag{22}$$

and gradients:

$$\begin{aligned} \nabla_{\hat{y}} J_u^{QC} &= \nabla_{\hat{y}} J_u^N (1-P_u) \\ \nabla_{\hat{y}} J_v^{QC} &= \nabla_{\hat{y}} J_v^N (1-P_v) \end{aligned} \tag{23}$$

2.6 Adjoint calculations

It is possible to avoid re-doing the quality control calculations in the adjoint. The adjoint calculations without VarQC start by computing the gradient, using Eq. (8). The gradient is subsequently used to force the adjoint model integration, as explained in Talagrand and Courtier (1987). With VarQC, however, the direct application of the corresponding Eq. (10) would require that the VarQC calculations be repeated at the start of the adjoint. It is possible to avoid this by storing, during the forward calculations, one extra value per observation. The most appropriate quantity to store (for reasons which will become apparent in the following section on correlated data) is the 'effective' normalized departure, z_e^{QC}:

$$z_e^{QC} = \frac{y-\hat{y}}{\sigma_o} W^{QC} \tag{24}$$

so that the adjoint forcing term is simply calculated by:

$$\nabla_{\hat{y}} J_o^{QC} = -z_e^{QC}/\sigma_o \tag{25}$$

which is provided to the adjoint observation operators. The adjoint calculations from then on proceed as normal, unaffected by VarQC. In the absence of VarQC we may similarly define $z_e^N = (y-\hat{y})/\sigma_o$ giving $\nabla_{\hat{y}} J_o^N = -z_e^N/\sigma_o$.

3. Algorithm for correlated observations

3.1 Formulation

Assume $\mathbf{z}$ is the vector of n correlated observations with elements $(y_l-\hat{y}_l)/\sigma_{ol}$, where the subscript l indexes individual data in the correlated set of observations. The normal cost function based on a Gaussian pdf is then given by Eq. (1), or:

$$J_o^N = \frac{1}{2}\mathbf{z}^T\mathbf{R}^{-1}\mathbf{z} \tag{26}$$

and the gradient:

$$\nabla_{\hat{\mathbf{y}}} J_o^N = -\mathrm{Diag}(\sigma_{ol}^{-1})\mathbf{R}^{-1}\mathbf{z} \tag{27}$$

where $\mathbf{R}$ is the $n \times n$ matrix of observation error correlations and $\mathrm{Diag}(\sigma_{ol}^{-1})$ is a diagonal matrix with elements $1/\sigma_{ol}$. The only type of observation for which the correlation of observation error is taken into account currently at ECMWF is radiosonde geopotential height data. The quality control of radiosonde height data is therefore more complex than for other data types. The exact calculation of the VarQC probability density function for these data involves the explicit evaluation of the probabilities of 2^n events of all possible combinations of rejection/non–rejection (Ingleby and Lorenc, 1993). We write:

$$p^{QC} = \sum_{i_e=1}^{2^n} p_{i_e} = \sum_{k=0}^{n} \sum_{i_p=1}^{n_{p,k}} p_{i_p,k} \tag{28}$$

where p_{i_e} is the pdf of event i_e. Subscript k indicates the number of correct data in each term (or event) and i_p is a counter from 1 to $n_{p,k}$ over all possible permutation of k correct and $n-k$ incorrect data in a set of n data, with $n_{p,k}$ given by the binomial frequency distribution as

$$n_{p,k} = \frac{n!}{(n-k)!k!} \tag{29}$$

Note that the subscripts i_p, k refer to a different selection of k data (from the set of n data), for each of the $n_{p,k}$ permutations i_p.

In the following we will denote the $k = n$ term ($p_{1,n}$) as the 'zero-order' term, as it represents zero rejections, i.e. all data correct, and the $k = n-1$ terms ($p_{i_p,(n-1)}$, $i_p = [1,n]$) will be called 'first-order' terms, as they each represent rejection of one datum, *et cetera* for higher order terms. In Eq. (28) each term is of the form (Ingleby and Lorenc, 1993, their Eq. 29):

$$p_{i_p,k} = \left(\prod_{l=1}^{k} (1 - A_l) \right)_{i_p,k} N_{i_p,k} \left(\prod_{l=k+1}^{n} A_l F_l \right)_{i_p,k} \tag{30}$$

where $N_{i_p,k}$ is a multi–dimensional Gaussian of k data:

$$N_{i_p,k} = \zeta_{i_p}^{k} \left[(2\pi)^k \left| \mathbf{R}_{i_p,k} \right| \right]^{-1/2} \exp\left(-\frac{1}{2} \mathbf{z}_{i_p,k}^{\mathrm{T}} \, \mathbf{R}_{i_p,k}^{-1} \, \mathbf{z}_{i_p,k} \right) \tag{31}$$

with $\zeta^k = \prod_{l=1}^{k} \frac{1}{\sigma_{ol}}$.

The matrix $\mathbf{R}_{i_p,k}^{-1}$ is the inverse of the $k \times k$ elements of $\mathbf{R}_{i_p,k}$ and $\left| \mathbf{R}_{i_p,k} \right|$ is its determinant. The calculation of $\mathbf{R}_{i_p,k}^{-1} \mathbf{z}_{i_p,k}$ in Eq. (31) is performed by solving the linear system of equations $\mathbf{R}_{i_p,k} \tilde{\mathbf{z}} = \mathbf{z}_{i_p,k}$ for $\tilde{\mathbf{z}}$ using Choleski decomposition. The determinant is obtained as the product of the diagonal elements of the decomposed matrix.

Given the expression for the combined pdf p^{QC}, the observation cost function is obtained as in Eq. (6):

$$J_{\mathrm{o}}^{\mathrm{QC}} = -\ln p^{\mathrm{QC}} + c \tag{32}$$

where c can be arbitrarily chosen such that $J_{\mathrm{o}}^{\mathrm{QC}} = 0$ for $\hat{\mathbf{y}} = \mathbf{y}$, i.e. $c = \ln p^{QC}(\hat{\mathbf{y}} = \mathbf{y})$.

Since $N_{i_p,k}$ contains the product ζ^k and F_l is proportional to $1/\sigma_{ol}$, Eq. (5), we have that every term of p^{QC} and $p^{\mathrm{QC}}(\hat{\mathbf{y}} = \mathbf{y})$ contains the product ζ^n, which therefore vanishes from the expression for $J_{\mathrm{o}}^{\mathrm{QC}}$. We see that the VarQC modification of the cost function, also in the correlated case, is a function of the input parameters (A and d) and the normalized departures, only (compare with Eq. (9)).

The gradient of $J_{\mathrm{o}}^{\mathrm{QC}}$ is obtained by differentiation of Eq. (32), which yields:

$$\nabla_{\hat{\mathbf{y}}} J_{\mathrm{o}}^{\mathrm{QC}} = -\mathrm{Diag}(\sigma_{ol}^{-1}) \left[\left\{ \sum_{i_e=1}^{2^n} \mathrm{Diag}(1 - \Gamma_l)_{i_e} p_{i_e} \mathbf{R}^{-1}{}_{i_e} \mathbf{z}_{i_e} \right\} / p^{\mathrm{QC}} \right] \tag{33}$$

where $\Gamma_l = 1$ for those terms which represent rejection of level l, and $\Gamma_l = 0$ otherwise. The terms representing rejection do not contribute to the gradient (since $1 - \Gamma_l = 0$), due to F being flat. The expression within curly brackets can be interpreted as a probability weighted average of the vector of observed normalized departures $\mathbf{z}$ projected on the inverse correlation matrix, $\mathbf{R}^{-1}$.

The *a-posteriori* probability, P_l, that the observation at level l is affected by gross error is given by the sum over all those terms in Eq. (28) that represent rejection of that level, normalized by the total probability, p^{QC}:

$$P_l = \sum_{i_e=1}^{2^n} (\Gamma_l)_{i_e} p_{i_e} / p^{\mathrm{QC}}. \tag{34}$$

3.2 Evaluating all terms

The full flexibility to allow the rejection of any combination of levels within a radiosonde ascent can only be achieved by calculating all the terms as in Eq. (28) above. This is not difficult but may be too expensive. It needs to be done once per iteration of the variational minimisation. The main cost is in the Choleski decomposition of the different $\mathbf{R}$-matrices involved. Some considerable savings could be done by pre-calculating all the matrix inverses for the fix set of standard pressure levels. However, at ECMWF non-standard level data may be used in cases when standard level data are missing.

3.3 Omitting some terms

It is feasible to simply neglect some of the terms, thereby limiting the procedure to reject only certain combinations of data. In particular it is possible to merely retain the possibility of rejecting the entire report - not individual levels within the report - by keeping only the first and the last terms of Eq. (28). This would be justifiable if one source of error is likely to affect the whole report (i.e. wrong station height or biased temperature sensor). The probability would then refer to the chance of having a gross error affecting the report at one or more levels.

Incidentally, this form of VarQC is the best solution for the quality control of wind data, as we have seen in section 2.5.

3.4 Diagonalising the problem

A further possibility is to assume that the gross errors themselves are correlated in the same way as the random errors, i.e. according to the matrix $\mathbf{R}$. One could then diagonalise the $\mathbf{R}$-matrix: $\mathbf{R} = \mathbf{E}\Lambda\mathbf{E}^{\mathrm{T}}$, where $\mathbf{E}$ is the matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues. We can define 'rotated' departures as $\mathbf{E}(\mathbf{y} - H\mathbf{x})$. We are now back to the uncorrelated case where the quality control could be applied independently to the individual 'rotated' departures. The adjoint would include a multiplication by $\mathbf{E}^{\mathrm{T}}$ to come back to the observed quantities, and to distribute the QC-weights to the observed data.

3.5 Preferred method

As already mentioned, the implementation of VarQC for uncorrelated observations is straight forward and computationally inexpensive. For correlated data it is too expensive and probably unnecessary to evaluate the probability for each combination of events (rejection/non-rejection), as described in 3.2 above. On the other hand, it seems un-

physical to assume that the gross errors have the same correlations as the random errors, as suggested in 3.4 above. The gross errors are more likely to be uncorrelated.

The preferred solution is to allow for a limited number of outcomes. The thoughts in 3.3 can be elaborated to include not only the zero-order term and the n-order term, but also the terms representing the rejection of any one, two or more data per profile - i.e. the first and second order terms, *et cetera*.

In our application n is the number of used levels from a radiosonde report, currently maximum 15. The number of terms to evaluate could be up to $2^{15} = 32768$, which was found to be prohibitively expensive. An approximate method was found in which only those terms which represent rejection of up to n_r levels are computed. If more levels than that are found likely to be in error the whole report will be rejected. The approximation is achieved by restricting k in Eq. (28) to run from $k = n - n_r$ to n. With $n = 15$ and $n_r = 3$ for example, there are then (using Eq. (29)) $n_{p,15} + n_{p,14} + n_{p,13} + n_{p,12} = 576$ terms to be evaluated. The term representing rejection of all data in the report, i.e. $p_{1,0}$ is given by the accumulated probability of all neglected terms, or equivalently one minus the probability of all included terms:

$$p_{1,0} = \left[1 - \sum_{k=n-n_r}^{n} \sum_{i_p=1}^{n_{p,k}} \left\{ \left(\prod_{l=1}^{k} (1 - A_l) \right)_{i_p,k} \left(\prod_{l=k+1}^{n} A_l \right)_{i_p,k} \right\} \right] \prod_{l=1}^{n} F_l \tag{35}$$

Introducing $\hat{p}^{QC}$ for the approximate form of p^{QC} of Eq. (28), we have:

$$p^{QC} \approx \hat{p}^{QC} = \sum_{k=n-n_r}^{n} \sum_{i_p=1}^{n_{p,k}} p_{i_p,k} + p_{1,0} \tag{36}$$

The gradient corresponding to this approximate form is obtained similarly to Eq. (33) by differentiation of J_o^{QC}, Eq. (32), with the approximate $\hat{p}^{QC}$ instead of p^{QC}:

$$\nabla_{\hat{\mathbf{y}}} J_o^{QC} = -\mathrm{Diag}(\sigma_{ol}^{-1}) \left(\left\{ \sum_{k=n-n_r}^{n} \sum_{i_p=1}^{n_{p,k}} \mathrm{Diag}(1 - \Gamma_l)_{i_p,k} \mathbf{R}_{i_p,k}^{-1} \mathbf{z}_{i_p,k} \right\} / \hat{p}^{QC} \right) \tag{37}$$

The value of n_r is set to 3 currently. In order to reduce the cost further, there is first a preliminary calculation in which $n_r = 1$, i.e. terms representing rejection of zero or one level (zero-order and first-order terms) are computed. If a gross error is found likely, the more accurate calculation (up to $n_r = 3$) is performed, otherwise it is not. If more than three levels are likely to be in error, P_l for all levels of that report becomes large and the weight of the whole report is reduced. The expression for P_l is:

$$P_l = \left[\sum_{k=n-n_r}^{n} \sum_{i_p=1}^{n_{p,k}} (\Gamma_l)_{i_p,k} p_{i_p,k} + p_{1,0} \right] / \hat{p}^{QC} \tag{38}$$

with Γ_l as defined in Eq. (33).

3.6 Adjoint calculations, correlated data

In section Section 2.6 we introduced an 'effective' normalized departure, z_e and z_e^{QC}, which can be computed in the forward observation operators, and saved as one datum per observation to be reused at the start of the adjoint. The same technique applies here, for the correlated data, with $\mathbf{z}_e^{QC}$ (now a vector) given by the summations within the curly brackets in Eq. . The gradient is then computed like Eq. (25) i.e. $\nabla_{\hat{\mathbf{y}}} J_o^{QC} = -\mathrm{Diag}(\sigma_{ol}^{-1})\, \mathbf{z}_e^{QC}$.

For completeness we may add that the effective departure for correlated data *without* VarQC is $\mathbf{z}_e^N = \mathbf{R}^{-1}\mathbf{z}$ which, if it has been stored, saves the cost of recomputing the observation error correlations in the adjoint.

3.7 Illustration

An example of VarQC applied to correlated data is shown in Fig. 2. The observation error correlation matrix is here given by a continuous vertical function $ae^{-b(\chi_1-\chi_2)^2}$ where a is 0.8, b a tuning constant close to 1 and χ_1 and χ_2 are transformation values, based on a sixth degree polynomial in logarithm of pressure, of the two pressures involved. The remaining fraction $1-a$ of the variance is assumed un–correlated (as documented in ECMWF Research Manual 1). The correlation function is thus $(1-a)\delta_{\chi_1-\chi_2} + ae^{-b(\chi_1-\chi_2)^2}$, with δ being the Kronecker delta.

The full line in Fig. 2 shows the normalized departure $\mathbf{z}$, for a profile of radiosonde geopotential data at the 15 standard pressure levels. The dashed and dash-dotted lines show $\mathbf{z}_e^N$ and $\mathbf{z}_e^{QC}$, respectively and the dotted line is P_l, the *a-posteriori* probability of gross error as defined in the previous section. Comparing $\mathbf{z}$ and $\mathbf{z}_e^N$ we see that the effect of the observation error correlation is to reduce the importance of the bias in the observed profile and to accentuate the gradient between levels. From the dotted curve P_l, we see that VarQC in this case has rejected the lowest level of the observed profile, having estimated the probability of gross error to be 80%. All other levels have been found to be correct, i.e. $P_l \approx 0$. Because of the vertical correlation, the rejection of the lowest level also affects the levels immediately above (Eq.), as can be seen by comparing the curves for $\mathbf{z}_e^N$ and $\mathbf{z}_e^{QC}$ (dashed and dash-dotted).

4. Implementation and calibration

The first ECMWF implementation of 3D-Var (Courtier *et al.* 1998; Rabier *et al.* 1998; Andersson *et al.* 1998) relied on the Optimal Interpolation (OI) analysis system (Lorenc 1981) for quality control of data. All data which passed the OI quality controls (OIQC) were then used in the 3D-Var analysis. The OIQC compared each datum in turn against an analysis based on surrounding data, excluding the datum being checked. Data were flagged and subsequently rejected if the observation departure from the analysis exceeded a certain multiple of the estimated analysis error. VarQC was developed to replace OIQC, in order to have the variational system independent from the OI codes.

4.1 Background check

Both OIQC and VarQC rely on there being a pre-analysis screening of data, such as a check against the background fields - in our case a six hour forecast from the previous analysis. The BgQC (Järvinen and Undén 1997) has effectively remained unchanged during the transition from OI to 3D and 4D-Var. It checks that the normalized departure from the background $z_b = (y - H\mathbf{x}_b)/\sigma_b$ is less than a factor α times its estimated error variance, i.e.

$$z_b^2 \leq \alpha\,(1 + \sigma_o^2/\sigma_b^2) \tag{39}$$

The factor α may be different for different observation types and variables (Järvinen and Undén 1997).

4.2 Replacement for OIQC

At first VarQC was tuned, as closely as possible, to reproduce the results of OIQC. Histograms of OI rejections were studied to determine its effective rejection limits. The A and d parameters of VarQC were then set such that similar rejection limits were achieved, using Eq. (17). It turned out that the d parameter could be set to 5 for most data types, as the rejection limit is a much stronger function of A than of d.

Figure 3 shows an example of such histograms. It shows all rejected SYNOP pressure data in a one-month trial period as a function of the departure from the background, comparing OIQC (open bars) and VarQC (shaded). The diagram verifies that the two quality control methods performed similarly. The A and d parameters for all data types except TOVS and SCAT data (for which there was no OIQC), were set in a similar manner. Data assimilation and forecast experiments confirmed that the two quality control methods (tuned in this way) gave equivalent forecast performance (not shown). On that basis VarQC became operational at ECMWF in September 1996.

4.3 Revision of VarQC parameters

In May 1997 the background term of the ECMWF 3D-Var was redesigned (Bouttier *et al.* 1997) and the estimation of background errors was improved (Fisher and Courtier 1995). These changes had a pronounced effect on the quality control and it was found necessary to adjust the BgQC and the VarQC settings. This was done by study of histograms of observation departures from the background and the analysis, respectively. The histograms were transformed according to $\hat{f} = \sqrt{-2\ln[f/\max(f)]}$, where f is the number of data in each bin of the histogram (Hollingsworth 1989). In such a transformed histogram the Normal part of the distribution appears as a pair of straight lines (in the vicinity of the origin), with a slope equal to the inverse of the standard deviation of the population. The BgQC rejection limits and the VarQC parameters A and d were then determined such that rejection would occur for data values with significantly higher frequencies (lower $\hat{f}$) than is compatible with an assumption of Normal distribution.

An example constructed from two weeks of AIREP temperature data is shown in Fig. 4. Fig. 4a shows the histogram f of observation departures from the background and Fig. 4b shows the transformed histogram, $\hat{f}$. The slope of the straight lines in Fig. 4b gives the standard deviation of the Normal distribution - in this case 1.4 K, i.e. $\sqrt{\sigma_o^2 + \sigma_b^2} = 1.4\text{K}$. The corresponding Gaussian distribution is shown with a dashed line in Fig. 4a. The rejection limit is set some distance beyond the point were $\hat{f}$ separates from (falls below) the straight lines in Fig. 4b, in this case 6.0 K was chosen for the

BgQC. The VarQC parameters were determined in a similar fashion, from transformed histograms of normalized departures from the analysis, $z_a = (y - H\mathbf{x}_a)/\sigma_o$.

The distributions of rejected data (by BgQC and VarQC) before and after the revision are shown in Fig. 5. The example again shows AIREP temperature data, for the same two week period as in Fig. 4. The vertical axis has been chosen such that the tails of the distributions can be clearly seen. Before retuning (Fig. 5a) the BgQC was too tight, and data partly within the Normal distribution were rejected. Only very few data (17) had therefore to be rejected by VarQC. After retuning (Fig. 5b) we see that BgQC rejections coincide better with the tails of the distribution, leaving VarQC to reject those data that have large departures from surrounding data in the analysis.

The performance of VarQC is however best monitored in plots of normalized departures from the analysis, z_a. Fig 5c shows a histogram of data with probability of gross error between 0 and 25% (thin), 25 to 50% (bold), 50 to 75% (heavy bold) and 75 to 100% (filled), as a function of z_a. These four categories can be labelled 'good', 'suspect', 'possibly wrong' and 'rejected', repectively. We can see that the separation lines between these four categories are a function of z_a only, as expected from Eq. (16).

4.4 TOVS and SCAT data

At the time of the revision, VarQC was extended to apply also to TOVS radiance data and SCAT ambiguous winds. The check of radiance data is like the check of any other uncorrelated data, i.e. according to Eqs. (9) to (11). For SCAT data, however, a non-standard observation cost-function $J^{\mathrm{N}}_{\mathrm{SCAT}}$ is used in order to enable the variational analysis to carry out the ambiguity removal (Stoffelen and Anderson 1997):

$$J^{\mathrm{N}}_{\mathrm{SCAT}} = \left(\frac{J^4_{uv1} J^4_{uv2}}{J^4_{uv1} + J^4_{uv2}} \right)^{1/4} \tag{40}$$

where subscripts 1 and 2 refer to each of the two ambiguous winds.Using the VarQC formulation for conventional wind data (Section 2.5) with J^{N}_{uv} replaced by $J^{\mathrm{N}}_{\mathrm{SCAT}}$ we obtain a scheme which will reject both of the two ambiguous SCAT winds if neither wind can be fitted by the analysis.

With the inclusion of TOVS and SCAT, VarQC is now applied to all data used by 3D/4D-Var. For a description of 3D/4D-Var data usage, see Courtier *et al.* (1998) with the addition that since that paper, the remaining usage of TOVS thickness data (in the stratosphere above 100 hPa, and in the North Pole area) has been replaced by TOVS cloud cleared radiances. Development of VarQC for TOVS retrieved data (thickness and precipitable water content) has therefore not been required.

5. Results

5.1 Number of rejected data

The performance of the quality control was studied in a three week period (19971009-19971029) during which 3D-Var and 4D-Var had been run in parallel, and both schemes were presented with the same data. The results, in terms of

rejection counts, are shown in Table 1. Both assimilation systems very nearly use the same number of data - only the 4D-Var sample sizes are given in the table. The observation errors are also the same in the two systems, as documented in the appendix of Courtier *et al.* (1998). The observation types are referred to as SYNOP: Manual or automatic surface observations from land or ships; AIREP: Aircraft observation; SATOB: Cloud drift winds from geostationary satellite imagery; DRIBU: Drifting buoys; TEMP: Radiosonde balloon ascents from land or ships; PILOT: Vertical wind profiles; TOVS: Polar orbiting satellite data; PAOB: Surface pressure pseudo observations derived from satellite imagery for the Southern Hemisphere oceans and SCAT: Radar backscatter measurements from polar orbiting platforms. (See Strauss (1996) for a general presentation of the global observing system.)

Table 1: Rejection counts and approximate rejection limits for BgQC and VarQC, in 3D-Var and 4D-Var, for all data types used in the analyses. From 971009-00 to 971029-18 UTC.

Obs. Type	*Obs. quantity*	*Sample size*	*No. of BgQC rejections*		*No. of VarQC rejections*		*Approximate rejection limits*		*Rejection ratio (%)*
			3D-Var	*4D-Var*	*3D-Var*	*4D-Var*	*BgQC*	*VarQC*	*4D-Var*
SYNOP	pressure	399614	5267	5061	2063	2156	9.0 hPa	3.8 hPa	1.81
SYNOP	height	11582	5021	4883	295	287	70 m	38 m	45.12
SYNOP	wind	35032	145	141	98	110	24.5 m/s	23.0 m/s	0.72
SYNOP	rel. hum	223171	303	271	542	578	55%	36%	0.38
AIREP	temp	589716	5478	5220	1943	2068	6 K	4.8 K	1.24
AIREP	wind	666734	9195	6730	749	523	5-30 m/s	26 m/s	1.09
SATOB	wind	426897	24678	22865	840	879	5-12 m/s	8.1 m/s	5.56
DRIBU	pressure	18141	937	940	237	217	8.0 hPa	4.2 hPa	6.38
DRIBU	wind	4693	86	72	15	13	13 m/s	13 m/s	1.81
TEMP	height	276337	2447	2336	883	1139	40-100 m	16-100 m	1.26
TEMP	wind	262878	1342	1328	684	753	20 m/s	16-29 m/s	0.79
TEMP	spec hum	139972	2505	2409	77	119	varying	$3.9*\sigma_o$	1.81
PILOT	wind	71969	633	600	184	244	16-22 m/s	15-23 m/s	1.17
TOVS	radiance	2768318	n/a	n/a	2100	2474	n/a	Table 2	0.14
PAOB	pressure	6874	52	44	130	129	14 hPa	9 hPa	2.52
SCAT	wind	189018	n/a	n/a	864	637	n/a	3.0 m/s	0.34

From the table we see that the number of BgQC rejections is generally similar in 3D and 4D-Var, except for those data types which have a large proportion of asynoptic (off-time) data, e.g. AIREP temperature and wind data. In 3D-Var there can be up to a three hour discrepancy between the valid time of the background and the time of the observation, whereas in 4D-Var the background is available hourly. The better 4D-Var background should affect the specified estimates of its accuracy (σ_b), although, currently, the same σ_b 's are used in 3D and 4D-Var. The 4D-Var σ_b are likely to be lowered in a future revision of the system. The TOVS and SCAT data are currently not subjected to BgQC. TOVS radiance data

are however quality controlled by carrying out a 1D-Var retrieval (Eyre *et al.* 1993) which uses the background. Radiance data failing the 1D-Var checks are not used in 3D/4D-Var.

The 3D and 4D-Var VarQC rejection counts are also similar for most data types (the 6^{th} and 7^{th} columns of Table 1). There are, however, fewer AIREP wind data rejections in 4D-Var - because of using the data at the correct time - and larger numbers of 4D-Var rejections of TEMP height and TEMP and PILOT wind data. It appears that the dynamics of 4D-Var helps detect inconsistencies in these vertical profiles of data.

The rejection limits shown in the table were determined as described in Section 4.3. The rejection limits could more accurately have been given in terms of normalized quantities z_b and z_a, but we think it is more useful to present the physical quantities. The AIREP BgQC rejection limits may be quite low (as low as 5 m/s) at lower levels of the atmosphere around important airports, where the high concentration of aircraft data leads to very low estimates of the background error, in our system. SATOB BgQC limits are tight for observations which are indicating weaker winds than the background (Tomassini *et al.* 1997); this is to minimize the effect of a slow bias in the data. The ranges of rejection limits given for TEMP and PILOT data are indicative of their vertical variation. The lower (higher) values apply in the lower (higher) parts of the atmosphere. TOVS VarQC limits are different for each channel, as listed in Table 2.

Table 2: VarQC rejection limits for used TOVS channels in brightness temperatures (K). HIRS=High resolution Infrared Sounder, MSU=Microwave SOunding Unit and SSU=Stratospheric Sounding Unit.

HIRS	1	2	3	4	5	6	7	8	9	10
limit	3.5	1.5	1.3	1.0	1.1	1.7	1.85	2.2	n/a	2.1
HIRS	11	12	13	14	15					
limit	3.1	4.7	1.4	1.4	1.5					
MSU	1	2	3	4						
limit	n/a	1.5	1.1	1.25						
SSU	1	2	3							
limit	1.8	2.2	4.0							

The last column of Table 1 shows the 4D-Var rejection ratio, i.e. the sum of BgQC and VarQC rejections divided by the sample size. TOVS and SCAT have the lowest rejection ratios: 0.14 and 0.34%, respectively. This is partly because these data are not subjected to BgQC (but screened instead in separate modules prior to the analysis, as discussed above), and partly because they tend to be horizontally consistent within the subsatellite tracks, and there are few independent data providing redundancy for effective VarQC over the oceans.

At high elevations some SYNOP stations report the height of the nearest standard pressure level, instead of reporting the observed pressure. The very high rejection ratio for SYNOP height data (45%) can be explained by the fact that these data are almost exclusively over the most mountainous regions of the world. Large departures from the background and analysis, and a large number of rejections occur in these circumstances because of lack of representativeness of the stations or difficulties extrapolating the model fields to the level of the data.

5.2 Influence of dynamics in 4D-Var

It has been demonstrated by Thépaut *et al.* (1996) that the structure functions implied by 4D-Var are flow dependent. The influence of model dynamics is stronger for longer assimilation periods (12 and 24 hours), but even in the current 4D-Var (with a 6-hour assimilation period) we can expect that the effective background error is increased in dynamically active (low-pressure) areas, compared to 3D-Var. The expected effect on VarQC is that rejections should be less likely in the most active areas.

To get an indication of the importance of this process, we subdivided the rejection counts for SYNOP-ship and DRIBU surface pressure data in three categories depending on the observed pressure: High (above 1020 hPa), Intermediate (between 1000 and 1020 hPa) and Low (below 1000 hPa), and extended the study period to include 19971009 to 19971112, for more significant results. The results are shown in Table 3. We see that the rejection counts are almost the same in 3D and 4D-Var for the High and Intermediate categories, but there are 22% less rejections in 4D-Var in the Low-pressure category. This result is a clear indication that the dynamics in 4D-Var has an influence on VarQC.

Table 3: Rejection counts for SYNOP-ship and DRIBU surface pressure data, 19971009 to 19971112, subdivided in three categories based on observed pressure.

Category	3D-Var	4D-Var
High pressure(>1020 hPa)	272	271
Intermediate (1000-1020 hPa)	726	733
Low pressure (< 1000 hPa)	176	136

5.3 Forecast impact

The forecast impact of VarQC was tested in a two week period: 19970419 to 19970502. A 3D-Var assimilation with the revised quality control (BgQC and VarQC) was compared with a 3D-Var assimilation without VarQC (but with BgQC). The results in terms of 500 hPa geopotential anomaly correlation (Fig. 6) is very slightly positive, or neutral, in the Northern Hemisphere (Fig. 6a), and it is more clearly positive (up to day five) in the Southern Hemisphere (Fig. 6b). Both experiments have been verified against operational analyses. The scatter diagrams (Fig. 7) of 500 hPa geopotential root mean square error, for the 48h forecasts, confirm that the Northern Hemisphere impact of VarQC is very small indeed (Fig. 7a). The scatter for the Southern Hemisphere indicates a fairly systematic positive impact of VarQC.

The sensitivity to quality control will depend on the level of synoptic activity in the study period. The difference in scores beyond about day five are not reliable for a two-week sample. Two earlier two-week experiments of the same type, performed with older versions of the 3D-Var system (before J_b -revisions etc.) had indicated similarly small impact of VarQC on Northern Hemisphere forecast performance (not shown). VarQC rejects data which deviate strongly from surrounding data. The impact on analyses is often large, but in the most data dense areas we found that the analysis very effectively filters out the negative effect of the occasional incorrect datum, even without VarQC. In such circumstances the analysis perturbation caused by an incorrect datum, may well be within the normal accuracy of the analysis, and randomly lead to better, as well as worse, forecast performance. An example of AIREP rejections, 19970205-12 UT, that lead to a forecast improvement, is shown in Fig. 8. Figure 8a shows the observed winds and 8b shows the departures

from the background for the temperature and wind AIREP data rejected by BgQC (marked FG) and the departures from the analysis for the AIREP data rejected by VarQC (marked 3D). We can see that the rejected wind data clearly deviate from the surrounding observations, and are inconsistent with the analysed height field. The rejected temperature data have departures in excess of 5 K. The diagnostic singular vector methodology of Buizza *et al.* (1997) was applied to this and several other cases of large short-range forecast impact of VarQC: The 48-hour forecast difference (between forecasts with and without VarQC) was projected onto the +48 hour evolved singular vectors. The amplitudes of the projection were then applied to the singular vectors at initial time to obtain the initial perturbation corresponding to a linear evolution of the forecast differences over the 48-hour interval, in the unstable subspace spanned by the singular vectors. This procedure can be very useful to help localize the important analysis differences which have contributed the most to specific forecast differences in the short range. It was found that the wind rejection at (58 N, 50 W) coincided with an unstable area between Newfoundland and Greenland (Fig 9a), which evolved to contribute to the large negative +48 hour forecast difference to the North of Scotland and the large positive forecast difference over the Baltic (Fig. 9b). The forecast with VarQC verified better than the forecast without VarQC, in this case (not shown).

The forecast impact of the QC revision (i.e. comparing forecasts from two assimilations, one before and one after the revision) on the same two week period indicated a slight positive impact of the revision (not shown). The impact of the BgQC has not been tested, as many of the BgQC rejected data are so obviously in error that an assimilation without BgQC would seem unrealistic.

6. Summary and conclusion

A new quality control mechanism (VarQC) suitable for variational data assimilation schemes has been developed and presented. The scheme is based on the theory set out by Ingleby and Lorenc (1993). We have shown that VarQC is an efficient and adequate replacement for Optimum Interpolation quality control (Lorenc 1981).

VarQC is operational as an integral part of the ECMWF four-dimensional variational data assimilation scheme (4D-Var). The quality control takes place through a modification of the observation cost function, which takes into account the probability of gross errors in the data. The probability density function (pdf) of gross error has here been modelled as a flat distribution. This is convenient since a flat pdf corresponds to the assumption that those data provide no useful information to the analysis.

The scheme provides an estimate of the probability that each observation is in error, given the observed value and the analysis. The weight given to the observation becomes smaller as the probability of gross error increases. The weight is recalculated at each iteration of the minimisation, which means that data can regain influence on the analysis if supported by surrounding data. The calculations are performed in terms of the observed quantities (e.g. wind components, surface pressure, geopotential height, temperature, radiances and specific humidity etc.) and all data from the global set of observations are used and quality controlled simultaneously. The background constraint and, in the case of 4D-Var, the model constraint also influence the quality control. In 4D-Var there is the potential for an efficient time continuity check since observations are used at their correct time with respect to the model's evolution over the assimilation period (currently six hours).

The implementation of VarQC is relatively straight forward. It requires prior values for the probability of gross error for all observations (determined from historical data), a functional form for the probability density of gross errors and the

setting of reasonable parameter values of that chosen function. The extension to correlated data is more complicated. It involves computing probabilities for all possible combinations of events of rejection/non-rejection amongst the correlated set of data (Ingleby and Lorenc 1993). For a set of n correlated data 2^n terms would need to be evaluated for the combined probability density function. In this paper we have used an approximate form which gives acceptable results for radiosonde height data, the only data which we currently assume to have correlated observation errors. The approximate form can detect up to three incorrect levels in the radiosonde reports. If more than three levels are likely to be affected by gross errors the whole report is rejected.

Quality control is a very non-linear process which could slow down the convergence of the variational minimisation. VarQC introduces multiple minima in the cost function, corresponding to different quality control decisions. It is therefore important to have a good starting point for the VarQC minimisation and to remove obviously incorrect data in the pre-analysis check against the background (BgQC). Moreover, the method assumes that the observation equivalent is correct at each iteration (equal to truth), which makes the starting point doubly important. In our implementation a good starting point for VarQC is achieved by carrying out 40 iterations of the 4D-Var minimisation without quality control before switching on VarQC for the remaining 30 iterations.

The current BgQC is performed separately, prior to the analysis. It rejects observations for which the departure from the background is in excess of a certain multiple times its expectation, taking specified observation errors and background errors into account. This implementation, although found to perform satisfactorily in most cases, has a drawback. In extreme cases of rapid development the background itself can be so much in error that correct observations are mistakenly rejected. Currently, the BgQC rejections are irrevocable. It may be advantageous in the future to allow some of the data rejected by BgQC to have a second chance to influence the analysis during the iterations with active VarQC. If these data are still far from the preliminary analysis their weight will remain small or zero. If however, other surrounding data have brought a sufficient correction to the analysis (during earlier iterations), these 'dormant' data may be reinstated, by VarQC gradually increasing their weight.

The development of the 4D-Var system on longer assimilation periods (12 and 24 hours) and the development of a Simplified Kalman Filter, will make the estimated background error at observation points more dependent on the model dynamics (Thépaut *et al.* 1996; Rabier *et al.* 1997). This will make VarQC less likely to reject data in dynamically active areas, such as developing or rapidly moving cyclones and troughs. In this paper we reported the encouraging result that 4D-Var, already on a six hour assimilation period, is less prone to reject SHIP and DRIBU surface pressure observations in low pressure areas than 3D-Var - a consequence of the flow-dependent structure functions implied by 4D-Var.

Acknowledgements

We are grateful to J. Eyre and P. Courtier, early advocates of variational quality control. We are also grateful to B. Ingleby and A Lorenc at UKMO and P. Unden, G. Kelly and other ECMWF colleagues for valuable discussions. We thank A. Hollingsworth, A. Simmons and F. Bouttier for helpful comments improving the manuscipt. R. Buizza kindly providied the singular-vector tools for diagnosing forecast differences (used in Fig. 9). Rob Hine skillfully improved the graphics.

References

Andersson, E., Pailleux, J., Thépaut, J-N., Eyre, J., McNally, A. P., Kelly, G., and Courtier, P., 1994: Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **120**, 627-653.

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasiljevic, D., Brankovic, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Thépaut, J-N. and Viterbo, P., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part III: Experimental results. To appear in *Q. J. R. Meteorol. Soc.*

Atkins, M. J., 1984: Quality control, selection and processing of observations in the Meteorological Office's operational forecast system.ECMWF workshop on "The Use and Quality Control of Meteorological Observations", Reading, 6-9 November 1984, 255-290.

Bouttier, F., Derber, J. and Fisher, M., 1997: The 1997 revision of the J_b -term in 3D/4D–Var. ECMWF Tech. Memo. 238.

Buizza, R., Gelaro, R., Molteni, F. and Palmer, T. N.: 1997: The impact of increased resolution on predictability studies with singular vectors. *Q. J. R. Meteorol. Soc.*, **123**, 1007-1033.

Cardinali, C., Andersson, E., Viterbo, P., Thépaut, J-N., and Vasiljevic, D., 1994: Use of conventional surface observations in three-dimensional variational data assimilation. ECMWF Tech. Memo. 205.

Cohn, S. E., Da Silva, A., Guo, J., Sienkiewicz, M. and Lamich, D., 1997: Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System. Submitted to *Mon. Wea. Rev.*

Collins, W. G. and Gandin, L. S., 1995: Comprehensive hydrostatic quality control at the National Meteorological Centre. *Mon. Wea. Rev.*, **18**, 2754-2767.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. To appear in *Q. J. R. Meteorol. Soc.*

Dharssi, I., Lorenc, A. C., and Ingleby, N. B., 1992: Treatment of gross errors using maximum probability theory. *Q. J. R. Meteorol. Soc.*, **118**, 1017-1036.

DiMego, G. J., Phoebus, P. A. and McDonell, J. E., 1984: Data processing, quality control and selection for optimum interpolation analyses at the National Meteorological Centre. ECMWF workshop on "The Use and Quality Control of Meteorological Observations", Reading, 6-9 November 1984, 325-368.

ECMWF: Research Manual 1. ECMWF data assimilation scientific documentation. Available from ECMWF.

Eyre, J.R., Kelly, G. A., McNally, A. P., Andersson, E., and Persson, A., 1993: Assimilation of TOVS radiance information through one-dimensional variational analysis. *Q. J. R. Meteorol. Soc.*, **119**, 1427-1463.

Fisher, M. and Courtier, P., 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation, ECMWF Tech. Memo. 220.

Gandin, L. S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.* **116**, 1137-1156.

Gauthier, P., Fillion, L., Koclas, P. and Charette, C., 1996: Implementation of a 3D variational analysis at the Canadian Meteorological Centre. XI AMS Conference on "Numerical Weather Prediction", Norfolk, Virginia, 19-23 August 1996.

Gelman, A., Carlin, J. B., Stern H. S. and Rubin, D. B., 1995: *Bayesian data analysis*. Texts in Statistical Science, Chapman and Hall.

Gustafsson, N., Lönnberg, P. and Pailleux, J., 1997: Data assimilation for high resolution limited area models. *J. Meteorol. Soc. Japan*, **75**, 367-382.

Hollingsworth, A., Shaw, D. B., Lönnberg, P., Illari, L., Arpe, K. and Simmons, A., 1986: Monitoring of observations and analysis quality by a data assimmilation system. *Mon. Wea. Rev.*, **114**, 861-879.

Hollingsworth, A., 1989: The role of real-time four-dimensional data assimilation in the quality control, interpretation, and synthesis of climate data. *Oceanic Circulation Models: Combining data and dynamics*. Ed. D. L. T. Anderson and J. Willebrand. Kluwer Academic Publishers.

Huber, P. J., 1977: Robust statistical methods. Soc. for Industrial and Applied Mathematics, 27.

Ingleby, N. B. and Lorenc, A. C., 1993: Bayesian quality control using multivariate normal distributions. Q. J. R. Meteorol. Soc., **119**, 1195-1225.

Järvinen, H. and Undén, P., 1997: Observation screening and first guess quality control in the ECMWF 3D-Var data assimilation system. ECMWF Tech. Memo. 236.

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701-721.

Lorenc, A. C., 1984: Analysis methods for the quality control of observations. Proc. ECMWF workshop on "The use and quality control of meteorological observations for numerical wether prediction', 6-9 Nov. 1984. 397-428.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177-1194.

Lorenc, A. C., 1995: Development of an operational variational assimilation scheme. *J. Meteorol. Soc. Japan*, **75**, 415-420.

Lorenc, A.C. and Hammon, P., 1988: Objective quality control of observations using Bayesian methods. Theory, and a practical implementation. *Q. J. R. Meteorol. Soc.*, **114**, 515-543.

Parrish, D.F. and Derber, J. C., 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

Purser, R. J., 1984: A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis. 10th Conf. on Weather Forecasting and Analysis, American Meteorological Society, 102-105.

Rabier, F., Mahfouf, J-F., Fisher, M., Järvinen, H., Simmons, A., Andersson, E., Bouttier, F., Courtier, P., Hamrud, M., Haseler, J., Hollingsworth, A., Isaksen, L., Klinker, E., Saarinen, S., Temperton, C., Thépaut, J-N., Undén, P. and Vasiljevic, D., 1997: Recent experimentation on 4D-Var and first results from a Simplified Kalman Filter. ECMWF Tech Memo 240.

Rabier, F., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A., and Bouttier, F., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part II: Structure functions. To appear in *Q. J. R. Meteorol. Soc.*

Sarukhanian, E. I., 1989: Data quality control. WMO/ECMWF workshop on "Data quality control procedures", Reading, 6-10 March. 13-24.

Stoffelen, A. and Anderson, D., 1997: Ambiguity removal and assimilation of scatterometer data. To appear in *Q. J. R. Meteorol. Soc.*

Strauss, B., 1996: The global observing system. Proc. ECMWF Seminar on "Data assimilation", Reading 2-6 September 1996. 7-29.

Talagrand, O. and Courtier, P., 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I. Theory. *Q. J. R. Meteorol. Soc.*, **113**, 1321-1328

Thépaut, J.-N., Courtier, P., Belaud, G., and Lemaître, G., 1996: Dynamical structure functions in a four-dimensional variational assimilation: a case study. *Q. J. R. Meteorol. Soc.*, **122**, 535-561.

Tomassini, M., Kelly, G., and Saunders, S., 1997: Use and impact of satellite atmospheric motion winds on ECMWF analyses and forecasts. EUMETSAT/ECMWF Fellowship Programme, Report No. 6. pp 31.

Woolen, J.S., 1991: New NMC operational OI quality control. Preprints of Ninth AMS Conference on "Numerical Weather Prediction", Denver, CO, 24-27.

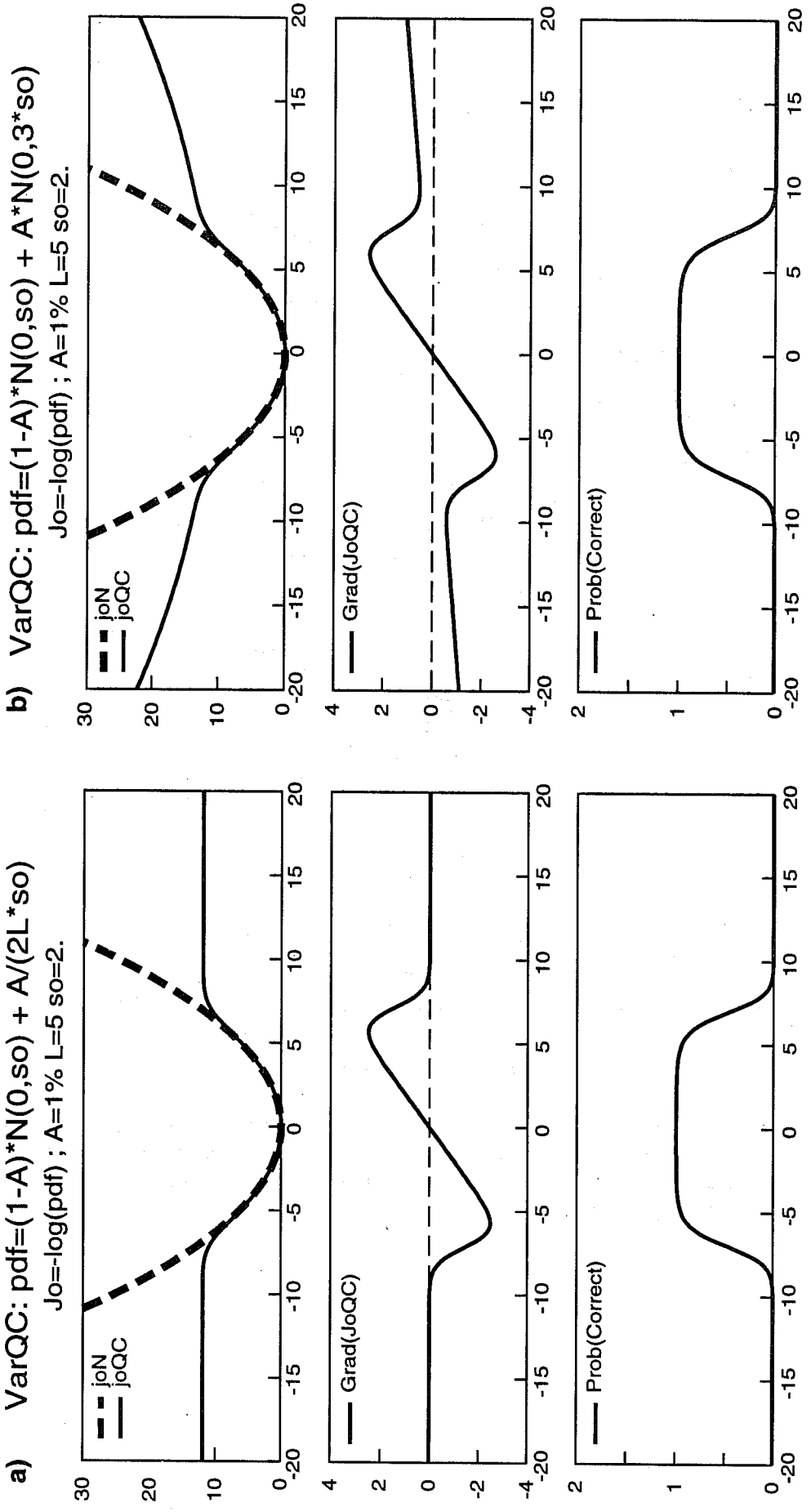


Fig. 1 Top panels show cost function without QC (dashed) and with VarQC (full line). Middle panels show gradient of VarQC cost function and lower panels show the a-posteriori probability that the observation is correct. The distribution of gross errors is flat in a) and a Gaussian with three times the observation error in b).

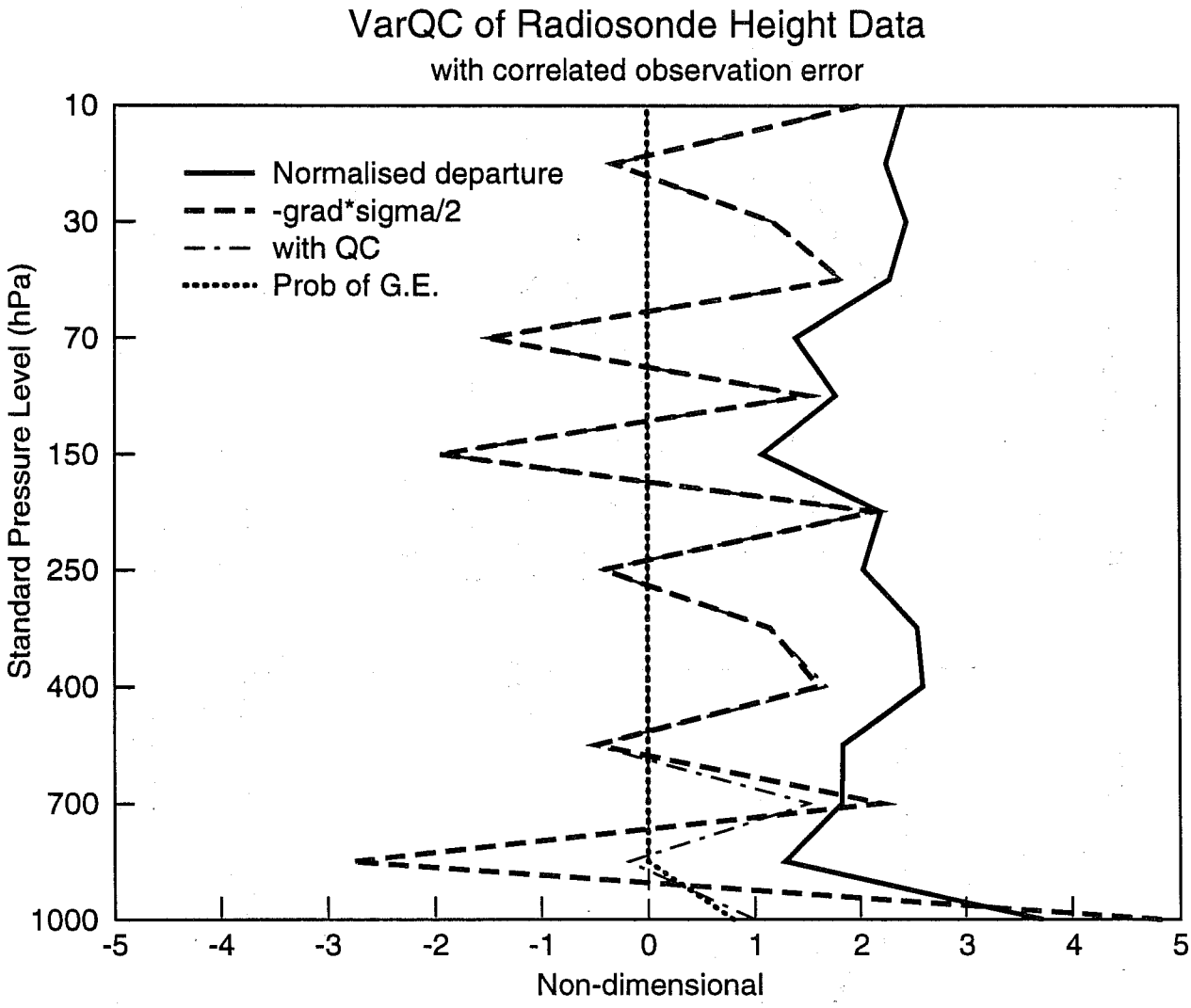


Fig. 2 Normalised analysis departure for a radiosonde profile of height data, against standard pressure level. The full line is observation minus analysis divided by observation error. The long dashed and the dashed lines are the 'effective' normalized departures (as defined in the text) with and without VarQC, respectively. The dotted line is the a-posteriori probability of gross error.

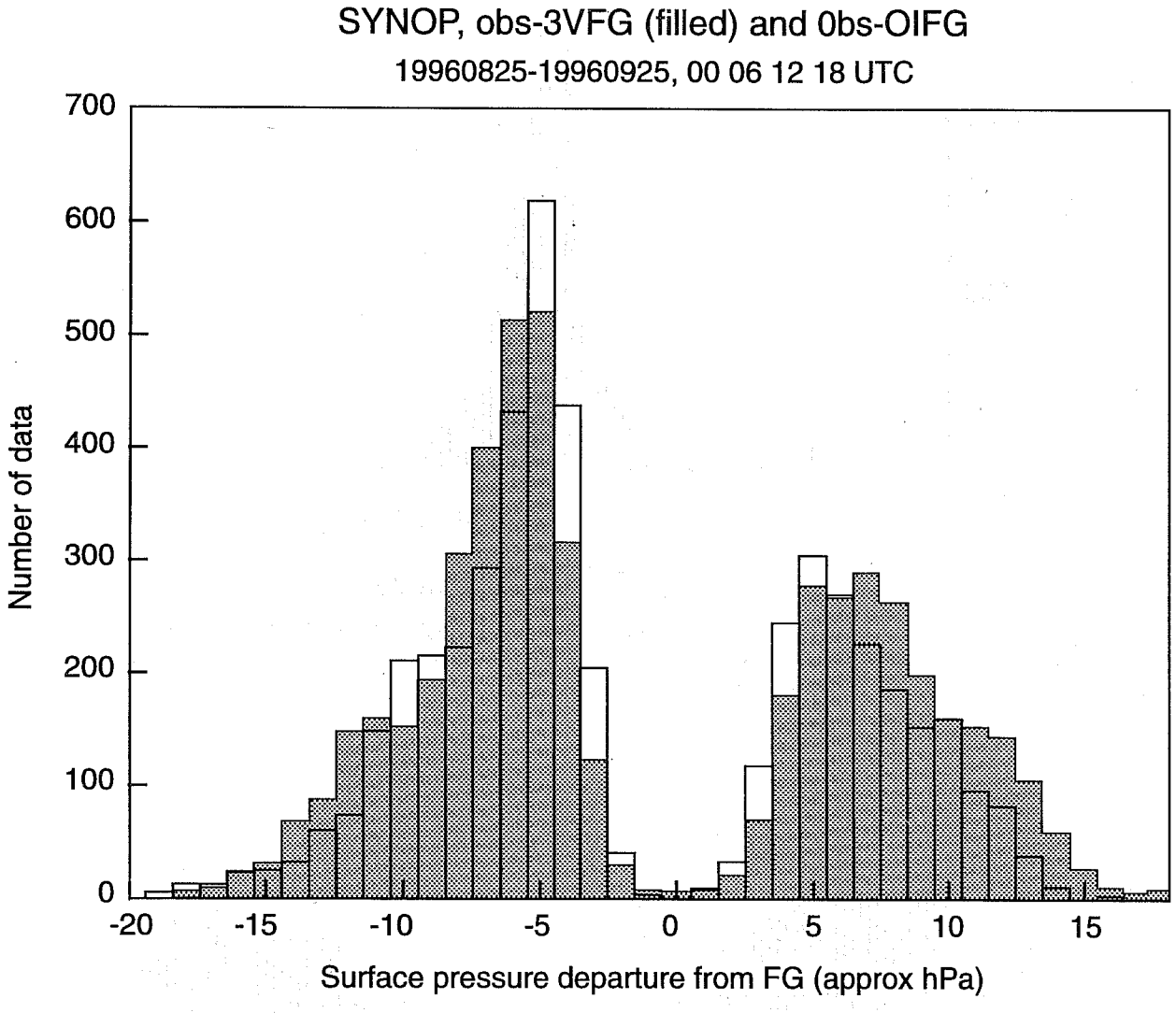


Fig. 3 Histogram of observation minus background departures for rejected SYNOP pressure data for OIQC (black outline) and VarQC (filled grey), in a 30 day period.

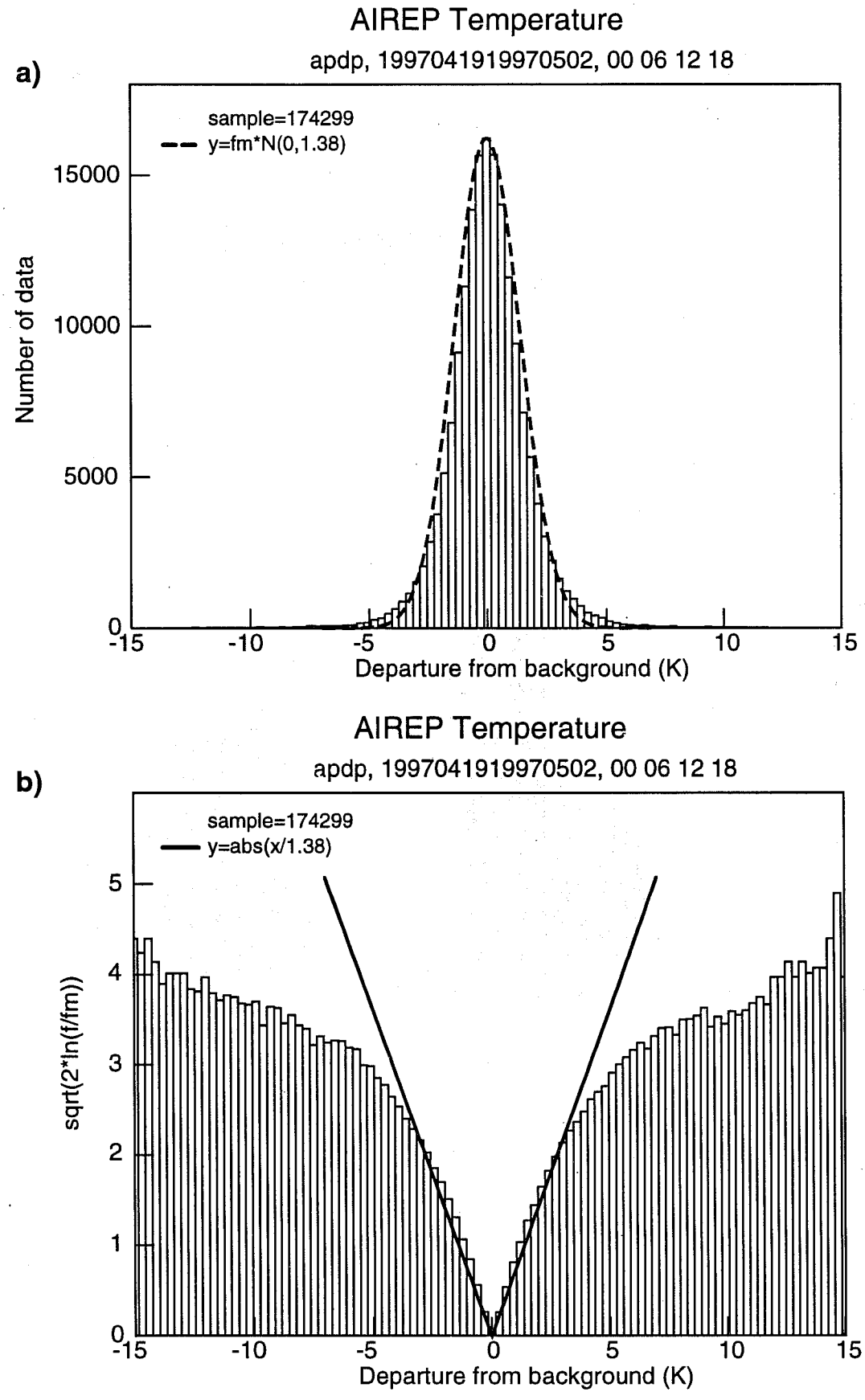


Fig. 4 Histogram (a) and transformed (see main text) histogram (b) of AIREP temperature departures from the background. The slope of the straight lines in b) defines the standard deviation of the Gaussian curve (dashed drawn in a). Data were collected from a 14-day assimilation, 19970419-00 to 19970502-18 UTC.

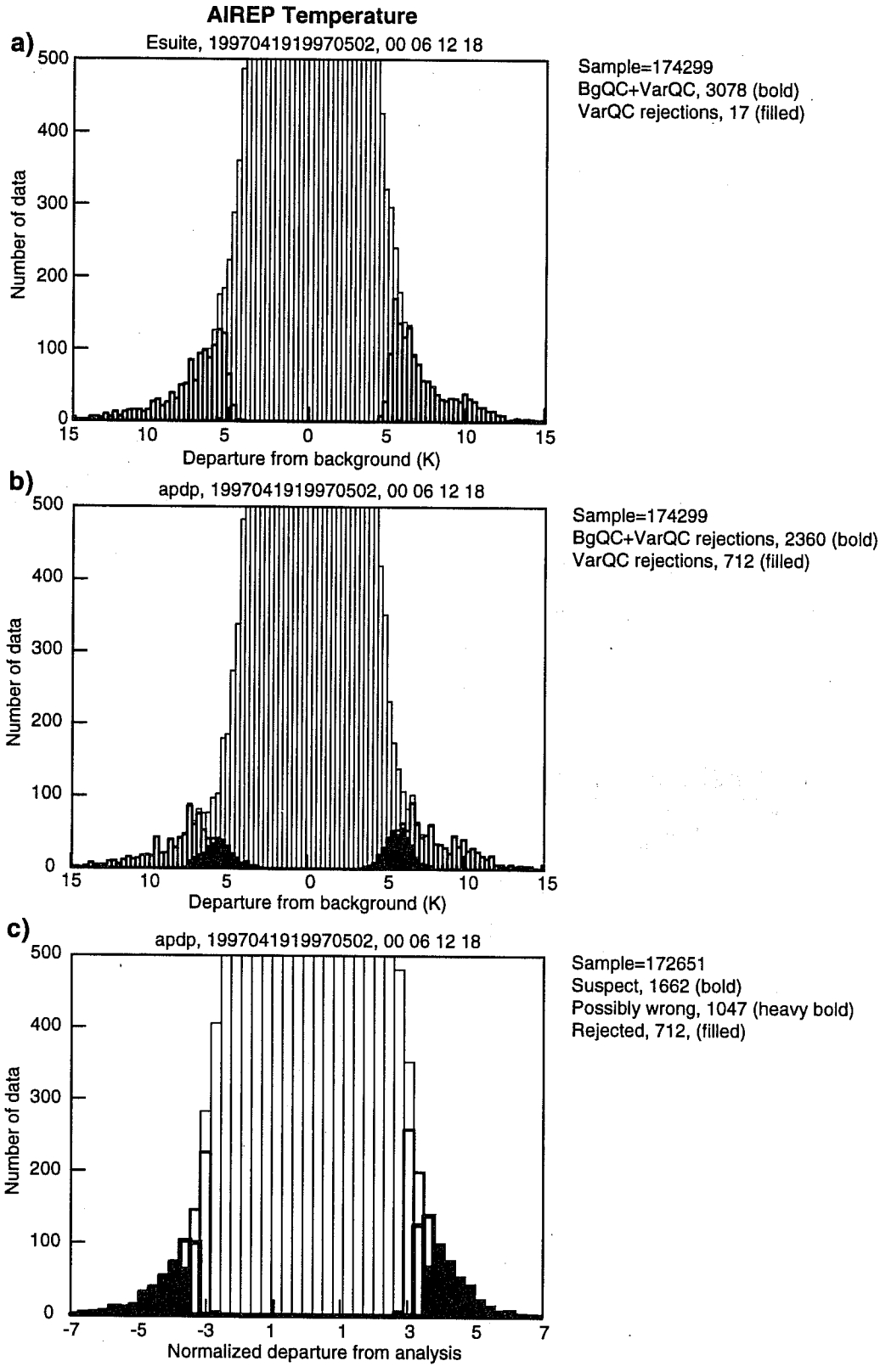


Fig. 5 Histograms of departures from the background, for AIREP temperature data, showing rejections before (a) and after (b) revision of VarQC. Non-rejected data (thin outline), data rejected by either BgQC (bold outline) and VarQC rejected data (filled) are shown. Panel (c) is a histogram of normalized departures from the analysis, showing the result of VarQC classified according to the a-posteriori probability of gross error: 0-25% (correct, thin outline), 25-50 % (suspect, bold outline), 50-75 % (possibly wrong, heavy bold) and 75-100 % (rejected, filled). The period is the same as in Fig. 4.

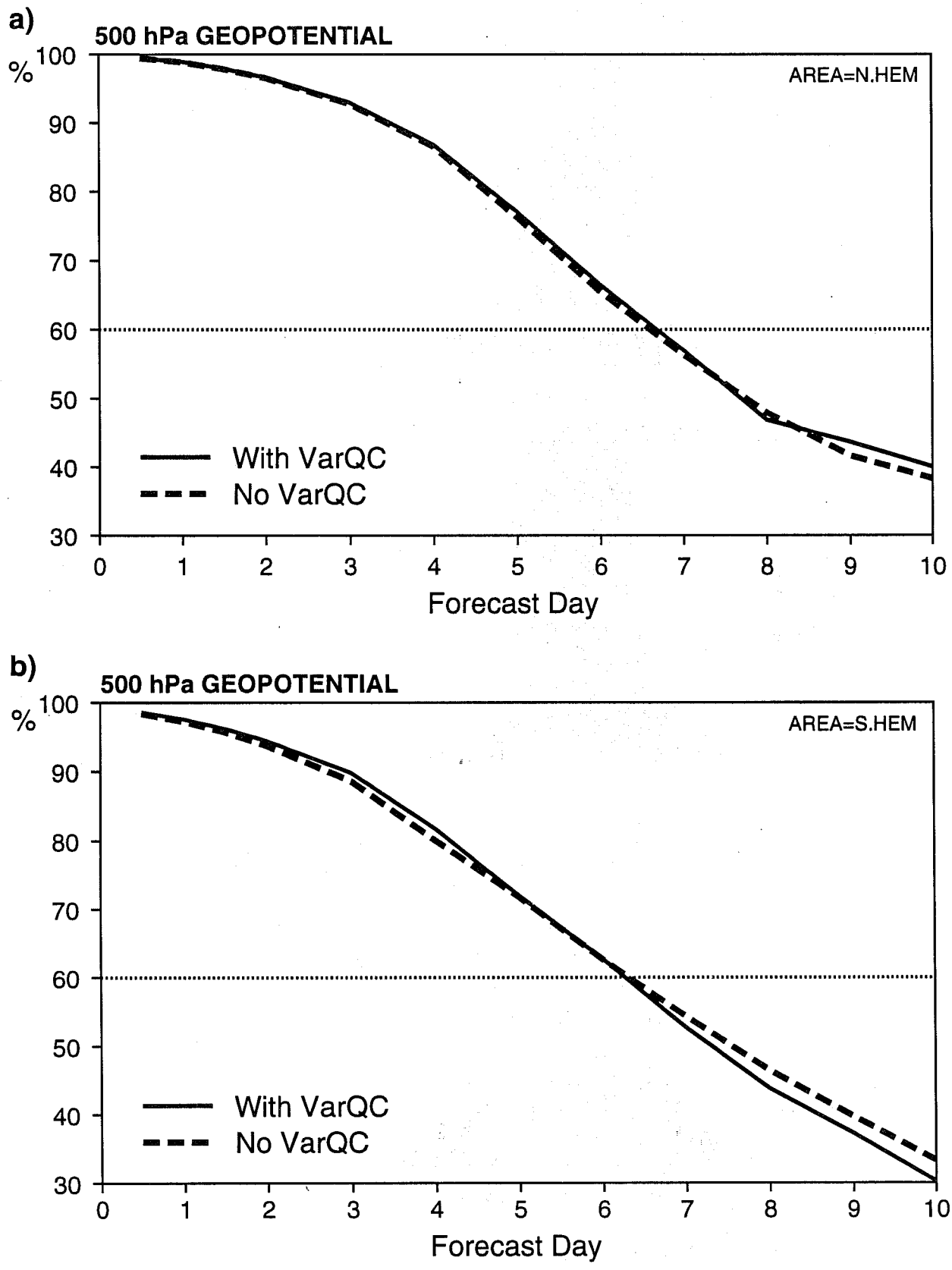


Fig. 6 Average 500 hPa geopotential anomaly correlation (%) forecast scores for two two-week data assimilation and forecast experiments with (full line) and without (dashed) VarQC, for the Northern (a) and Southern Hemispheres (b). The period is 19970419-12 to 19970502-12 UTC.

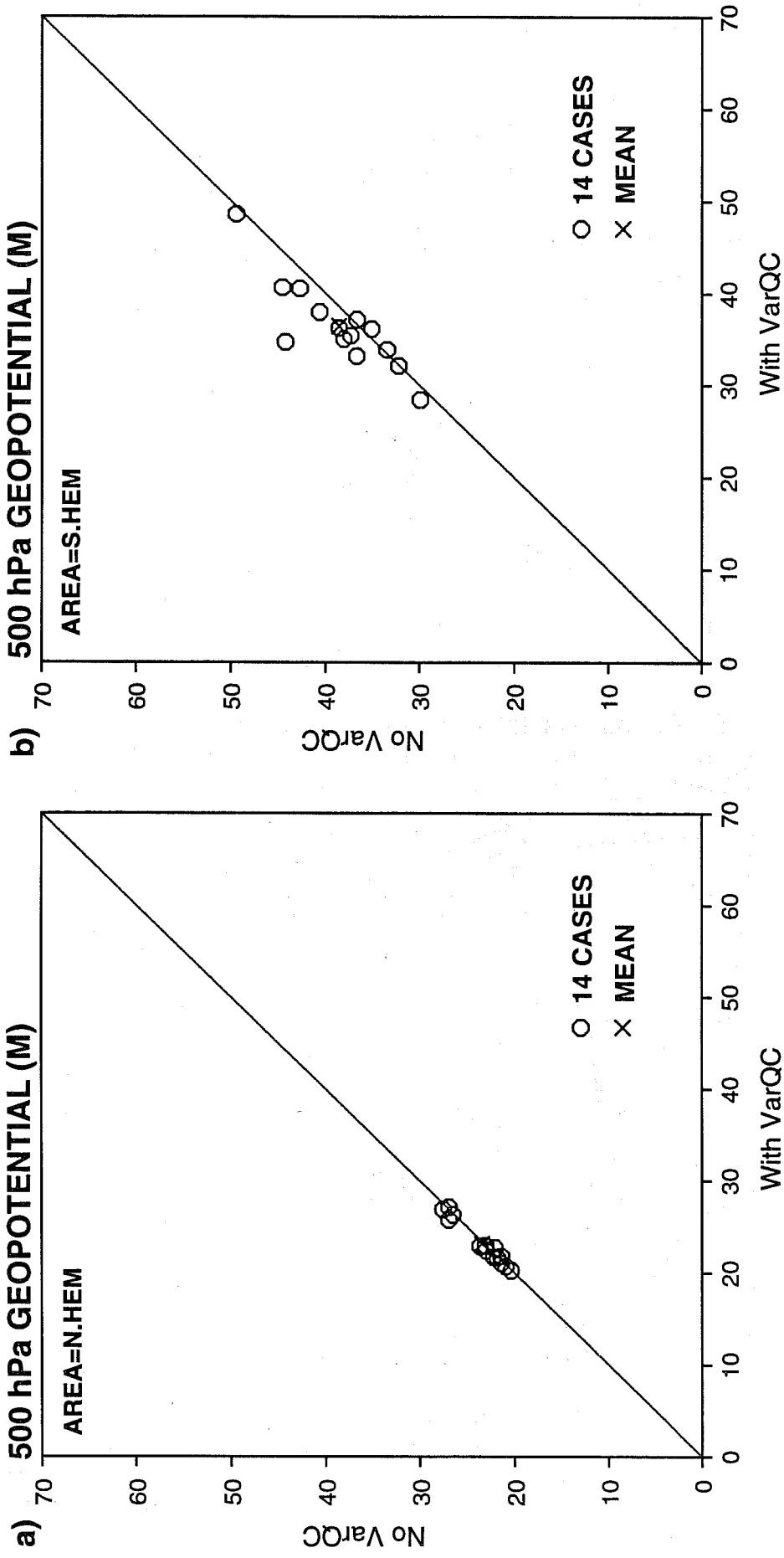
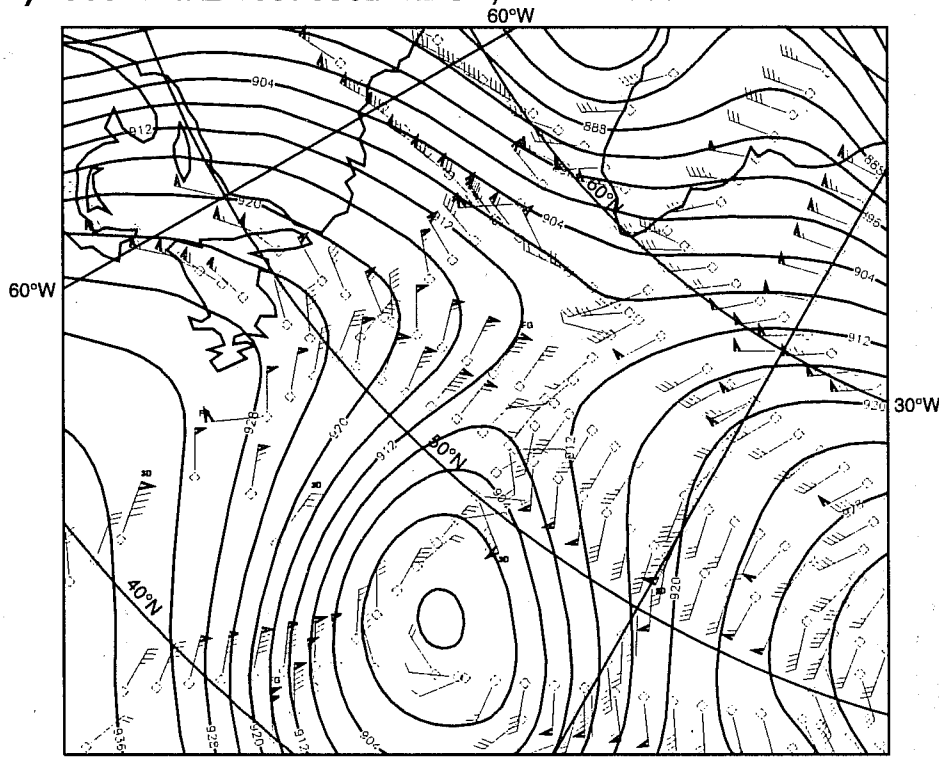


Fig. 7 Scatter diagram of the rms of 500 hPa geoptential 48-hour forecast error (m) from the same two-week experiments as in Fig. 6: The results with VarQC are plotted along the x-axis and without VarQC along the y-axis,for the Northern (a) and Southern (b) Hemispheres.

a) 300 hPa Z 19970502-12 UT, AIREP obs winds



b) AIREP rejections, temperature and wind departures.

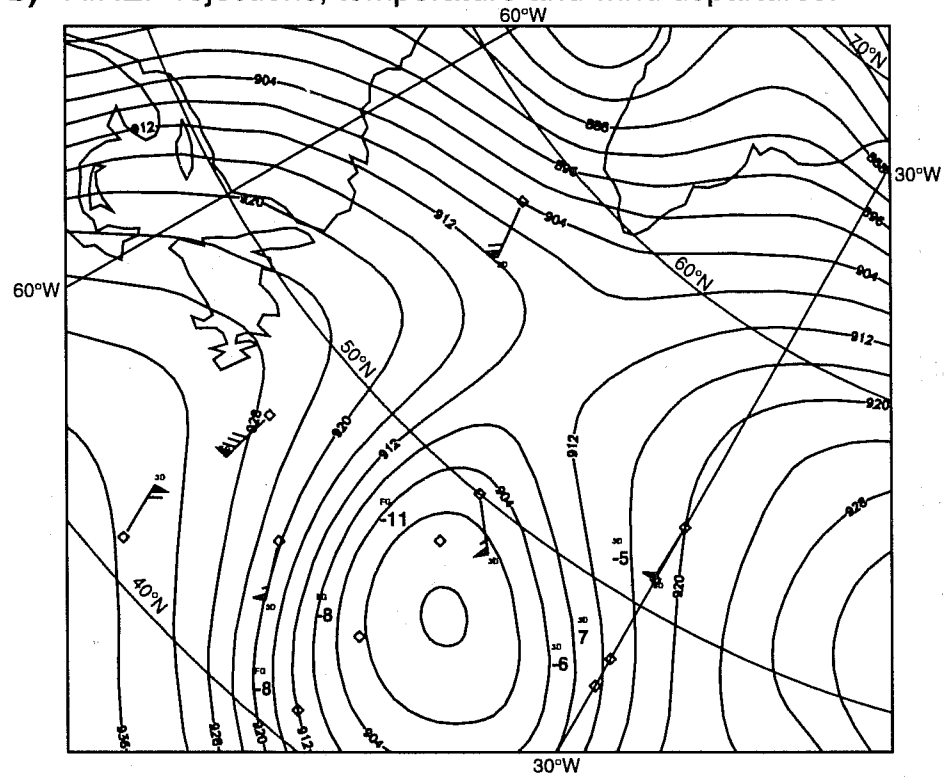
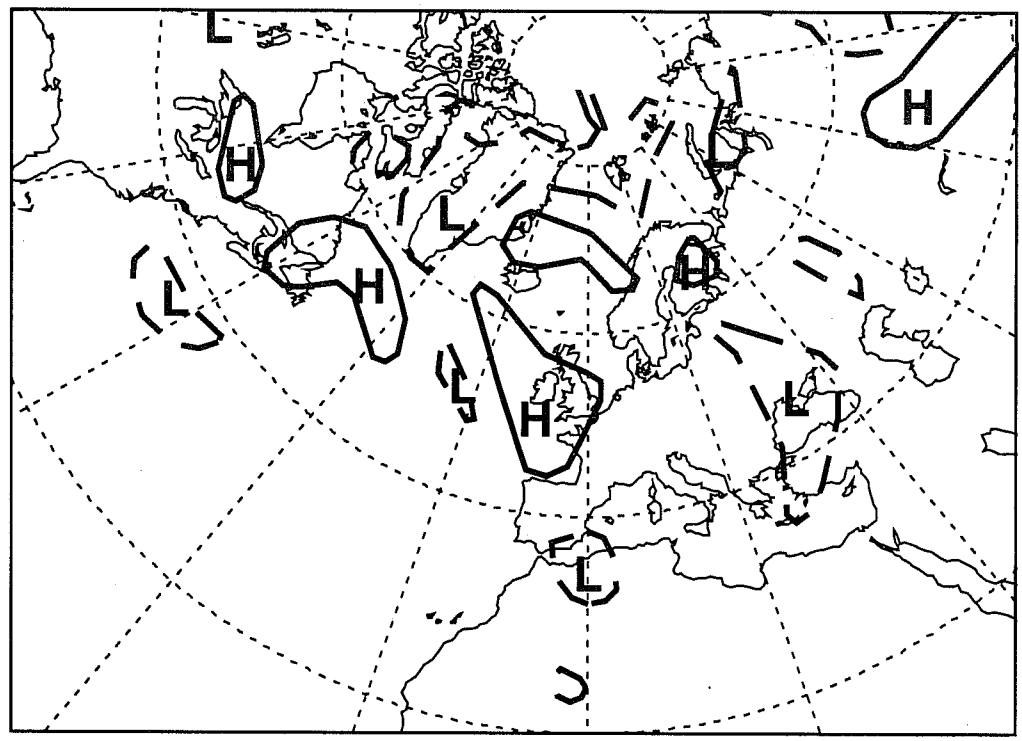


Fig. 8 (a) Analysed 300 hPa height field (decameters) and observed AIREP winds (wind flags in knots), 19970502-12 UT. Multiple AIREP winds at the same location are not plotted. (b) Like Fig. 8a for rejcted data. The data rejected by BgQC are marked FG and the wind arrow shows the departure from the background. The data rejected by VarQC are marked 3D and the wind arrow in that case shows the departures from the analysis. The plotted numbers show AIREP temperature rejections, similarly.

a) 97050212 Stream Function, Level 14



b) 97050212 Stream Function, Level 14

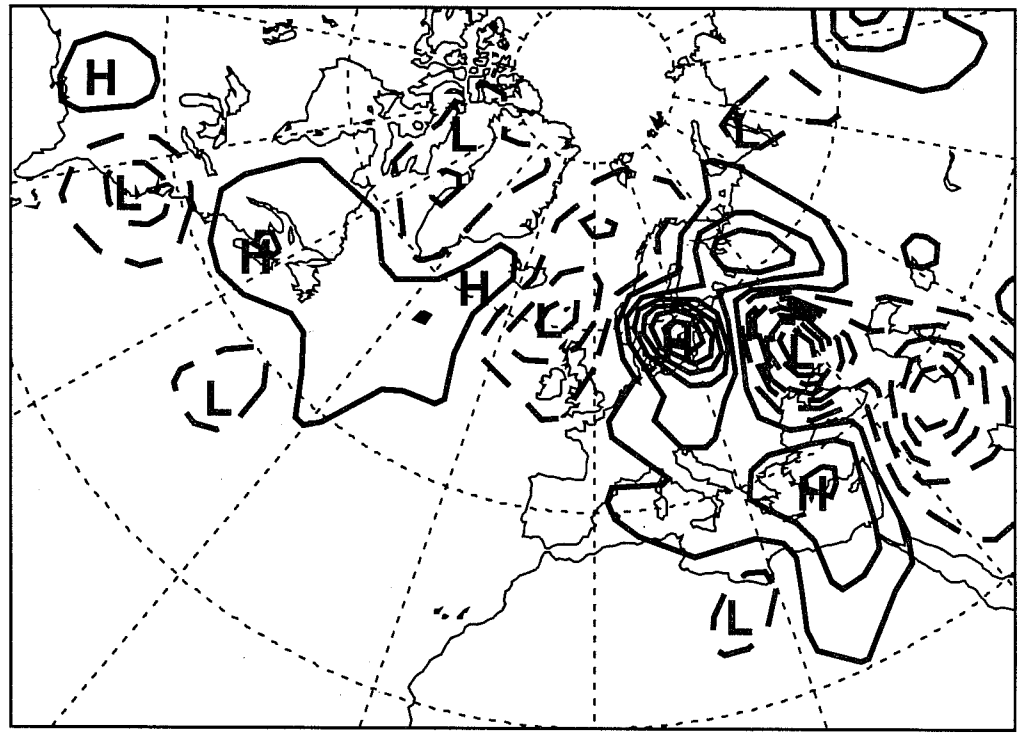


Fig. 9 The 48-hour forecast difference, 19970502-12 UT, between a forecast from a data assimilation experiment with VarQC and one without VarQC, in terms of streamfunction at model level 14 (270 hPa), projected onto the first 12 singular vectors, at initial time (a) and at +48 hours (b). As described by Buizza et al. (1997).