Humidity assimilation experiments with the Hirlam 3D-var

Loïk Berre[1] **and Ole Vignes**[2]

[1]*Météo France,* [2] *met.no*

1 Introduction

This paper summarizes two works that are related to the treatment of moisture in the background error term of the HIRLAM 3D-Var (a description of the HIRLAM 3D-Var can be found in Gustafsson et al. (2001) and Lindskog et al. (2001)). The first one deals with the study of background error cross-covariances involving humidity in J_b. The second work concerns the use of the logarithm of specific humidity instead of specific humidity itself. Another important work related to humidity in the HIRLAM 3D-Var is moreover the assimilation of GPS observations, which is described by Gustafsson (2002).

2 Moisture cross-covariances in Jb

2.1 Motivations

The specification of the background error covariance matrix B implies to determine not only the error auto-covariances of the mass, wind and humidity variables, but also cross-covariances between errors on these different variables.

Cross-covariances between mass and wind reflect atmospheric properties such as geostrophy and processes such as surface friction. Taking these cross-covariances into account allows to extract the maximum of information from the wind observations, in order to help correcting also the background field of mass, and vice-versa. In addition to this extraction of information from the observations, another motivation is the wish to obtain an analyzed state with as much physical consistence as possible between the different meteorological variables, knowing the involved atmospheric processes (and uncertainties on the involved processes should themselves be accounted for preferably).

Cross-covariances involving humidity are usually neglected by lack of knowledge. The use of observation operators involving e.g. some radiative transfer modelling or the atmospheric model itself (in 4D-Var) allows however to take into account some links between moisture, the other historical variables, and the observed variables; they help therefore to achieve a multivariate analysis of all variables. This is related to the fact that observation operators are implicitly combined with the matrix B, in order to model in particular cross-covariances between background errors on the observed variables and background errors on the historical variables. As the analysis results from this combination, the estimation of cross-covariances involving moisture in the matrix B is thus as wishable as the definition and use of some accurate multivariate observation operators.

2.2 Formalism based on linear regressions

A usual approach for the specification of the matrix B is to express this matrix as the product of several sparse matrices. In order moreover to precondition the minimization, this can be achieved by defining a sequence of operators that transforms the variable containing the background errors on the historical model variables into a variable whose covariance matrix is the identity matrix.

One of the first steps is to take into account cross-covariances between errors on wind, mass and moisture, in order to obtain some decorrelated variables. A natural way to obtain this is to define linear regressions between the different variables, in order to determine e.g. the part of mass and moisture errors that are decorrelated with wind errors.

If we express wind errors in terms of vorticity and divergence, this can lead to the following set of regressions (as in Parrish et al (1997) and Derber and Bouttier (1999), including the notations, but with an additional regression involving specific humidity errors q (Berre 2000) ; we will refer to this last equation as the "humid balance" in the rest of the text):

$$\begin{array}{ccl} P_b & = & \mathscr{H}\zeta \\ \eta & = & \mathscr{M}P_b + \eta_u \\ (T,P_s) & = & \mathscr{N}P_b + \mathscr{P}\eta_u + (T,P_s)_u \\ q & = & \mathscr{Q}P_b + \mathscr{R}\eta_u + \mathscr{S}(T,P_s)_u + q_u \end{array}$$

where $\mathscr{H}$ is called the horizontal balance operator, which is close to the geostrophic relationship ($\mathscr{H} = \beta f \Delta^{-1}$), and which provides the so-called balanced geopotential P_b. $\eta_u, (T,P_s)_u, q_u$ correspond to the residuals of the regressions. One may note also $q_b = \mathscr{Q}P_b + \mathscr{R}\eta_u + \mathscr{S}(T,P_s)_u$ as a notation for the "balanced" part of q. $\mathscr{M}, \mathscr{N}, \mathscr{P}, \mathscr{Q}, \mathscr{R}, \mathscr{S}$ are the matrices containing the linear regression coefficients. They are simply deduced from the cross-covariance matrix between the predictor and the predictand, and from the auto-covariance matrix of the predictor. For instance, the matrix $\mathscr{M}$ is given by:

$$\mathscr{M} = \overline{\eta P_b^T}(\overline{P_b P_b^T})^{-1}$$

An important property of these linear regressions is that the residuals are decorrelated with the predictors. This can be illustrated by expressing the covariance between η_u and P_b for instance:

$$\begin{array}{ccl} \overline{\eta_u P_b^T} & = & \overline{(\eta - \mathscr{M}P_b)P_b^T} \\ & = & \overline{\eta P_b^T} - \mathscr{M}\overline{P_b P_b^T} \\ & = & \overline{\eta P_b^T} - \overline{\eta P_b^T}(\overline{P_b P_b^T})^{-1}\overline{P_b P_b^T} \\ & = & 0 \end{array}$$

This illustrates the fact that transforming the vector $\varepsilon = (\zeta, \eta, TP_s, q)$ into the vector $\varepsilon' = (\zeta, \eta_u, (TP_s)_u, q_u)$ allows to obtain a variable ε' whose covariance matrix is block-diagonal (while the covariance matrix of ε is full). As previously mentioned, this is attractive for the preconditioning of the minimization, but it should be kept in mind that what matters for the analysis remains essentially the implied covariances for the variable ε itself (in other words, specifying a decorrelated control variable is relevant only if initially existing cross-covariances have been accounted for).

It may be of interest to mention that some analytical information can be incorporated into the set of regressions. As cross-covariances and auto-covariances are to be approximated, this can be useful indeed in order to define the best possible predictors and regressions for each variable.

This is the case for instance for $\mathscr{H}$: calculating a geostrophic geopotential $\phi_g = f\psi$ in physical space allows to take into account some of the latitudinal variations of the mass/wind coupling through the variations of the Coriolis parameter.

Conversely, using some statistical information in the form of a regression between TP_s and P_b allows to represent the variations of the mass/wind coupling as a function of height, and to relax geostrophy for the small horizontal scales. The challenge is indeed not only to determine what some physically-derived balance coefficients may be, but also how much the balance equation should be relaxed, knowing the involved uncertainties: this is the information that can be represented by covariances and linear regressions, in contrast with a purely

"analytical" approach. The degree of relaxation is determined not only by the variance of the residuals of the regressions, but also by the cross-correlations and variance ratios that are contained in the regression coefficient matrices.

The bi-Fourier approach that is used here for the statistics allows to study in an economical way the variations of the couplings involving humidity as a function of height and horizontal scale. One may also mention that the linear regressions are based here on some covariances which are estimated with the NMC method, and which are averaged in space and time (as the auto-covariances in many analysis schemes). This is of course restrictive, but it also means that any progress in the study and development of covariances (related e.g. to the estimation method, heterogeneities and flow-dependencies) may translate into similar achievements for the linear regressions.

2.3 Cross-covariances and explained variances

Examining cross-covariances is interesting as these are the basic quantities involved both in the matrix B and in the linear regression coefficients, and as it reveals the underlying multivariate couplings that are represented.

The decorrelation between the predictors and the residuals of the regressions allows moreover to obtain a decomposition of the q auto- and cross-covariance matrices as the sum of a few contributions. For instance:

$$\overline{qq^T} = \mathscr{Q}\overline{P_b P_b^T}\mathscr{Q}^T + \mathscr{R}\overline{\eta_u \eta_u^T}\mathscr{R}^T + \mathscr{S}\overline{(T,p_s)_u (T,p_s)_u^T}\mathscr{S}^T + \overline{q_u q_u^T}$$

and

$$\overline{q(T,p_s)^T} = \mathscr{Q}\overline{P_b P_b^T}\mathscr{N}^T + \mathscr{R}\overline{\eta_u \eta_u^T}\mathscr{P}^T + \mathscr{S}\overline{(T,p_s)_u (T,p_s)_u^T}$$

This allows in particular to calculate the percentages of the humidity error variance that are explained by the different predictors, which gives an information on the strength of the corresponding couplings.

We will present the corresponding results that were obtained with the SMHI version of the HIRLAM model (whose geographical domain is shown in Fig. 1), and compare them with what was obtained on a much smaller domain with the ALADIN model in France (Berre 2000).

The largest cross-covariance values between P_b and q are negative, and they concern the low levels (Fig. 2, panel (a)): they correspond to a link between e.g. the deepening of a low and some moistening due to low-level convergence (which is itself induced by surface friction). As expected, the percentages of explained variance indicate that this coupling is more important in the large scales (with a maximum of 27 %) than in the small scales (Fig. 3, top left).

The structure of the covariance between q and η_u (Fig. 2, panel (b)) corresponds to a coupling between e.g. a local moistening, some low-level convergence and high-level divergence. This coupling tends to be more important in the small scales (with a maximum of 12 % (Fig. 3, top right)).

The most complex error cross-covariances are between humidity and (full) temperature (Fig. 2, panel (d)). Positive values tend to predominate, which corresponds probably to the increase of saturation vapour with temperature. They result mostly from the contribution of the large-scale coupling with the balanced geopotential P_b. Values along the diagonal correspond to the "local" covariances (i.e. between q and T at the same level), and they appear to be smaller than some of the off-diagonal covariances. This is due in particular to the contribution of the negative covariances between q and T_u (Fig. 2, panel (c)), which correspond to diabatic processes such as evaporation and condensation. These diabatic couplings become more important in the small scales (with a maximum of 22 % (Fig. 3, bottom left)).

Many features of the cross-covariance structures and of the vertical and spectral variations of the explained

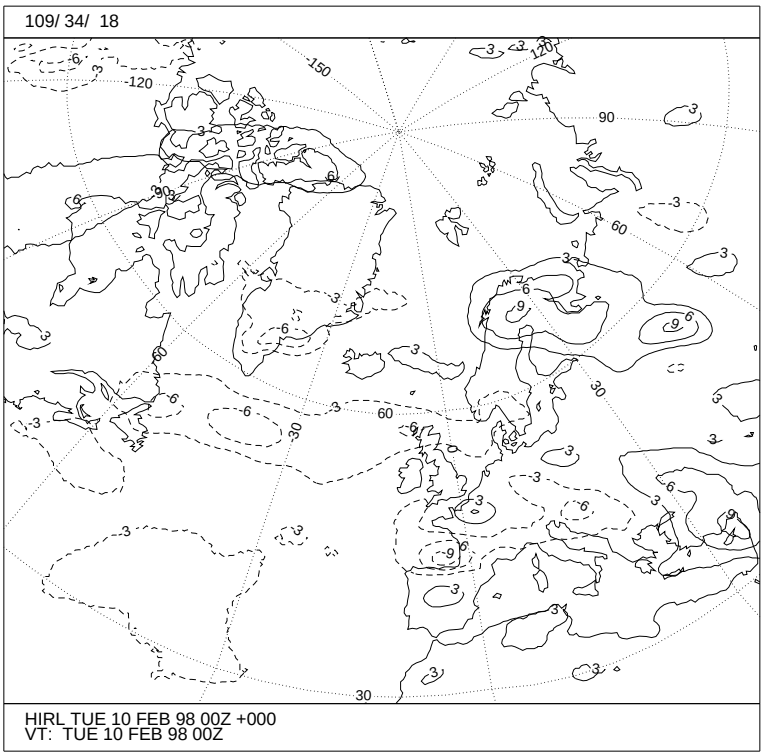

Figure 1: Geographical domain of the SMHI version of the HIRLAM model.

variances appear to be consistent with what was obtained with ALADIN-France, despite the large differences between the HIRLAM and ALADIN models. One may mention that the absolute values of the covariances appear to be larger for HIRLAM, which corresponds to the larger size of the HIRLAM domain: this translates into a larger contribution of the large scales, whose variance tends to be larger than the small scales ; it also explains the larger importance (in HIRLAM) of large-scale couplings (e.g. between q and P_b).

2.4 Assimilation experiments

Some parallel assimilation and forecast experiments were conducted with the HIRLAM 3D-Var during one month, in order to evaluate the impact of the humid balance on the forecast quality.

In order to evaluate the specific impact of the humid balance, one would like to set cross-covariances involving q equal to zero, while keeping the same auto-covariances in the two compared configurations. The "multivariate" and "univariate" configurations will be called "MULTI" and "UNI" in the rest of the text.

Setting $\mathscr{Q},\mathscr{R},\mathscr{S}$ equal to zero allows to zero the corresponding cross-covariances, but it also implies that the q variance is reduced (it becomes equal to $var(q_u)$). Having similar q variances in the two configurations is not straightforward, for instance because there is an implicit latitudinal dependence in $var(q_{b,Pb} = \mathscr{Q}P_b)$. The average estimate of $var(q)$ provided by the NMC method was used in UNI, but this is still imperfect: q variances are larger in UNI than in MULTI, because $var(q)$ tends to be larger than $var(q_b) + var(q_u)$. This may explain the larger analysis fit to humidity observations in UNI than in MULTI, and also the slightly better 06 and 24 hour humidity forecasts in UNI than in MULTI. Some additional sensitivity experiments would be therefore interesting with respect to background and observation error standard-deviations.

The impact on the mass and wind analysis is less ambiguous to interpret, as the mass and wind auto-covariances are exactly the same in UNI and MULTI: the only differences concern cross-covariances between $T,\vec{u}$ and q. The impact of the humid balance on the temperature and wind forecasts appears to be neutral to positive. The

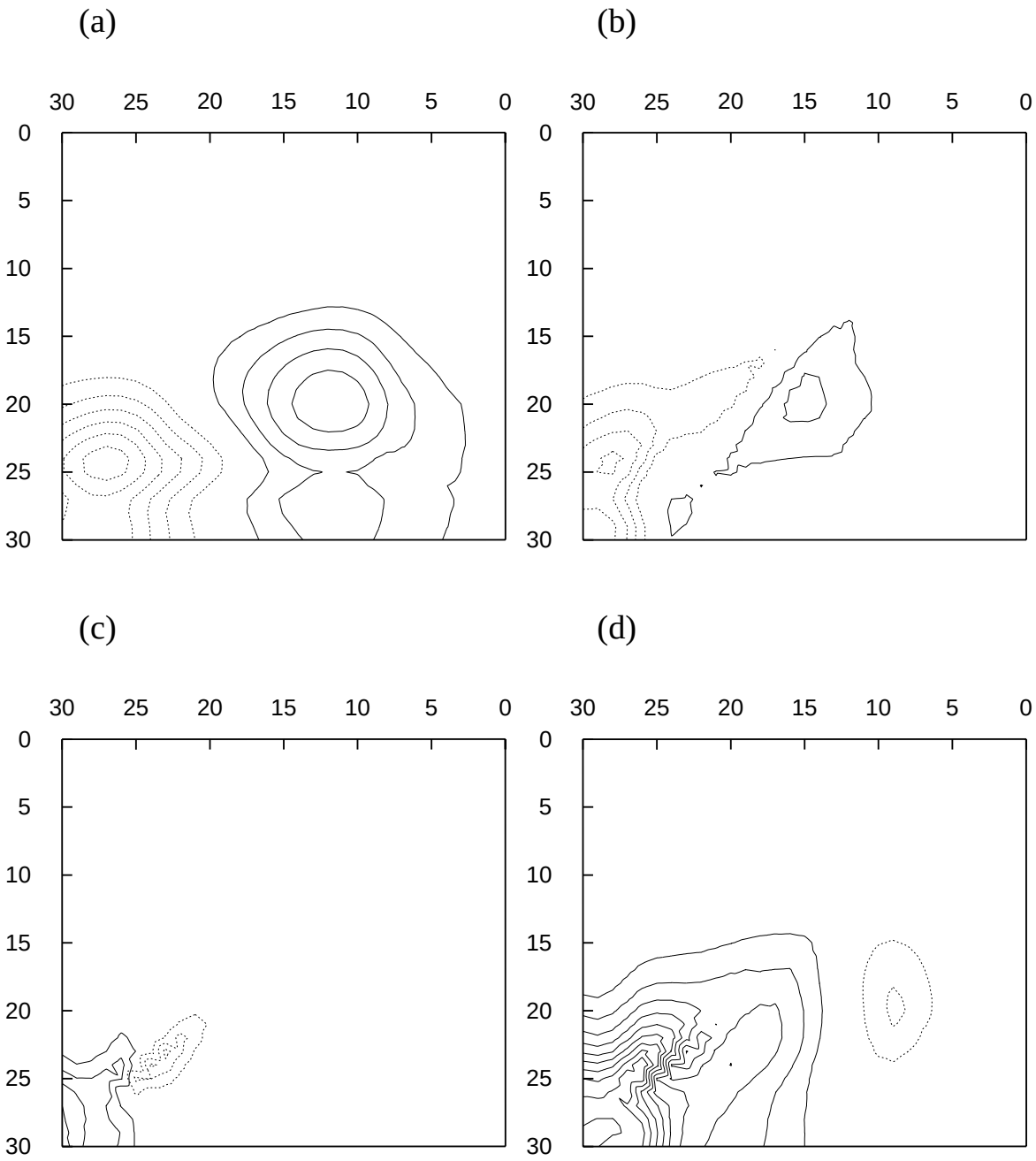

Figure 2: Average vertical cross-covariance matrices between specific humidity and balanced geopotential (panel (a) ; isoline spacing = 5.10^{-3} kg kg^{-1}J kg^{-1}), unbalanced divergence (panel (b) ; isoline spacing = 4.10^{-10} kg kg^{-1} s^{-1}), unbalanced temperature (panel (c) ; isoline spacing = 4.10^{-5} kg kg^{-1}K) and full temperature (panel (d) ; isoline spacing = 4.10^{-5} kg kg^{-1}K). Positive (resp. negative) values are indicated by full (resp. dashed) isolines, and the zero contour has been omitted. The ordinate axis corresponds to the humidity vertical levels, and the abscissa axis corresponds to the vertical levels of the other involved variable (predictor or full temperature). The index 0 corresponds to the highest level, and the index 30 corresponds to the level that is closest to the surface.

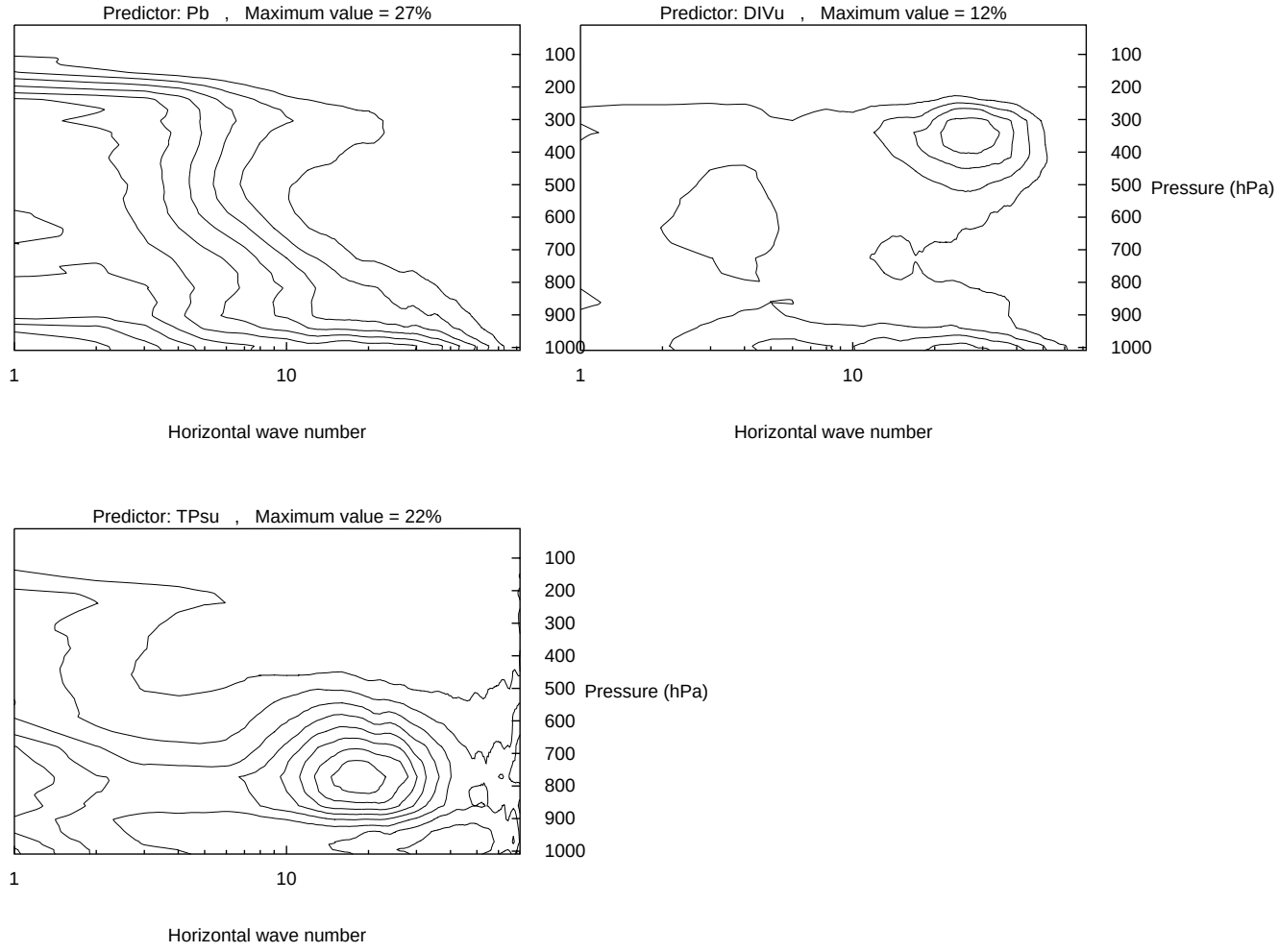

Figure 3: Percentage of the variance of specific humidity that is explained by balanced geopotential (top left), unbalanced divergence (top right) and unbalanced temperature (bottom left), as a function of pressure (unit:hPa) and horizontal wave number. The isoline spacing is 2.5 % and the zero contour has been omitted. The maximum value is indicated at the top of each panel.

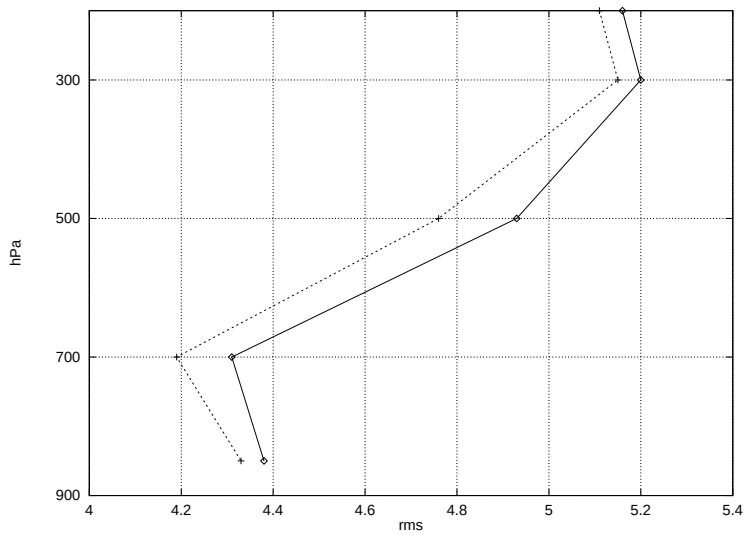

Figure 4: RMS vertical profiles of 6H forecasts for wind (in m/s). The full line corresponds to the configuration UNI (i.e. without the humid balance), and the dashed line corresponds to the configuration MULTI (i.e. with the humid balance). Observations were used for the verification.

positive impact can be illustrated by the observation verification scores of the six-hour wind forecasts (Fig. 4). This indicates that some useful information has been extracted from the humidity observations to influence the mass and wind analysis.

3 Experiments with the logarithm of q

In variational assimilation, the analysis estimate has maximum probability of being "true" if errors in analysed variables and observations have a gaussian probability distribution. In HIRLAM there has been a feeling that specific humidity errors are not gaussian, and that the analysis instead should be performed on e.g. the logarithm of specific humidity or on relative humidity.

The only alternative variable that has been tried so far is $\ln(q)$, mostly because it is the easiest one to implement. HIRLAM's variational assimilation is based on an incremental formulation and a diagonalizing transformation on the background error term:

$$J(\delta x) = \frac{1}{2}\delta x^{\mathrm{T}} B^{-1}\delta x + \frac{1}{2}[d - H\delta x]^{\mathrm{T}} R^{-1}[d - H\delta x],$$

where $\delta x = x - x_b$ and $d = y - H(x_b)$. The control variable is $\chi = U\delta x$, where $B^{-1} = U^{\mathrm{T}}U$. The transformation U consists of a series of simpler transformations, and $\ln(q)$ is one of these. Since it is U^{-1} that is actually needed in the minimization, the inverse humidity transformation becomes, in its tangentlinear version:

$$\delta q = q_b\, \delta \ln q.$$

However, a linear transformation in model (or observation) space alone will not change the result of the minimization. Therefore new background error statistics have been generated by the NMC method, and in this procedure differences in $\ln(q)$ (of +48h and +24h forecasts valid at the same time) have been used. In fact, 4 different sets of statistics have been generated; for the analytical balance as well as the statistical balance described above, and with and without the $\ln(q)$ transformation on the data. The data used were 99 pairs of forecasts from the norwegian 0.5° HIRLAM model, spanning almost one year (every 3.5 days, i.e., a mixture of 00utc and 12utc forecasts).

The HIRLAM model was run on the norwegian 0.5° area for a 30 days trial period (Jan 2 – Jan 31, 2002), once for each of the 4 sets of statistics generated. The results showed little impact of the transformation on forecast verification scores. An example can be seen in Fig. 5, which shows bias and RMS scores for relative humidity (averaged over all EWGLAM stations in the model area) at +0 and +24 hours. The figures are for statistically balanced errors, but the figures for the analytical balance look virtually the same.

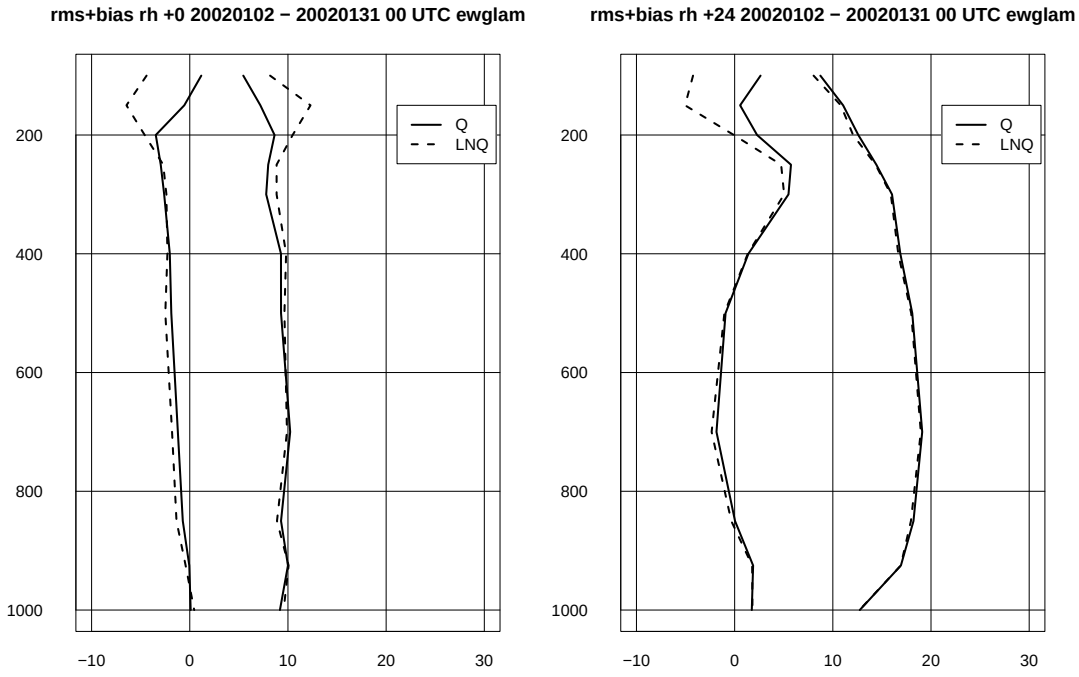

Figure 5: Bias and RMS errors for relative humidity with and without $\ln q$ *transformation.*

We find the largest deviations at initial time, i.e., in the fit to the observations. Background errors based on q seem to give a closer fit to the observations in areas where the specific humidity is small, i.e., in the upper layers. This must mean that bakground errors based on q are somehow larger than those based on $\ln q$. The reason for this is not yet clear. Humidity error variances at the various levels are shown in Fig. 6. We see that for the q-based statistics errors decrease much more with height than they do for the $\ln q$-based statistics.

The better balanced vertical humidity error distribution for $\ln q$ at least seems to have a positive impact on the conditioning of the minimization problem. Over the test period (118 analyses) the average number of iterations for convergence was 20% lower for the $\ln q$ cases, both for analytically and statistically balanced errors.

For other variables the differences in forecast scores are small, at analysis time as well as longer forecast times. It should be noted however that the only humidity data assimilated in these tests were radiosonde humidities.

Finally we mention that we have also briefly tried to use the $\ln q$ transformation on the observations, i.e., to assimilate $\ln q$ instead of q as is presently done for radiosondes. With hindsight it would have been more natural to assimilate relative humidity, even if this value is not present in the TEMP codes, since this is what most radiosondes actually measure. In theory, if observation errors are largely caused by random measurement errors, relative humidity errors should have approximately gaussian distribution. If there are also representativeness errors, which is not unlikely for humidity measurements, the situation is less clear.

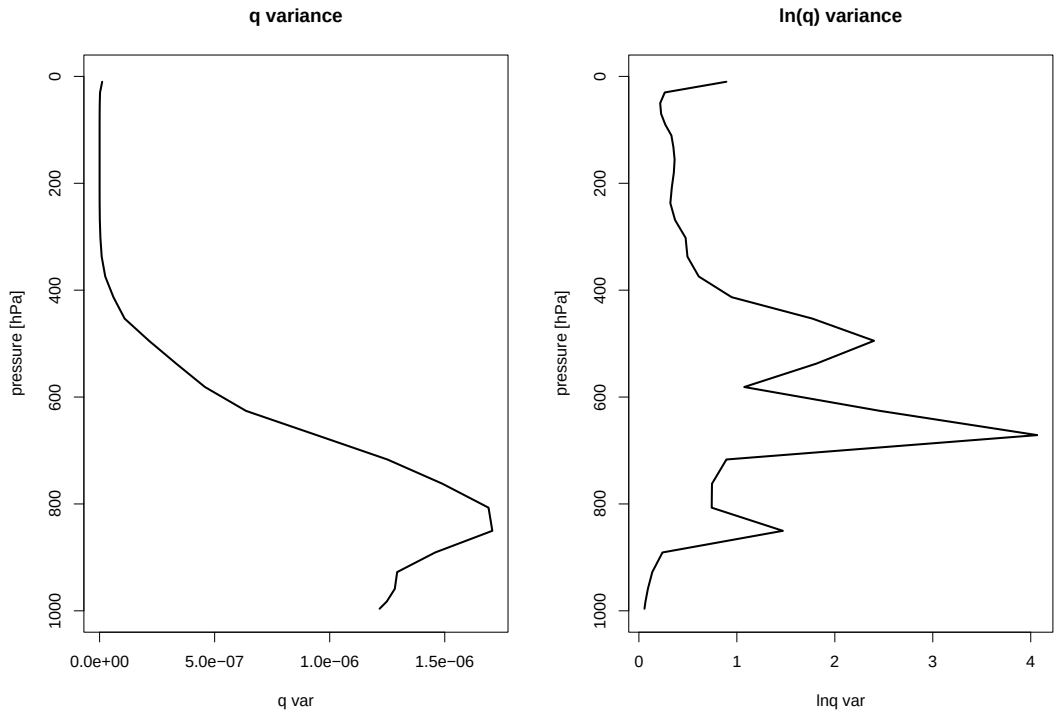

Figure 6: Vertical distribution of humidity error variances, q/$\ln q$.

4 Conclusions

Two efforts on the handling of humidity in the J_b of the HIRLAM 3D-Var have been here described. Representing cross-covariances involving humidity in J_b was found to be feasible and to have a neutral to positive impact on temperature and wind forecasts. The use of $\ln q$ instead of q had a positive impact on the conditioning of the minimization and a neutral impact on the forecast scores.

Further work could deal with the use of an ensemble approach to calculate the statistics, the study and representation of space and time dependences, and the test of other humidity variables.

5 References

Berre, L., 2000: Estimation of synoptic and meso scale forecast error covariances in a limited area model. *Mon. Wea. Rev.*, **128**, 644-667.

Derber, J. and Bouttier, F., 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A**, 195-221.

Gustafsson, N., Berre, L.,Hörnquist, S., Huang, X.-Y., Lindskog, M., Navascués, B., Mogensen, K.S., and Thorsteinsson, S., 2001: Three-dimensional variational data assimilation for a limited area model. Part I: General formulation and the background error constraint. *Tellus*, **53A**, 425-446.

Gustafsson, N., 2002: Assimilation of ground-based GPS data in HIRLAM 3D-Var. Proceedings of the HIRLAM workshop on variational data assimilation and remote sensing, FMI, Helsinki, 21-23 January 2002, 89-96.

Lindskog, M., Gustafsson, N., Navascués, B., Mogensen, K.S., Huang, X.-Y., Yang, X., Andræ, U., Berre, L., Thorsteinsson, S. and Rantakokko, J., 2001: Three-dimensional variational data assimilation for a limited area model. Part II: Observation handling and assimilation experiments. *Tellus*, **53A**, 447-468.

Parrish, D.F., Derber, J.C., Purser, R.J., Wu, W.-S., Pu, Z.-X., 1997: The NCEP global analysis system: recent improvements and future plans. *J. of Met. Soc. of Japan*, **75**, No. 1B, 359-365.

6 Acknowledgements

We would like to thank all people involved in the HIRLAM 3D-Var.