Met Office

Foundation & Weather Science

Improved variational analyses using a nonlinear humidity control variable: Formulation and trials

Technical Report No. 558

Bruce Ingleby, Andrew Lorenc, Keith Ngan, Rick Rawlins and David Jackson

email: nwp_publications@metoffice.gov.uk

Table of contents

Abstract

A nonlinear transformation of the humidity control variable has been successfully introduced into the Met Office's variational data assimilation system. The transform is now operational in both global and limited area systems. The forecast improvement is largest in the southern hemisphere and there is a better fit of both background and analysis to humidity sensitive satellite channels. These results suggest that the transformation is particularly beneficial for the use of satellite data. The transform reduces the problem of negative humidities in the analysis which were more prevalent over the ocean.
The transformation has two parts: a nonlinear normalisation and a link with temperature increments that is a function of background humidity. The normalisation (which seems to be the more important aspect) makes the new variable's errors more symmetrical. Thus the error distributions can be better represented by covariances in the variational cost function. It is combined with the use in the assimilation of a total water variable which combines water vapour and cloud.

1. Introduction

For users of short range weather forecasts atmospheric humidity is important particularly when it forms cloud, fog or precipitation. Also latent heat release is a major driver of the atmospheric circulation. Despite widespread effort over many years, and clear impacts on short-range "mesoscale" forecasts (Anderson et al., 2000) there have been few demonstrations of a large positive impact from humidity assimilation on synoptic-scale forecasts. The traditional view is that in most cases humidity adjusts to the dynamical fields (Lorenc and Tibaldi 1980, Bengtsson et al. 2004). However Andersson et al. (2005, 2007) demonstrated that with recent improvements in observations, modelling and assimilation (including the humidity transform of Hólm et al, 2002) humidity observations can have a significant impact. That work however did not separate how much of the improvement was due to the assimilation changes, as we do in section 3 below.

Saturation and precipitation prevent large values of relative humidity, while negative values are unphysical, so the distribution of humidity values in the atmosphere is bounded. The probability density functions (PDFs) are broad and locally can be skewed and highly non-Gaussian (Sherwood et al, 2005). The distributions of forecast errors which we are trying to correct in data assimilation inherit many of these awkward properties. The computationally efficient data assimilation methods which are necessary for large atmospheric models require the errors to have Gaussian distributions (Lorenc 2003), usually also assumed to be unbiased. Because of its short-scale variability, the errors in humidity in the forecasts used as background in the data assimilation can be very large, for instance cloud can be in the wrong place. The coupling of humidity with other atmospheric variables (especially temperature) depends on the closeness to saturation. Using ideas from Hólm et al (2002), we introduce a transformed humidity variable whose errors are more symmetric. The different coupling with temperature near saturation is also allowed for, while preserving the Met Office system's ability to analyse cloud and represent in a simple way the links between cloud and humidity.

This paper describes the design (section 2) and testing (section 3) of a method to address, as far as possible, the above problems, in the Met Office's operational four-dimensional variational data assimilation (4D-Var) system. There is further discussion in section 4 and a summary in section 5. First there is a brief review of humidity observations and previous work on humidity data assimilation

1.1 Humidity observations and their use at the Met Office

Radiosondes provide in situ humidity measurements with high vertical resolution but poor horizontal and temporal resolution. In general the measurements are more closely related to relative humidity (RH) than to specific humidity. At low temperatures the slower response of the sensors limits their performance, but the best radiosonde humidities are now usable in the upper troposphere. Aircraft humidity data are not currently used at the Met Office. Humidity from surface stations has been assimilated globally at the Met Office since April 2008 (Ingleby, 2012, in preparation).

Microwave and infrared 'soundings' from various satellites have good horizontal coverage but also very deep weighting functions. Traditional soundings only contain two or three independent pieces of humidity information; hyper-spectral infrared soundings have more. Recently, changes have been made to assimilate infrared channels that peak above the diagnosed cloud (Pavelin et al, 2008) and to use clear

soundings from Meteosat. GPSRO – Global Positioning System Radio Occultation - (Buontempo et al, 2008) profiles are sparser than satellite soundings but they are less sensitive to calibration issues. GPSRO provides information on vertical refractivity gradients – these are related to temperature gradients in the stratosphere and upper troposphere and mainly to humidity gradients in the lower troposphere with a transition in between. Ground based GPS techniques give vertically integrated water content, these data were not assimilated in the global trials reported here, but were introduced later in 2011. The vertical distribution of humidity analysis increments produced from satellite data is very dependent on the background error model used. If we count every pixel then satellite imagery observations are more numerous than all others, and provide information on clouds. Despite a human's ability to forecast weather from cloud imagery (Bader et al, 1995), the use of these observations in numerical weather prediction (NWP) has been limited mainly to derived atmospheric motion vectors, and to attempts to improve short-period cloud forecasts (e.g. Macpherson et al, 1996).

Thornton et al (2009) describe upper atmosphere measurements from limb-sounding and other satellites and attempts to assimilate these data. They found that an experimental version of the Met Office stratospheric humidity analysis system performed relatively poorly, due to problems in modelling the vertical error covariances. The focus of the current paper is on tropospheric performance, but the work should also benefit stratospheric performance in the future.

1.2 Previous work on humidity assimilation

An early attempt to include synoptic dependence in a humidity analysis was reported by Atkins (1974) and Riishøjgaard (1998) demonstrated a related method for ozone concentrations. They preferentially spread increments along fronts, or along contours of constant relative humidity (or ozone) – whilst interesting, this method is empirically based, and would be difficult to implement in our variational system. Lorenc et al (1996) showed that the preservation of RH in the absence of humidity observations can be important in certain meteorologically important situations. Dee and Da Silva (2003) compared several different humidity variables and chose to use pseudo-relative humidity (with scaling provided by the background saturation humidity) and showed that this was equivalent to a mixing ratio analysis with flow-dependent variance specifications. They found that correlations between specific humidity, q, and temperature were smaller in magnitude than those between RH and temperature (an argument for analysing q rather than RH), but noted that a multivariate analysis taking humidity-temperature error correlations into account would be better. Hólm et al (2002) introduced a non-linear transform at ECMWF (European Centre for Medium-Range Weather Forecasts) – see next section. Gustafsson et al (2011) tried a similar transform in a limited area model but found little impact over Europe (also seen in our system, section 3 below). Berre (2000) calculated covariances of forecast differences for a limited area model in northern mid-latitudes and found quite complex relationships between humidity and other variables. Montmerle and Berre (2010) have produced separate background error statistics for precipitating and non-precipitating points which are significantly different in some respects.

Negative humidity values from undershoots are an unfortunate feature of most humidity analysis systems and are generally removed in a final initialisation stage. (The resetting to zero represents a source of humidity in the system, but data assimilation doesn't conserve overall humidity anyway.) Figure 1 shows how they can arise in the vicinity of sharp humidity gradients. There is a fundamental problem in that the climatological background error covariances are inappropriate in the vicinity of such gradients. It is possible to avoid negative values by analysing the logarithm of specific humidity, however this effectively gives very large weights to very small values of background or observed humidity (Dee and Da Silva, 2003). Lorenc (2007) used differences between radiosonde and background values to examine background error covariances in the vicinity of cloud layers. He found that the neglect of correlations between RH and temperature errors is not usually justified and that complex, non-linear relationships – rather than simple covariances - are necessary to provide a good description. He also suggested that background error variances are approximately doubled near cloud-topped inversions.

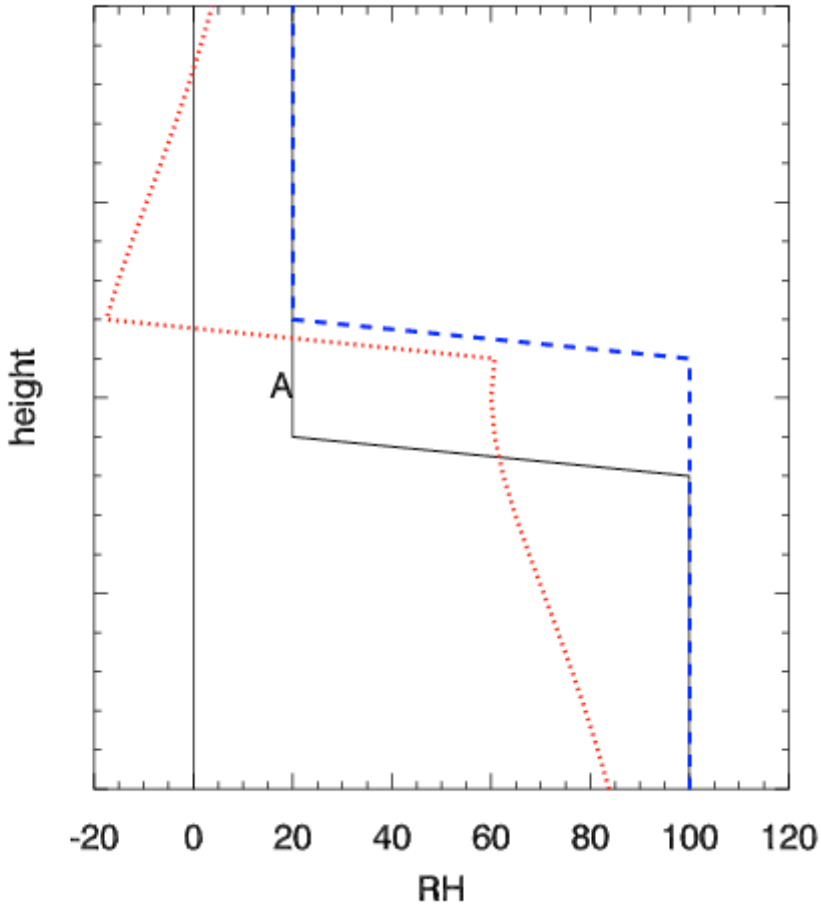

Figure 1. Schematic illustration of negative humidity analysis using typical error covariances in the presence of a sharp gradient (such as at the top of the boundary layer, although it could also apply to a horizontal gradient). Solid line is the true humidity profile, dashed blue line is the background and dotted red line the analysis using a single observation at A. If there was a profile of observations (from a radiosonde) then the negative values would be much reduced.

2. Method

2.1 Met Office humidity analyses – previous system

Forecasts are produced by the Met Office Unified Model - Milton and Earnshaw (2007) provide a description with emphasis on the water cycle. Cloud water and cloud ice are independent variables in the forecast model. A prognostic cloud fraction and prognostic condensate scheme (PC2, Wilson et al, 2008, Morcrette et al, 2011) was implemented in 2010.

The original 3D-Var system (Lorenc et al. 2000) ignored cloud and used RH as the control variable. Since then, we have moved to 4D-Var (Rawlins et al. 2007) and the system has been extended to use total relative humidity. We define total water as

$$q_T = q + q_c \tag{1}$$

where q is specific humidity (vapour) and q_c is cloud water (both liquid and solid). Statistically q and q_T are very similar as vapour makes up more than 99% of water in the atmosphere, however q_T can exceed q_s the saturation specific humidity. The total relative humidity is

$$rh_T = \frac{q_T}{q_s(T)} = q_T \left(\frac{p}{\varepsilon e_s(T)} \right) \quad (2)$$

T is temperature and *p* pressure, ε is the ratio of molecular weights of water and dry air and $e_s(T)$ is the saturated vapour pressure of water (solid below 0°C and liquid above). There are two minor approximations in (2), the "enhancement factor" (e.g. Buck, 1981) is taken to be 1 and $p+e_s(T)$ is replaced by *p* in the numerator.

The "simplifying operator", S, which converts from model to analysis variables simply adds vapour and cloud water to define total water, and its generalised inverse, the "incrementing operator" partitions analysed increments of total humidity into increments for each model variable (Lorenc 2003, section 4(c)). The nonlinear incrementing operator is based on diagnostic relationships such as that of Smith (1990), regularised to be smooth and differentiable (Sharpe, 2007). The partition of each humidity increment depends of the presence of cloud and the closeness to saturation of the model state being incremented. The linearised operator (S^{-I}) defines a situation dependent covariance for the full set of model humidity variables:

$$B(x) = S^{-I} B(w) S^{-T} \quad (3)$$

where $B(w)$ describes covariances for the total humidity. This approach allows the assimilation of satellite microwave radiances sensitive to cloud liquid water (Deblonde and English, 2003).

The climatological background error covariances form an important part of a variational analysis system and they are built up using a series of transforms and statistical information from a training data set (Lorenc et al 2000, Ingleby, 2001). Fisher (2003) introduced the use of ensembles of analyses in the estimation of background error covariances. Covariances based on the training data shown in this paper have been used in the operational system since November 2010. Analyses and analysis ensemble perturbations (at T399 resolution) were obtained from ECMWF for 1-17 October 2006, there are 10 members in the ensemble available at 00 and 12 UTC (330 samples in total). The control and perturbed analyses were reconfigured onto our 70 level grid, with cloud water/ice set to zero. From the perturbed analyses 6-hour forecasts were run using the Unified Model at N320 resolution (about 40 km grid spacing). These 6-hour forecasts were subtracted from the relevant 06 and 18 UTC control analyses, with two modifications: 1) mean pressure differences were removed (without this the pressure forecasts were noisy, possibly because the ECMWF forecast model is hydrostatic, whereas the Unified Model is non-hydrostatic) 2) the mean cloud water/ice fields from the forecasts were used to give the 'analysis' cloud fields (Piccolo, pers. comm., 2010). The previous training data (using differences of T+30 and T+6 forecasts) had larger variances and length scales, but gave qualitatively similar coefficients for the humidity transform.

The Met Office 4D-Var system is unusual in having a double inner loop (Rawlins et al, 2007), other systems have an inner loop and an outer loop (e.g. Rabier et al, 1998). The current operational global configuration is that 30 linearised iterations are performed at lower resolution and then 30 at higher resolution with iterations 31, 41 and 51 being non-linear updates. (Some tests use analyses with 50 or 60 iterations at a single resolution with every tenth one nonlinear.) The linearised iterations are performed with an efficient version of the minimisation algorithm and a quadratic penalty function. In the nonlinear updates radiative transfer calculations are re-linearised, the weights given to different scatterometer aliases are updated and if in use the humidity transform is updated. There is only one outer loop with a full non-linear forecast. The double inner loop has reduced the need for multiple outer loops. The 30 iterations at lower resolution give a preliminary analysis, but they also help to precondition the higher-resolution iterations.

2.2 The principle of symmetry

Let us assume that our forecast model is unbiased, in that its distribution of model humidity values is the same as that of the atmosphere mapped into model space. This "truth" state is the goal of our assimilation process and background errors are in principle measured from it. As we do not know it, we have to study background errors using a proxy – the training data. Figure 2 shows the joint distribution of background and analysis (proxy true) values. It is close to symmetric about the diagonal (at least below 100%), showing that the assumption of zero bias is reasonable. Yet the distribution of true/analysis values conditional on any particular background value is slightly biased, with mean value given by the dash-dot line. (By comparison with Lorenc (2007, his figure 30 shows the joint PDF of background and radiosonde humidity) it seems that this ensemble underestimates the magnitude of background humidity errors and that the dash-dot line in figure 2 may be too close to the diagonal.) Without assimilating any observations, the minimum variance

best estimate of the true RH would be obtained by bias correcting the background to this line. The resulting overall distribution underestimates the extremes. The only way to obtain a realistic humidity PDF in the analysis is to abandon the minimum-variance best estimate for RH, i.e. to allow for non-Gaussian error distributions.

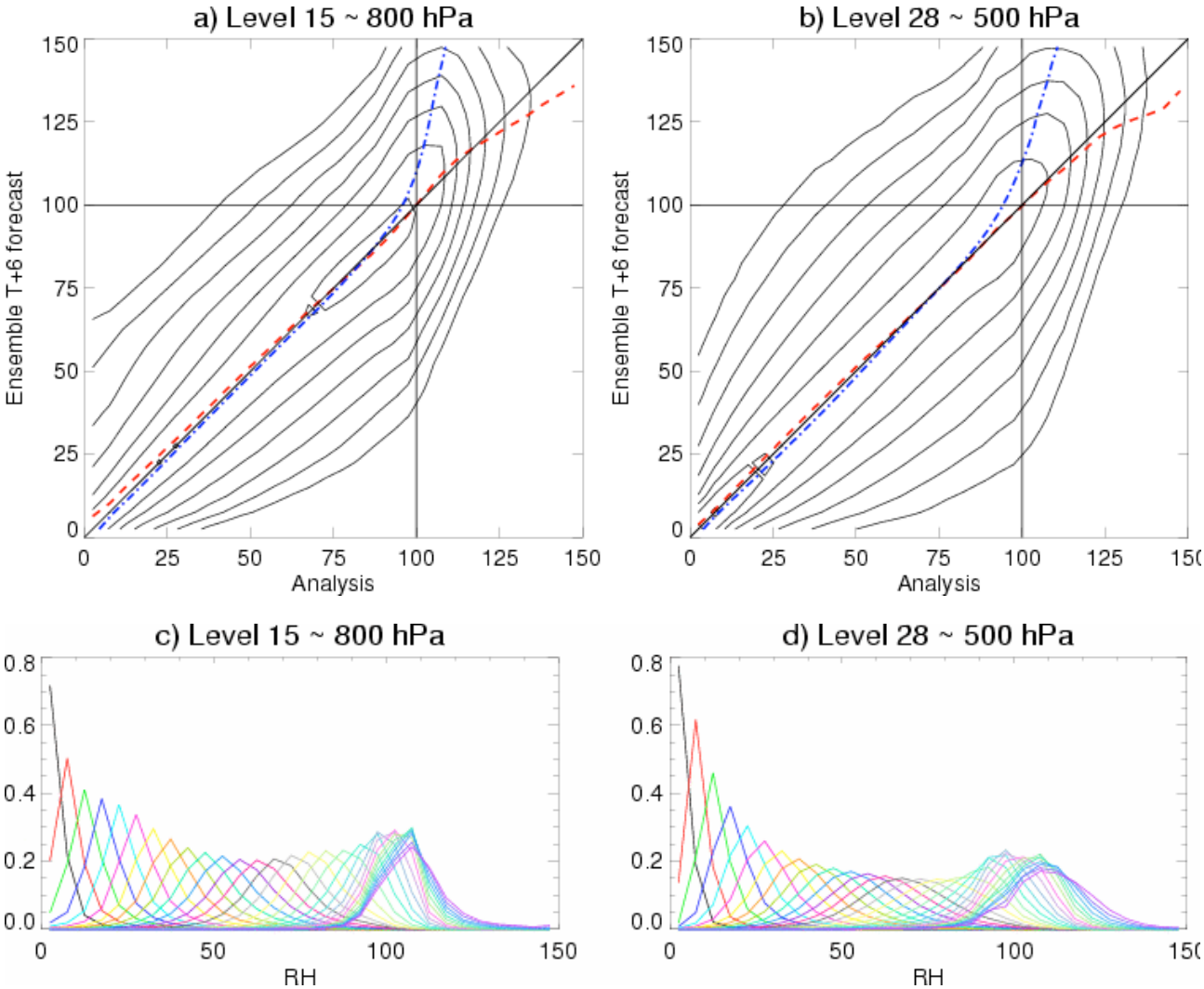

Figure 2. Top: joint PDFs of rh_T from analyses and forecasts in the training data at about 800 and 500 hPa. The contours are logarithmically spaced such that two intervals represent a factor of 10. (The asymmetry above 100% is because analysis cloud water was set to the mean of the forecast cloud water for each case.) The dash-dot line (blue) shows the mean analysis rh_T for each background rh_T bin, and the dashed (red) line shows the mean background rh_T for each analysis rh_T bin. The mode of the distribution is at near-saturated values of rh_T at low levels but in the mid- and upper-troposphere the mode is at low values of rh_T. Bottom: distribution of T+6 RH values for different values of analysis RH – an alternative way of viewing the joint PDFs. The PDFs for near-zero background values are truncated/asymmetric as expected, the PDFs for large background values are still reasonably symmetric because of the use of total RH.

The method suggested by Hólm et al (2002) was to use probability distributions conditional on $\overline{rh} = (rh_b + rh_a)/2$ (where subscripts a and b refer to analysis and background). We could illustrate the effect by plotting the joint PDFs of the difference in RH and the mean RH (not shown since they are equivalent to figure 2 a/b rotated by 45°). This is unbiased. However having the assumed probability distribution dependent on the analysed value makes the problem implicit, requiring an iterative solution method. The method constructs a Bayesian prior distribution whose mode is the background value and which is skew about this mode. We rely on the variational method returning the mode, not the mean, as its "best estimate" so that the analysis is equal to the background in the absence of observations. The transform is a device to allow us to specify a symmetric function which implicitly defines asymmetric PDFs. We note that the same symmetry principle can apply to observational error PDF (Geer and Bauer 2010); if we consider PDFs conditional on the observed value we get a similar apparent but harmful bias as shown by the dashed line in figure 2. In this paper we are only concerned with the background.
The main results of this symmetric approach are:

- There is no spurious bias.
- The background error standard deviation is small when the background RH is near zero and the analysis increment is negative, making negative analysed values unlikely, while for positive increments the standard deviation can be larger, making it possible to change near zero background RH to any positive value.

2.3 Humidity transform

In our notation Hólm et al (2002) introduced the transform:

$$\mu' = rh' / \sigma(rh' \mid rh_b + 0.5rh') = rh' / \sigma(rh' \mid \overline{rh}) \tag{4}$$

They also looked at correlations between humidity and temperature as a function of the reference humidity (their figure 31) as a preliminary to using this relationship in their analysis scheme. Our nonlinear humidity transform and its inverse uses the ideas of Hólm et al but some of the details are different. This section provides an overview of the transform with particular attention given to the novel features.

We linearise equation (2), and use primes to denote perturbations (or increments) giving our old humidity transform

$$rh_T{}' = \left(\frac{q_T{}'}{q_s} - \frac{q_T}{q_s} \frac{\partial \ln e_s}{\partial T} \Pi \theta' - \frac{q_T}{q_s} \frac{p'}{p} \left(\kappa \Pi \theta \frac{\partial \ln e_s}{\partial T} - 1 \right) \right) \tag{5}$$

$\kappa=R/Cp$, the ratio of the gas constant for dry air to the specific heat of dry air at constant pressure. θ is potential temperature $T = \theta\Pi = \theta(p/100000)^{\kappa}$. We take (5), add a normalisation factor *a* and scale the last two terms to give our new control variable

$$\mu' = a \left(\frac{q_T{}'}{q_s} - h_1 \frac{q_T}{q_s} \frac{\partial \ln e_s}{\partial T} \Pi \theta' - h_2 \frac{q_T}{q_s} \frac{p'}{p} \left(\kappa \Pi \theta \frac{\partial \ln e_s}{\partial T} - 1 \right) \right) \tag{6}$$

The *p'* term is very small and is neglected, giving

$$\mu' = am' \quad where \quad m' = \frac{q_T{}'}{q_s} - h_1 c\theta' \quad and \quad c = \frac{q_T}{q_s} \frac{\partial \ln e_s}{\partial T} \Pi \tag{7}$$

The coefficients *a* and h_1 are determined statistically (section 2.4) [1]; as discussed above they should be symmetric functions of background and analysed values. This is the forward transform, we also need the inverse transform to convert from *μ'* to $q_T{}'$. To be more precise given *μ'* and virtual potential temperature increments we have to find $q_T{}'$ and *θ'*. Since we use this transform within our inner loop (Hólm et al. 2002 only put their similar transform in the outer loop), we need this inversion to be robust for all background and analysed values. This can only be ensured if the Jacobian of the forward transform is positive everywhere. With our large set of statistically derived coefficients we found this condition could not be guaranteed without modifying the transform. So we dropped the symmetry principle for h_1, making it a function of the background only, and made $a=a(rh_{Tb}, rh_{Tb}+m')$ instead of the exactly symmetric function $a=a(rh_{Tb}, rh_{Tb}+rh_T')$. This gives us a single nonlinear function for *μ'* in terms of *m'*, so the condition for invertibility becomes simply one of monotonicity for this function, i.e.:

$$\frac{\partial \mu'}{\partial m'} = a + \frac{\partial a}{\partial m'} m' > 0 \tag{8}$$

We tabulate a look-up table for *a* in terms of all possible values of $(rh_{Tb}, rh_{Tb}+m')$. It is possible then to check the full table and modify it so that (8) is always obeyed (using a tolerance of 0.05 rather than 0 because of the approximations involved). Inversion is done iteratively using a secant method (slightly modified so that the two points retained bracket the solution whenever possible). In the code there is an option to keep *a* fixed at $a(rh_{Tb}, rh_{Tb})$ - a linear version of the transform. Further details of some aspects can be found in Lorenc (2010; VSDP11_2, unpublished manuscript).

[1] We could have combined h_1 with *c* but with this derivation h_1=1 gives a RH-like variable, whereas h_1=0 gives a q-like control variable.

Monotonicity problems occurred mainly for larger magnitudes of rh_T' and were more common at levels with bimodal humidity distributions, but were not restricted to them. Smoothing the humidity distributions did not give a satisfactory solution. The first guess for the table is $a(rh_{Tb}, rh_{Tb} + rh_T') = 1/\sigma(m' \mid \overline{rh_T})$ where $\sigma(m' \mid \overline{rh_T})$ is the standard deviation of m' evaluated at $\overline{rh_T} = rh_{Tb} + 0.5rh_T'$ (the table is extended to cover small negative values as unphysical states can be tried during the minimisation). Condition 8 is then checked along rows of the table (constant rh_{Tb}) starting from the diagonal ($rh_T'=0$) and modifying the a value further from the diagonal if necessary (this makes the table non-symmetric). In principle, given the joint PDF of rh_{Tb} and $rh_{Tb}+rh_T'$ we could derive a directly. Although this would be less *ad hoc*, in practice certain areas of the PDF would be poorly sampled. Note also that a only gives an approximate normalisation, i.e. the variance of μ' is approximately rather than exactly 1. In the linearised iterations we replace $\mu' = am'$ (from (7)) with its linearization $\mu' = \hat{a}m' + \mu_c$ where $\hat{a} \triangleq \partial\mu'/\partial m'$ and $\mu_c = \mu' - \hat{a}m'$. These are held constant until the next nonlinear iteration.

ECMWF also had problems with invertibility of their transform and in response split the transform into two regimes, subsaturated and supersaturated, in each regime a curve was fitted to the data (Hólm, pers. comm.). We prefer to work with total humidity and to recognise that the normalisation should be a function of both rh_{Tb} and $rh_{Tb}+rh_T'$ rather than just their midpoint.

The computational cost of the humidity transform is minimal compared to other aspects of 4D-Var. It was found that with the RH control variable there was a large case to case variability of the largest eigenvalue of the Hessian matrix, with the new control variable there was much less variability and a significant reduction in magnitude of the largest eigenvalue (also true with h_1=0 and h_1=1 so it seems to be related to the normalisation). In one set of tests the control had leading eigenvalues between about 900 and 4000 whereas the trial had values between about 460 and 520 (with one value of 900). When the number of iterations is allowed to vary then the humidity transform sometimes gives convergence in 10 or 20 fewer iterations than the control.

Figure 3 shows the minimum and maximum rh_T values produced in a single analysis. The background does not contain negative humidities, all the analyses do but as expected the non-linear transform performs best at reducing undershoots – both in terms of the minimum values and the percentage of negative points. In all the analyses the percentage of negative points is higher than average over the ocean and lower in the tropics. Overshoots (relative to the background) are mainly confined to the lower troposphere. Behaviour near zero is affected by the a values there – increasing a would further reduce negative analysed values, but might increase the likelihood of some points being inappropriately stuck at very small rh_T (Gustafsson et al, 2011).

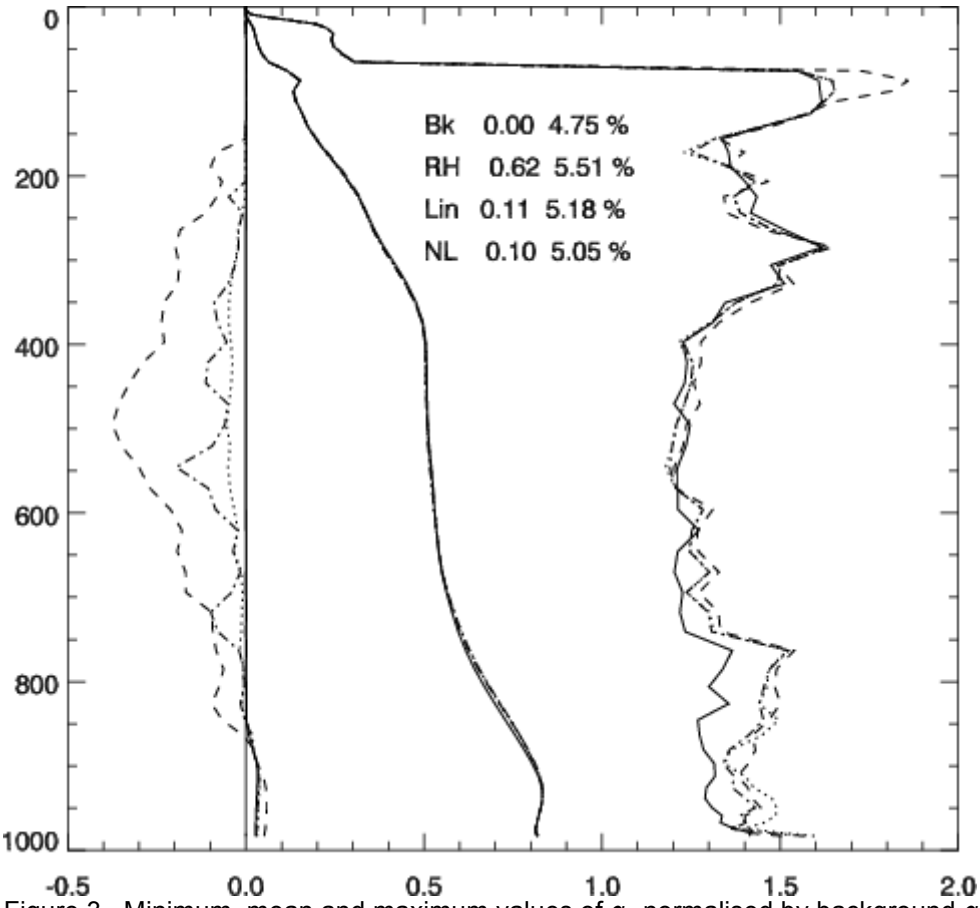

Figure 3. Minimum, mean and maximum values of q_T normalised by background q_s. Values are calculated on model levels and plotted at the mean pressure of that level. Solid line – background, dashed line - total RH analysis, dotted (dash-dotted) line – analysis with non-linear (linear) humidity transform. Percentages of negative and supersaturated points are given. Values have been calculated on model levels and plotted at the mean pressure (hPa) of each level (also true of subsequent figures).

2.4 Calculation of coefficients

To calculate h_1 and *a* we take the training data (described in section 2.2) and stratify it by values of the reference humidity (either rh_{Tb} or $\overline{rh_T}$) using 30 bins from 0.0 to 1.5 at each vertical level. The (co)variances of q'_T/q_s and $c\theta'$ were calculated (but without removing mean values). Slight smoothing $v_i^{new} = 0.25(v_{i-1} + 2v_i + v_{i+1})$, was applied over the *rh* bins, where v_i is the sample (co)variance for bin *i*, e.g. $v_i = <(c\theta')^2>_i$. Two estimators for h_1 have been tried:

$$h_{1,i}^{c} = <(q'_T/q_s)c\theta'>_i \Big/ \sqrt{<(q'_T/q_s)^2>_i <(c\theta')^2>_i} \qquad (9a)$$

$$h_{1,i}^{rr} = <(q'_T/q_s)c\theta'>_i \Big/ \left(<(c\theta')^2>_i + k\right) \qquad (9b)$$

Eqn 9a gives the correlation coefficient, shown in figure 4a as a function of pressure and $\overline{rh_T}$. The correlation is positive for $\overline{rh_T}$ above about 0.7 or 0.8 (slightly less near to the ground) and negative below, this behaviour is similar to that in figure 31 of Hólm et al (2002). Note that Hólm et al (2002) describe an early version of their transform, before they introduced an equivalent to h_1, their σ is approximately equivalent to our *1/a*. The negative correlations in the subsaturated regime are stronger in the tropics and weaker at high latitudes (not shown). Over land comparison with observations suggest a generally negative correlation between temperature and humidity errors at low levels – this does not show up in figure 4 either due to the influence of sea points, or because the training data has insufficient soil moisture perturbations. Eqn 9b gives the regression coefficient if *k=0.* This is larger in magnitude in both regions and in particular

has values less than -2 for small *rh*, these seem unphysical. We also tried ridge regression with $k = 0.5 < (c\theta')^2 >$ (with the variance taken over all *rh* bins) to give more robust regression coefficients – shown in figure 4b. This has the required effect of making the less reliable coefficients smaller whilst having little effect on the coefficients around saturation. Correlations as a function of rh_{Tb} (not shown) are broadly similar but generally reduced in magnitude. Correlations as a function of background vertical velocity (not shown) were found to be even smaller in magnitude. Hólm (2010, pers comm.) also recognised a problem with large negative regression coefficients at low $\overline{rh}$ and set negative values to zero. We suspect that the problem is related to the constraints on rh_a and rh_b at low $\overline{rh}$ We have trialled various options (section 3) and are using the correlation coefficient as h_1 – this is only feasible because the use of $c\theta'$ rather than θ' makes h_1 non-dimensional and O(1).

The variance of m' is

$$\sigma_i^2 = < (q_T'/q_s - h_1 c\theta')^2 >_i = < (q_T'/q_s)^2 >_i - 2h_{1,i} < (q_T'/q_s) c\theta' >_i + h_{1,i}^2 < (c\theta')^2 >_i \qquad (10)$$

Figure 5a shows σ, which necessarily becomes rather small for $\overline{rh_T}$ <0.2 because rh_{Tb} and rh_{Tb} +rh_T' both have to be small in this case. At most levels there are two broad maxima in σ round 0.5 and 1.3. (Both features are largely absent from plots as a function of rh_{Tb} – not shown.) The PDFs (figure 5b) show rather broad distributions, falling off rapidly above about 1.1. There seems to be some seasonal/latitudinal variation (not shown) linked to boundary layer and tropopause depth as expected. This is not taken into account and averaged values of h_1 and σ (as shown in these figures) are used.

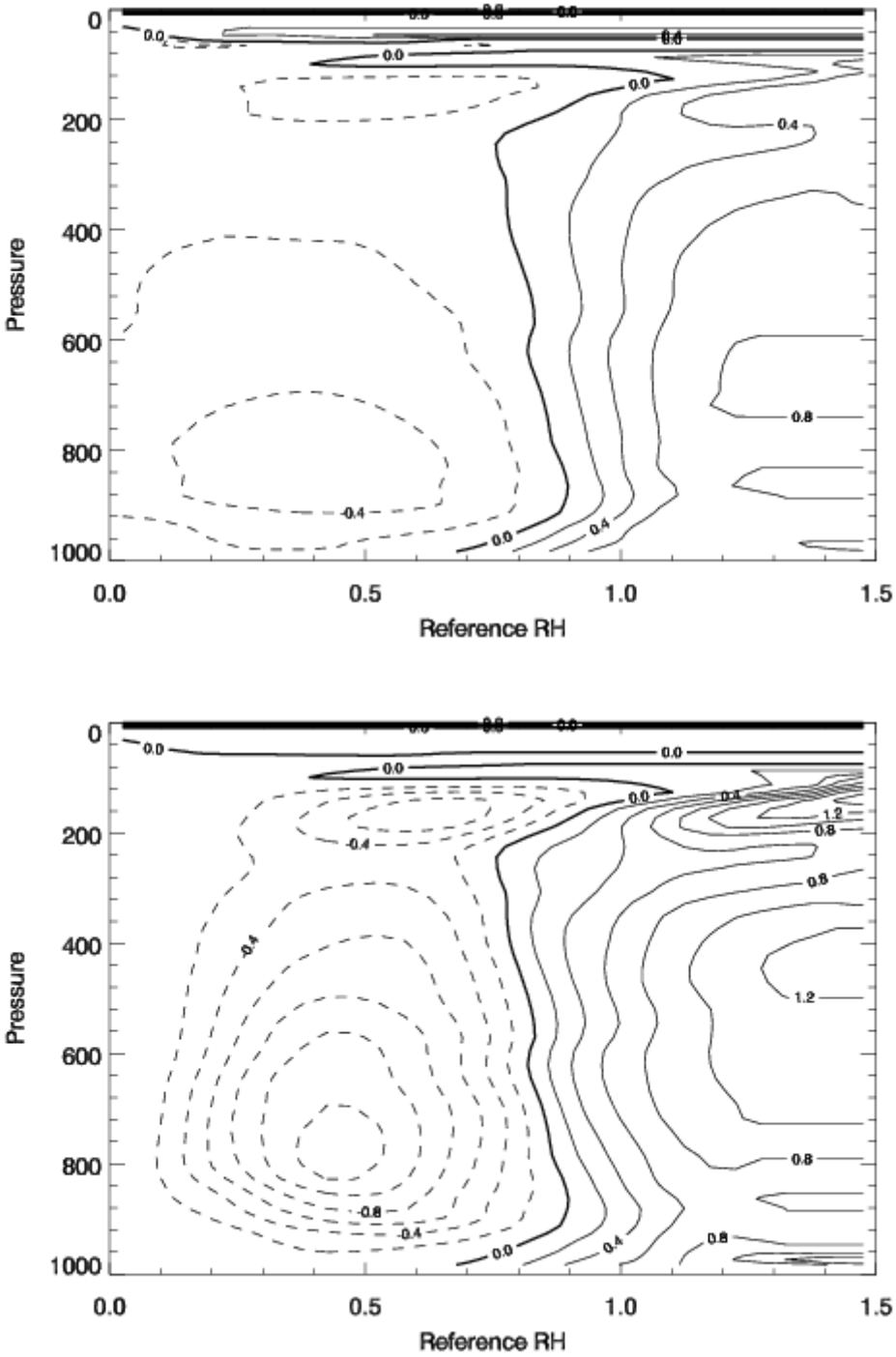

Figure 4. a) Correlation coefficient and b) robust regression coefficient as a function of level and $\overline{rh_T}$. See text for details. The correlation coefficient has range [-0.55,0.87], the robust regression coefficient has range [-1.30,1.51], the standard regression coefficient (not shown) has range [-3.13,1.57].

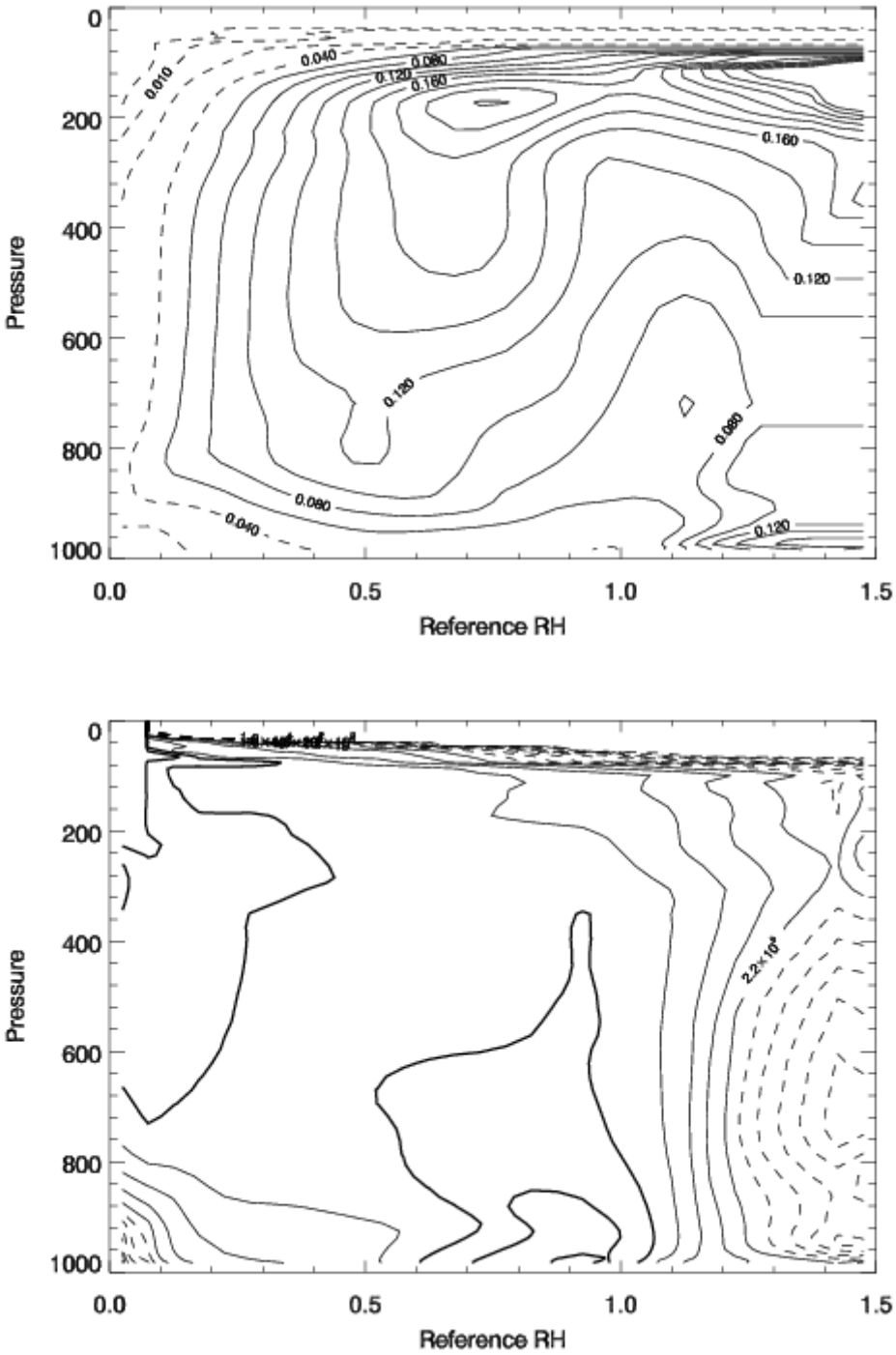

Figure 5. a) σ and b) sample density as a function of level and $\overline{rh_T}$. b) uses logarithmic contours, with three contours implying a factor of 10.

The non-linear nature of the humidity transform means that an extra pass through the training data is necessary. The first pass calculates the statistics conditional on $\overline{rh_T}$ as above. The second pass calculates vertical covariances and horizontal spectra for the various control variables including μ'. The vertical transform uses modes calculated as eigenvectors of $P_V B_V P_V$ (eqn 19 of Lorenc et al, 2000) where B_V is the vertical covariance matrix for μ' and P_V is an inner product to make the modes more physical and reduce the dependency on spacing of the layers. For most variables the inner product is proportional to the average mass of the layer, for humidity this has been multiplied by the variance of q_T' since 2001, this was done to put more emphasis on low level humidity and gave a slight improvement in forecast skill. During the testing of the new humidity transform we changed to multiply by the mean q_T which effectively put more

weight on low levels and gave a minor improvement in results. For humidity the vertical modes need not extend to the top of the model.

The calculation and use of the coefficients is not entirely consistent. Arguably the initial iterations should use statistics with rh_{Tb} as a reference and later iterations should use $\overline{rh_T}$ (except for h_1) but for simplicity we use statistics with $\overline{rh_T}$ as a reference throughout the iterations.

Figure 6 shows the standard deviation (SD) of μ' (these statistics, and those used in the trials, were calculated using the linear transform due to an oversight[2]). At levels that are completely in the troposphere the μ SDs are relatively uniform (except for some high near-surface values over Antarctica, where the PDF is very peaked near saturation). There are modest maxima in mid-latitudes – becoming higher and larger in the tropics (to some extent this reflects the RH SDs – not shown). Around 200 hPa the SDs go from 0.4 or less to over 2 reflecting the differences in stratospheric and tropospheric distributions (this could be reduced by using a latitude dependent version of the transform). Above 100 hPa there are some very large SDs, but at the highest levels (table 1) the transform isn't used and q increments are zero.

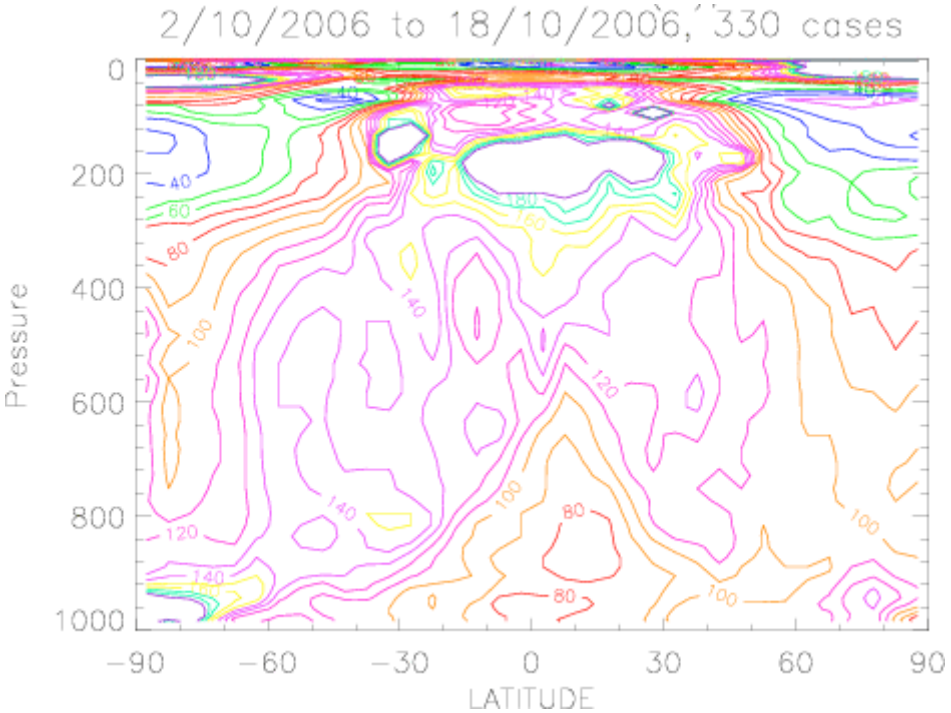

Figure 6. Standard deviation (x100) of transformed humidity variable.

2.5 Higher moment statistics

Skewness and flatness were calculated from the training data in order to give a measure of how non-Gaussian each PDF is. (Note however that, as discussed in section 2.2, the key feature of our model for the PDF is that its mode is equal to the background, with the distribution and hence the mean allowed to be skew from this mode. The skewness and flatness are defined about the mean and do not measure this in a simple way.) For each sample and level skewness and flatness were calculated in boxes of 5° latitude by 10° longitude, they were then averaged horizontally and over the different samples (times). Normalising the higher moments by the standard deviation within each box should be more appropriate than normalising by the global standard deviation, but of course the box boundaries are rather arbitrary. The results are shown in figure 7. When calculated using the nonlinear transform μ' appears more Gaussian than rh_T', but this is reversed when using the linear transform. There are spikes in the μ' flatness around the tropopause.

[2] A later trial was run using statistics generated from the nonlinear transform. This gave some additional improvement in verification scores.

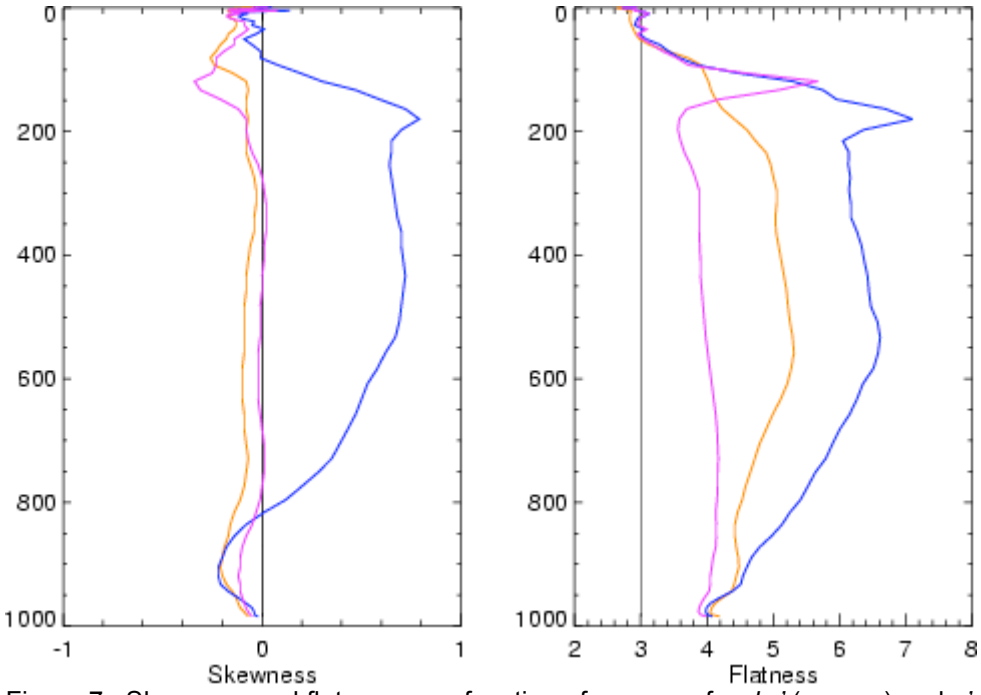

Figure 7. Skewness and flatness as a function of pressure for rh_T' (orange) and μ' using linear (blue) and nonlinear (purple) transform. Gaussian distributions have zero skewness and a flatness of 3. See text for further details.

3. Results

3.1 Trials performed

Period	Dates	PS	Analyses	Forecasts	Top of μ modes
1	24.06.2009 – 24.07.2009	23	N108 (~ 120 km)	N320 (~ 40 km)	70 (~ 0.01 hPa)
2	15.12.2009 – 15.01.2010	25	N108/N216 (~ 60 km)	N320 (~ 40 km)	55 (~ 32 hPa)
3	01.06.2010 – 30.06.2010	25	N108 (~ 120 km)	N216 (~ 60 km)	53 (~ 50 hPa)
4	01.12.2010 – 31.12.2010	27	N108 (~ 120 km)	N216 (~ 60 km)	53 (~ 50 hPa)
Op	02.11.2010 – 19.07.2011	25/26	N108/N216 (~ 60 km)	N512 (~ 25 km)	43 (~ 172 hPa)

Table 1. The periods used for global trials. Columns give the dates, the baseline parallel suite (PS), the analysis and forecast resolution (Nn indicates 2n grid points along a latitude row, approximate grid spacing in brackets) and the number of levels used for humidity modes. The bottom row gives the operational settings prior to implementation of the transform.

The period 1 trials were based on PS23 (implemented operationally in March 2010) and forecast difference training data. Later trials were based on PS25 (implemented in November 2010) using the statistics described in section 2. It was necessary to perform the forecasts (and sometimes the analyses) at lower horizontal resolution than used operationally. The control humidity analyses used RH modes (section 2.1) up to level 43, "Control2" used RH modes up to level 55 and gave slight improvements over the original control. All the trials of the transform used sufficient modes to cover the tropical tropopause. There were no noticeable problems using all 70 levels in period 1, but this was reduced as a precaution in later periods.

Analyses are performed every six hours and forecasts out to T+120 hours are run from 00 and 12 UTC. Forecasts are verified against both observations (radiosondes and surface stations) and against analyses from the same trial. Root-mean-square (rms) 'errors' for individual parameters are available and also the "NWP index" a basket of scores normalised by persistence and weighted, with most weight on T+24 performance (Appendix A of Rawlins et al, 2007). The 22 scores used in the NWP index are for pressure at mean sea level, height at 500 hPa and wind at 850 and 250 hPa and so measure the large scale dynamics of the situation and not the humidity distribution. The trials run and their scores are summarised in table 2, this also gives the average (unweighted) change in rms for the same 22 variables/ranges.

A preliminary experiment was run including cloud water in the 4D-Var linearization state (LS) and extra terms in the advection of cloud water in the perturbation forecast model (already used in the regional NWP system). The results were neutral or marginally negative but around the noise level. It was decided on principle to include this change in the tests of the humidity transform.

Period	Trial	Verification against observations		Verification against analyses	
		Change in index	Change in rms	Change in index	Change in rms
1	Cloud in LS	-0.09	+0.16%	-0.26	+0.19%
1	Linear, h_1^c [(a)]	+0.31	-0.20%	-0.18	+0.14%
1	Non-linear, h_1^c [(a)]	+0.39	-0.65%	+0.36	-0.54%
2	Control2 (L55 rh)	+0.02	0.00%	+0.25	-0.14%
2	Linear, h_1^c	+0.16	-0.50%	+1.38	-1.02%
2	Non-linear, h_1^c	+0.04	-0.39%	+1.63	-1.21%
2	Non-linear, h_1=0	+0.01	-0.55%	+1.84	-1.40%
3	Non-linear, h_1^c	+0.21	-0.37%	+0.99	-1.20%
3	Non-linear, h_1^c [(b)]	+0.02	-0.13%	+0.67	-0.95%
3	RH*0.8	+0.09	-0.07%	+0.31	-0.21%
4	NL stats [(c)]	+0.34	-0.06%	+0.44	-0.17%
4	NL stats, lat var [(c)]	-0.01	+0.59%	-1.01	+1.48%

Table 2. Summary of main global trials and their verification (see text for details). All verification is compared to the appropriate control run without the humidity transform (except for period 4). All subsequent trials include "Cloud in LS". Improvements correspond to positive changes in the index and negative changes in rms. [(a)] In the period 1 trials there were errors such that increments were smaller than they should have been in boundary layer humidity and in soil moisture, the linear trial had an additional error in the normalisation factors. [(b)] Includes a correction to a minor error in the weighting used in estimates of a and h_1 in the other trials. [(c)] The period 4 trials are compared with a control that includes the non-linear humidity transform.

Several features were seen in all the non-linear trials: most rms differences were improved or near-neutral. The improvements were particularly marked in the tropics and southern extratropics and when measured against analyses, a point made in table 3. (In period 2 the index measured against observations would have been more positive but for changes to tropical wind persistence.) In period 2 the northern/southern extratropical Pmsl (pressure at mean sea level) rms values changed by +0.05/-1.09% at T+24 and +0.26/-0.96% at T+120 vs. surface observations and -1.09/-3.45% at T+24 and +0.06/-1.04% at T+120 against analyses. Short range surface temperature rms against observations was changed by -0.29/-0.39/-0.38% at T+12 in the northern extratropics/tropics/southern extratropics, by T+48 the figures were -0.23/-0.27/-0.10%. Verification against in situ observations mainly samples land areas, and verifying short-range forecasts against their own analyses is flawed because the two are not independent. The comparison of background fields with all types of observations (section 3.2) gives confidence that there is a real improvement in the forecasts.

	Period 1		Period 2		Period 3	
	Ob	An	Ob	An	Ob	An
20°-90°N	0/0	0/3	0/0	2/5	0/0	0/ 6
20°S-20°N	2/0	21/1	2/2	49/2	1/1	54/10
20°-90°S	7/3	6/7	7/2	61/8	8/1	23/ 2

Table 3. Nonlinear trials, verification against observations and analyses: number of rms values better/worse by 2% or more (this is from an extended set of 123 values for each region). Some of the "worse" values are for RH at T+24 against analyses. Many of the improvements are for height, Pmsl and temperature fields (arguably less important than winds in the tropics).

A nonlinear trial using robust regression coefficients (h_1^{rr}, section 2.4) gave worse results and was stopped after 21 days. Relative to the relevant h_1^c trial the change in NWP index was -0.166 and -0.717 and the change in rms was +0.22% and +0.58% against observations and analyses respectively. In contrast using h_1=0 gave very similar results to the nonlinear trial (Table 2). Setting negative values of h_1 to zero (as done at ECMWF) was also tried but this gave worse results.

The period 1 linear trial was compromised by an error. For period 2 the linear trial performance is similar to that of the non-linear trial, it was marginally better measured against observations and marginally worse against analyses.

As discussed below (section 3.3) the transform gives smaller humidity analysis increments in the lowest 200 hPa. To see if smaller increments by themselves gave a similar improvement we ran a period 3 trial with the old (RH) control variable but with its standard deviations scaled by 0.8. This gives some improvement in the fit to analyses at short range (as seen in the NWP index, table 2) but we conclude that this improvement is more apparent than real because there is no improvement in the fit of background fields to observations, and in particular the fit to satellite humidity channels (cf figure 10 below).

Following operational implementation there were two trials testing detailed aspects of the humidity transform. These are the period 4 trials in table 2, in this case the control and trials all use the non-linear humidity transform (with statistics generated from "unpacked" training data). As mentioned above a second pass through the training data is needed to generate covariance statistics for the transformed humidity variable. All earlier versions of this second pass used the linear version of the transform, the "NL stats" versions use the non-linear transform on the training data – this gives a modest improvement in the forecast scores. Additionally including latitudinal variation (using 15° bands) in the humidity transform gave quite a strong negative signal. The reasons for this aren't clear, but the effective sample size for the transform statistics will be smaller perhaps giving more noise.

The humidity transform (using the same coefficients as the global model) has also been tested in the North Atlantic and Europe regional model. A summer trial (1-31 July 2010) gave neutral results. A winter trial (14 December 2010 to 13 January 2011) gave slightly improved forecasts, especially of precipitation and surface temperature (Renshaw, pers. comm., 2011). Impact in the UK model was essentially neutral (Dow, pers. comm.., 2011).

3.2 Fit to observations

Figure 8 shows the analysis fit to radiosondes in the northern extratropics for period 2. Throughout the troposphere the nonlinear trial has a slightly improved rms fit (in period 1, not shown, the improvement was particularly marked in the upper troposphere). Below 925 hPa the analysis is about 1% RH wetter than the observations but there is little bias in mid-troposphere. From 400 to 250 hPa the analysis is 1-4% RH wetter than the radiosondes (daytime radiosondes tend to be too dry at these levels, e.g. Sun et al, 2010). From 200 hPa upwards the analyses are about 4% RH drier than the radiosondes that are not rejected at these levels. As the forecast proceeds the biases become larger (by T+120 almost +3% RH, -2% RH and +4% RH at 925, 700 and 200 hPa respectively) but with little difference between the control and trial. The background fields fit radiosonde humidities about 1% better in rms at most levels (and temperatures about 0.5% better in the troposphere). The trial analysis has a slightly better rms fit to surface RH and the bias is reduced from +1.4 %RH to +1.2 %RH in the northern extratropics.

Errors in the Met Office analyses near the tropopause are assessed by comparison with Aura Microwave Limb Sounder (MLS) water vapour retrievals (which were not assimilated). Errors were calculated by interpolating Met Office water vapour in space to observation locations. No time interpolation was carried out, but only Met Office analyses at 12 UTC and MLS data in the window between 09 and 15 UTC were used. MLS data quality flags were used to reject unreliable MLS retrievals. Read et al (2007) report that, for mixing ratios less than 500 ppmv, the biases of the retrievals are less than 25% between 316 and 147 hPa and rms errors are 25% at 147 hPa and 10-20% from 121 to 83 hPa. Figure 9 shows that the nonlinear trial has a generally closer fit to MLS than control2: standard deviation mainly 10-30% better at higher levels and latitudes and about 5% better at 150 hPa in the tropics. Both sets of analyses are too dry in the stratosphere compared to MLS (not shown), this is also seen in comparison with IASI and radiosondes.

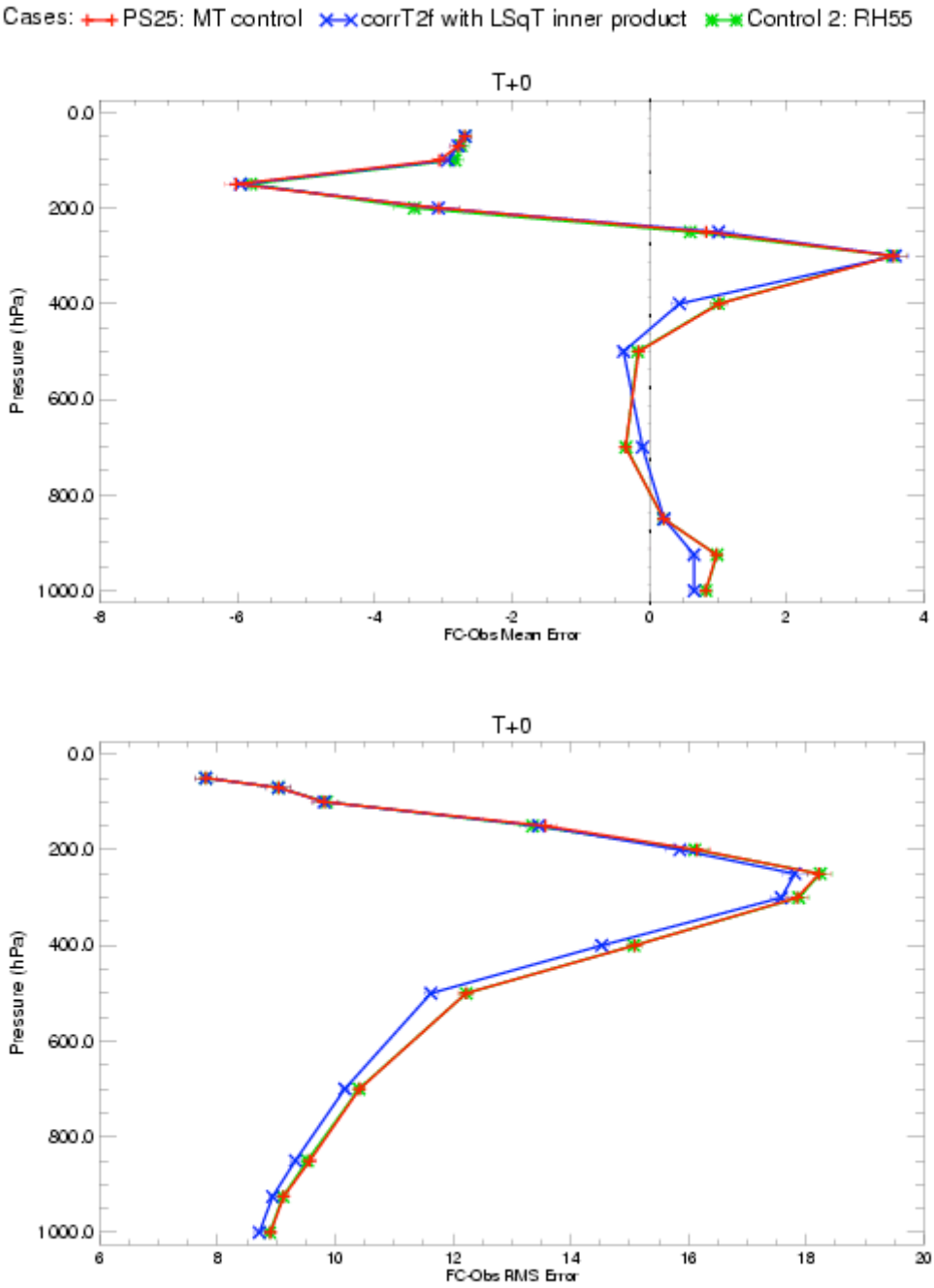

Figure 8. Analysis fit to radiosondes, 20-90°N. Top mean (A-O) and bottom rms in % RH for control (red), control2 (green) and nonlinear trial (blue).

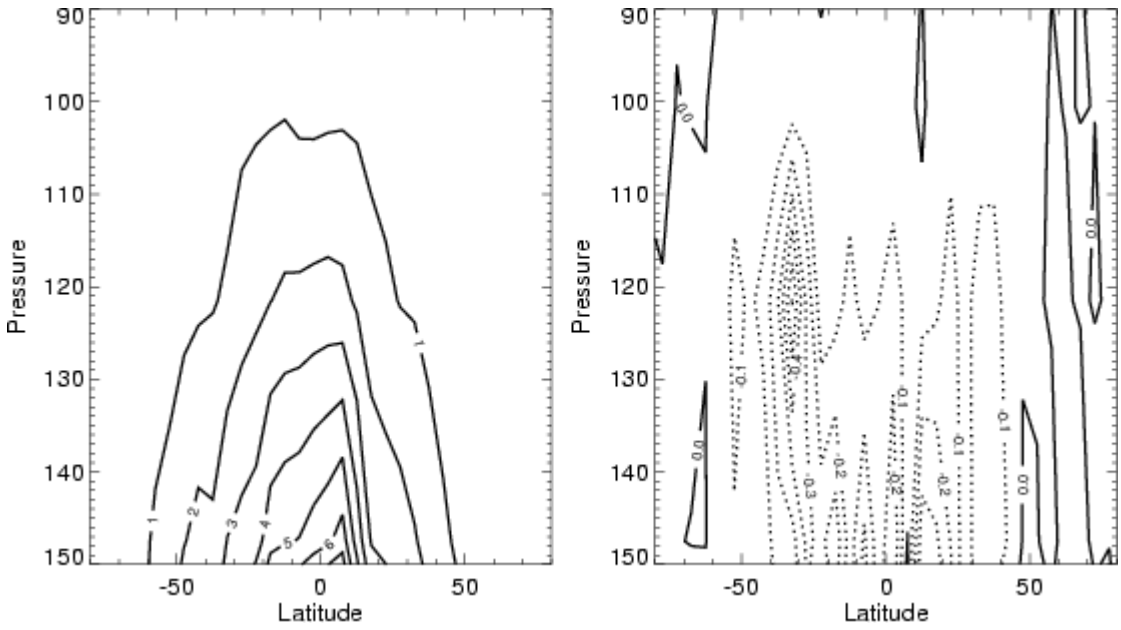

Figure 9. Standard deviation of non-linear trial analyses – MLS water vapour (ppmv) (left) and difference from Control2 – MLS standard deviation (right). Differences have been averaged in 5° latitude bins and pressure bins centred on the fixed pressure levels at which MLS retrievals are produced.

The fit of background fields to observations of all types showed a consistent improvement with the initial observation penalty improved by 0.2% in period 2 (0.6% in period 1). The largest improvements were for humidity sensitive satellite channels, shown in figure 10 (see caption for satellite acronyms). The statistics are for assimilated soundings. Except for SEVIRI these channels are only assimilated over the ocean. AMSUA channels 1 and 2 are significantly sensitive to cloud liquid water, these show slightly worse fits. AMSUA-4 has some sensitivity to cloud liquid water, but is mainly a temperature channel, peaking at about 800 hPa. For water vapour channels the improvement is largest, about 3%, for higher-peaking channels and least for the channels that have significant sensitivity to the surface. Most of the humidity sensitive channels also have a modest increase (typically 0.5%; 1.5% for the SSMI-S channels) in the numbers of reports passing the quality control checks. Similar changes (not shown) are also found in the other periods studied and represent a relatively large improvement in the quality of background humidity fields over the ocean. The assimilation of humidity sensitive radiances always has the potential to change the temperature analysis directly because all channels are temperature-sensitive. With an explicit link between temperature and water vapour increments in the humidity transform this potential is even greater. The diagnostics of analysis increments (below) also suggest a direct effect on the temperature analysis – without this it is unlikely that the impact on Pmsl and height forecasts would be as large as it is.

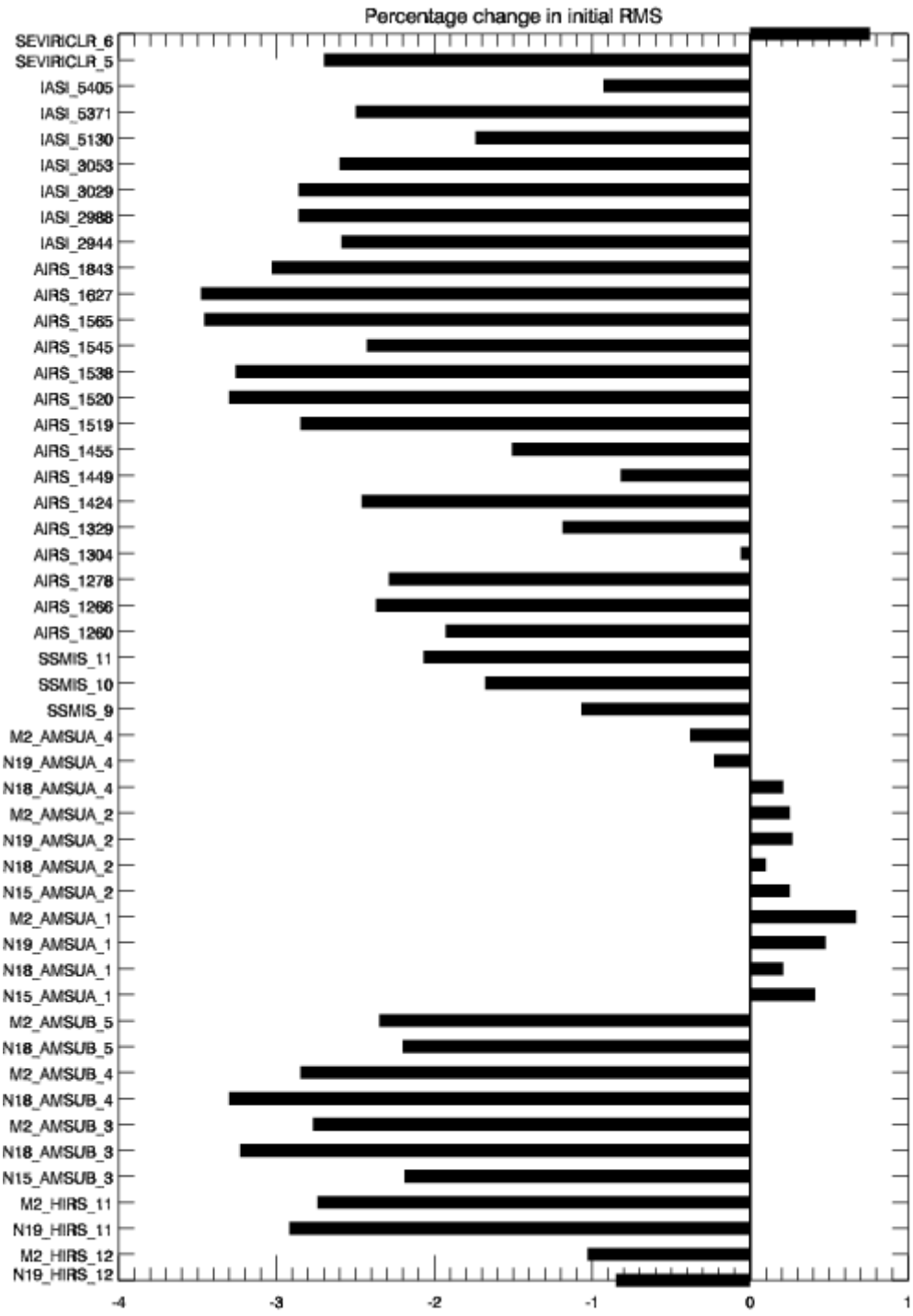

Figure 10. Percentage change in rms(O-B) for assimilated humidity sensitive satellite channels (period 2 nonlinear trial vs. control2). The instrument acronyms used are: AIRS – Atmospheric Infrared Sounder, AMSU – Advanced Microwave Sounding Unit, HIRS – High-resolution Infrared Radiation Sounder, IASI – Infrared Atmospheric Sounding Interferometer, SEVIRI – Spinning-Enhanced Visible and InfraRed Imager, SSMI – Special Sensor Microwave Imager. AMSU and HIRS are available from METOP2 and several NOAA (National Oceanic and Atmospheric Administration) satellites.

As suggested by figure 1 and noted by Dee and da Silva (2003, section 7) the details of covariance models are less important where there is an abundance of observations. Satellite soundings effectively vertically integrate water vapour over very thick layers and because of this the vertical distribution of the analysis increments produced is very dependent on the background error model used. The largest increments will tend to be at the level of maximum background error variances (see section 4). The analysis fit to a radiosonde profile of temperature and humidity observations will show less sensitivity. The importance of the interaction between the humidity transform and satellite soundings is consistent with the large impacts in the southern extratropics (also found by Hólm, pers comm., 2010) where forecast quality is heavily

dependent on satellite data. Also the verification against (mainly land-based) observations shows less impact than the verification against analyses.

3.3 Analysis increments

Figure 11 shows q analysis increments averaged over trial period 2. The nonlinear trial has rms increments that are 5-10% larger in the mid-troposphere, but the most marked feature is that increments below about 900 hPa are 20% smaller. Near the surface the trial rms temperature increments (not shown) are up to 7% larger. In period 1 the rms at 800 hPa was about 0.5 g/kg (not shown), this has been reduced to about 0.34 g/kg in period 2 – thought to be mainly due to the use of the analysis ensemble statistics. At most levels the minimum and maximum increments (Fig 11b) are approximately symmetric, but below 950 hPa the control has a bulge in the minimum values absent in the trial, which is more symmetric.

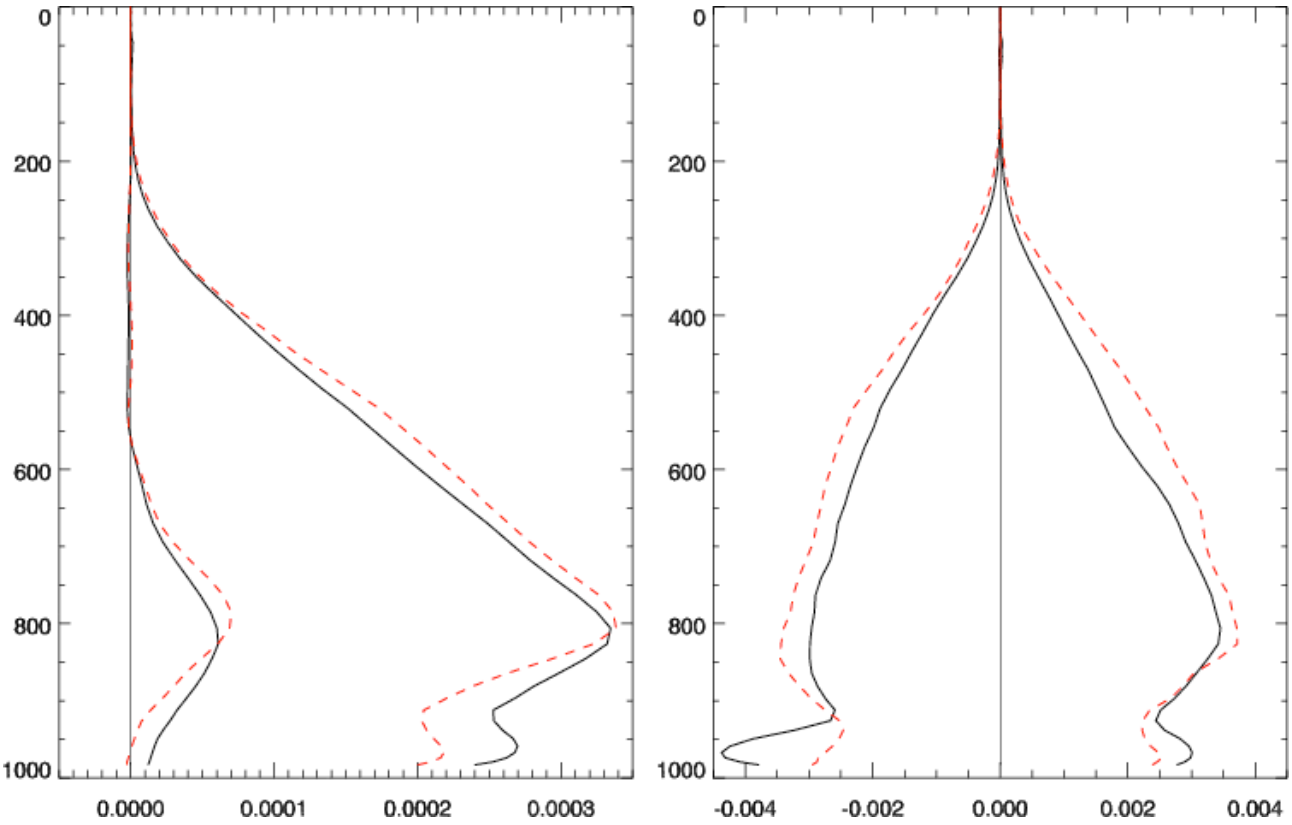

Figure 11. Mean and rms (left) and minimum and maximum (right) q increments from control2 (solid line), and the nonlinear trial (dashed - red). Axes are pressure (hPa) and q (kg kg^{-1}).

3.4 Evolution of the hydrological cycle

The model precipitation depends on both the humidity fields and convergence/vertical motion and so is sensitive to imbalances introduced by the analysis. There is often a 'spin-up' or 'spin-down' of precipitation with forecast range. Milton and Earnshaw (2007) found that Met Office short range forecasts have too much precipitation and evaporation over both land and ocean.

Hourly precipitation, evaporation and cloud from T-3 (initialisation of forecast) to T+24 were averaged over the trials, combining 00 and 12 UTC forecasts and large regions. Global mean precipitation is excessive over the first hour at about 4.3 mm day^{-1}, it then drops dramatically to about 3.5 mm day^{-1} and then declines gradually to about 3.2 mm day^{-1} by T+24 (Figure 12). The same general pattern is seen in different regions of the globe (and in the other periods – not shown). The tropics and the summer extratropics have two humps, due to the diurnal cycle in convection over land being sampled twice. In contrast the winter extratropical curve is much smoother and shows a decline up to about T+6 and then relatively flat values.

The nonlinear trial reduces the spin-down between the first and second hours of the forecast by about 20% in the tropics. According to the model parameterizations 70-80% of the precipitation is convective, and this dominates the signals seen. There is an initial spin-down in the large-scale precipitation as well (not shown) - very similar between the control and non-linear trial - but after T-1 the large-scale rates are almost constant. For period 1 the excessive initial precipitation was reduced more by the nonlinear trial.

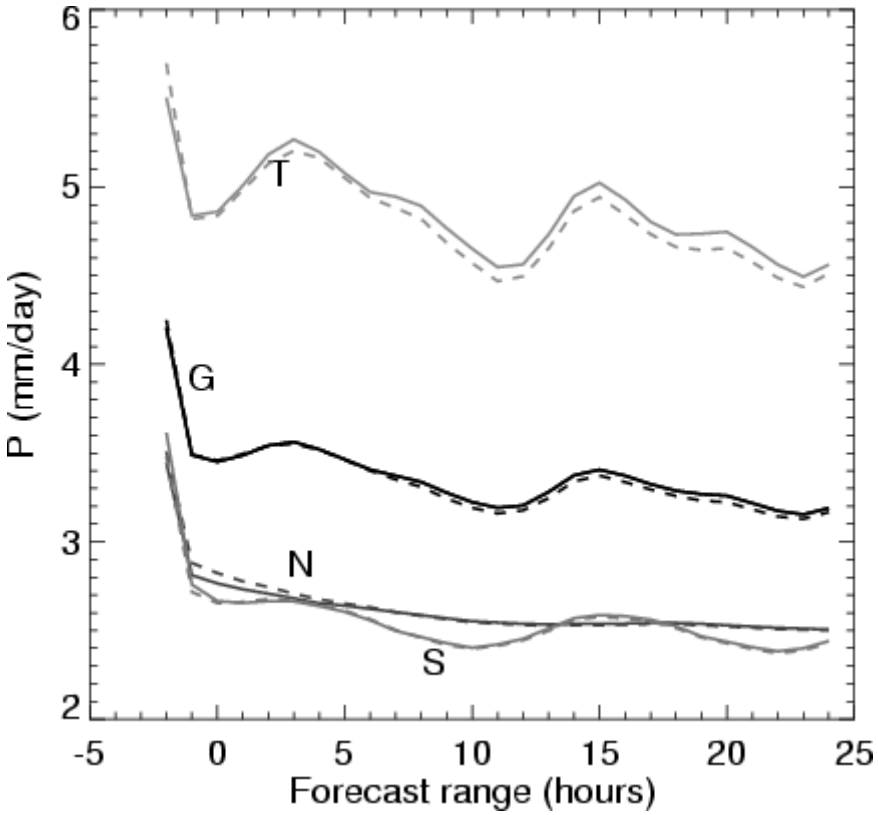

Figure 12. Hourly precipitation rates over the first day of the forecasts (period 2). The regions used are: S - 90°-20°S; T – 20°S-20°N; N – 20°-90°N; G - global. For the nominal 12 UTC analysis observations from 9 to 15 UTC (T-3 to T+3) are used and 4D-Var provides initial conditions at the start of window: T-3. Similarly for 00 UTC analyses. The lines are: dashed – control2 and solid – nonlinear trial.

There is a very slight spin-up in evaporation (not shown) over the forecast range considered, with diurnal cycles superposed. Global mean evaporation varies between 3.0 and 3.1 mm day^{-1} so it is rather less than global mean precipitation over this period. The trials slightly increase evaporation in the extratropics.

In the tropics total cloud (figure 13) increases for 15 hours and then declines, this seems to come mainly from convective and high cloud. In the extratropics there is a slight increase of cloud with forecast range (more marked in the northern extratropics in period 3 - 0.025 over 24 hours – not shown) apart from a slight reduction in the winter extratropics over the first few hours. The trial reduces total cloud in the tropics but has less effect in the extratropics in period 2. (In period 3 there was a reduction in the northern extratropics and an increase in the southern extratropics.)

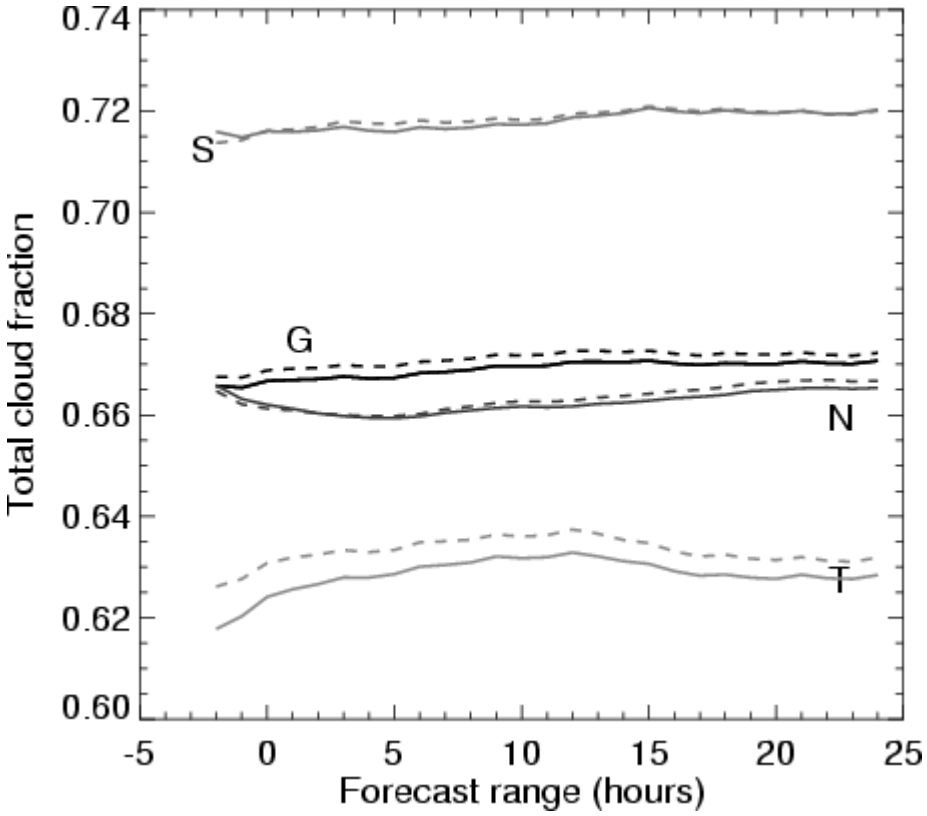

Figure 13. As fig. 12 but for total cloud. 1 corresponds to completely overcast conditions.

4. Discussion

4.1 The distribution of humidity in the atmosphere

Peixoto and Oort (1996) give a climatology of RH derived from radiosonde data (only up to 300 hPa due to the historical limitations of radiosonde humidity measurements). Their figure 4 shows zonal mean RHs of 70-80% near the surface decreasing to 30-50% at 300 hPa. The vertical decrease is fairly gradual, whereas the operational ECMWF analyses for 1986-89 (their figure 15) show a sharp decrease between 850 and 700 hPa, presumably associated with the top of the boundary layer. Mahrt (1991, section 1 and references) notes that in the real boundary layer "specific humidity is often not well mixed even when potential temperature is well mixed". Particularly in anticyclonic conditions there can be a sharp decrease in humidity at the top of the boundary layer (for example figure 4 of Andersson et al, 2005), but this will be blurred in mean values by changes in boundary layer depth and by the more gradual decrease in cyclonic conditions. The sharpness of the decrease at the top of the boundary layer (BL) has implications for data assimilation because the training data (used to represent forecast errors) have their maximum variance round the top of the model boundary layer. This is true for q as well as RH, despite the fact that actual q values are largest near the surface. The rms O-B q values shown by Dee and da Silva (2003, their fig 2) have a maximum at 925 hPa, or 1000 hPa for extratropical winter statistics. Arguably boundary layer humidity is better constrained over the ocean than over land.

The Unified Model has a model of the boundary layer (Lock et al, 2000) that can generate characteristics similar to a zeroth order model: potential temperature and specific humidity are effectively well mixed within the BL and there is a step decrease of humidity at the top of the BL. Relative humidity will increase with height within the BL, which may be capped with layer cloud as a result. The variance of q will tend to have a maximum at levels that are sometimes within the BL and sometimes within the free atmosphere. In the morning over land the BL scheme may mix too aggressively, because in reality the turbulence should take time to spin up (Lock, pers. comm. 2011).

Above the BL, especially in the tropics and sub-tropics, the "advection-condensation" model (Sherwood et al, 2006, and references) is useful. In this paradigm the specific humidity of an air parcel is conserved during diabatic subsidence, until it is resaturated by randomly entering a convective system. This "wet up, dry down" overturning is asymmetric in that the updraughts tend to be quite intense and only cover a small fraction of the total area whereas gentle descent covers a much larger area. Sherwood et al (2006) use this model to develop a distribution law for free-tropospheric RH. They note that RH distributions are very broad and look nothing like the Gaussian distribution often found for other geophysical quantities. Pierrehumbert et al (2006) discuss distributions in the extratropics. In the updraughts the more buoyant the parcel the wetter it will tend to be (positive T-q covariance) whereas in the subsidence areas stronger subsidence will tend to give warmer but drier conditions (negative T-q covariance). We use RH rather than vertical velocity to distinguish these conditions. Use of RH gives better statistical relationships, partly because the properties of a particular air parcel may be more related to vertical motion (convection) a few hours earlier rather than to the current vertical motion.

The most obvious cause of non-linearity in humidity fields is the cut-off at zero, and also at saturation if just considering water vapour. Using total water replaces the cut-off at 100% with a rather fuzzy upper bound. The importance of the zero cut-off can be overstated, in the lower troposphere typical minima are 5-10% RH, and as seen in figure 5 the humidity PDF is relatively flat across much of its range (bimodal at some levels). In data assimilation we are concerned with the distribution of humidity errors rather than the distribution of humidity itself; the two are linked even if the exact relationship isn't clear. The supersaturated part of the q_T distribution is not well known, but it seems likely that the Unified Model cuts off too quickly above saturation (especially for vapour supersaturation wrt ice). Tompkins et al (2007) introduced a parameterization for ice supersaturation into the ECMWF system**.**

4.2 Why the transform works

There are two possible reasons why the transformation gives an improvement in both forecast skill (section 3) and convergence of the minimisation (section 2.3). 1) It makes the humidity control variable more symmetric or 2) it constrains the analysis in model space to be closer to the attractor. These ideas are discussed in turn; in practice both aspects probably give benefit.

Hólm et al (2002) presented their transform as a means of making the transformed humidity error statistics more Gaussian and their figure 23 supports this (albeit it only applies to one level). In general we assume that background errors are unbiased with Gaussian errors. Prior PDFs, especially when conditioned on the background RH, are not unbiased (figure 2) and it may be that the bias is more important than the skewness and flatness about the mean (figure 7). It could be argued that showing an improved PDF when applied to the training data introduces some circularity – the training data could be divided into two subsets and one used to derive the transform and the other to test it (in practice this may not make much difference). Of course if there are deficiencies in the training data (e.g. in the treatment of cloud water in the analyses, sections 2.2 and 2.4) then these should be remedied. We briefly looked at the PDFs of humidity control variables within sample analyses. For both RH and the transformed variable these looked approximately exponential with an extra peak around zero at some levels (not shown). The extra peak probably reflects regions where there is effectively no observed humidity information.

The atmosphere, and the forecast model, have an approximate attractor and we want the analysis to stay close to the attractor (e.g. Lorenc, 2003). Clearly negative humidities represent a deviation from the attractor and penalising them (as the transform does) is desirable. (If this is the main issue then a somewhat simpler transform could be used.) There are other more subtle deviations from the attractor – resulting in the excessive precipitation early in the forecast (figure 12) for example. In the ECMWF system Andersson et al (2005, table 1) found that increased use of satellite data increased short range precipitation to excessive values. One possibility is that where background integrated humidity is too low because the mixed layer is too shallow the analysis will tend to increase the mixed layer humidity rather than its depth. In general the smaller low level humidity analysis increments (figure 11) will help keep the analysis close to the attractor (and make some of the linearisations more accurate). As described in section 3.1 smaller humidity increments produced by reducing SDs for the RH control variable gave no real benefit. Gustafsson et al (2011) treat saturation as a strict limit, whereas we don't by working with total humidity. Using total humidity seems to work quite well and allows use of observations of total humidity, the main issue is the operator used to diagnose cloud water and ice from q_T. An alternative approach (Hólm and Gong, 2011) is to have a separate cloud water analysis variable – this simplifies the observation operator but complicates the relationship with the water vapour analysis variable.

5. Summary and future work

A new humidity control variable has been introduced into the Met Office 4D-Var system via a nonlinear transform, based on previous work by Hólm et al (2002). A 'balanced' specific humidity increment is calculated from the temperature increment (as a function of background humidity), the residual 'unbalanced' specific humidity is subject to a nonlinear normalisation (a function of both background humidity and the current analysis increment). Where the background RH is small the distribution of q or RH errors is skewed and not taking this into account can result in negative analysed humidity. Using the transformed humidity variable the error distribution is reasonably symmetric even in this case and the problem of negative humidities is reduced. Throughout we work with total specific humidity, i.e. including cloud water and ice. Specific humidity increments in the lowest 100 hPa of the atmosphere are about 20% smaller, near-surface temperature increments are slightly increased. In the first few forecast timesteps after the analysis there is excessive precipitation. The humidity transform slightly reduces this, particularly in the tropics. This is probably related to slightly smaller low level humidity increments in general (and perhaps to a reduction of excessive analysed humidities by the normalisation).

Trials have been performed for three different periods, using two different baseline configurations of the Met Office global system including different training data for the representation of background errors. The results are encouraging and broadly consistent. Some forecast verification scores are improved (especially for mass fields), notably in the southern extratropics, with verification against analyses improved more than that against *in situ* observations. Impact in the northern extratropics is generally either neutral or slightly positive, impact in the tropics is mainly positive but more so for mass than for winds. Short range humidity forecasts show some improvements compared to radiosonde data, but larger improvements compared to humidity-sensitive satellite radiances. The rms fit to background is improved by about 3% for upper tropospheric humidity-sensitive channels, but by less for lower tropospheric channels. The evidence suggests that the assimilation of humidity-sensitive satellite data has been particularly improved - probably both the vertical distribution of humidity increments, and the balance between humidity and temperature increments. This is consistent with the larger improvement seen in regions that are predominantly oceanic.

The linear transform (in which the normalisation factors only depend on the background humidity) gave fairly similar results in a trial (slightly improved when measured against in situ observations, slightly worse against analyses). It is slightly less effective at alleviating the problem of negative analysed humidities. Nonlinear transform trials were run with different settings for h1, the coefficient that links temperature and humidity increments. Using a value of 0.0 (normalised q analysis) the results are similar to our default of using correlation coefficients. Using regression coefficients gave worse results, as did setting negative values of the correlation coefficients to zero. These results, and some of the other sensitivities in the system, are not fully understood. However it appears that the normalisation factor is more important than the link between temperature and humidity increments. The coefficients used in the transform are a function of vertical level but not of horizontal position – a trial of latitudinally varying coefficients gave worse results. Arguably the link between temperature and humidity increments should be different over land and sea at low levels.

The nonlinear humidity transform become operational on 20 July 2011 in the Met Office global, regional and UK systems, it represents the first nonlinearity in our model of background errors. At the same time a hybrid system in which ensemble perturbations add "errors of the day" and replace 20% of the background error covariances was introduced in the global NWP system (Clayton, et al, 2012, in preparation). The hybrid method was designed to combine perturbations transformed using the parameter transforms built into the variational method, including this humidity transform, so it inherits its properties such as reducing negative analysis values. The improved fit to certain satellite channels as in figure 10 was still quite clear when the changes were tested together. It is unlikely that the humidity transform will be developed much further, but there is almost certainly scope for other improvements to the background error representation. One aspect is whether the training data is correct in having a peak in humidity variation at the top of the boundary layer or whether this is partly due to deficiencies in the forecast model. The training data used in this paper did not properly represent errors in land surface and model parameterisation, as well as the effect of sub-grid-scale variability on cloud processes. Detailed verification against surface cloud observations (Mittermayer 2011, personal communication) show the model has little skill in forecasting partial cloud amounts; this contradicts the skill implied by our diagnostic relationship in (3), applied to the very diagonal training distribution shown in figure 3(a). This and better representations of the humidity "errors of the day" should improve the use of observations, especially satellite measurements. Comparison with radiosonde humidity profiles suggests that on occasion the analysis is over-smoothing vertical detail. Longer-term, better assimilation of clouds and hence better use of satellite imagery and soundings promises bigger improvements.
Other possible future changes include:

- The Met Office processing of satellite soundings has a separate 1D-Var step used for quality control and cloud detection (Pavelin et al, 2008), this uses log(q) to avoid negative humidities. Replacing this with the new transform and statistics would improve consistency.
- Currently radiosonde and surface humidity data are presented to the analysis as RH, some of the measurements are effectively RH (from capacitive sensors), and some are dew point. Near the surface RH has a marked diurnal variation whereas q is much more constant, we could consider converting surface humidity to q for assimilation. Some of the new aircraft measurements are of q (or mixing ratio) rather than RH.
- Geer and Bauer (2010) suggest the use of a symmetrising transform in other aspects of data assimilation.
- In the medium term assimilation of stratospheric humidity data is desirable.
- Higher resolution assimilation providing better discrimination of cloudy areas.

Acknowledgements

We have greatly benefitted from the ideas of Elías Hólm and various discussions with him. Within the Met Office Richard Renshaw, Chiara Piccolo, Marek Wlasak and Mike Thurlow have contributed in various practical ways and Brett Candy, John Edwards, John Eyre and Dale Barker amongst others have made useful comments.

7. References

Anderson SR, Graham RJ, Bader MJ. 2000. The impact of observations on Mesoscale Model forecasts of three-hourly rainfall accumulations. *Met. Apps*., **3**: 193-203

Andersson E, Bauer P, Beljaars A, Chevallier F, Hólm E, Janisková M, Kållberg P, Kelly G, Lopez P, McNally A, Moreau E, Simmons AJ, Thépaut J-N, Tompkins AM. 2005. Assimilation and Modeling of the Atmospheric Hydrological Cycle in the ECMWF Forecasting System. *Bull. Amer. Meteorol. Soc*., **86**: 387-402

Andersson E, Hólm E, Bauer P, Beljaars A, Kelly GA, McNally AP, Simmons AJ, Thépaut J-N, Tompkins AM. 2007. Analysis and forecast impact of the main humidity observing systems. *Quart. J. Roy. Met. Soc.* **133**: 1473-1485

Atkins MJ. 1974. The objective analysis of relative humidity. *Tellus*. **26**: 663-671

Bader MJ, Forbes GS, Grant JR, Lilley RBE, Waters AJ, eds. 1995. Images in weather forecasting: a practical guide for interpreting satellite and radar imagery. Cambridge University Press, pp.xxiii+499

Bengtsson L, Hodges KI, Hagemann S. 2004. Sensitivity of large-scale atmospheric analyses to humidity observations and its impact on the global water cycle and tropical and extratropical weather systems in ERA40. *Tellus*, **56A**: 202-217

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

Buck AL. 1981. New equations for computing vapour pressure and enhancement factor. *J. Appl. Meteorol*. **20**: 1527-1532

Buontempo C, Jupp A, Rennie M. 2008. Operational NWP assimilation of GPS radio occultation data *Atmos. Sci. Lett*., **9**: 129 – 133

Deblonde G, English SJ. 2003. One-dimensional variational retrievals from SSMIS-simulated observations. *J. Appl. Meteorol*. **42**: 1406-1420

Dee, DP, da Silva AM. 2003. The Choice of Variable for Atmospheric Moisture Analysis. *Mon. Wea. Rev*., **131**: 155–171

Fisher M. 2003. Background error covariance modelling *ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean*, 8-12 September 2003. Available via http://www.ecmwf.int/publications/library/do/references/list/17111

Geer AJ, Bauer P. 2010. Enhanced use of all-sky microwave observations sensitive to water vapour, cloud and precipitation. ECMWF Technical memorandum 620.

Gustafsson, N., Thorsteinsson, S., Stengel, M. and Hólm, E. 2011. Use of a nonlinear pseudo-relative humidity variable in a multivariate formulation of moisture analysis. *Quart. J. Roy. Met. Soc*. **137**: 1004-1018. doi: 10.1002/qj.813

Hólm E, Andersson E, Beljaars A, Lopez P, Mahfouf J-F, Simmons AJ, Thepaut J-N. 2002. Assimilation and Modelling of the Hydrological Cycle: ECMWF's Status and Plans. ECMWF Technical memorandum 383.

Hólm EV, Gong J, 2011: Use of cloud condensate in the background error formulation. ECMWF-JCSDA Workshop on Assimilating Satellite Observations of Clouds and Precipitation into NWP Models 15-17 June 2010, pp 111-119, available from http://www.ecmwf.int/publications/library/do/references/list/201009

Ingleby NB. 2001. The statistical structure of forecast errors and its representation in the Met Office global 3-D variational data assimilation system. *Quart. J. Roy. Met. Soc*. **127**: 209-231

Lock AP, Brown AR, Bush MR, Martin GM, Smith RNB. 2000. A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Wea. Rev* . **128**: 3187-3199 doi: 10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2

Lorenc AC, Tibaldi S. 1980. The treatment of humidity in ECMWF's data assimilation scheme. Atmospheric Water Vapour ed. A. Deepak T. Wilkinson and L.H. Rubnke, 497-512 Academic Press, New York.

Lorenc AC. 2003. Modelling of error covariances by four-dimensional variational data assimilation. *Quart. J. Roy. Met. Soc.*, **129**: 3167-3182

Lorenc AC. 2007. A study of o-b monitoring statistics from radiosondes, composited for low-level cloud layers. Forecasting Research Technical Report No. 504. Available at http://www.metoffice.gov.uk/archive/forecasting-research-technical-report-504

Lorenc AC, Barker D, Bell RS, Macpherson B, Maycock AJ. 1996. On the use of radiosonde humidity observations in mid-latitude NWP. *Meteorol. Atmos. Phys*., **60**: 3-17

Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW. 2000. The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Met. Soc*. **126**: 2991-3012

Macpherson B, Wright BJ, Hand WH, Maycock AJ. 1996. The impact of MOPS moisture data in the UK Meteorological Office mesoscale data assimilation scheme. *Mon. Wea. Rev*., **124**: 1746-1766

Mahrt L. 1991. Boundary-layer moisture regimes. *Quart. J. Roy. Met. Soc*., **117**: 151-176

Milton SF, Earnshaw P. 2007. Evaluation of surface water and energy cycles in the Met Office global NWP model using CEOP data. *J. Meteorol. Soc. Japan*. **85A**: 43-72

Montmerle T, Berre L. 2010. Diagnosis and use of heterogeneous background error covariances at mesoscale. *Q. J. Roy. Meteorol. Soc*., **136**: 1408-1420

Morcrette CJ, O'Connor EJ, Petch JC. 2011. Evaluation of two cloud parametrization schemes using ARM and Cloud-Net observations. *Q. J. Roy. Meteorol. Soc* doi: 10.1002/qj.969

Pavelin EG, English SJ, Eyre JR. 2008. The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction. *Q. J. Roy. Meteorol. Soc*., **134**, 739-751

Peixoto JP, Oort AH. 1996. The climatology of relative humidity in the atmosphere. *J. Climate*, **9**: 3443-3463

Pierrehumbert RT, Brogniez H, Roca R. 2006. On the relative humidity of the atmosphere. The Global Circulation of the Atmosphere, T. Schneider and A. Sobel, Eds., Princeton.

Rabier F, Thépaut J-N, Courtier P. 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Q. J. Roy. Meteorol. Soc*., **124**: 1861-1887. doi: 10.1002/qj.49712455005

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ. 2007. The Met Office Global 4D-Var Scheme. *Quart. J. Roy. Met. Soc*., **133**: 347-362

Read WG, and coauthors. 2007. Aura Microwave Limb Sounder upper tropospheric and lower stratospheric H2O and relative humidity with respect to ice validation. *J. Geophys. Res*. **112**, D24S35, doi:10.1029/2007JD008752

Riishøjgaard LP. 1998. A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus*, **50A**: 42-57

Sharpe M. 2007. Incrementing liquid and frozen cloud water and grid-box cloud fractions for VAR observation operators and UM initialisation. VAR Scientific Documentation Paper 61, Met Office

Sherwood SC, Kursinski ER, Read WG. 2006: A distribution law for free-tropospheric relative humidity *J. Climate*, **19**: 6267-6277

Smith RNB. 1990. A scheme for predicting cloud layers and their water content in a general circulation model. *Quart. J. Roy. Met. Soc*., **116**: 435-460

Sun B, Reale A, Seidel DJ, Hunt DC. 2010: Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics, J. Geophys. Res., 115, D23104, doi:10.1029/2010JD014457.

Thornton HE, Jackson DR, Bekki S, Bormann N, Errera Q, Geer AJ, Lahoz WA, Rharmili S. 2009. The ASSET intercomparison of stratosphere and lower mesosphere humidity analyses. *Atmos. Chem. Phys*., **9**: 995-1016

Tompkins AM, Gierens K, Radel G. 2007. Ice supersaturation in the ECMWF integrated forecast system *Quart. J. Roy. Met. Soc*., **133**: 53-63

Wilson DR, Bushell AC, Kerr-Munslow AM, Price JD, Morcrette CJ, 2008. PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Quart. J. Roy. Met. Soc*., **134**: 2093-2107

Appendix 1 Reduction of upper tropospheric dry bias (November 2009)

In recent years the Met Office global analysis/forecast system had an upper tropospheric dry bias as evidenced by comparison with IASI observations and other operational centres (Hilton et al, 2012, Milton et al, 2008). In November 2009, along with the move to 70 levels, the dry bias was addressed by allowing use of humidity data from some radiosonde types higher in the atmosphere and by adjusting the resetting of stratospheric humidity. Previously no radiosonde humidities were used if the temperature was below -40°C. For Vaisala RS80 sondes this was relaxed to -60°C and for RS90/92 -80°C was used based on Nash (2003) and further investigations. In the model stratosphere humidity values from the data assimilation were reset if they fall outside the climatological range, taken as $1 < q < 3$ parts per million and RH < 10%. This was applied above the 2.5 PVU surface (provided that it lay between 400 and 100 hPa). From November 2009 no humidity increments are applied above the 5 PVU surface. These changes increased extratropical RH at some levels by up to 20% RH, and at T+24 the forecast at 300 and 250 hPa changed from being drier than the radiosondes to being wetter. In daytime most radiosondes are too dry in the upper troposphere, so the mean analysis/forecast RH at these levels is probably about right, or possibly slightly too high.

Hilton, F.I., S.M. Newman and A.D. Collard, 2012: Identification of NWP humidity biases using high-peaking water vapour channels from IASI. Atmospheric Science Letters, 13: 73–78. doi: 10.1002/asl.366

Milton, S.F., F. Hilton, P. Earnshaw, S. Newman and S. Clough, 2008: Dry bias in Met Office upper tropospheric humidities: Analysis or Model problem? *Unpublished manuscript*

Nash J, 2003: Review of test results on the accuracy of radiosonde relative humidity sensors. ECMWF/GEWEX Workshop on Humidity Analysis, July 2002. pp 117-123