Numerical Weather Prediction

A study of o-b monitoring statistics from radiosondes, composited for low-level cloud layers.

Forecasting Research Technical Report No. 504
Met R & D Technical Report No. 504

Andrew C. Lorenc

email: nwp_publications@metoffice.gov.uk

A study of o-b monitoring statistics from radiosondes, composited for low-level cloud layers.

Andrew C. Lorenc. September 2007.

Abstract

A common complaint in forecast case studies is that the assimilation misrepresents inversions and stratocumulus layers, because of inappropriate background error covariances. This study looked at statistics from a large number of such cases, to see if any systematic patterns were apparent. It used the operational global archive of model soundings at radiosonde positions, with the observations pre-processed to the model levels. Soundings were classified using a simple algorithm based on relative humidity and stability, and composited according to cloud layers in the model forecast background and/or the observation. Results are interpreted in terms of recommendations for the development of the Met Office's operational variational assimilation scheme (4D-Var). The main conclusions are:

- It is common for the model to have a plausible cloudy inversion structure in the wrong place. This leads to error distributions which are non-Gaussian, with the mean unrealistically smooth and apparently biased. This is a fundamental problem of minimum-variance best-estimate methods; it cannot be cured by better covariances within such methods. The 4D-Var fitting of a model forecast to observations might help.
- The background error variances used by 4D-Var are about right for average conditions. But they should be about doubled near cloud-topped inversions.
- The background vertical correlations used by 4D-Var are about right for humidity, but too broad for temperature, for average conditions. The correlations are too large across cloud-topped inversions.
- The operational boosting of sonde humidities (introduced to correct an apparent bias in 1996) should be stopped.
- The 4D-Var assumption that temperature and relative humidity background errors are not correlated is not usually justified. A more complicated statistical relationship is needed, including a nonlinear component to take account of the non-Gaussian distributions near saturation.
- The model's cloud layers are on average half a model level lower than observed.

Document History

March 2007	Study started.
25 July 2007	Preliminary version circulated for comment.
6 Aug 2007	Draft FRTR passed for review by Adrian Semple and others.
7 Sep 2007	Revised version, addressing minor review comments, released as Technical Report 504. (Approved for issue by Andrew Lorenc on advice from Adrian Semple.)
18 Sep 2007	Met R&D Tech. Rep. added to title page and internal reference links removed, for publication on Met Office web site.

A study of o-b monitoring statistics from radiosondes, composited for low-level cloud layers.

Andrew C. Lorenc. September 2007.

1 Introduction

Met Office NWP forecast users have as their highest priority problem "*Poor assimilation of temperature and dewpoint, primarily in the lowest 100mb (items 10.1, 15.2). This is dramatically reducing the utility of our high resolution model suite*" (NWP Problem Group, 16th meeting, 18/05/07). Probably the most common complaint is the misrepresentation of inversions and stratocumulus layers. Such cases can have very high impact, for instance, in December 2006, 3 days of poor visibility at Heathrow caused over a thousand flights to be cancelled, disrupting hundreds of thousands of Christmas travel plans.

Case study investigations usually attribute the failure to fit the inversion structure in radiosonde soundings to inappropriate background error covariances in the variational assimilation system (e.g. Semple 2006). This study set out to look at statistics from a large number of such cases, to see if any systematic patterns were apparent. The data source was the MetDB archive of radiosonde soundings with the corresponding model background and analysis values, from our 50-level global NWP assimilation.

The study looks mainly at observation minus model background differences ("o-b"), without trying to partition them into observational and model errors. Hollingsworth and Lonnberg (1996) did partition similar statistics by using a dense homogeneous network of sounding over North America and by making the assumptions that observation errors had a constant vertical covariance and zero horizontal correlation. In sub-samples of soundings based on the presence of layer cloud, this technique is not appropriate: observational errors are probably affected by the cloud, and a dense network with the same cloud is not available. So this study has only a limited, descriptive, ambition, in presenting o-b statistics which reflect the sum of these error components. Nevertheless they present a useful upper-bound on each component, as well as a description of the relative changes when there are cloud layers.

2 Data and Processing

This study used merged model-level values from the global assimilation, retrieved from the MetDB database. Unless stated otherwise, all 740328 available global soundings for the 50-layer version of the model were used, from December 2005 to July 2007 (a few dates were missing from the archive). The global dataset was chosen in order to get statistically significant samples in the sub-classes studied, and also to simplify interpretation of the statistics in terms of their impact on the variational assimilation (VAR: Lorenc *et al*. 2000, Rawlins *et al*. 2007). The limited-area configurations, such as the NAE, supplement VAR

by “MOPS nudging” of cloud data (Macpherson *et al*. 1996); this prevents us attributing observed analysis increments to the variational scheme.

The model’s grid point values should represent reality filtered to the resolution of the model, so the Met Office assimilation pre-processes radiosonde soundings to produce average values for each model layer (OSDP5). An exception is relative humidity (RH), because such averaging reduces the occurrence of near-saturation values, and hence of diagnosed cloud amounts; instead linear interpolation is used (Lorenc *et al*. 1996). As these are the data presented to the assimilation, representing the “best achievable” with this model vertical resolution, they were used in this study. The effect of the averaging can be seen by comparing the red curves on the left of Figure 1 with the black curves on the right; these show the same sounding unprocessed and using the layer-averaged data of this study. (Note that the plots use different packages, and the other curves show different model soundings.)

As well as this averaging, the sonde RH values are “boosted” near saturation, before assimilation. OSDP5 equation (5.14) says the boosting of values near saturation is given for the global model by

$\Delta RH = RHmaxboost\ (RH-80)/(90-80)$ for $80 < RH \leq 90$

$\Delta RH = RHmaxboost$ for $90 < RH \leq 95$

$\Delta RH = RHmaxboost\ (100-RH)/(100-95)$ for $95 < RH < 100$

OSDP5 says that RHmaxboost=3%, but the actual value used for the global model is 4.6%. The bulk of the results in this paper are for the observations without the boost. The above equation was inverted to recover the original values from the archived boosted values. The historical reason and (bad) current effect of the boost are discussed in section 5.

Of course the radiosonde sounding must not be taken as truth. In particular the dew point maximum just above the inversion in Figure 1 left is probably a spurious artefact of moistening of the humidity sensor. The pre-processed sounding does not show it, but it is capable of showing the inversion, at model level 5, albeit with an increase from the moist adiabatic lapse rate of the original sounding, caused by the inclusion of some above-inversion temperature data in the layer 5 average.

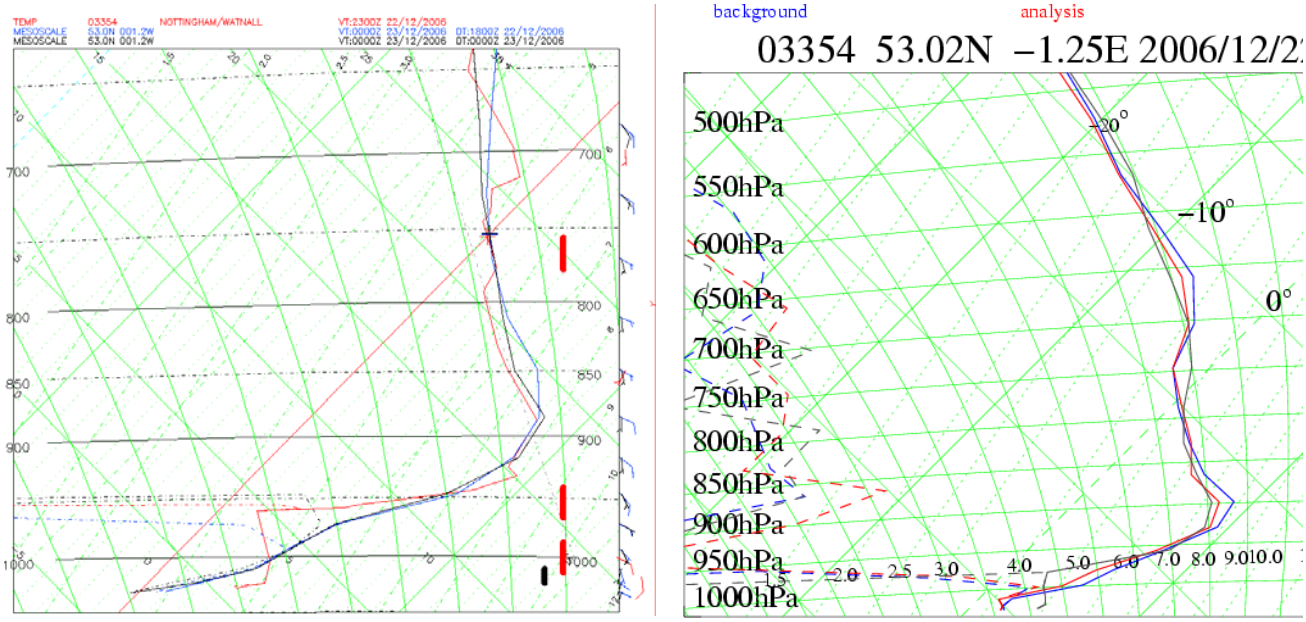

Figure 1 (LEFT) Tephigram plot of sounding (red) and mesoscale model background (blue) and analysis (black) as plotted and reported as a problem by forecaster Tim Hewson. The case is at 00Z on 23 Dec 2006, during the pre-Christmas disruption at

Heathrow due to poor visibility. (RIGHT) Tephigram plot of same layer-mean sounding (black) and global model background (blue) and analysis (red), as used in this study.

In order to classify cases by the presence of cloud layers, a test for a layer top was devised which could be applied to the layer-averaged data of both model and sondes. The following all had to be true to classify level k as a cloud top:

- necessary values not missing and not flagged;
- rh(k)>rhc;
- rh(k+1)<rhi OR rh(k+2)<rhi;
- $\frac{\partial\theta}{\partial z}$(evaluated from k to k+1)> dthdzi;
- a cloud top NOT detected at level k+1.

An additional criterion separated "stratocumulus (SC)" clouds as those layers with evidence of vertical mixing:

- $\frac{\partial\theta_w}{\partial z}$(evaluated from k-1 to k)< dthwdzc;

The limits were set to allow some leeway for the effect of averaging, by trial and error:

rhc = 97.0 ; limit for cloud indicator
rhi = 80.0 ; limit for inversion indicator
dthwdzc = 0.0032 ; limit for cloud indicator
dthdzi = 0.004 ; limit for inversion indicator

Note that rhc is larger than the model's rh_crit (~80%), the value at which the model starts diagnosing partial cloud cover, since I was interested mainly in solid cloud sheets in sonde ascents. In the model, rhc=97.0 only selects grid-boxes with large cloud fractions. The effect of lower values on the frequency of cloud is discussed in sections 4 and 6. They did not have a significant impact on other aspects like the correlations.

Cloud layers which failed the final test were called "stable" clouds, and statistics were also calculated for the combined sample of all layer clouds. In the example of figure 1, the observed profile was classified as SC with top at level 5, while the background was classified as "stable" cloud with top at level 3.
The extra test on rh 2 levels above the cloud was added to avoid the effects of spurious wetting of the humidity sensor.

Three variables were studied: temperature, relative humidity, and pseudo-relative humidity (i.e. 100 times the specific humidity divided by the saturated specific humidity at the background temperature and pressure). Because high level humidity observations from sondes are either missing or unreliable, plots here only show the lower 23 levels. The average pressures for these levels were: 971, 966, 956, 942, 926, 905, 879, 852, 820, 786, 748, 710, 671, 629, 586, 544, 500, 458, 416, 375, 335, 297, 261, 228 hPa.

So-called covariances were calculated as the average product of model minus observed values. N.B. mean errors were not subtracted first, so these are not true covariances. However they are arguably the correct values to consider in an assimilation which takes no account of mean errors.
So-called correlations were calculated by dividing the covariances by the square root of the product of the appropriate so-called variances. (The variances were based on all values for each level and variable, whereas the covariances were estimated as averages of those cases where values for both variables were available, so it is possible to get so-called correlations greater than 1.0.)

3 Background minus observation statistics

In this section I concentrate on the combined results for layer cloud; those for the sub-classifications of SC and stable were little different. The legends in the figures below describe the classification of the sample and its size, e.g. 5Clb is the sample with cloud top at level 5 in the background. There is a choice as to which humidity variable to use in assimilation; Dee and da Silva (2003) discuss the obvious candidates relative humidity (RH) and specific humidity (q). Because of the wide range of values of q, I follow them and show instead statistics for "pseudo-RH" (defined above); this has the same correlations as q but variances similar to those of RH.
I plot b-o rather than o-b since it is common to interpret these biases as "model error".

3.1 Mean errors

It is apparent from Figure 2 and Figure 3 that mean errors are not negligible in the presence of layer cloud. Figure 4 mean errors are smaller, showing that much of the bias error comes from errors in the presence or not of cloud, rather than its structure. These biases are a significant part of the total RMS errors (below). This contradicts an assumption of our variational assimilation scheme, that error distributions are unbiased and Gaussian, so they can be characterised by covariances alone. When there is an inversion and cloud layer in the background, it is much more likely that the true values have a weaker, or different level inversion, than that the true inversion is at that level, but stronger. So if a cloud layer is detected in the background, and if we want to minimise the RMS errors from the true sounding, we should correct the background by the biases plotted in Figure 2. This would give a smoother, probably unrealistic sounding, but one with lower RMS errors. (This bias correction could by independent of whether a radiosonde sounding exists at that location.)
This is an example of a more general weakness of our variational assimilation: where there are physical structure like inversions, or fronts, or cyclones, or convective cells, which tend to have position errors but otherwise similar structures, then error distributions in normal coordinates are non-Gaussian. The theoretical basis of most data assimilation methods, which rely on finding a least-squares best fit, using covariances to characterize PDFs, breaks down. This was discussed for convective scales in Lorenc and Payne (2007). The problem can be ameliorated by adding an additional constraint that the "best" estimate must be capable of being generated by the model in "spun-up" forecast mode. As long as the model represents the physical processes generating inversions, fronts etc., then this will give a preference for these model generated structures. However this approach reduces the ability of the assimilation to correct model errors in these structures – a serious problem for the layer clouds we are considering.

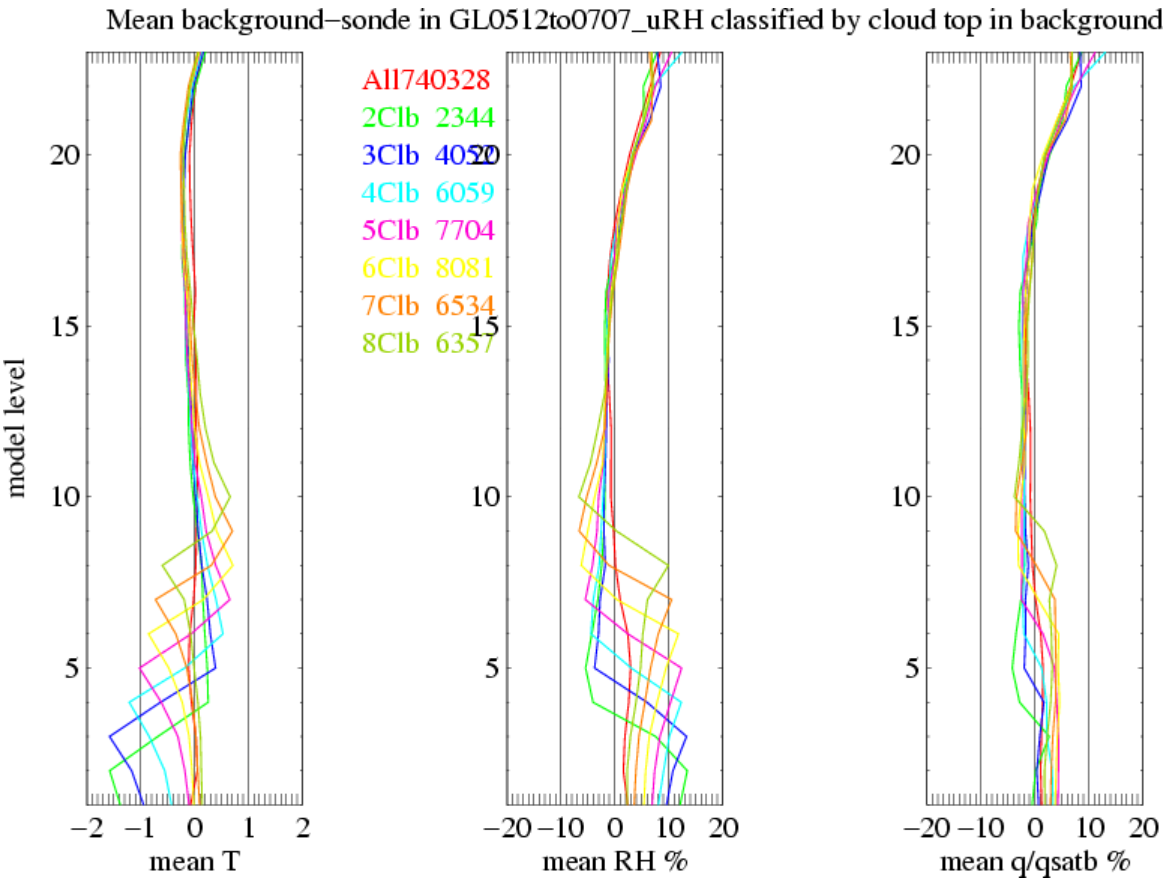

Figure 2 Mean radiosonde observation minus global model background, for T, RH and q/qsatb. The composites are classified according to the layer-cloud top in the background. The numbers in each class are shown in the legends.

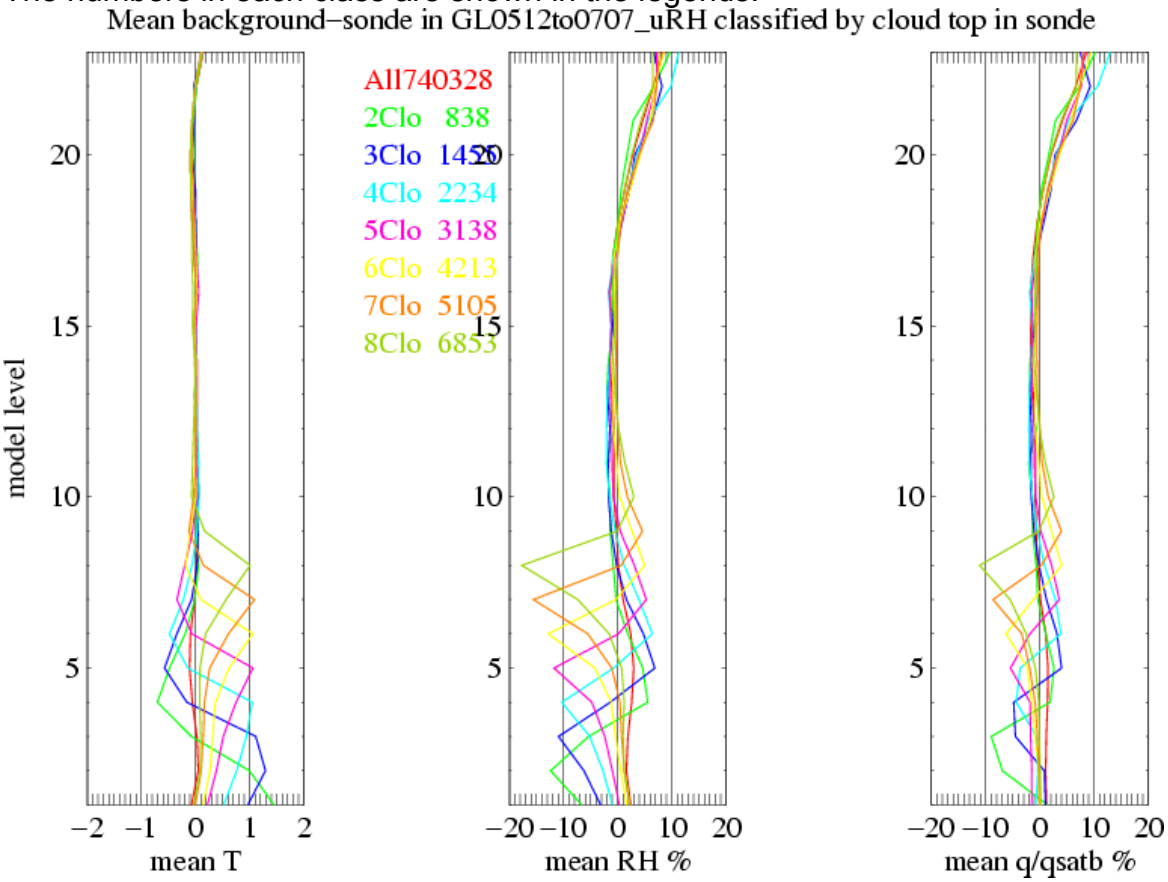

Figure 3 As figure 2, classified by layer cloud top in the sonde.

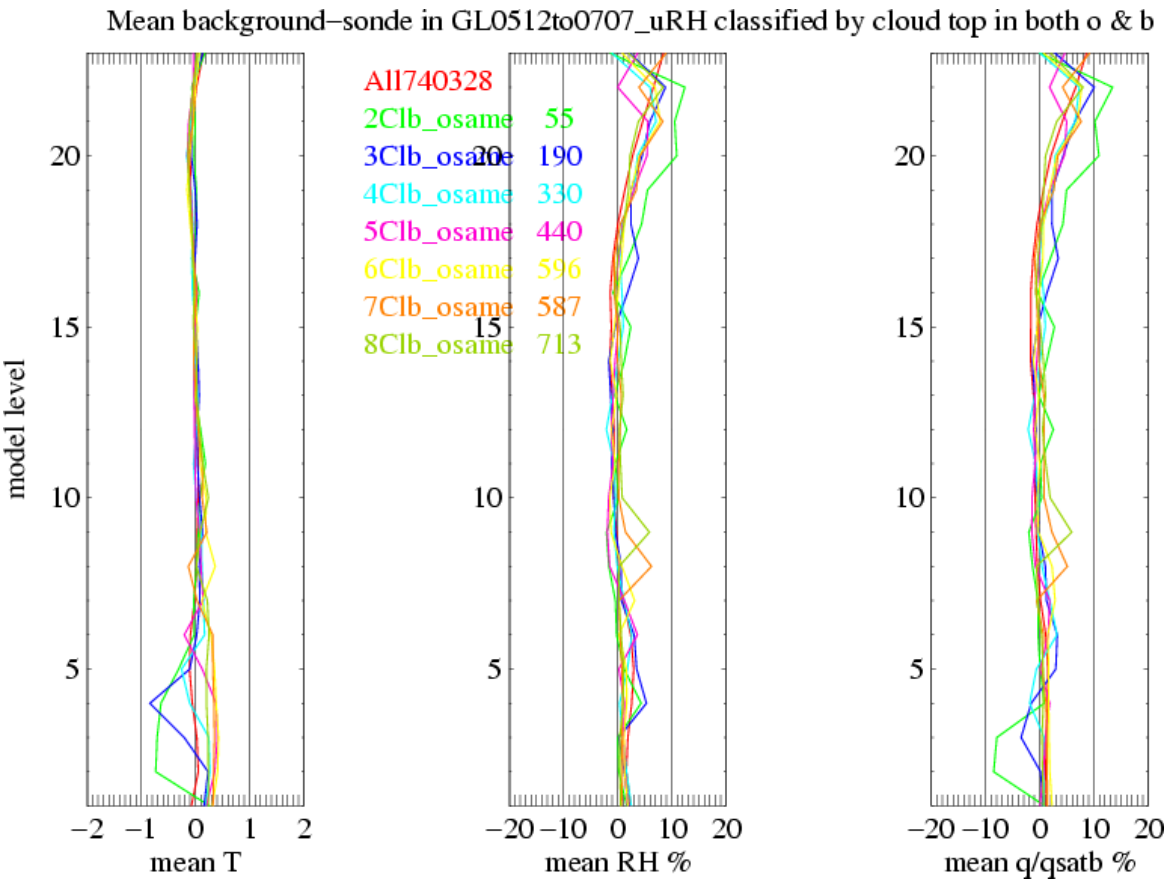

Figure 4 As figure 2, for classes with identical layer cloud tops in background and sonde.

3.2 RMS Errors

Figure 5, Figure 6 and Figure 7 show the root mean square errors corresponding to the mean errors discussed in section 3.1. In most cases, there is a maximum at or near the cloud top. This is not entirely due to the large mean errors, since similar plots of the s.d. (not shown), which excludes this term, still have peaks about a factor 1.4 greater than the red “all data” curves.

It is the reciprocal of the background error variance which determines the weight given to the background in data assimilation, so as long as this increase is not due to observational error, this implies that the weight given to the background near cloud layers should be about half normal.

The actual background errors used in current VAR, for a mid-latitude winter case such as Figure 1, and the assumed observational errors, are shown in the dotted curves on these figures. Ignoring for the moment correlations, these values squared should determine the relative weights given to the observation and the background in VAR. We see from Figure 1 (right) that this is about the case: in the free troposphere the weights are about equal, and so the analysis curve (red) lies between the others. Near the surface the observational error is larger, so the background gets more weight. This simple view is modified by the assumed vertical correlation of background errors (shown in the next section). These have the effect of “preferring” corrections to the background which have a similar structure. So observational information which tries to add detail near the cloud layer gets less weight than implied by the variances.

The exception to the remark that RMS errors are larger near cloud layers is seen in Figure 7: when cloud is present in both background and sonde, both have relative humidity near 100% so errors will be small. The figure shows that even with this match of clouds,

temperature errors are large, hence so too are specific humidity errors. And the counts in this and previous figures show that a good match of model and observed cloud is sufficiently rare that their low RH errors do not dominate in Figure 5 and Figure 6.

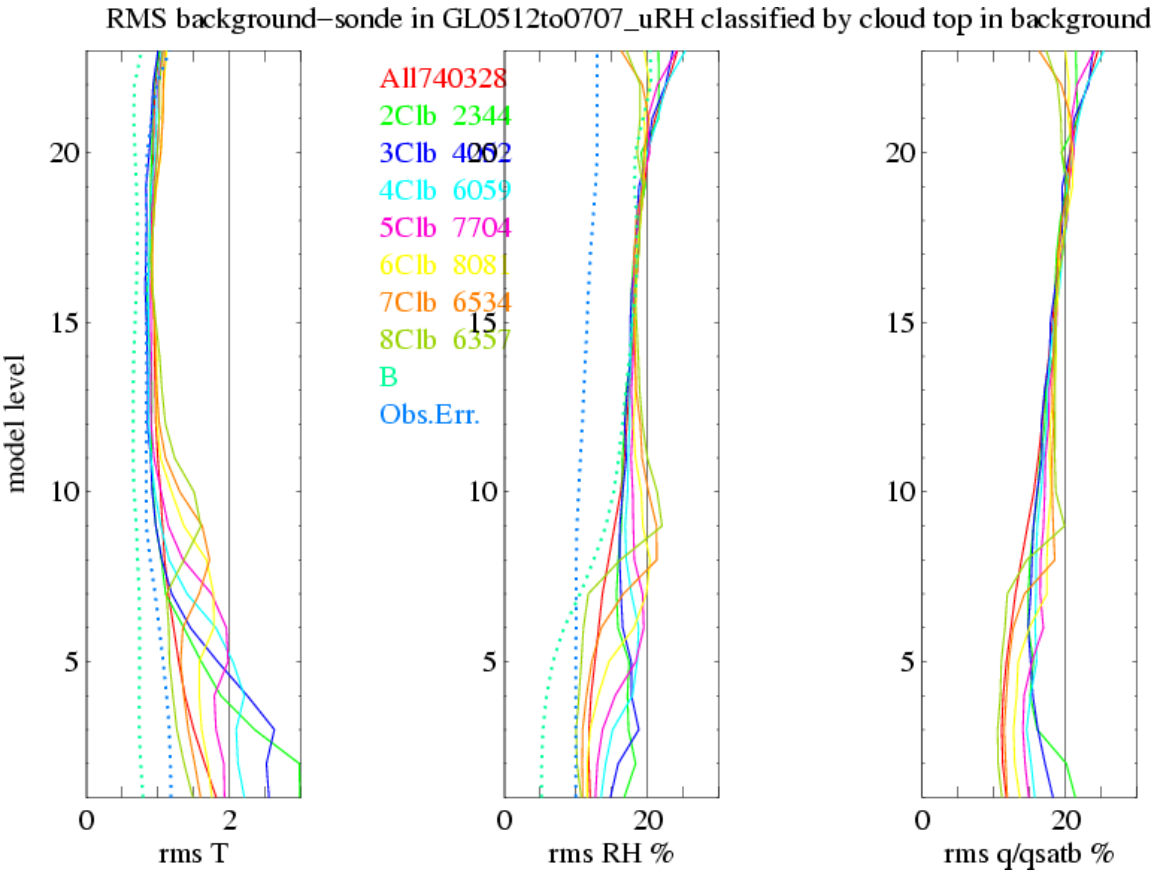

Figure 5 As figure 2, for RMS radiosonde observation minus global model background. Also shown for T and RH are the assumed background errors (cyan dots) and observation errors (blue dots), as use in the variational assimilation.

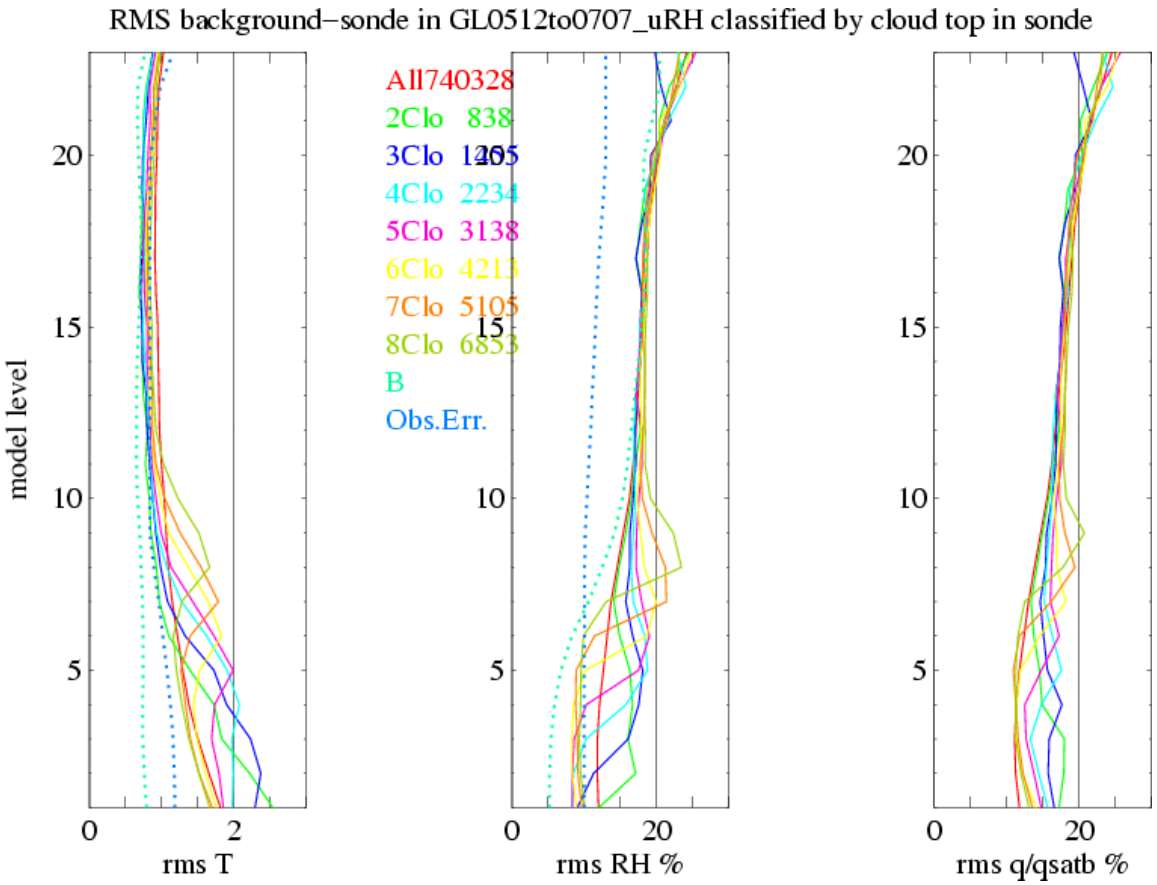

Figure 6 As figure 5, classified by cloud top level in sonde.

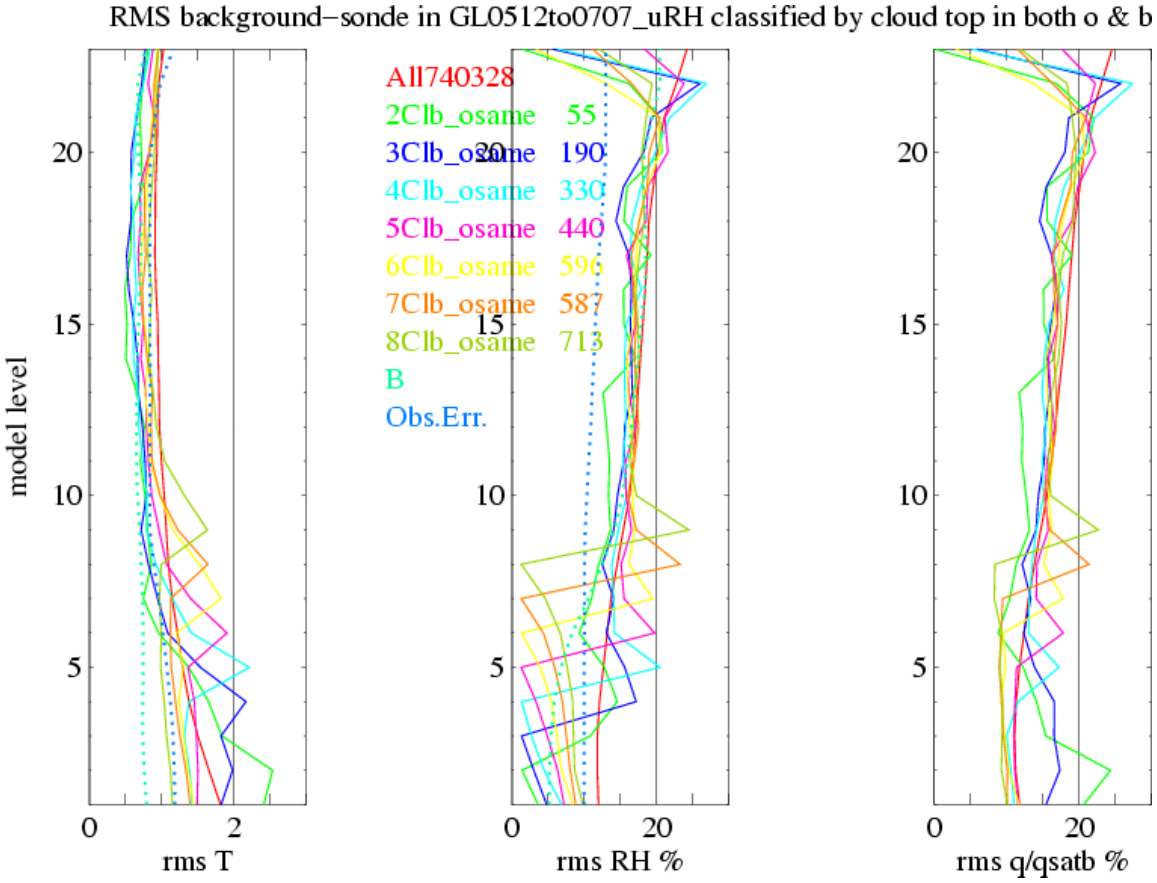

Figure 7 As figure 5, for classes with identical layer cloud tops in background and sonde.

3.3 Vertical correlations

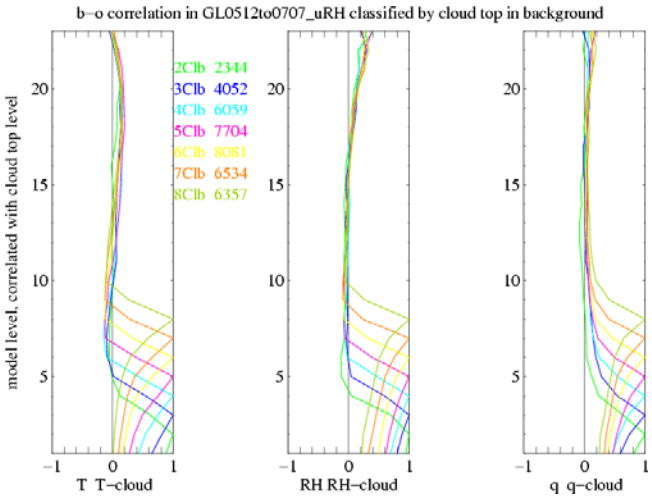

Figure 8 o-b correlations with the cloud-top level in background.

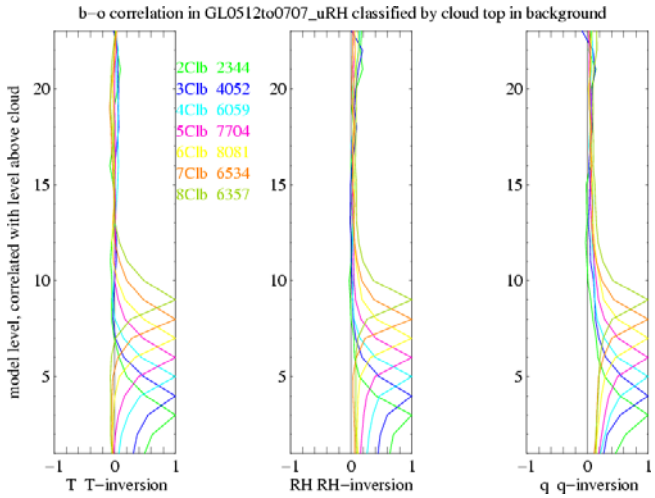

Figure 9 As Figure 8, for correlations with the layer above the cloud top.

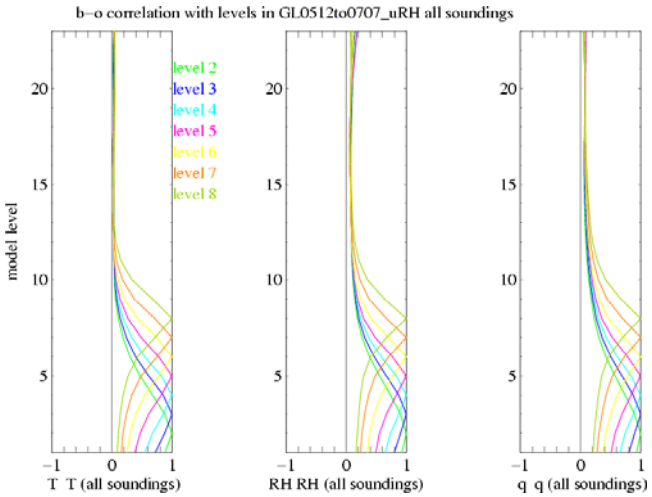

Figure 10 As Figure 8, for o-b correlations in all soundings, with the same levels and colours as Figure 8.

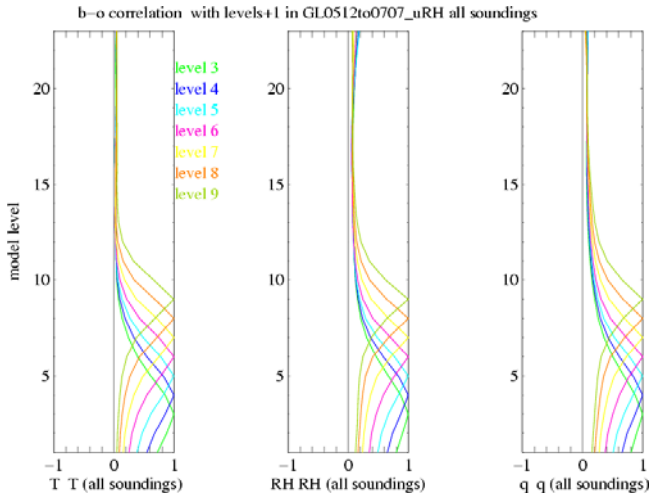

Figure 11 As Figure 9, for o-b correlations in all soundings, with the same levels and colours as Figure 9.

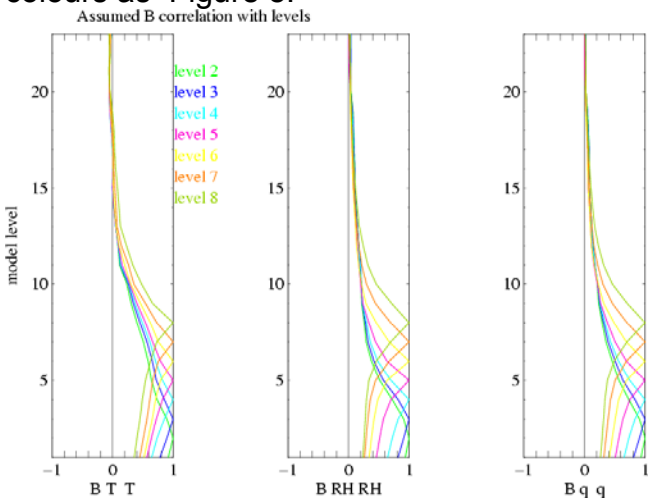

Figure 12 As Figure 8, for the assumed VAR correlations, with the same levels and colours as Figure 8.

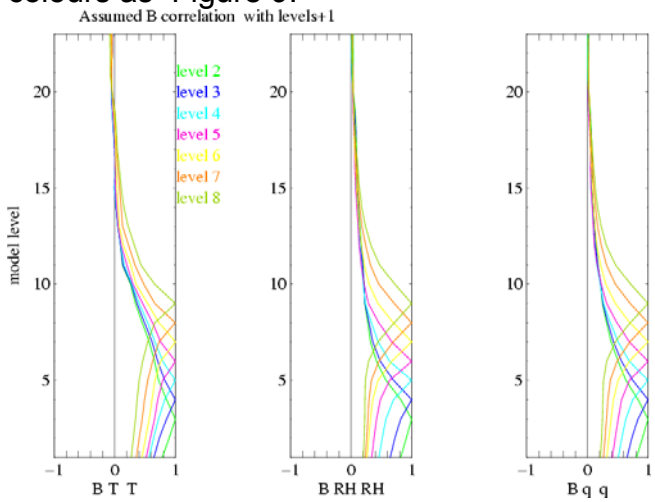

Figure 13 As Figure 9, for the assumed VAR correlations, with the same levels and colours as Figure 9.

Figure 8 and Figure 9 show the composite vertical correlations with errors at the cloud top, and in the inversion above the cloud top, respectively. For comparison, Figure 10 and Figure 11 show the "all soundings" correlations and Figure 12 and Figure 13 show the assumed VAR correlations, for the same levels plotted in the same way. They show a very small vertical correlation of levels in the cloud with levels above the cloud (Figure 8), and a quite small correlation of levels above the cloud with levels in and below the cloud (Figure 9). Of course the average behaviour (Figure 10 and Figure 11) does not show these reduced correlations across the cloud-top inversion. Comparing Figure 12 and Figure 13 with Figure 10 and Figure 11, we see that the assumed VAR correlations are not too bad a representation of the average behaviour for RH, but somewhat too broad for temperature, especially with too large correlations with lower levels.

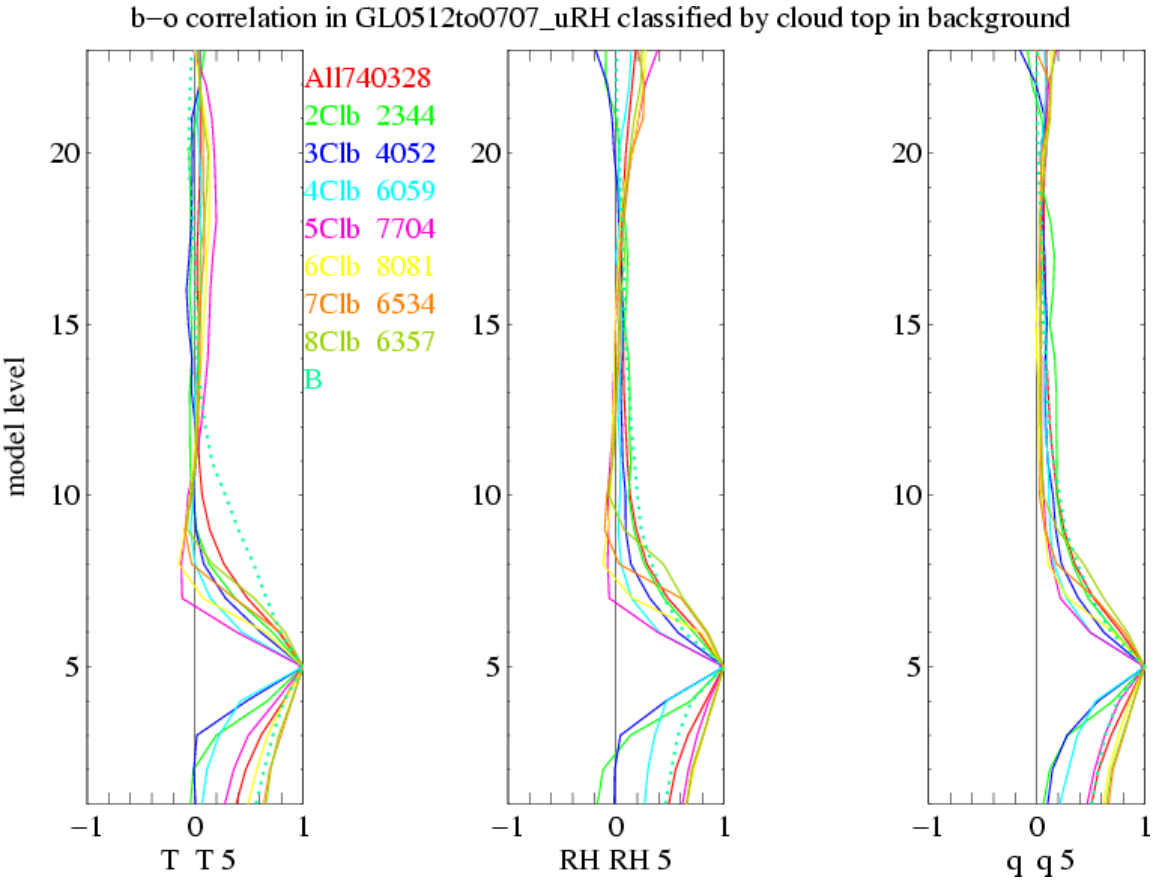

Figure 14 Vertical o-b correlation with level 5, for composites with background cloud tops at different levels. The dotted line shows the assumed VAR background error correlation.

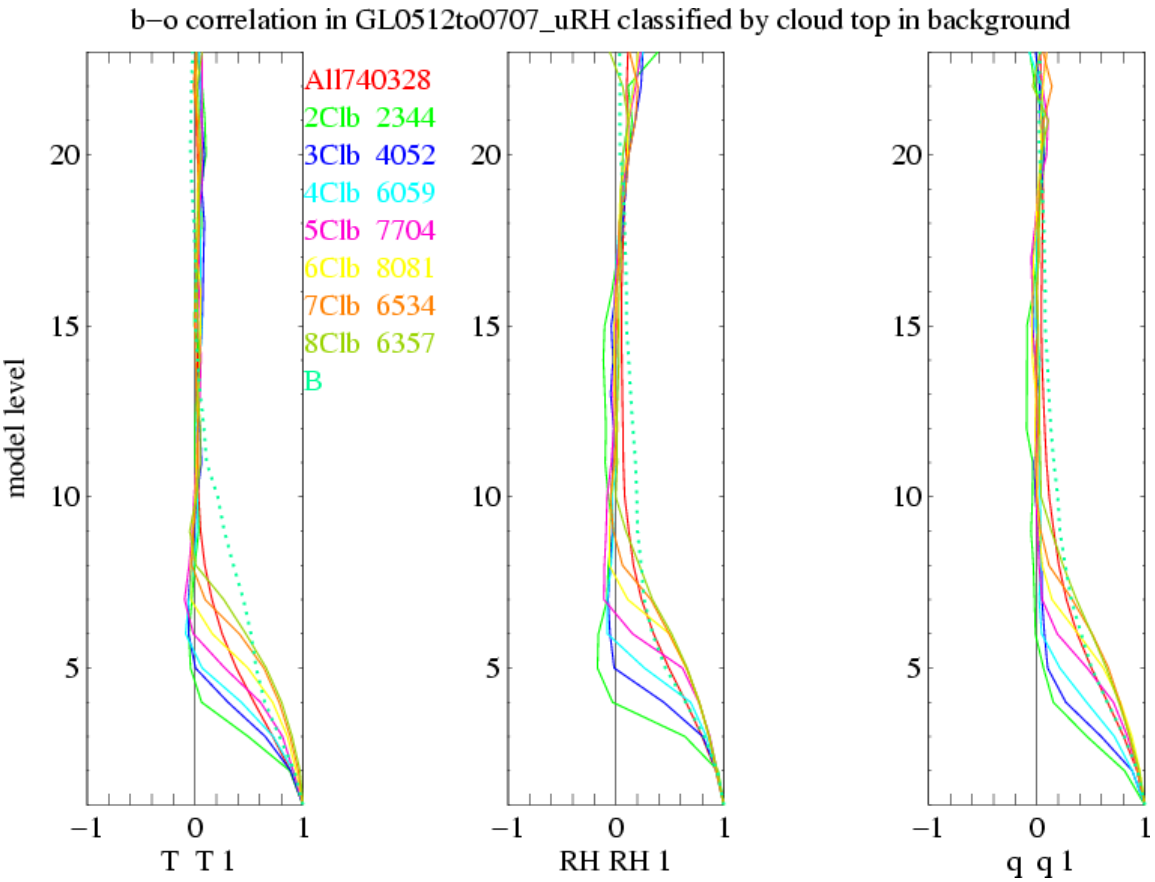

Figure 15 As Figure 14, for correlations with level 1.

Figure 14 and Figure 15 allow a more detailed comparison. The dotted "VAR" correlations are very similar to the red "all soundings" correlations for RH, but too broad for T. In Figure 14 cases with cloud layers below level 5, correlations of level 5 with levels in the boundary layer are nearly zero, while in cases with a cloud layer above, correlations of level 5 with other levels in the boundary layer are significant. Figure 15 clearly shows that vertical correlations with level 1 drop to zero just above the cloudy boundary layer tops. This correlation is important, since it determines the optimal vertical spreading of information from near-surface observations. (Synop observations of T and RH are not currently used in the global model, but developments to include them are underway.)

In the VAR system, vertical correlations are specified as a function of model levels (with some latitudinal dependence). The scheme cannot represent the reduced correlations across the cloud-top inversion. One method other schemes use to do this is to specify vertical correlations as a function of potential temperature. The large potential temperature change across an inversion makes the levels there further away in potential temperature coordinates, giving lower covariances. One way to get this effect in a VAR scheme like ours, constructed using a vertical transformation operator, is to insert an additional transform which interpolates to different levels whose spacing reflects their potential temperature differences. We can get an impression whether potential temperature coordinates give more uniform correlations, by replotting Figure 14 and Figure 15 with this vertical coordinate. For each classification the mean potential temperature differences of each level from that of level 5 and level 1 were calculated, and used as vertical coordinate in Figure 16 and Figure 17. The spread of correlations between the different curves is reduced, but not removed. If anything, the transformation to potential temperature goes too far, with the "distance" across low-level inversions now too large for the observed correlations. (The interpolations in such a transform would introduce extra

smoothing, which would reduce the ability to fit grid-scale detail. It remains to be tested whether this outweighs the better spread of correlations.)

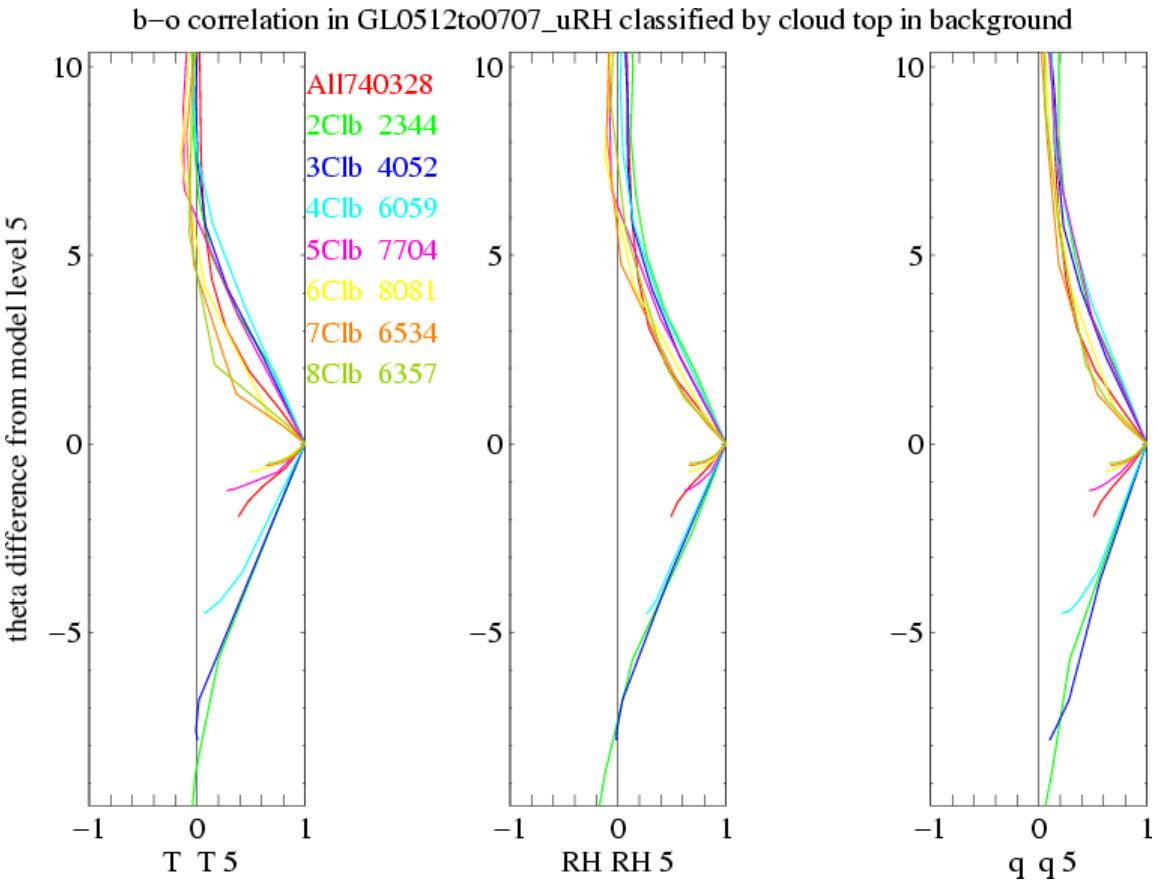

Figure 16 As Figure 14, with potential temperature as a vertical coordinate.

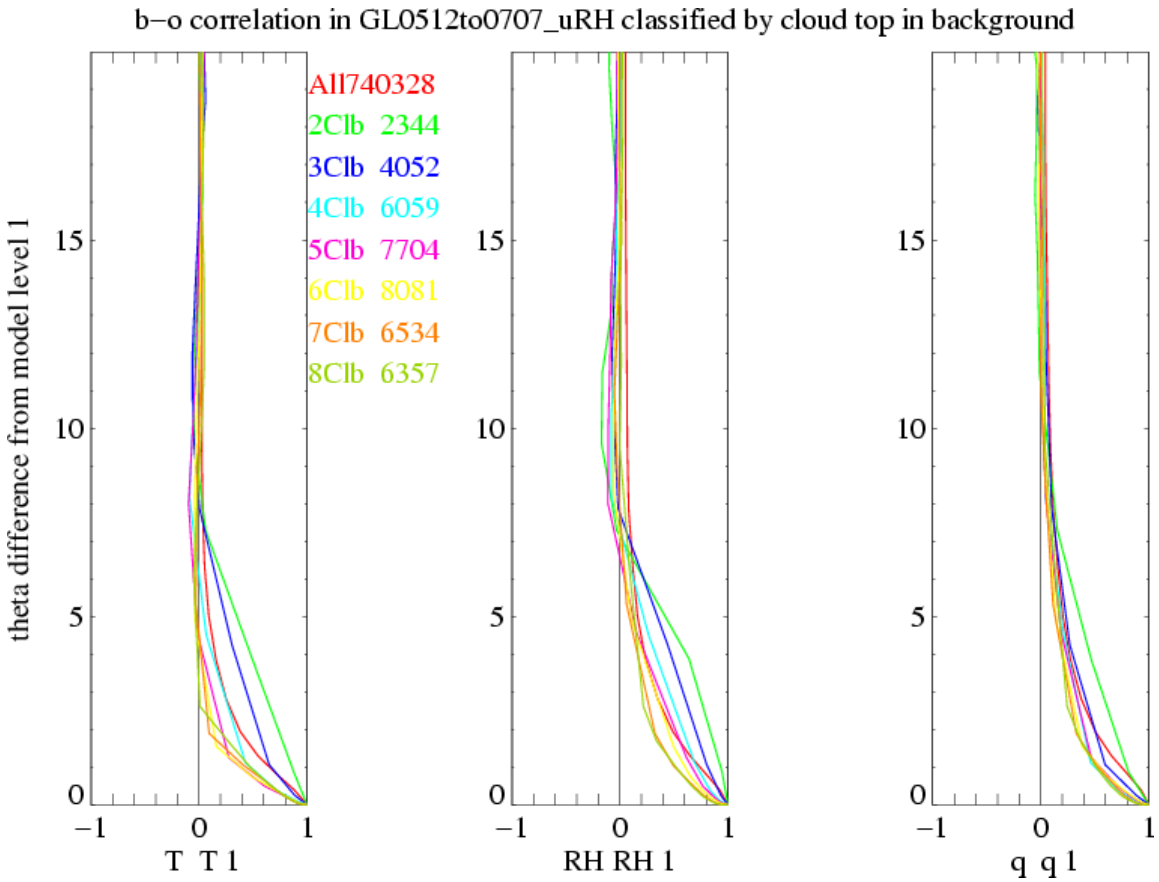

Figure 17 As Figure 15, with potential temperature as a vertical coordinate.

3.4 Cross-correlations

In 1994 the Analysis Correction data assimilation scheme was changed to assume that relative humidity errors (rather than specific humidity errors) are independent of temperature. This change was made partly to address forecast bias (Lorenc *et al*. 1996), and partly based on the argument that in important, cloudy, situations, relative humidity errors are more likely to be independent. This assumption was carried over into the VAR system (Lorenc *et al*. 2000). Figure 18 demonstrates that in situations with cloud layers correctly positioned in model and observations, this assumption might be justified: the T-RH correlations at the level of the cloud in each classification are about -0.2: quite near zero. The RH errors are small in these cases (Figure 7), so T-q correlations are near 1. However it is more common that cloud layers are incorrectly positioned. Stratified solely on the background cloud layers, T-RH correlations are not near zero in the cloud layers, and not very different from the all-soundings correlations (Figure 19). A similar conclusion applies to the mean errors shown in Figure 2 and Figure 3: the mean errors in T in cloud layers are not matched by a mean error in q, leading to corresponding errors in RH with opposite sign.

In the free troposphere, above the inversions, we see a slight negative correlation between q and T. This is probably because incorrect vertical motion is a significant cause of both T and q errors in this region; excessive descent will usually increase temperatures and decrease q. At upper levels, and in the stratosphere, there is no evidence of (nor reason to expect) any correlation between T and q errors.

Clearly, as suggested by Dee and da Silva (2003), our current assumption of zero RH-T correlation is not generally valid. Work is underway to generate a better humidity correlation model (VTDP10). The plan is to construct a “normalised unbalance q” variable

using background error covariance statistics, such that it is uncorrelated with temperature and has a uniform variance.

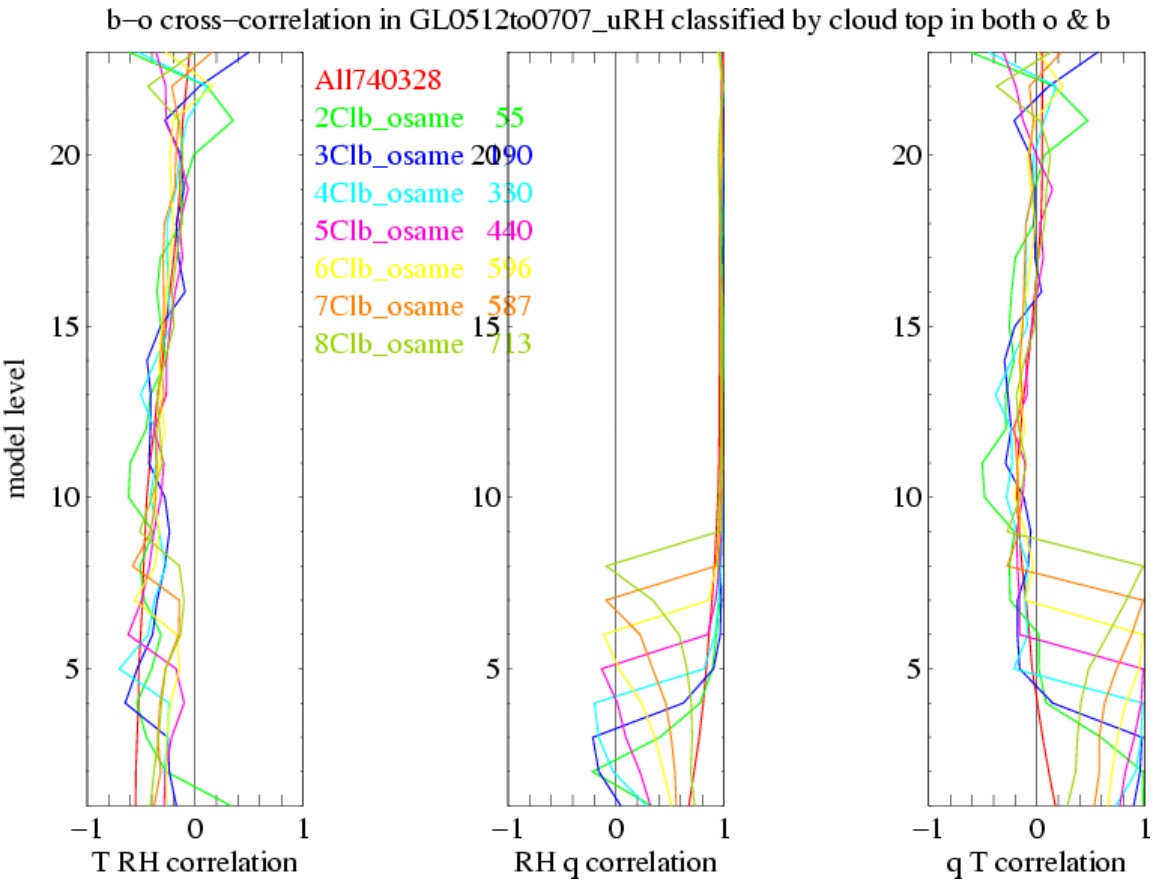

Figure 18 Cross-correlations between T, RH and q, for cases where observation and model have the same cloud layer.

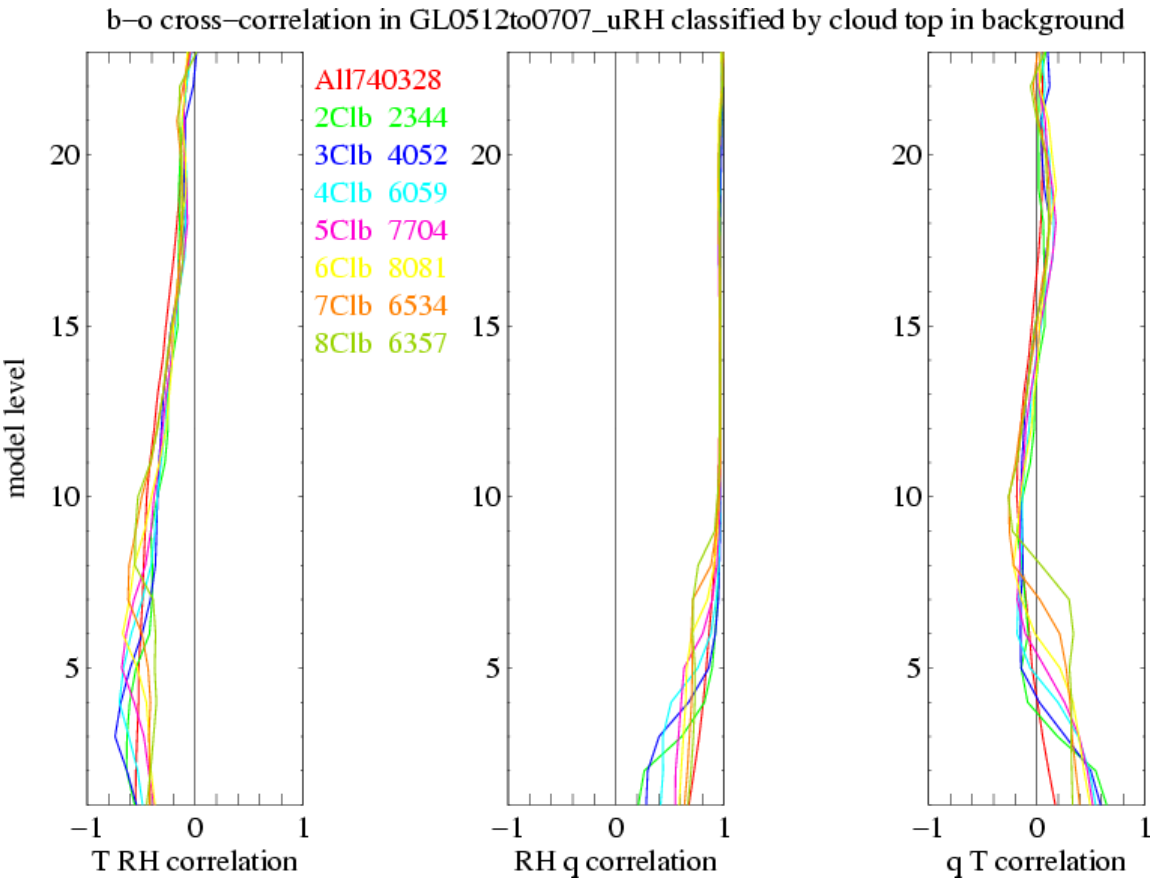

Figure 19 As Figure 18, classified by background cloud layers.

4 Model errors in cloud prediction

The layer-cloud classifications in background and observation can be used to verify the background 6-hour forecast. Usual categorical verification of hits, misses and false alarms is complicated by the identification of “stable” and “stratocumulus (SC)” cloud, and by the possibility of cloud being at the wrong level. To cover the latter, cases when the observed cloud was within plus or minus three levels of the background cloud were also counted. Global results with rhc=97.0 are shown in Figure 20 to Figure 27. For comparison with Figure 20, but with a change of scale, Figure 28 shows the frequencies with rhc=90.0, and Figure 29 shows the frequencies with the boosted RH that are assimilated operationally. Boosting the sonde RH is about equivalent to reducing rhc to 92.4%, so the sonde frequencies in these two curves are similar. Other scores for different rhc or boosted RH (not shown), compared to Figure 21 to Figure 27, show proportional changes, but the same patterns. Remember that it is my diagnosis of cloud layers from T and RH profiles that is being verified, not the model’s actual prediction, nor visual observation of cloud, so results may differ from operational cloud verification statistics. The figures show that:

- The global model generally predicts boundary layer cloud too low (Figure 20 and Figure 28), with the observed cloud layers being over a wider range. (Because of deficiencies in humidity sensors at low temperature, sondes do not detect many upper level cloud layers.)
- Integrated over the lower levels, the modelled cloud amounts seem about right. It is hard to be more precise because of the uncertainties in setting rhc. These global statistics use the same rhc for background and sonde. For the model, where RH represents a grid-box average, a lower value is appropriate (Figure 28) to take account of the partial cloud layers. But this argument does not apply to sondes. This question is studied more in section 6.

With the operational, boosted RH values, the diagnosed cloud amounts from sondes are probably too large (Figure 29).

- Correct prediction of layer cloud is not usual (Figure 22), even if an error in height is allowed for (Figure 24). Misses and false alarms are more common (Figure 26).
- Cases such as that reported as a problem and shown in Figure 1, where the model predicts a stable cloud layer while the observation has a SCu cloud layer (possibly at a different level), are very rare (Figure 25 red dashed line).
- When the model does correctly predict a cloud layer, it is on average about half a model level lower than the sonde (Figure 27).

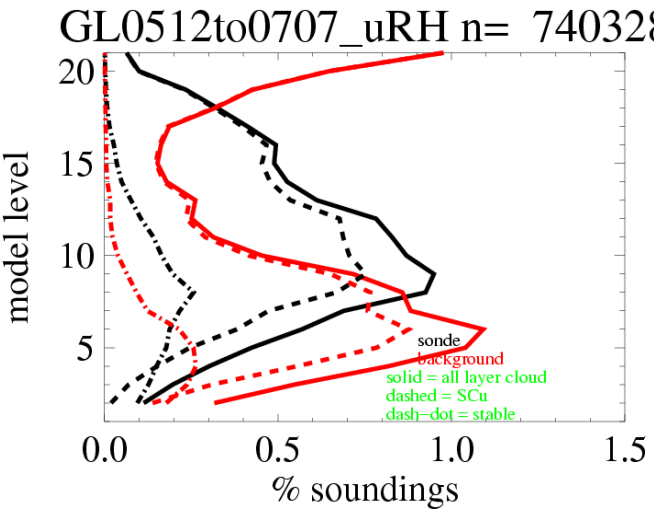

Figure 20 Frequency of cloud tops at each level, in sondes (black) and background (red).

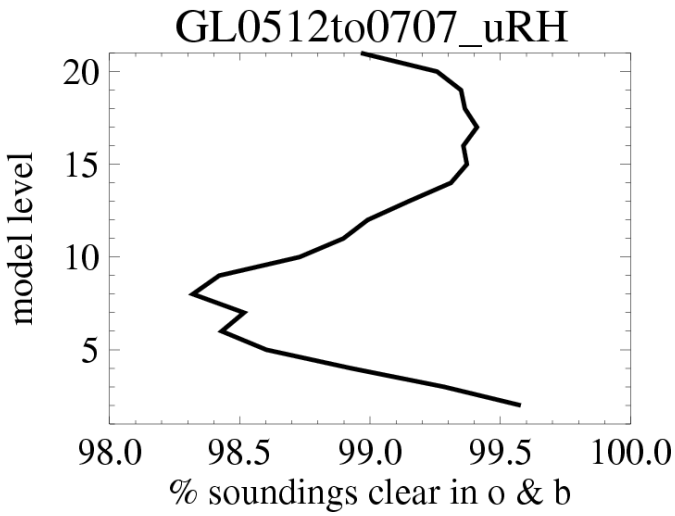

Figure 21 Correct predictions of no cloud.

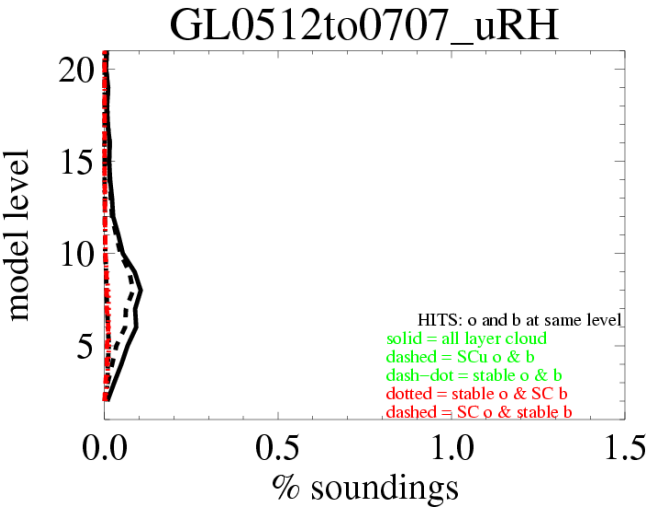

Figure 22 Correct predictions of cloud at same level (scale matching Figure 20).

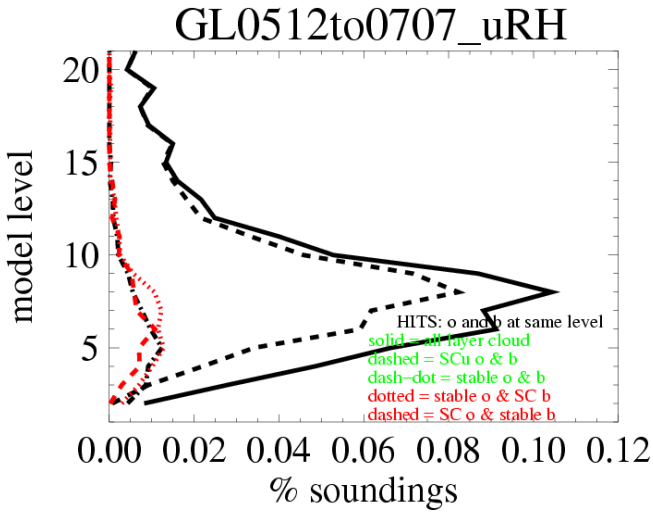

Figure 23 Correct predictions of cloud at same level.

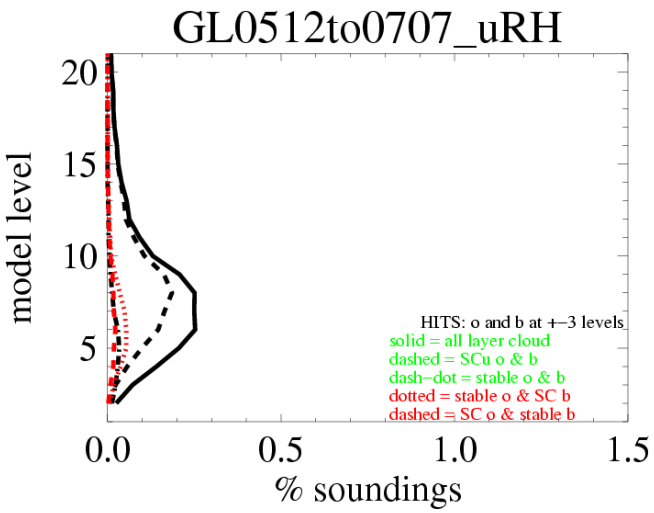

Figure 24 Correct predictions of cloud within +or-3 levels (scale matching Figure 20).

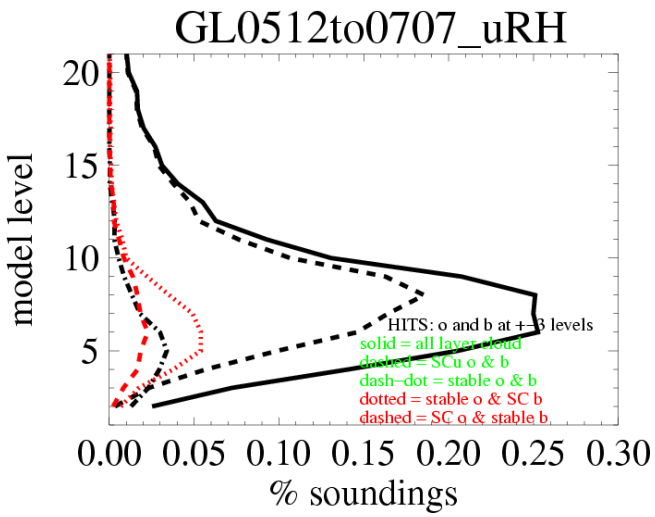

Figure 25 Correct predictions of cloud within +or-3 levels.

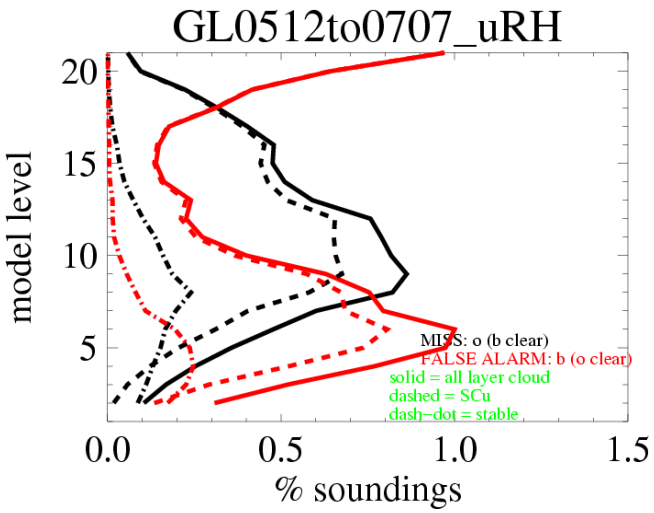

Figure 26 Incorrect predictions of cloud at same layer.

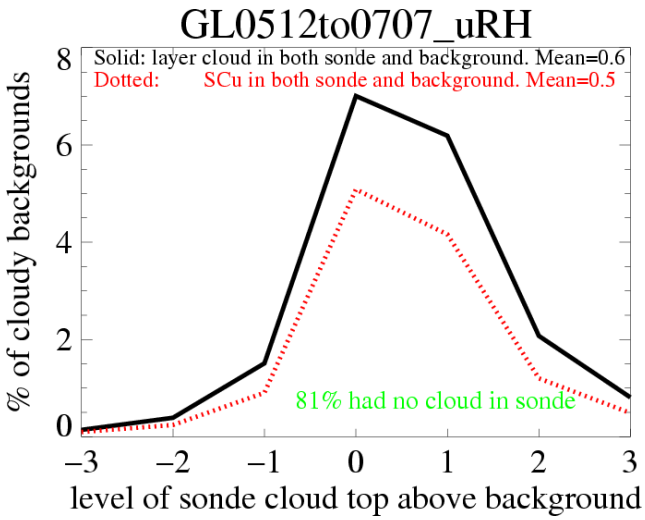

Figure 27 Histogram of offset of observed cloud top above models, as percentage of cases with a cloud layer diagnosed in background.

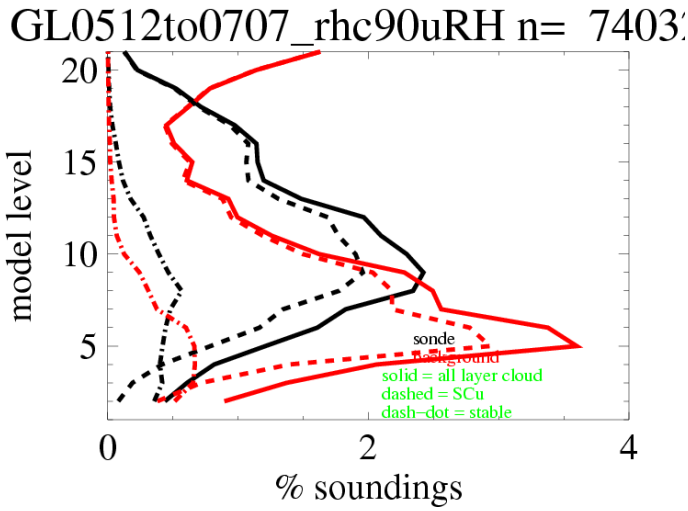

Figure 28 As Figure 20, with rhc=90%.

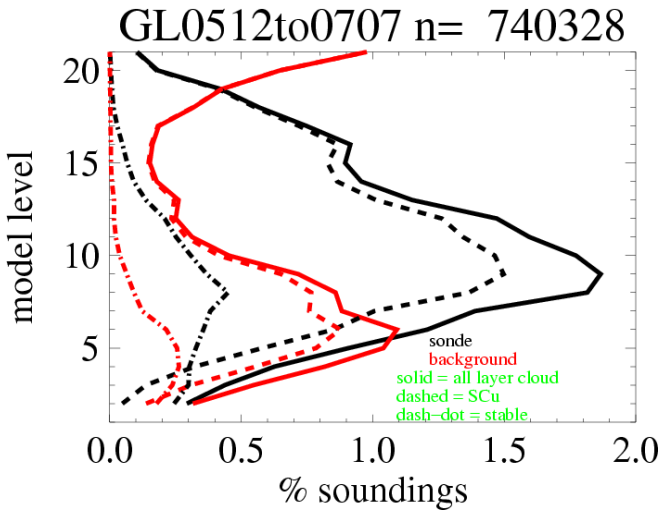

Figure 29 As Figure 20, with sonde RH boosted by up to 4.6%, as assimilated operationally.

5 Relative Humidity distributions

Relative humidity is limited by its definition to be ≥0% and by condensation processes to usually be ≤100%. These limits mean that the distribution of errors is non-Gaussian, and that near these limits mean-errors are non-zero. This was pointed out by Holm (2003); who presented a two-dimensional histogram of forecast and assumed true RH (his "truth" actually came from another model run). Figure 30 shows a similar diagram using sondes as "truth". It shows a similar result to Holm's. The blue line demonstrates the apparent bias of mean background RH, as a function of observed RH. The red line shows the apparent bias of mean observed RH, as a function of background RH. The fact that these are almost symmetrical shows they are not real biases, but rather due to the shape of the distribution. Holm suggested a nonlinear transform to model errors as a function of the mean of the background and analysis relative humidity. Taking the sonde as a surrogate for the "truth" that we are estimating by the analysis, this is equivalent to rotating Figure 30 by 45°. After this "Holm transform", the bias is near zero for most of the distribution, as seen from either the mean (red line) or the median (blue line) in Figure 31. Other aspects of the near-normality of the o-b distribution can be seen in the symmetry and closeness of the +-1 s.d.(red dots) and the 16 and 84 percentiles (blue dots); the red and blue lines almost coincide, as they should for a Gaussian.

Figure 31 demonstrates that there is no real dry bias in sonde humidities with respect to the model; in fact o-b is slightly positive near saturation. Despite this, for historical reasons, the sonde humidities are boosted as discussed in section 2 before they are assimilated. This boosting was based partly on the underreporting of humidity in cloud, particularly for the Vaisala radiosondes then used in the UK (Lorenc *et al.* 1996). The assimilation of UK sondes caused incorrect holes in cloud sheets in the operational mesoscale model. The humidity sensors have changed since. Another reason may have been a misinterpretation of the apparent bias seen in the blue line in Figure 30. The effect of the adjustment is to dramatically distort the distribution, as seen in Figure 32. The linear phasing in of the boost between 80% and 95% leads to a reduction in the pdf, with a corresponding increase above 99.6%. More seriously, the boost clearly introduces a bias in the o-b values. We have already seen in Figure 29 that the boost causes the frequency of layer cloud diagnosed from global sondes to be too large. There is a clear case for removing the boost. [1]

As discussed in section 3.4, we need to revise our humidity control variable assumptions. But a simple change of variable would not address the apparent bias seen in the blue line in Figure 30, for that we need a nonlinear Holm transform. It is clear that, if we go ahead with the planned development of a Holm transform, the humidity boost should be removed first. Note that introduction of the Holm transform will be complicated by the need to retain the benefits of our current RHtotal control variable, with a cloud and moisture incrementing operator designed to model behaviour near saturation, and by the desire to address at the same time the cross-correlation with temperature discussed in section 3.4.

[1] Bruce Macpherson commented: "I asked Mark Naylor to measure the impact of rhboost in an anticyclonic week in 12km Mes 3DVAR, because I knew it was so out of date. Switching off the boost gave a 0.3% improvement in the UK index."

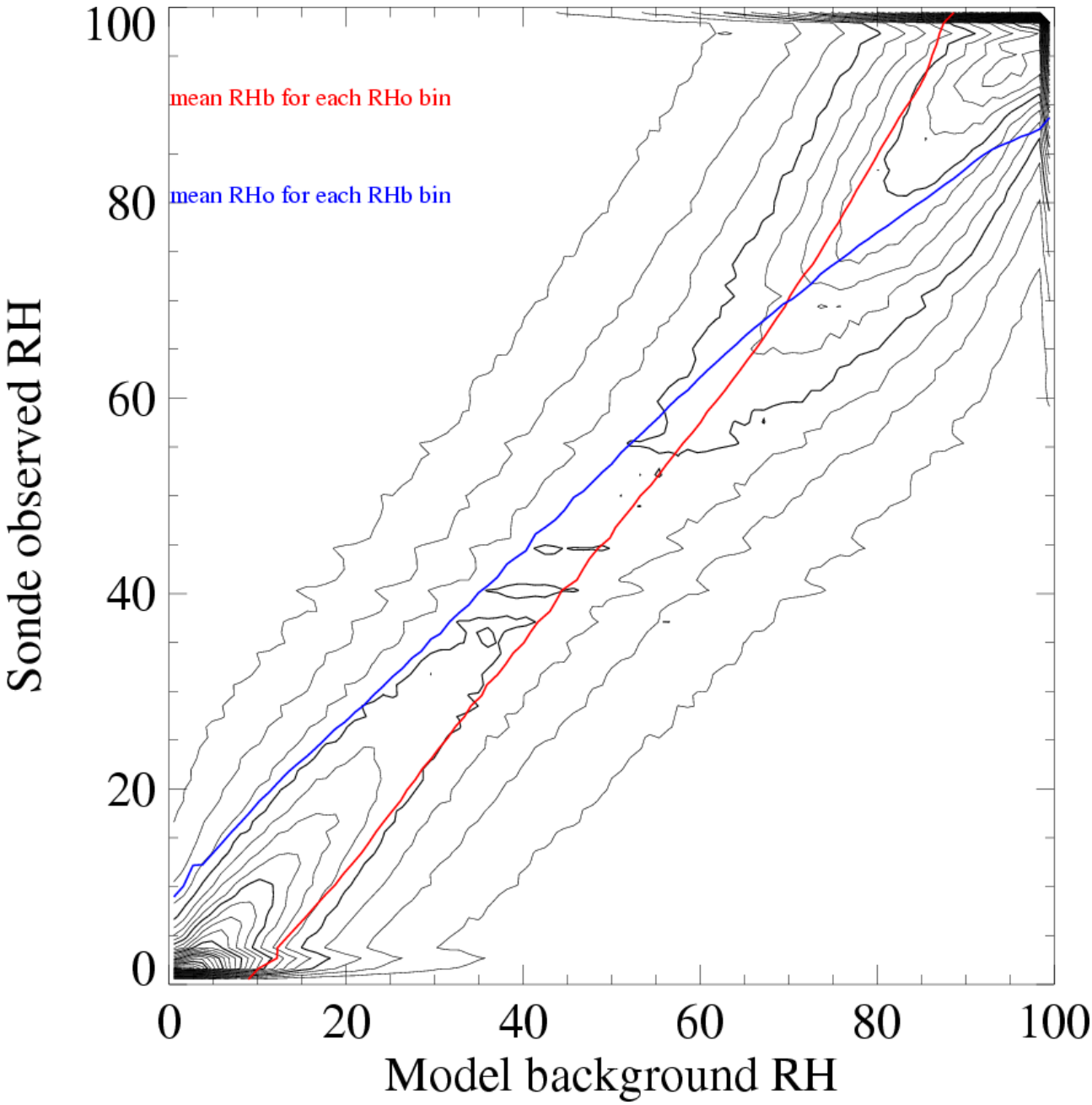

Figure 30 Two dimensional histogram of collocated model background and sonde RH values.

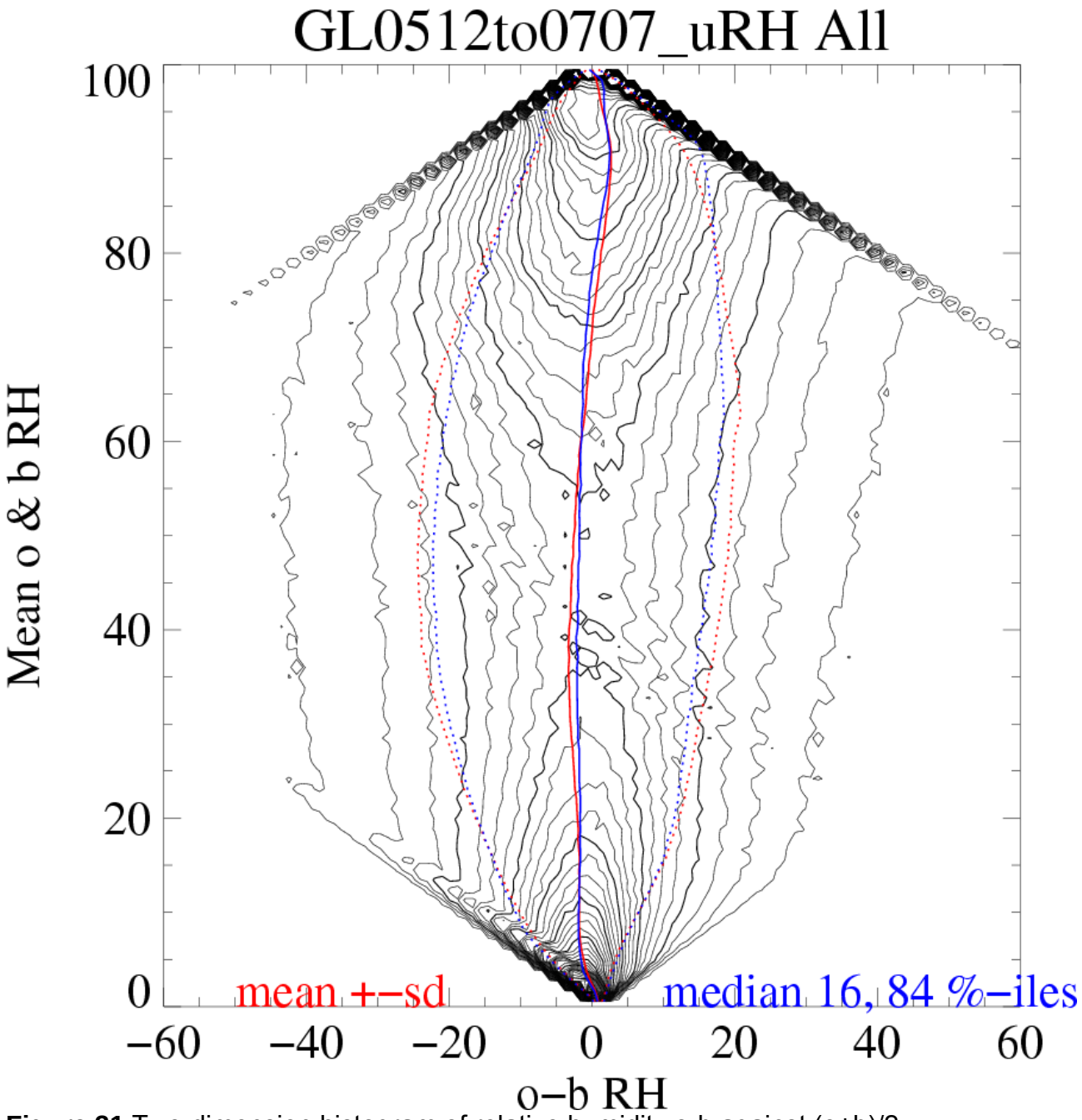

Figure 31 Two-dimension histogram of relative humidity o-b against (o+b)/2.

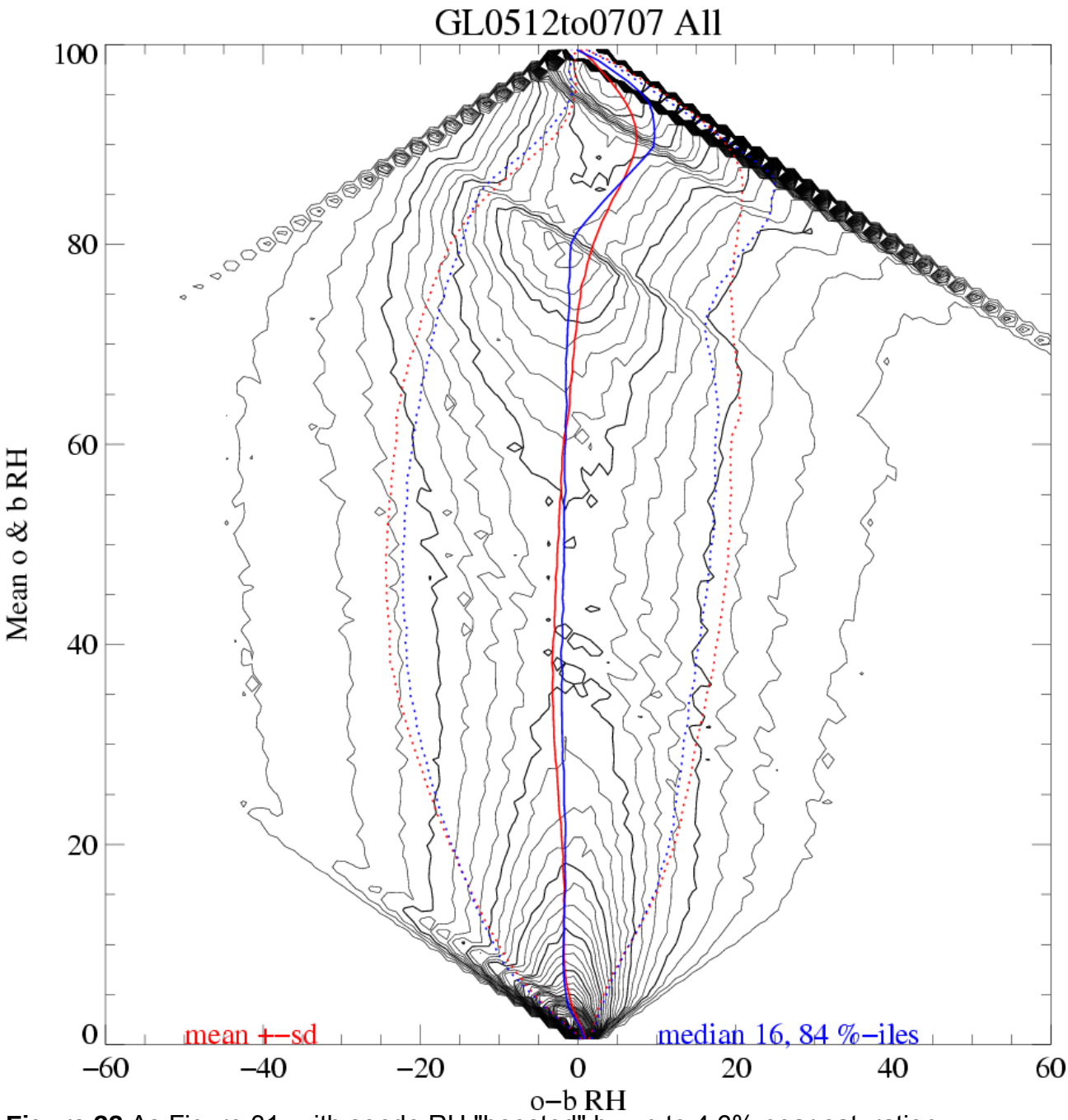

Figure 32 As Figure 31, with sonde RH "boosted" by up to 4.6% near saturation.

6 UK Results

The results presented thus far have been for all global radiosonde data. Here I present a selection of similar results for UK sondes only, in the same order as the corresponding global results. As well as different [frequencies of] weather patterns, and hence different background error statistics, differences might reflect observational error characteristics of the Vaisala sondes used in the UK.

For reasons discussed below, I used in this section the lower value rhc=94% for the background, keeping rhc=97% for the sondes. This affected the frequencies of detection of cloud in the background, but made very little difference to other statistics.

Both the mean errors, and the RMS errors, were less for the UK. The patterns of mean errors were similar to, but smaller than in Figure 2. Because the range of cases in the composites was much smaller, I show in Figure 33 an example mean profile instead of the differences. This shows that, for a composite based on a background cloud layer, the mean observed profile contains some cases without the inversion, and is hence smoother. The perceived bias in the background is no more of a bias than that of the blue line in Figure 30 – if we composite instead according to a sonde cloud layer, the perceived bias goes the other way, like the red line in Figure 30. This sonde-based composite is complicated by another effect, best shown by showing stratocumulus and "stable" cloud separately. Figure 34, for the SCu, shows a symmetrical effect to Figure 33; the background gives a smoothed and therefore biased estimate of the observed inversion. But Figure 35, for the sample diagnosed as "stable" cloud, shows an observed structure without an obvious meteorological explanation, with a maximum in wet-bulb potential temperature and in dew-point-temperature at the cloud top. Similar structures are much less common in model profiles. I do not know of a physical mechanism which would generate such a maximum. It is possible that this maximum is an observational error, due to the wetting and slow response of the humidity sensor, with the real cloud top at a lower level (as the model background indicates). However John Nash, a radiosonde expert, says; "The relative humidity sensors we use now are directly exposed to the atmosphere, with moistening driven off by heating the sensor [about every 40s near the ground]. So it is impossible for the relative humidity to read erroneously high for more than about 20s above the cloud and on average it would be only for about 10 seconds , i.e. about 50 m in the vertical, which is not long enough to give the distance between the two model levels that you see." So the maximum in the black dashed Td profile in Figure 35 remains unexplained.

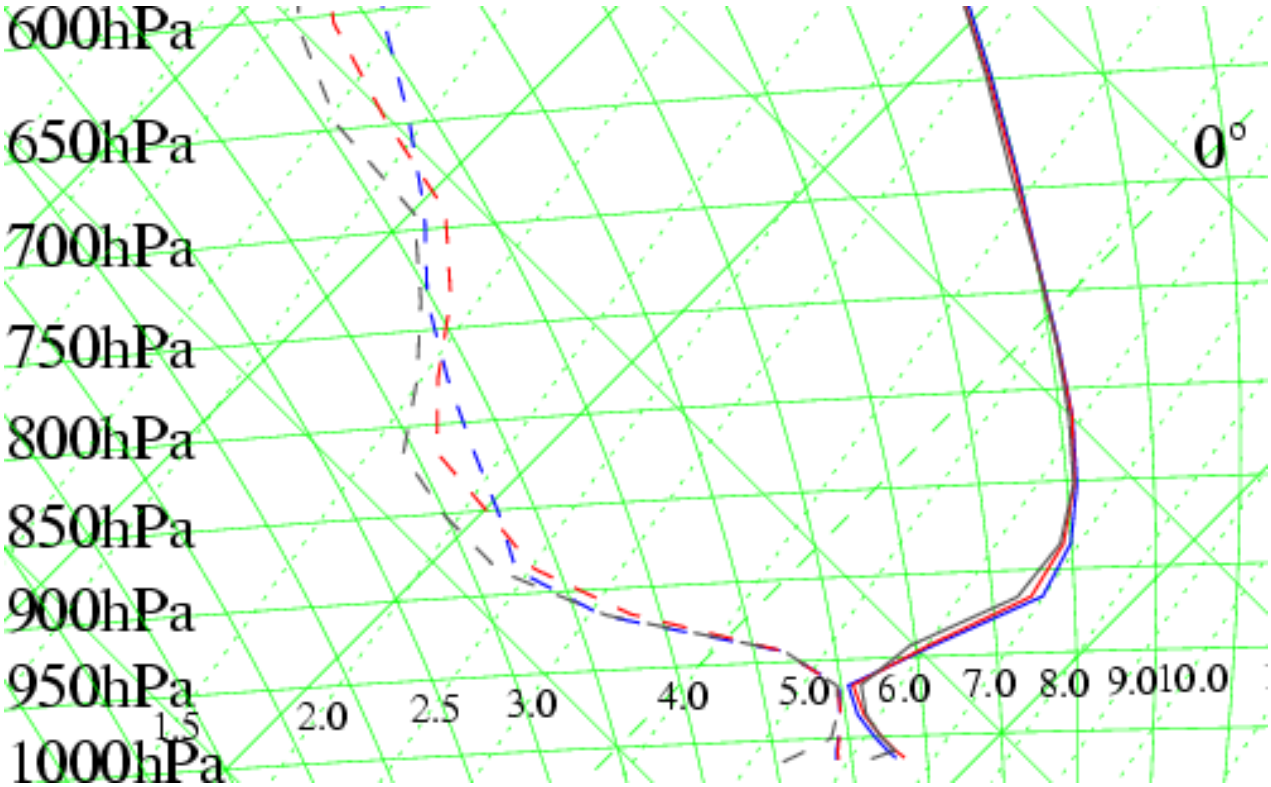

Figure 33 Tephigram of sounding (black), global model background (blue) and analysis (red), for the mean of 136 UK soundings with layer cloud top diagnosed at level 5 in the background.

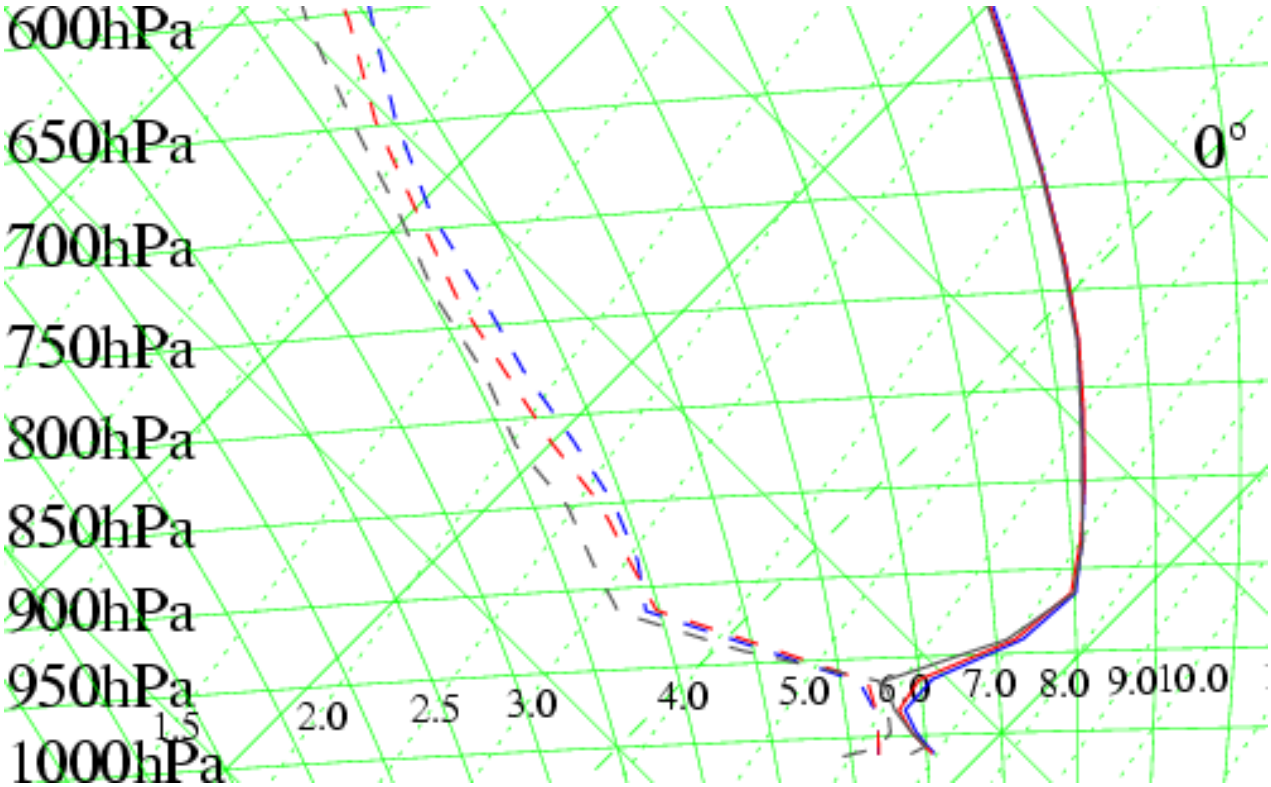

Figure 34 As Figure 33, for the mean of 140 soundings with stratocumulus tops diagnosed at level 5 in the sonde.

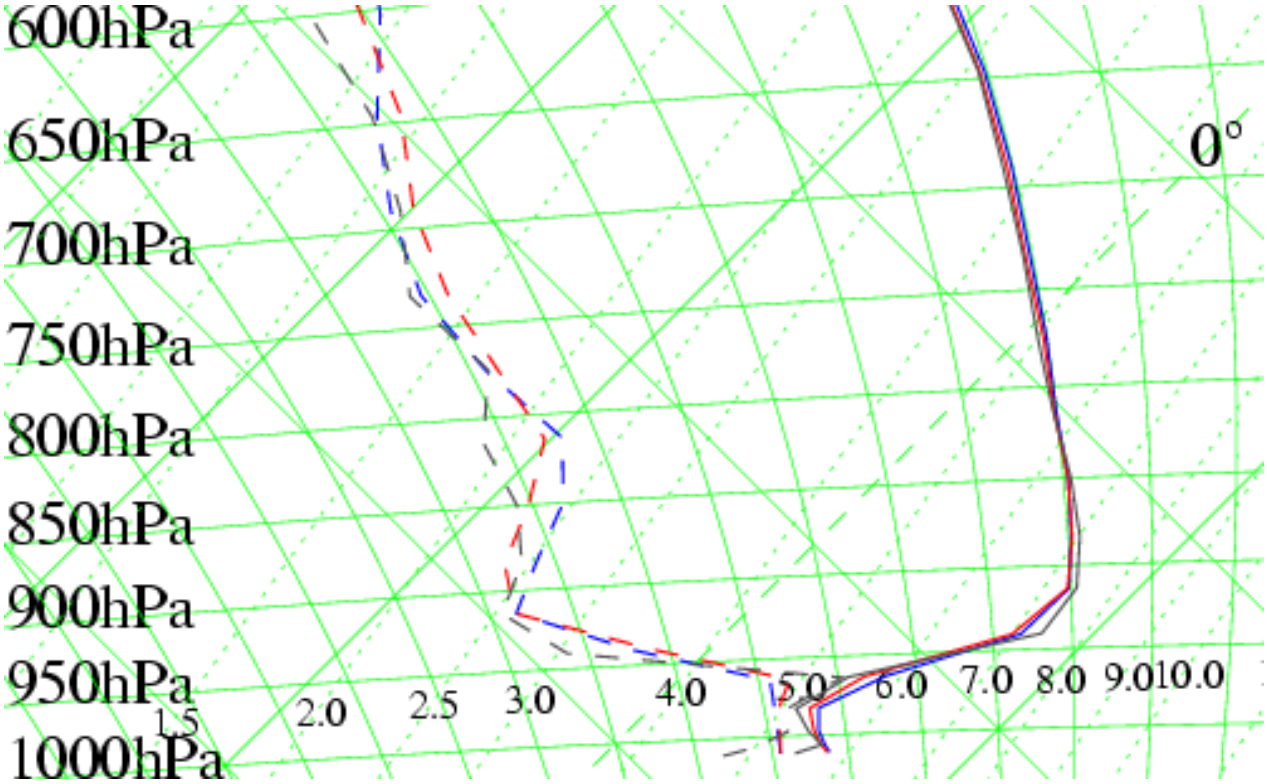

Figure 35 As Figure 33 for the mean of 50 soundings, with stable cloud tops diagnosed at level 5 in the sonde.

Figure 36 shows that RMS o-b for UK sondes are less than the global values (Figure 5). This allows us to draw conclusions from the inequality relationships with assumed observational and background variances: the assumed observational errors for temperatures are too large. John Nash says: "The actual random errors in UK radiosonde temperature sensor measurements have been lower than 0.3 K for many years, but the errors in the reported temperatures can be increased by the conversion into TEMP code. In our case the limits used in fitting the TEMP code are 0.5K in the troposphere, so the resultant temperature errors shouldn't be larger than 0.5K. Most countries will have similar sensor accuracy, but not the same accurate fitting limits, only fitting temperatures to 1K. So it is not surprising that you should find that the errors including representativeness errors are less than 1 K in the UK . Hopefully, if we could get the suitable BUFR code implemented then you should get the same performance across a much larger range of countries."

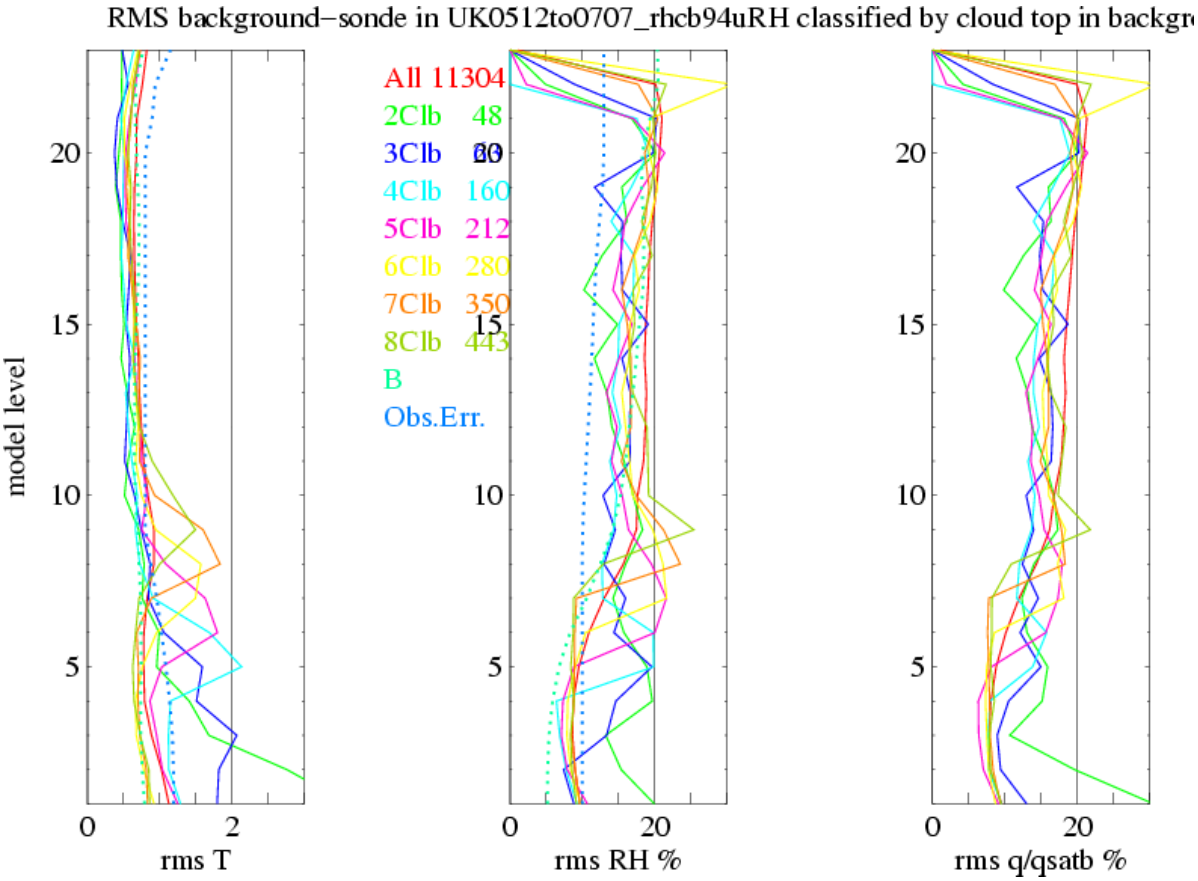

Figure 36 As Figure 5 for the UK.

The main reason for concentrating on the global results is the smaller UK sample size, which makes sub-samples with specific cloud levels too small for very stable statistics. This can be seen in Figure 37 and Figure 38, which exhibit often spurious variations in correlations as large as 0.2 due to their sub-sample sizes of order 100.[2] These figures do show a consistent pattern across all sub-samples however, which is similar to or stronger than that seen in the global sample: there are larger correlations within the boundary layer

[2] This inaccuracy of correlations from small samples is a general result, illustrated also in Lorenc (2003). It illustrates a major weakness of ensemble data assimilation methods which in their simplest form use correlations from an ensemble of this size directly in data assimilation. In the VAR covariance estimation system, we currently average in space as well as over realisations, giving a bigger sample. If we develop a system with more localised statistics, we will need more realisations in the training set.

(Figure 37) than across the boundary layer top (Figure 38). These patterns, and the corresponding all data correlations (Figure 39 and Figure 40) are slightly narrower than their global equivalents. (Note that the assumed correlations plotted in Figure 12 and Figure 13 are actually valid for the UK.)

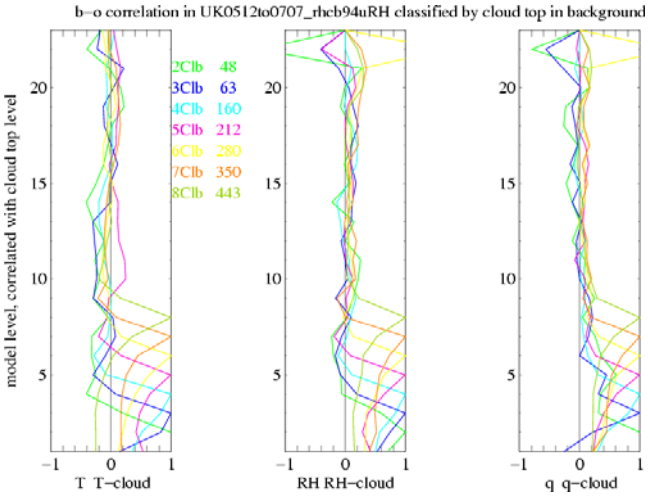

Figure 37 As Figure 8 for the UK.

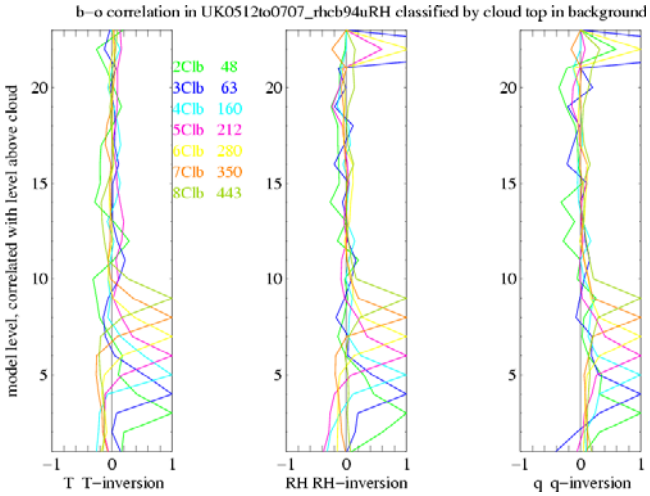

Figure 38 As Figure 9 for the UK.

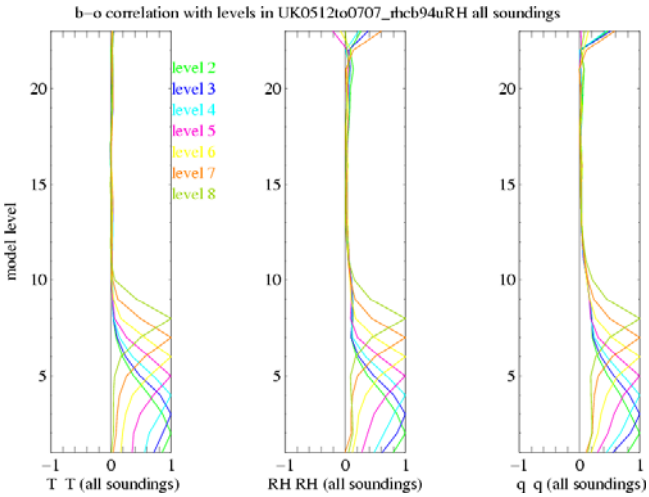

Figure 39 As Figure 10 for the UK.

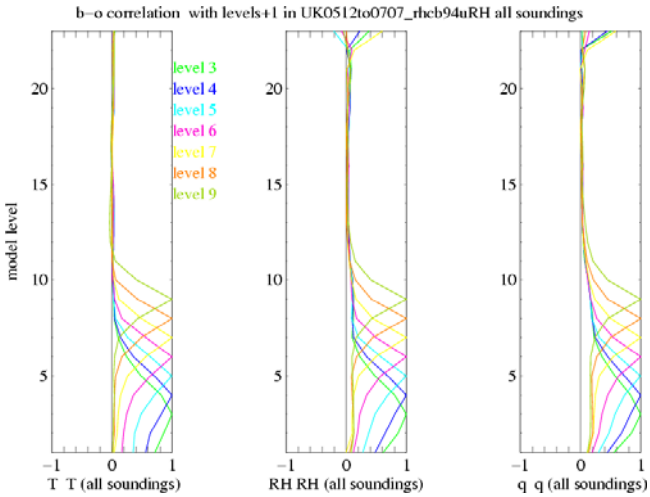

Figure 40 As Figure 11 for the UK.

Figure 41 shows the cloud diagnosis frequency statistics for the UK, using the same rhc as Figure 20. There is a much better match at each level. This, and the related increased accuracy seen in Figure 36, makes it worthwhile following up the discussion of rhc in section 4. Because of the grid-box average nature of the model RH, the model's internal formulation starts diagnosing partial cloud cover at 80% at most levels, although 50% cover is not reached until much higher RH. Radiosondes in a partially cloudy box might ascend through cloud, or clear air, and presumably will be diagnosed accordingly by the tests with rhc=97%. So there is a case for reducing rhc just for the model. Trial and error found that background rhc=94% gave a reasonable fit to the frequencies (Figure 43), so this value was chosen for the results in this section.

The verification statistics in Figure 43 to Figure 50 reflect the greater frequency of layer cloud in the UK. There is less bias in the vertical distribution of cloud (Figure 43). In order to assess the effect of the observational bias due to wetting seen in Figure 35, Figure 50 also shows offsets for cases where both background and sonde have a near-constant

theta_w layer showing a vertically mixed layer cloud (on the assumption that this pattern in the sonde means that wetting was not a problem). The background stratocumulus tops are still about half a level lower on average. This, and the much bigger bias seen in global frequencies (Figure 20), implies there is a significant model bias as well as the possible observational bias.

I also investigated whether there were any significant signals in classifications of winter layer cloud, but, apart from the expected increase in frequency (Figure 42), results were similar.

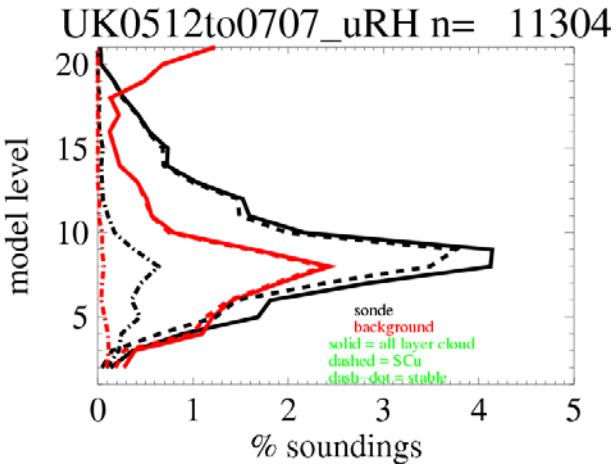

Figure 41 As Figure 20, for UK sondes.

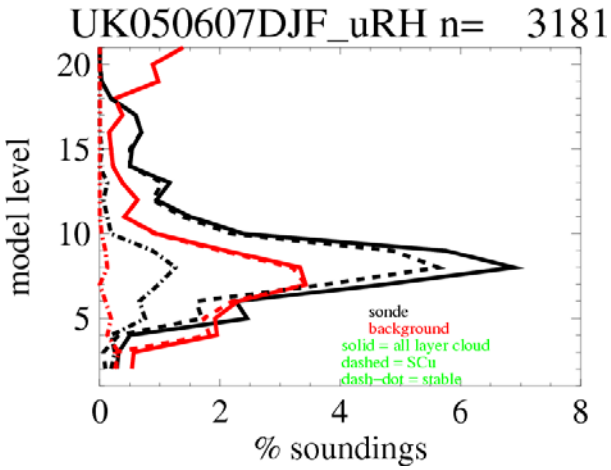

Figure 42 As Figure 20, for UK sondes in December, January and February.

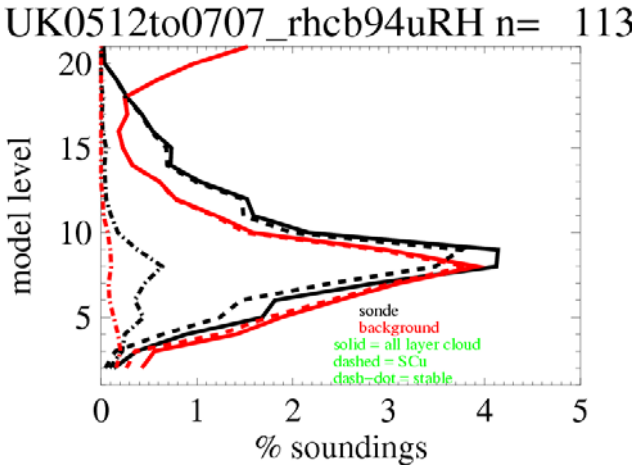

Figure 43 As Figure 20 for UK sondes, with background rhc=94%.

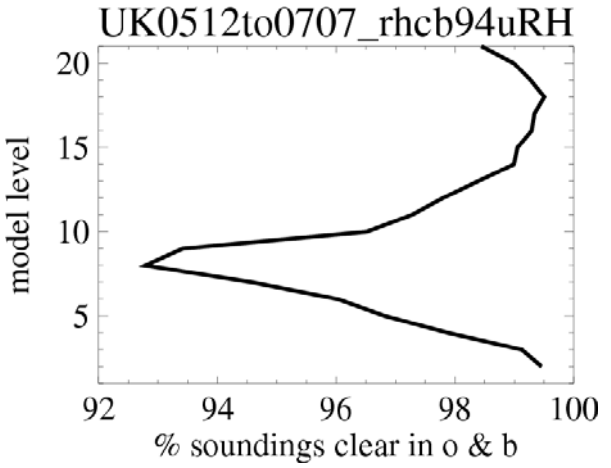

Figure 44 As Figure 21 for UK sondes, with background rhc=94%.

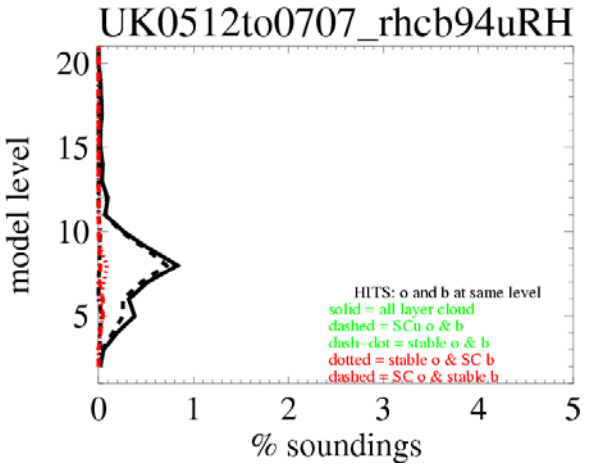

Figure 45 As Figure 22 for UK sondes, with background rhc=94%.

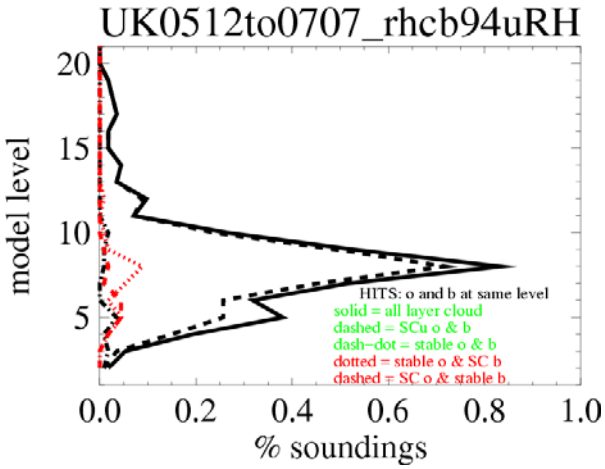

Figure 46 As Figure 23 for UK sondes, with background rhc=94%.

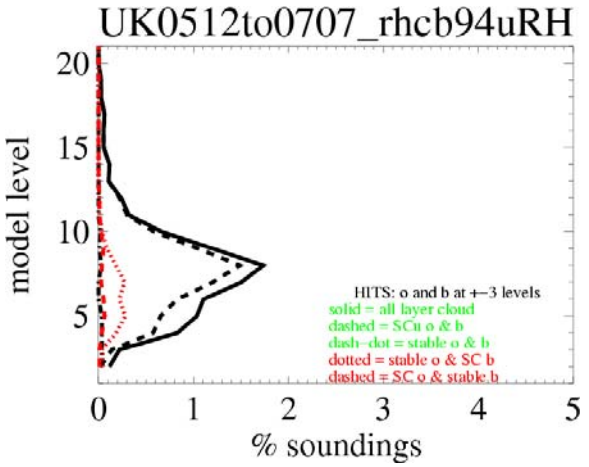

Figure 47 As Figure 24 for UK sondes, with background rhc=94%.

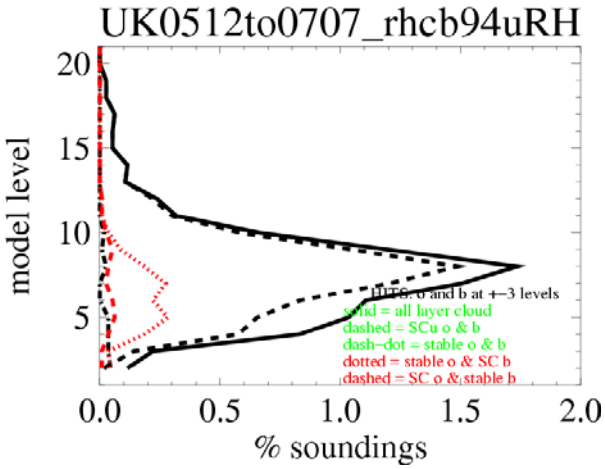

Figure 48 As Figure 25 for UK sondes, with background rhc=94%.

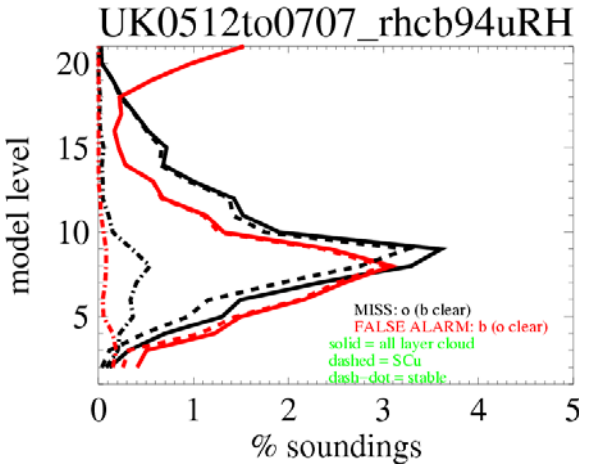

Figure 49 As Figure 26 for UK sondes, with background rhc=94%.

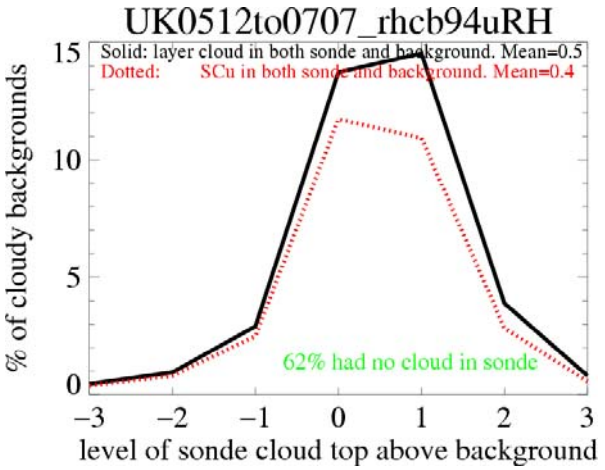

Figure 50 As Figure 27 for UK sondes, with background rhc=94%.

The above verification statistics for diagnosed cloud achieved a reasonable match using the un-boosted RH. Figure 51 confirms that for the UK sondes too, there is no justification from the RH distributions for the boost of RH near saturation. It also shows an interesting

difference from the global results: there is a clear bias in o-b for values of relative humidity between about 70% and 0%. It is not clear whether this is due to observational or model bias.

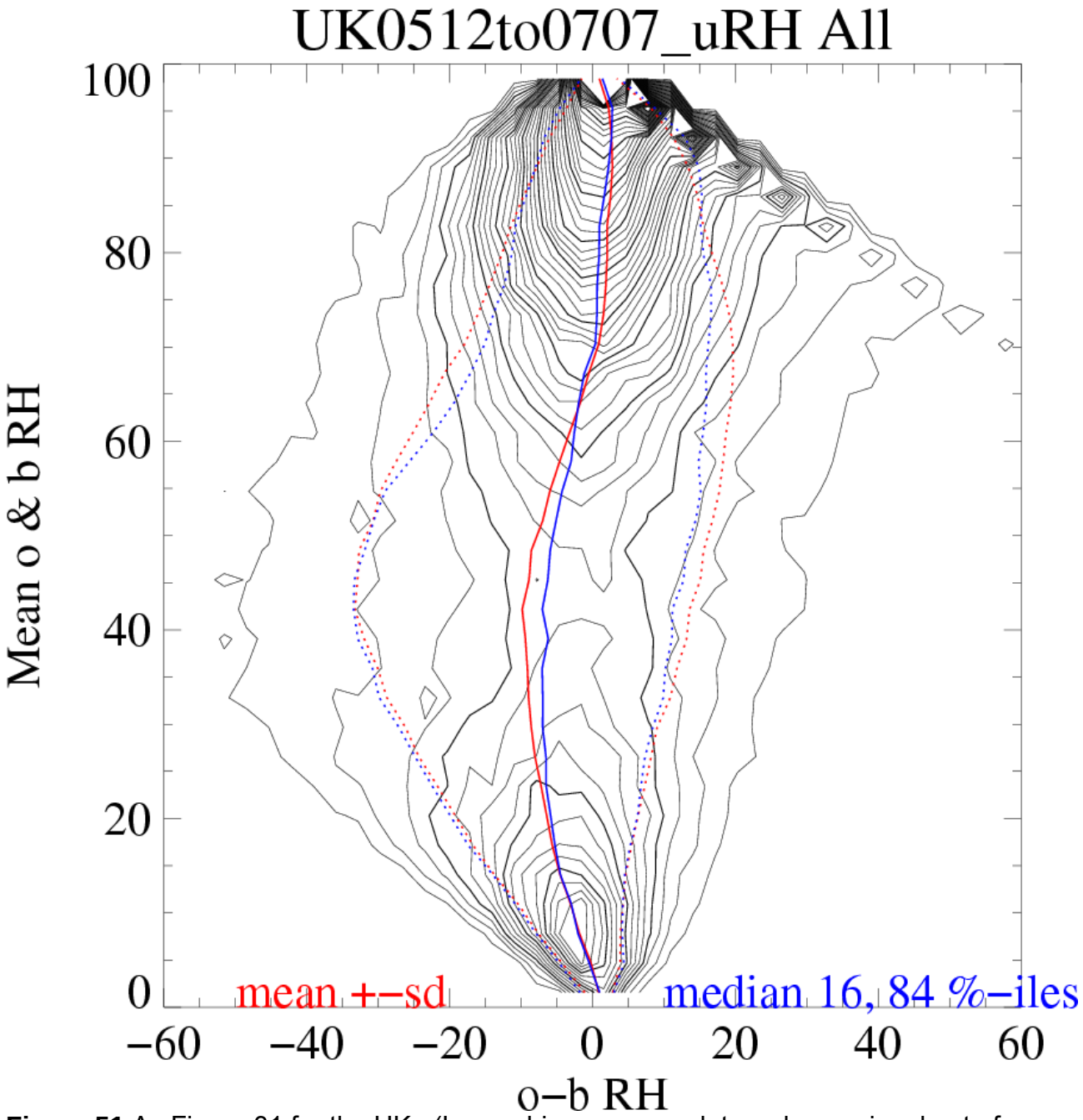

Figure 51 As Figure 31 for the UK. (Larger bins were used, to reduce noise due to fewer data.)

7 Conclusions and implications for further work

The 50-level global model is capable of resolving the main features of cloud-topped inversion, but not the detail (Figure 1). The proposed 70-level configuration has 50% more levels in this region, so should improve the modelled inversions. It will be worth repeating this study for the new model.

Cloud layers are associated with an inversion structure; it is common for the model to have a plausible structure in the wrong place. This leads to error distributions which are non-Gaussian, and often biased (section 3.1). A purely statistical approach to data assimilation, expressed using best estimates, mean errors and covariances, cannot give realistic results. The minimum variance "best estimate" will smooth inversions, giving structures which are not useful for NWP since they are not sustainable in a numerical forecast. Therefore we should not expect "improved covariances" to solve our problems in analysing inversions (although they might help). A better hope is that 4D-Var fitting of an NWP model to observations will constrain the analysis to structures which can be generated by the model.

The background error variances used in VAR are about right for average conditions. However they should be about doubled near background cloud-topped inversions (section 3.2). Such an adaptive system may be difficult to implement; as a shorter-term "fix" we could consider reducing the assumed observation errors instead in the same areas. This would have a similar effect of increasing the weight given to observations.

The vertical correlations of background errors used in VAR are about right for average conditions for RH, but too broad for temperature (section 3.3). We should check the vertical correlations implied by the ensemble-based covariances which are being developed, to see if they correct this.
The correlations are too large across cloud-top inversions, reducing the ability to extract detail from soundings. This effect, combined with the too small variances discussed above, explains the failure of VAR to fit the observed inversion. Covariance models which can adapt correlations to the presence of an inversion should be researched (however the problems of doing this for just-resolved structures are large).

The boosting of sonde humidities should be stopped.

The current VAR assumption that the RH and T background errors are not correlated is not usually justified (section 3.4). Research to develop a new humidity control variable with more flexible cross-correlations is needed. This is not straightforward, since we need to retain current ability to couple cloud errors to humidity and also to add a "Holm transform" (section 5).

Cases like Figure 1, where the model has a low-level stable cloud layer instead of the observed stratocumulus, are very rare in the global model (Figure 25 shows at most 0.02% at any level). Results in sections 4 and 6 support the forecaster perception of poor analysis of cloud below inversions. But the detailed structure of model and sounding in the example given is probably not typical.

The model's cloud layers are on average half a level lower than the observed (Figure 27). Some of this may perhaps be due to an observational bias in the height of the cloud top, due to wetting of the sensor (Figure 35), but the bias also exists in stratocumulus cases,

which do not have this bias signature (Figure 50). This result should be checked after the vertical resolution has been enhanced.

The assumed observational errors for temperature from UK sondes are too large (Figure 36).
There is a negative bias in o-b humidity for UK sondes in the range 5-70% (Figure 51). This is not seen in the global statistics. It is not clear whether it is due to model or observation bias.

Covariance estimates need 1000s rather than 100s of independent cases to be accurate (Figure 37). This should be borne in mind when we move to a more localised covariance model.

8 References

Dee, Dick P. and Arlindo M. da Silva. 2003: The Choice of Variable for Atmospheric Moisture Analysis. *Mon. Wea. Rev.*, **131**, 155-171.

Hollingsworth, A. and Lonnberg, P. 1996: " The statistical structure of short-range forecast error as determined from radiosonde data. Part I: The wind field." *Tellus*, **38A**, 111-136.

Holm, E.V. 2003: “Revision of the ECMWF Humidity analysis: construction of a Gaussian control variable” ECMWF/GEWEX Workshop on Humidity Analysis, July 2002. http://www.ecmwf.int/publications/library/do/references/list/17000

Lorenc A. C., D. Barker, R. S. Bell, B. Macpherson, and A. J. Maycock 1996: "On the use of radiosonde humidity observations in mid-latitude NWP." *Met. and Atmos. Physics*, **60**, 3-17..

Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne and F. W. Saunders. 2000: The Met. Office Global 3-Dimensional Variational Data Assimilation Scheme. *Quart. J. Roy. Met. Soc.*, **126**, 2991-3012.

Lorenc, A. C., 2003: The potential of the Ensemble Kalman filter for NWP - a comparison with 4D-Var. *Quart. J. Roy. Met. Soc.,* **129**, 3183-3203.

Lorenc, Andrew C. and Tim Payne 2007: “4D-Var and the Butterfly Effect: Statistical four-dimensional data assimilation for a wide range of scales.” *Quart. J. Roy. Met. Soc.*,**133**, 607-614.

Macpherson, B., Bruce J. Wright, William H. Hand, and Adam J. Maycock, 1996: "The impact of MOPS moisture data in the UK Meteorological Office mesoscale data assimilation scheme", *Mon. Wea. Rev.*, **124**, 1746-1766

OSDP5: Radiosonde Processing. OPS Scientific Documentation Paper 5.

Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, DingMin Li, G. W. Inverarity, A. C. Lorenc and T. J. Payne 2007: “The Met Ofiice Global 4-Dimensional Data Assimilation System” *Quart. J. Roy. Met. Soc.*, **133**, 347-362.

Semple, Adrian T. 2006: Impact of the Aberporth Radiosonde Ascent on a Waving Frontal Zone, 18th April 2006. Met Office FTTR490. http://www.metoffice.gov.uk/research/nwp/publications/papers/technical_reports/2006/FRTR490/FRTR490.pdf

VTDP10: Normalisation statistics for the humidity control variable.