TECHNICAL MEMORANDUM

620

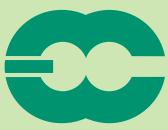
Enhanced use of all-sky microwave observations sensitive to water vapour, cloud and precipitation

Alan J. Geer and Peter Bauer

Research Department

Simultaneously released as EUMETSAT / ECMWF research report 20

April 2010



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

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Abstract

Cycle 36r3 will include a complete revision of observation errors, quality control, thinning and resolution-matching for the all-sky assimilation of microwave imagers. The new approach makes use of the symmetric nature of cloud and rain-affected first guess (FG) departures to predict the observation and background error for quality control purposes. All-sky FG departures are Gaussian when normalised by this model, allowing for the first time an effective quality control for cloud and rain-affected observations. The model is also used to provide observation errors that increase with the mean amount of cloud in model and observation.

In the previous approach, observation error was inflated as a function of distance from grid point, but this has been abandoned in the new system. In practice, the increase of FG departure errors with distance is not important over the 20 to 50 km distances involved. The spatial scale of the observations has also been examined. Instead of taking the nearest single all-sky observation to a grid point, we calculate an average or 'superob' of all observations falling into a grid box, prior to assimilation. Finally, the new approach screens out observations where the model shows 'cold sector' and 'heavy snowfall' biases. These biases are too difficult to deal with using a predictor-based bias correction scheme.

The new approach gives a substantial increase in the weight of all-sky observations in the analysis, with improved analysis and FG departure fits to radiosonde and dropsonde humidities, microwave sounder humidity channels and infrared sounder lower-tropospheric temperature channels. Short-range own-analysis forecast errors in humidity and vector wind are larger over the tropical oceans, but this is caused by bigger increments, rather than real errors. Vector wind forecast errors against radiosonde do not show the same effect. In summary, the new approach substantially increases the constraint of lower-tropospheric humidity over oceans in the analysis and first few days of forecast, but has no impact at longer ranges.

1 Introduction

All-sky assimilation of microwave imager radiances went operational with cycle 35r2 in March 2009 (Bauer et al., 2010; Geer et al., 2010). In the all-sky approach, radiance observations from Special Sensor Microwave / Imager (SSM/I, Hollinger et al., 1990) and Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E, Kawanishi et al., 2003) are assimilated in all conditions, whether clear, cloudy or rainy. At the frequencies used by microwave imagers, the atmosphere is semi-transparent except in heavy cloud and precipitation. Observations are only assimilated over oceans, where clear sky radiative transfer is dominated by water vapour absorption. Hence, the observations are sensitive to ocean surface properties (e.g. surface temperature and windspeed), atmospheric water vapour, cloud and precipitation. The intention is to use all of this information to improve analyses and forecasts.

All-sky radiance assimilation has not been tried before so initially we applied cautious observation errors and quality control. As a result, lower tropospheric humidities were constrained only half as well in cycle 35r2 as in previous cycles, where the old technique of 1D+4D-Var was used in cloudy skies (Bauer et al., 2006a,b; Geer et al., 2008) and direct radiance assimilation was used in clear skies. Now we have more experience with the new approach, it is time to increase the weight of data in the system. However, we have had to completely revise our approach to observation errors, quality control, thinning and resolution-matching. Much of this is based around a new 'symmetric' error model for cloud- and rain-affected data.

This new approach has been included in cycle 36r3. Two other all-sky developments were included in 36r3 but are not discussed here. First, it is now possible to use the microwave imagers on WindSat, Tropical Rainfall Measuring Mission (TRMM) and Special Sensor Microwave Imager / Sounder (SSMIS), which were not previously available in the all-sky system. This will help to fill a gap, as the longstanding SSM/I instruments are now starting to fail. Second, there has been a complete revision of the use of the observation database (ODB) for all-sky data, which brings significant performance benefits, as well as simplifying the code.

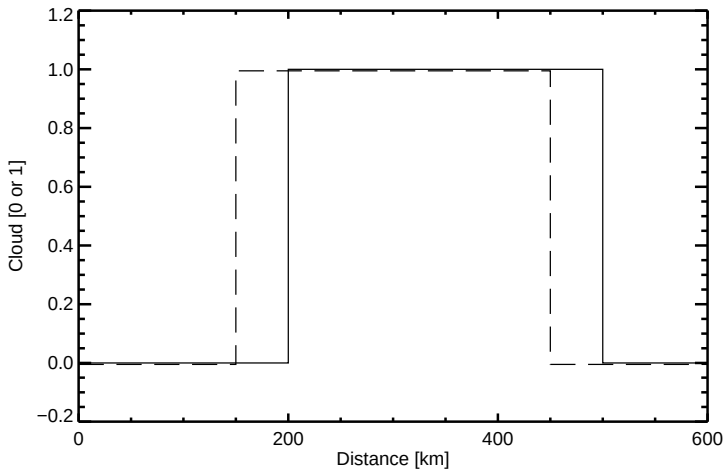


Figure 1: Displacement error between hypothetical first guess (FG, dashed) and observed (solid) binary clouds.

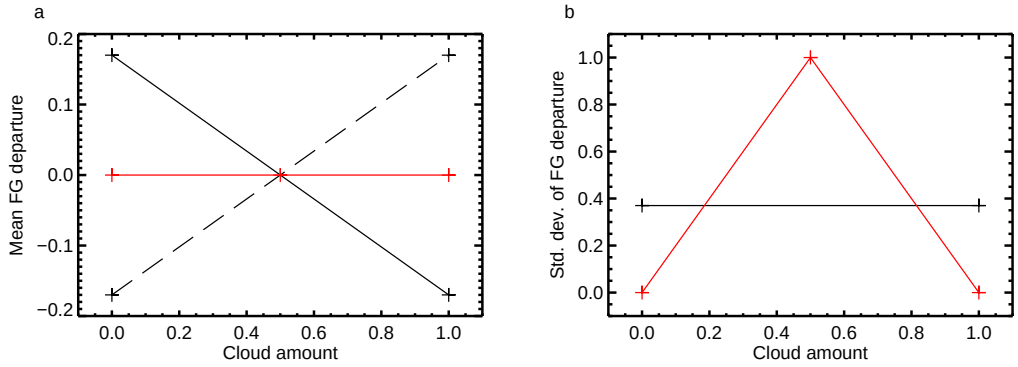


Figure 2: ***(a)*** *Mean and* ***(b)*** *standard deviation of FG departures of the cloud distribution shown in Fig. 1, as a function of FG cloud (black, solid), observed cloud (black, dashed) and mean cloud (red, solid)*

2 New developments

2.1 The importance of symmetrical sampling

Most of this work relies on a better understanding of the need for symmetrical sampling throughout a cloud-affected assimilation system. Differences between modelled and observed cloud and rain are typically large. Some of these differences may be due to displacement errors, but there are also errors in the structure and intensity of forecast cloud and rain. Figure 1 shows a hypothetical system in which the only difference between model and observation is a displacement error. Typically we may wish to make observation error and bias correction vary as a function of the cloud or rain amount. However, it is very easy to do this in the wrong way.

Fig. 2a shows the mean first guess (FG) departures (observation minus FG) of the hypothetical cloud in Fig. 1, as a function of cloud. In the area where the FG is clear, there are some cloudy observations, and where the FG is cloudy, part of the area is observed to be clear. This causes mean departures to be positive where the FG is clear, and negative where the FG is cloudy. For the mean departure as a function of observed cloud, the

opposite applies. This behaviour may seem trivial, but its consequences are very important.

When correcting observational bias as a function of cloud amount, the choice of predictor is crucial. In clear-sky assimilation it is typical to use the FG forecast to provide predictors for bias correction. The bias is calculated as a function of these predictors, and then removed from the observations. However, the bias as a function of FG cloud in Fig. 2 is simply a feature of the sampling, and not a real bias. The only difference between FG and observation is a displacement. Hence it would be incorrect to use the FG cloud as a bias predictor.

At ECMWF, FG rain was used as a bias predictor in the 1D+4D-Var system (Geer et al., 2007), and FG logarithmic cloud optical depth was used as a bias predictor in experimental assimilation of Moderate Resolution Imaging Spectroradiometer (MODIS) observations (Benedetti and Janisková, 2008). In particular, their Fig. 3 presents the mean FG departure as a function of FG cloud, and shows exactly the sampling bias we would predict using our simple model. The effect of treating the sampling bias as a real bias and removing it from the observations is to reduce the average size of the increments, i.e. it reduces the impact of the observations by removing real information. Hence, a bias correction based on cloud or rain in the FG is undesirable. The initial implementation of the all-sky assimilation of microwave imagers used only clear-sky quantities (e.g. water vapour, temperature) as bias predictors, so it does not suffer this effect. However, there are clearly biases in cloudy areas, so it would be useful to develop a better strategy for bias correction.

We need to find a way of binning FG departures that does not create a sampling bias. Against FG or observed cloud in Fig. 2a, the biases are equal and opposite, which suggests that binning by the average of FG and observed cloud may be suitable. This strategy is represented by the red line, and is indeed unbiased. In this binary example, the average cloud amount can take one of three values: when it is 0 or 1, both FG and observation agree; when it is 0.5, FG and observation disagree, but the sample includes equal numbers of positive and negative departures, so on average there is no bias. Another unbiased binning strategy, not shown here, is to take the maximum of cloud in the FG and observations. We will term the unbiased strategies as ‘symmetric’ and the biased ones (e.g. FG or observed cloud) as ‘asymmetric’. These strategies are applied to real observations in the next section.

It is not just bias correction that can be affected by asymmetric sampling. It is typical to prescribe observation error as a function of FG or observed cloud amount (e.g. Bauer et al., 2010), but that is also undesirable. Consider assimilating the observations in Fig. 1 with a large error for those which are cloudy, and a small error for those which are clear. If the background error were the same in each case, the cloudy observations would be relatively less influential than the clear observations. The clear observations have a sampling-induced bias of -0.17 and would cause the analysis to dry. The cloudy observations have a sampling bias of +0.17 which should counterbalance this, but it does not, because these observations have less influence in the analysis. The net effect would be a spurious drying of the analysis relative to the FG. Hence, just as for bias correction, if observation error is defined as a function of cloud or rain, the cloud or rain predictor must be symmetric.

It is also interesting to examine the standard deviation of FG departure as a function of cloud (Fig. 2b). In our example standard deviation does not vary with observed or FG cloud (this is because cloud makes up exactly half of the domain). However, as a function of mean cloud, standard deviations are zero where both model and FG agree, i.e. where mean cloud is either 0 or 1. Where model and FG disagree, i.e. where mean cloud is 0.5, the standard deviation is 1. Large departures are associated with areas where the model and observation disagree. We can use this effect to our advantage for the quality control of cloud and rain observations.

2.2 The symmetric behaviour of all-sky FG departures

The behaviour of all-sky FG departures is remarkably similar to what our conceptual model would predict. Figure 3 shows the mean and standard deviation of FG departures from the all-sky system, calculated as a function

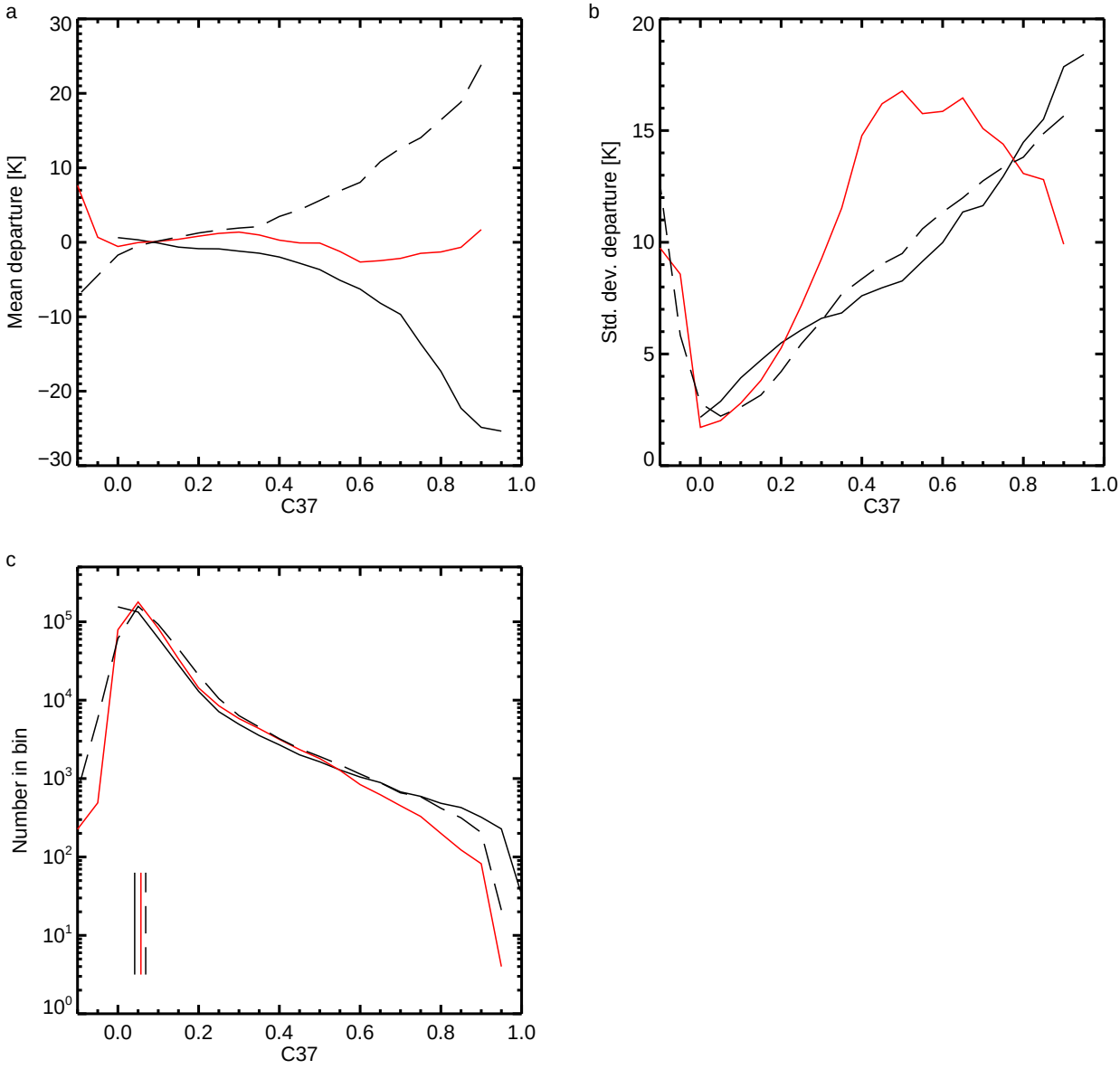


Figure 3: ***(a)*** *Mean and* ***(b)*** *standard deviation of SSM/I channel 19v FG departures binned as a function of 'cloud' derived from 37 GHz TBs (C37) and* ***(c)*** *number per bin, for a sample of 419159 observations from 1 - 10 October 2009. C37 is derived from the FG (black, solid), observations (black, dashed) or is the mean of the two (red, solid). Vertical lines on panel c show the medians. Bin size is 0.05 in C37. Standard deviations and means are only shown when there are more than 50 observations in a bin.*

of FG, observed, or symmetric 'cloud'. Of course, cloud is not a binary quantity here. More importantly we need to be careful with our definition of 'cloud', since we are working with a radiance observation. It would be easy to determine the FG cloud amount from the model, but cloud is only one of several quantities to which the radiances are sensitive. We need to use a measure of cloud that can be computed consistently for both FG and observations.

The solution is to work in radiance space, and to apply a simple approximate retrieval of the cloud amount. Here, we take the normalised 37 GHz polarisation difference $P37$ (Petty and Katsaros, 1990; Petty, 1994), which is roughly proportional to the square of the slant path cloud and precipitation transmittance at this frequency, τ_{37}:

$$P37 = \frac{T^v - T^h}{T^v_{clr} - T^h_{clr}} \simeq \tau^2_{37}. \tag{1}$$

Here, T^v and T^h are the vertically and horizontally polarised 37 GHz brightness temperatures at the top of the atmosphere, and T^v_{clr} and T^h_{clr} are the the brightness temperatures for the same profile but without clouds or precipitation. Since emission at the sea surface is highly polarised, but atmospheric absorption is in general unpolarised, $T^v - T^h$ is a measure of atmospheric opacity, with a completely opaque atmosphere giving no polarisation difference. Hence, $P37$ varies between 0, which represents a profile with opaque cloud, and 1, which represents clear sky. Note that cloud and precipitation have much lower optical depth in the microwave than in the visible or infrared, so only the most intense convection is opaque at 37 GHz. To be consistent with our earlier discussion, we use $C37$ as the x-axis in Fig. 3, where

$$C37 = 1 - P37 \tag{2}$$

and hence $C37$ increases with cloud amount.

Here, T^v and T^h come either from the observation or the bias-corrected first guess. T^v_{clr} and T^h_{clr} are always simulated using the FG profile, making the assumption that errors in modelled moisture and sea surface state are less important than the cloud signal. In any case, we are still working in observation space and just transforming T^v and T^h to a quantity that is a rough indicator of the cloud amount. We compute $C37_{FG}$ and $C37_{OBS}$ for the FG and observed values and then the 'symmetric' or mean cloud,

$$\overline{C37} = \frac{C37_{FG} + C37_{OBS}}{2}. \tag{3}$$

As expected, selecting a cloudy FG ($C37_{FG} > 0.8$) gives a positive mean departure and selecting a completely clear FG ($C37_{FG} = 0$) gives a negative mean departure (Fig. 3a). Mean departure as a function of observed cloud is almost a mirror image. The crossover point occurs at $C37 \simeq 0.09$. This is roughly consistent with the median $C37$ in our sample, which is about 0.05 though it differs slightly for $C37_{FG}$, $C37_{OBS}$ and $\overline{C37}$ (Fig. 3c). For a clear FG defined by $C37_{FG} < 0.05$, a randomly chosen observation would on average be cloudier; for $C37_{FG} > 0.05$ a randomly chosen observation would on average be clearer. Hence, we would expect the crossover point somewhere near the median cloud amount, though any true bias between model and observations could affect this. Against $\overline{C37}$, the mean departure is generally quite close to zero. Deviations from zero probably indicate true biases between modelled and observed cloud.

In contrast to the simple binary model, standard deviation increases as a function of FG or observed cloud (Fig. 3b). We would expect FG or observation error to be larger in cloudy situations than in clear sky because: (i) both the model and the observation operator are less accurate in cloudy situations and (ii) the dynamic range of brightness temperature (TB) is much larger in cloud and rain than in clear skies. Against mean cloud,

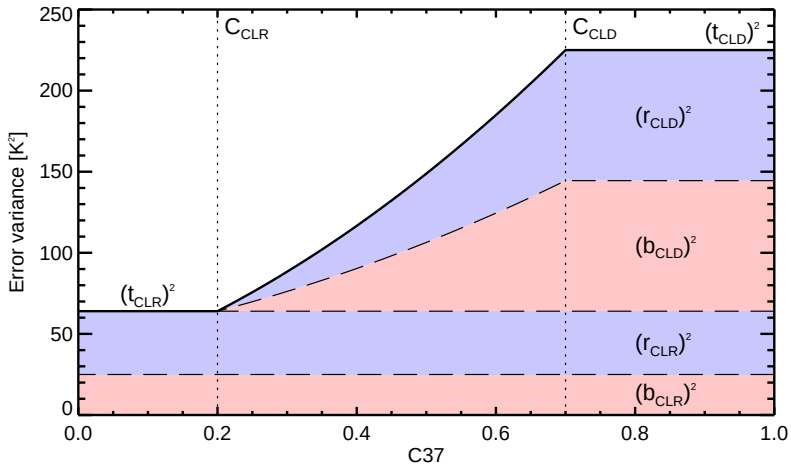


Figure 4: Error model for all-sky observations

standard deviation peaks at 17 K between 0.45 and 0.65, and declines for higher mean cloud amounts, where increasingly the FG and observation agree that cloud is present. For $\overline{C37} = 0$, standard deviation is 1.8 K, which is slightly smaller than in the other binning strategies. This would be expected since we are selecting on the basis that both observation and FG are clear.

In plotting Fig. 3 a and b, we have eliminated bins with fewer than 50 observations, though these are shown in Fig. 3c. So far we have ignored the bins with negative $C37$. This is a feature that occurs when calculating $C37_{OBS}$ when the FG estimate of atmospheric water vapour is too small, or surface windspeed too high, giving a $T^v_{clr} - T^h_{clr}$ smaller than observed $T^v - T^h$ (see Eqs. 1 and 2). The number of observations affected is only 5% of the total. However, the mean and standard deviation are larger when $C37_{OBS} < 0$ than in the bin where $C37_{OBS} = 0.0$. This is another sampling effect, in that $C37_{OBS} < 0$ implicitly selects for locations where there is no cloud or rain but the FG is still quite different from the observation. In the rest of this work, these locations are treated as if they had $C37_{OBS} = 0.0$.

Overall, Fig. 3 suggests that mean cloud, as represented by $\overline{C37}$, should be useful in all-sky assimilation as a symmetric predictor for bias correction and observation error.

2.3 Revised observation errors

2.3.1 A model for error in all-sky assimilation

The lesson from Fig. 3b is that the standard deviation of FG departures is well predicted by the mean cloud amount. We will use this to help estimate the observation error. Figure 4 shows how this is done. We start with the variance of the FG departures, which gives the 'total error'. This should be the sum of the observation and background error variance in observation space (see, e.g. Desroziers et al., 2005):

$$t^2 = r^2 + b^2, \tag{4}$$

where t, r and b are respectively the total, observation, and background error standard deviations. Here, as is conventional, observation error includes errors of representativity and of the observation operator.

We know how t varies with $\overline{C37}$ from Fig. 3b. To simplify, we parametrise as:

$$t(\overline{C37}) = \begin{cases} t_{CLR} & \text{if } \overline{C37} \leq C_{CLR} \\ t_{CLR} + \left(\frac{t_{CLD}-t_{CLR}}{C_{CLD}-C_{CLR}}\right)\overline{C37} & \text{if } C_{CLR} < \overline{C37} < C_{CLD} \\ t_{CLD} & \text{if } \overline{C37} \geq C_{CLD} \end{cases} \quad (5)$$

Here, t_{CLR} and t_{CLD} are the minimum and maximum total error standard deviation taken from Fig. 3b (roughly 2 K and 17 K), and C_{CLR} and C_{CLD} give the range over which the main increase in observation error takes place (roughly 0.05 to 0.4). We model this increase as a straight line and ignore the fall in standard deviations for very high values of $\overline{C37}$. This is done for simplicity but it is also a cautious approach to avoid giving small errors in very cloudy situations. In any case the number of observations with very high $\overline{C37}$ is small. The thick black line in Fig. 4 shows an example of the error variance t^2 that would be predicted by such a model, but for illustrative purposes, t_{CLR} and C_{CLR} are exaggeratedly large.

Next, we assume that observation and background error can be described as a sum of clear sky and cloudy components, i.e.

$$r^2 = r_{CLR}^2 + r_{CLD}^2, \quad (6)$$

$$b^2 = b_{CLR}^2 + b_{CLD}^2, \quad (7)$$

Based on the estimates of clear sky background error traditionally used in quality control in the ECMWF system, we simply assume that $b_{CLR} = 1$ K. Hence, clear-sky observation error r_{CLR} can be determined from Eq. 4, since we know the clear sky total error t_{CLR}. We then split the cloudy sky total error into background and observation parts using a factor α, so that:

$$r_{CLD}^2 = \alpha(t^2 - r_{CLR}^2 - b_{CLR}^2) \quad (8)$$

$$b_{CLD}^2 = (1-\alpha)(t^2 - r_{CLR}^2 - b_{CLR}^2) \quad (9)$$

This particular functional form for splitting the cloudy error has no objective basis but it is convenient because it allows us to tune the observation error in cloudy skies on the basis of a single number, α. However, the upper bound on observation error is objectively determined from the standard deviation of FG departures, which is valid under the assumption that observation errors are uncorrelated.

2.3.2 Inflation of observation error with distance

In the initial all-sky system, observation error was inflated according to the distance to the nearest model grid point:

$$r = r_n + (d/10)^2. \quad (10)$$

Here, d is the distance to the nearest grid point in km, r is the actual observation error and r_n the error before inflation, which is a function of channel and FG hydrometeor amount (Bauer et al., 2010). Error inflation with distance was intended to account for representativity errors, and to downweight the influence of all-sky observations in the first few inner-loop minimisations, which have very low resolution and hence typically large distances between observation and the nearest grid-point. However, this approach resulted in the assignment of very large observation errors (Fig. 4b, Bauer et al., 2010), and was the main reason that the constraint of lower-tropospheric moisture was only half as strong in the all-sky approach as in previous systems.

Figure 5 shows the standard deviation of FG departures in a control experiment at cycle 36r1 with a T799 resolution. Departures are binned according to distance from the T799 grid point. Ordinarily, each observation is matched to the closest grid point and none is further than 18 km from a T799 grid point. In Fig. 5, observations

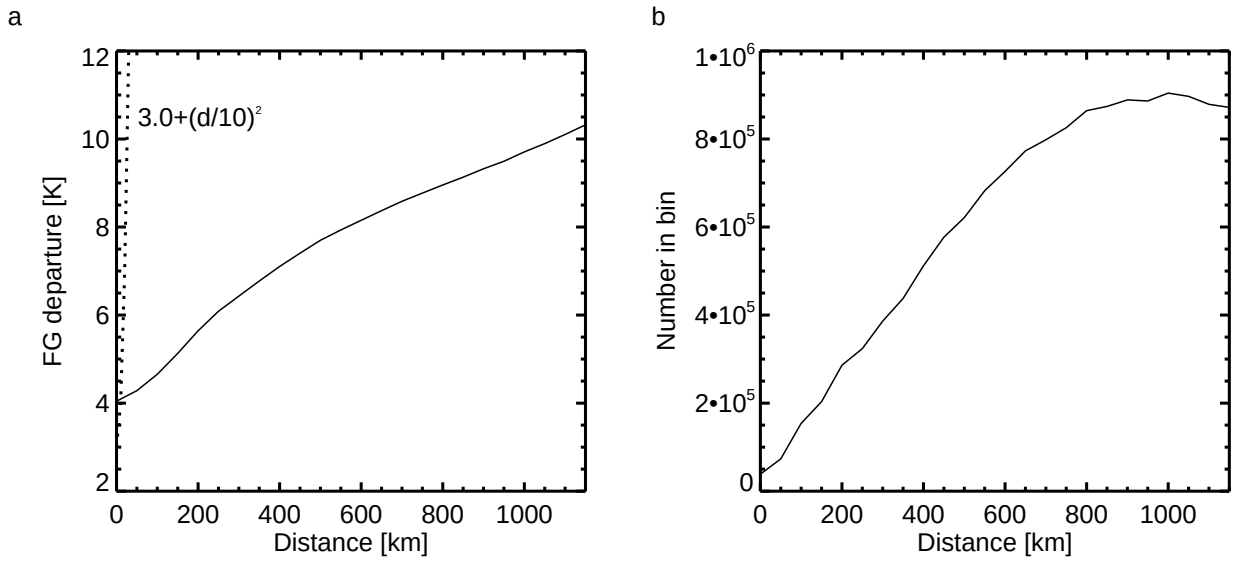


Figure 5: The effect of displacements between observation and model grid-point: ***(a)*** *Standard deviation of FG departure as a function of distance and* ***(b)*** *number per bin. This is based on a global sample of 38570 channel 19v SSM/I observations at 00Z 1 September 2009. The dotted line on panel a shows the observation error model used in the first all-sky implementation.*

have been matched with all available first guesses out to a radius of 1200 km. However, we have not recalculated FG brightness temperatures for every grid point, but instead we have just used the FG associated with the other assimilated SSM/I observations in the same satellite swath. Hence, standard deviations for distances greater than about 600 km are affected by a satellite-swath effect, which means that we get more matches in the N-S direction than in the E-W direction, which would likely cause standard deviations to increase compared to a more balanced sample. Nevertheless, we are mainly interested in smaller distances, where our approach is reliable.

The dotted line in Figure 5 shows the observation error model of Eq. 10. For channel 19v, r_n was set to 3 K in all sky conditions. The observation error increases far more rapidly with distance than the standard deviation of real FG departures. The maximum possible distance between observation and grid point is 150 km in the current operational ECMWF configuration. This occurs in the first inner-loop, which runs at T95. FG error standard deviations increase by only 27% over this distance. At the final inner-loop resolution of T255, the maximum distance is 55 km, which corresponds to only a 6% increase in standard deviation. Distance from grid-point appears to be only a minor factor in the quality of agreement between model and observation.

The old approach also had the disadvantage of modifying the cost-function as the model resolution changed, meaning that the shape of the cost-function was not consistent between high-resolution outer loop and low-resolution inner-loop. Future Integrated Forecasting System (IFS) cycles will have increasingly high inner-loop resolutions, and even at the current resolution the error associated with displacements between observation and model grid point is not great. Hence, we have abandoned the inflation of observation error with distance in the new approach. This is mainly to give more weight to the all-sky observations, but it is also much simpler to understand, and gives a consistent cost-function for 4D-Var to minimise.

Table 1: Parameters for the new observation error model.

Satellite	Channel	t_{CLR}	t_{CLD}	C_{CLR}	C_{CLD}
SSM/I	19v	2.0	15.0	0.05	0.45
	19h	3.5	30.0	0.03	0.53
	22v	3.0	8.0	0.05	0.45
	37v	3.0	18.0	0.03	0.4
	37h	99.0	300.0	0.02	0.45
	85v	3.0	18.0	0.0	0.5
	85h	99.0	900.0	0.0	0.15
AMSR-E	19v	2.0	18.0	0.05	0.45
	19h	3.5	36.0	0.00	0.55
	24v	3.0	10.0	0.05	0.45
	24h	5.0	20.0	0.03	0.5
	37v	3.0	16.0	0.03	0.4
	37h	99.0	300.0	0.0	1.0

2.3.3 Comparison of old and new observation errors

From early experiments, it seemed that the best results came with the cloud error tuning parameter α (Eq. 8) set to 1. In other words, apart from a near-token 1 K background error, all error in FG departures is assumed to come from the observations. This is what is implemented in cycle 36r3; section 4.1 justifies this choice with some a-posteriori testing. The parameters in the new total error model (Eq. 5) were determined by fitting $t(\overline{C37})$ by eye to the FG departures from the early test experiments, as illustrated in Sec. 2.3.1. Table 1 lists the resulting values. Errors in channel 37h and 85h were deliberately set to very large values to prevent assimilation. These channels are not assimilated in either the old or the new approach, because doubts remain as to the accuracy of the sea-surface emissivity model in these channels, and the sensitivity to cloud and rain is extremely large. Figures 6 and 7 show the SSM/I and AMSR-E observation errors in the original and revised approaches, binned as a function of mean cloud.

In the old approach, nominal errors were small in both clear and cloudy areas in the 19 GHz channels, but the effect of inflation with distance was to add around 8 K to the observation error. Hence, in clear-sky areas, the new approach reduces observation error from 11 K to 1.7 K in channel 19v and or 13 K to 3.5 K in channel 19h. However, in the areas most affected by cloud and rain, observation errors have become a little larger than before, reaching either 15 K or 30 K (v and h) in the new approach. In the SSM/I 22 GHz channel, where the water vapour signal is quite strong relative to that of the hydrometeors, the new approach gives smaller observation errors for all states of the atmosphere. In the old approach, extremely large observation errors where chosen in channel 37v and 85v in cloudy areas in order to allow assimilation only in clear areas. In the new approach, these channels are treated like the lower frequency channels, with clear sky errors of 2.8 K increasing to 18 K in the most cloud-affected areas (errors are identical in SSM/I 37v and 85v). The picture for AMSR-E is very similar, except that we do not use the 85 GHz channels. Overall, the new approach gives a much greater weight to the all-sky observations.

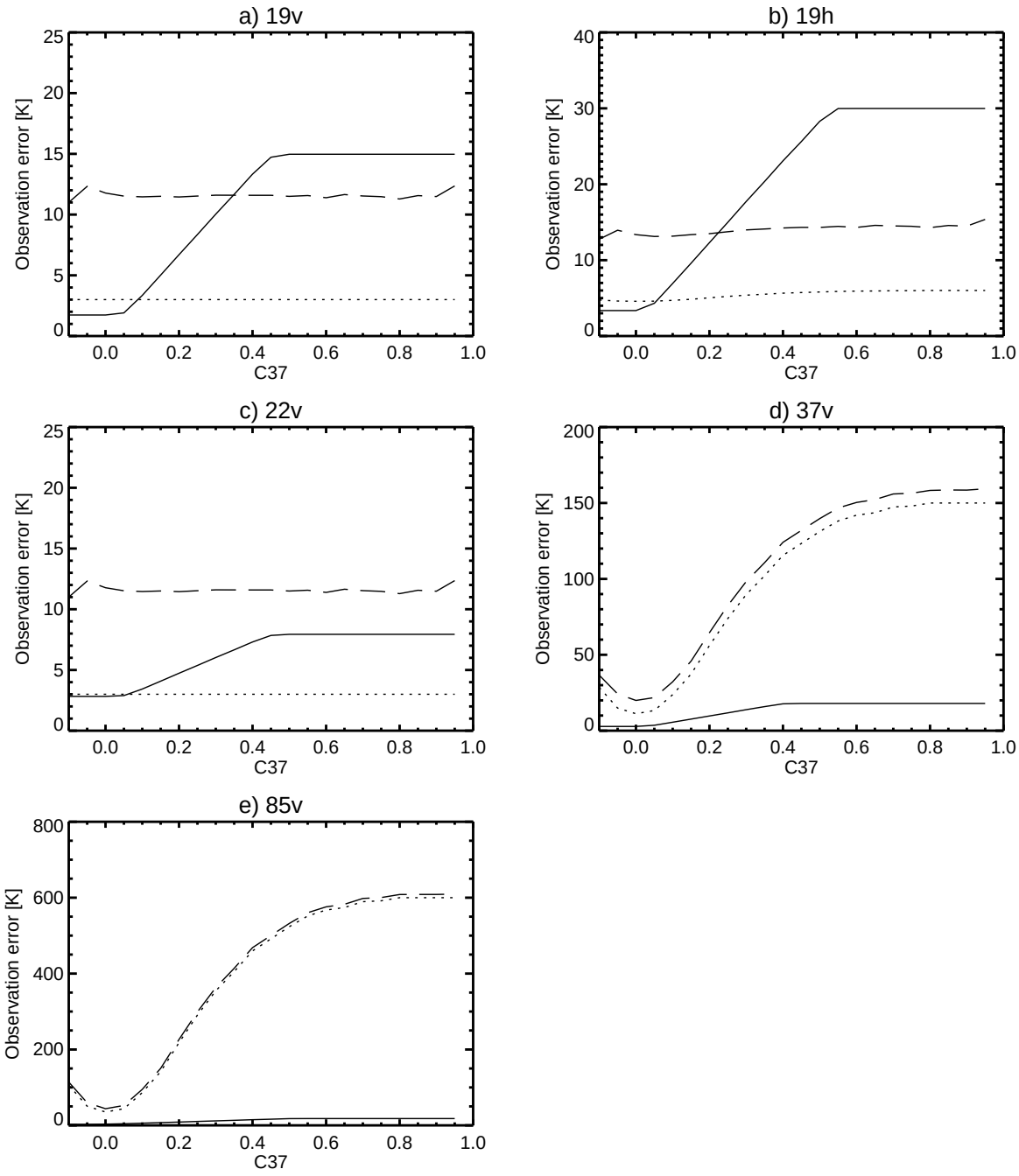


Figure 6: Original and revised observation errors for SSM/I, as a function of mean cloud amount, $\overline{C37}$. Dotted: original errors before inflation (nominal error, r_n); Dashed: original errors inflated according to distance from the T255 grid; Solid: revised errors, which are constant through the minimisation. Neither 37h nor 85h is assimilated in either version. Sample is from 1-2 September 2009. On panels d and e, minimum and maximum observation errors in the new approach, which are too small to see on the graph, are 2.8 K and 18 K.

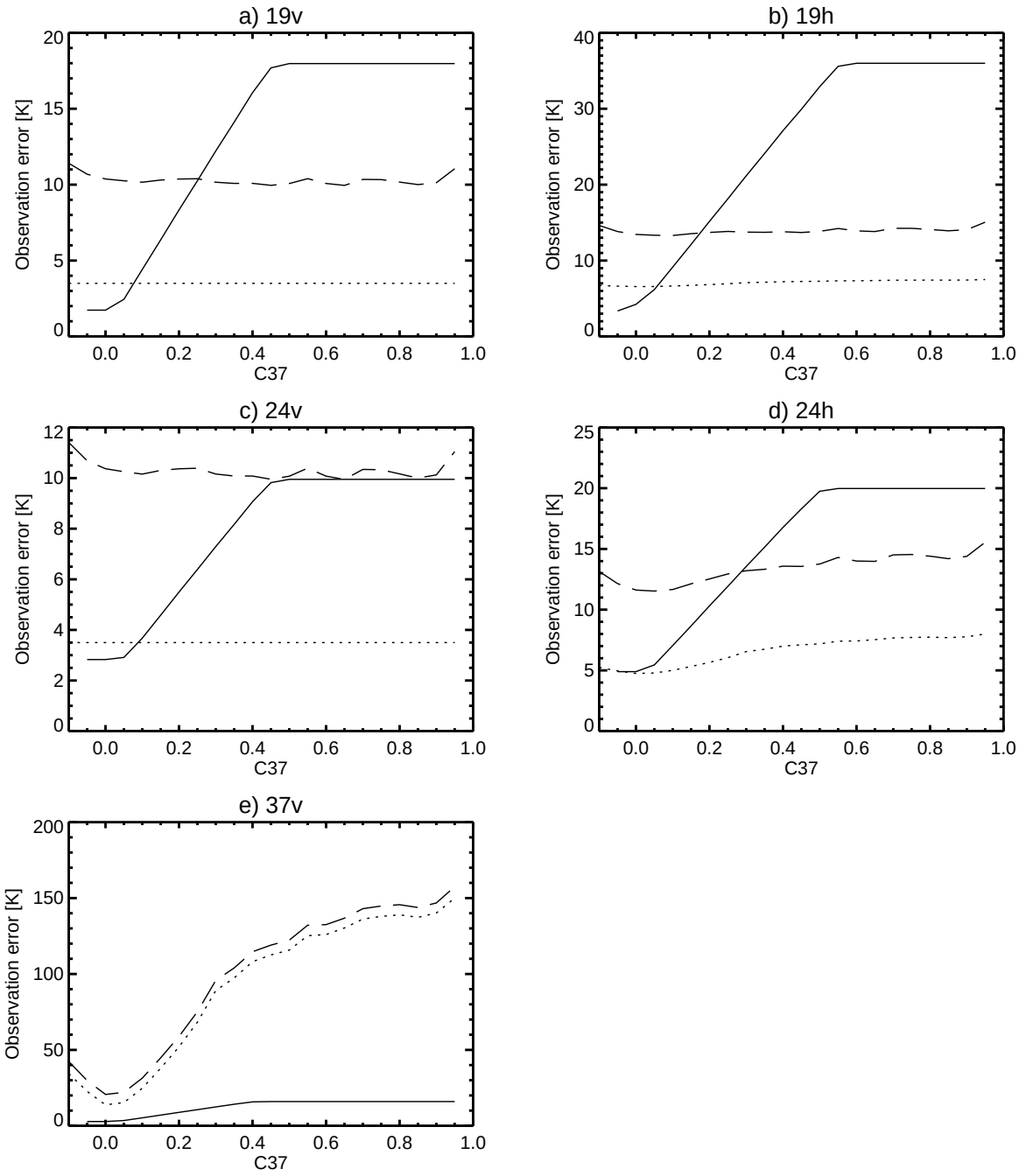


Figure 7: As Fig. 6 but for AMSR-E

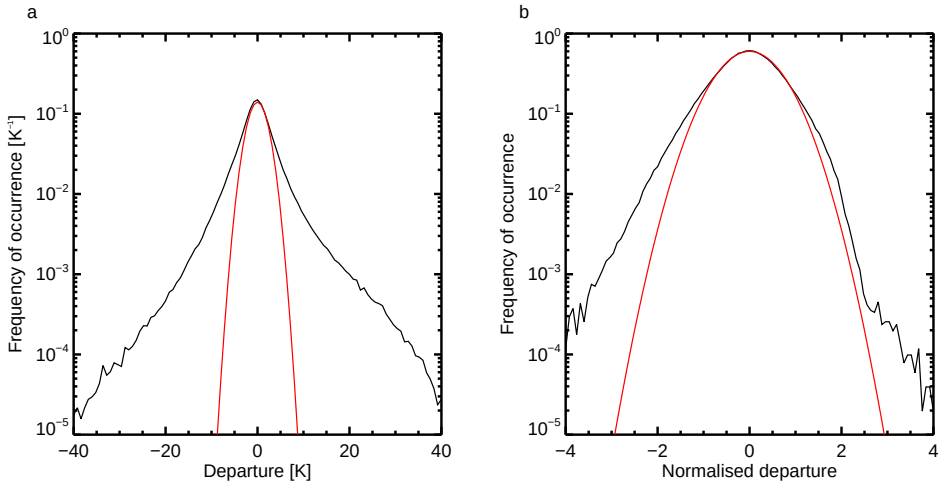


*Figure 8: Histograms of SSM/I channel 37v FG departures **(a)** as brightness temperatures **(b)** normalised by the symmetric error model. Sample is from 27 June to 6 July 2009. The red curves show Gaussians fitted by eye to the peak of the distribution.*

2.4 Quality control

In the IFS, background quality control (BgQC, Järvinen and Unden, 1997) rejects observations with large normalised departures, i.e. where

$$\frac{d}{\sqrt{r^2+b^2}} > \delta \tag{11}$$

Here, d is the bias-corrected FG departure. The rejection threshold δ is set to 2.5 for all-sky observations. In the initial all-sky system, b did not vary as a function of cloud and was appropriate only for clear-sky observations, giving an unrealistically small value between 0.04 K and 2 K. As a result BgQC rejected a large proportion of cloudy observations (Bauer et al., 2010). The new error model is used to provide BgQC with values of r and b consistent with the known variation of FG departure as a function of symmetric cloud.

Figure 8 shows histograms of actual and normalised FG departures for SSM/I channel 37v, along with roughly fitted Gaussians. The actual histogram emphasises how hard it is to do all-sky quality control in brightness temperature space. The distribution of departures is very non-Gaussian, with large departures (e.g. as much as 40 K) occurring where observation and model disagree in terms of the rain or cloud amount. It is impossible to use a threshold to distinguish erroneous observations from real information. When normalised, the departures become far more Gaussian. Normalised departures with magnitudes greater than 2.5 are infrequent and quite often associated with gross observation error. An example is shown in Fig. 9. SSM/I suffers from occasional bad scan-lines, such as those around 40N, 165W. However, only when the departures are normalised can the threshold check be used to identify the problem. The remaining red areas on Fig. 9b indicate the problematic scan lines. The large positive departures around 5N, 150W in Fig. 9a indicate a cloud system that is observed but not present in the first guess, and these observations would be rejected in the old approach but not in the new.

There are also normalised departures below -2.5 but these do not appear to be associated with gross observation error, as they correspond to widely and infrequently scattered observations. This may indicate a remaining inaccuracy in our error model, but the numbers affected are very small: just 0.25% of data are rejected by BgQC with a negative departure (an additional 0.1% of observations are rejected with a positive departure).

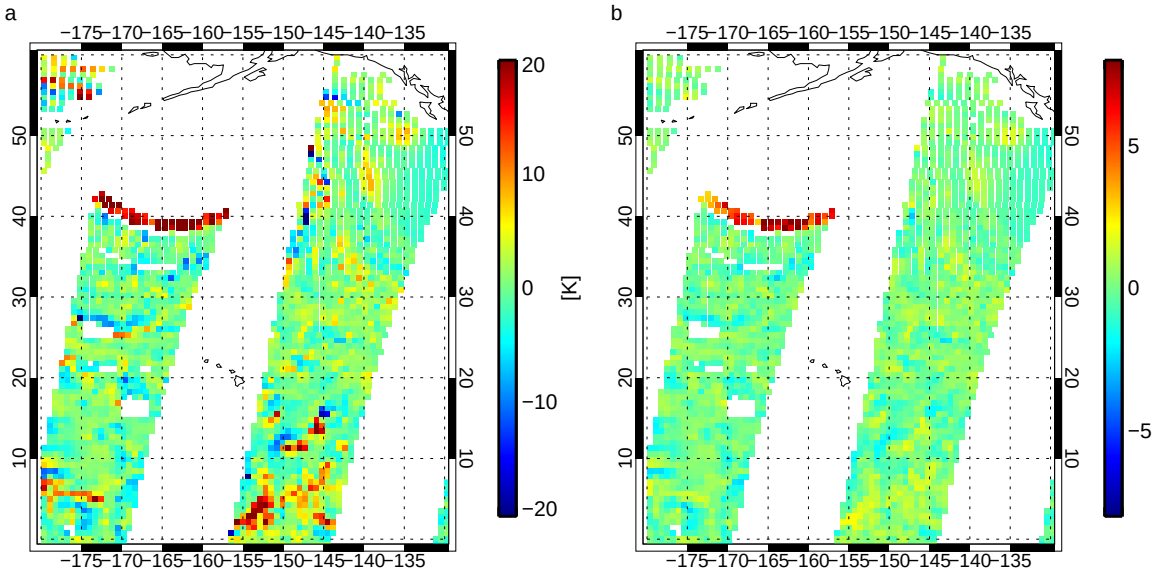


*Figure 9: Maps of SSM/I channel 37v FG departures **(a)** as brightness temperatures **(b)** normalised by the symmetric error model, on 1 July 2009, looking at observations from the Defense Meteorological Satellite Program (DMSP) F-13 satellite only.*

It is worth briefly going back to the issue of Gaussianity illustrated by Fig. 8. Data assimilation assumes Gaussian error statistics and it is common to check this with histograms of departures. However, as Fig. 8 illustrates and Eq. 11 implies, it is only correct to look at un-normalised departures if the standard deviation of background and observation error is constant. With temperature-related observations, this may be roughly true, but it is not true for all-sky observations. It has often been stated that cloud and rain assimilation will be difficult because errors may not be normally distributed (e.g. Errico et al., 2007). Here we see that with an appropriate model for observation and background errors, this is not the case. Along with the ability of incremental 4D-Var to assimilate the mildly non-linear all-sky observations (Bauer et al., 2010), this is more evidence that cloud and rain-affected observations can be used without violating the assumptions on which our assimilation system is based.

2.5 Bias correction

We have seen in Sec. 2.1 that only symmetric predictors should be used for the correction of cloud-related biases, and that up until now, asymmetric predictors have been used. With a symmetric approach, for the first time we have a hope of doing a proper bias correction for cloudy observations. In practice, it is still very difficult. A number of experiments were run with $\overline{P37}$ and other cloud- or rain-based bias predictors in Variational Bias Correction (VarBC, Dee, 2004). However, even with appropriate tuning the bias corrections took months to spin up. Without going into detail, the reasons for this are thought to be:

1. Bias as a function of symmetric cloud amount is small (see e.g. Fig. 3b), especially compared to the large standard deviations of FG departures encountered in cloudy areas.

2. ‘Cloudy’ biases end up being determined by a very small number of observations at the extreme cloudy end of the distribution (see Fig. 3c) which, on a day-to-day basis are too few to provide a representative sample.

3. The bias estimation is vulnerable to interactions with quality control (e.g. Auligné and McNally, 2007) because the estimated bias is based on a small number of observations that are often very close to the quality control threshold.

Even when the bias corrections had spun up, VarBC did not correct the main geographical patterns of all-sky bias. As will be mentioned in the next section, real biases exist in cloudy and rainy areas, but they are highly situation-dependent and it would be very difficult to come up with appropriate predictors. These biases are far more complicated than a simple function of cloud amount. The ultimate solution is to identify the real cause of the bias, be it a problem with model or observation operator. As a result, we decided to continue without any cloud or precipitation related predictors in VarBC. In the short term, situations affected by large uncorrected biases simply have to be screened out.

2.6 New screening criteria

There are two problems for which screening criteria have been developed. The first occurs in cold dry air moving equatorward behind cold fronts in the winter-hemisphere. We will describe this as the 'cold-sector' bias. Here, the model appears to have too little cloud, and in particular, too little water cloud. This results in positive FG departures of 5 to 10 K, which have been a feature of all-sky and 1D+4D-Var for a long time (see e.g., Geer et al., 2009). Initial comparisons to CloudSat appear to confirm the hypothesis that this is a model problem, but whatever the true explanation, we cannot assimilate the affected data, for we will simply be assimilating bias.

Figure 10 shows an example for channel 19v on SSM/I. There are large contiguous areas with positive departures near the bottom of the plot at 170W, 70W, 20E and 160E (panel a). These areas are identified (panel b) by a combination of three criteria:

$$\mathrm{TCWV} < 15 \text{ kg m}^{-2}; \tag{12}$$

$$\frac{\mathrm{LWP}}{\mathrm{LWP}+\mathrm{IWP}} > 0.5. \tag{13}$$

$$\mathrm{LWP} + \mathrm{IWP} + \mathrm{RWP} + \mathrm{SWP} > 0.01 \text{ kg m}^{-2}; \tag{14}$$

Total column water vapour (TCWV), cloud liquid water path (LWP) and cloud ice water path (IWP), rain and snow water paths (RWP, SWP) are taken from the FG model state. The TCWV criterion selects cold, dry, winter air masses. Biases seem to occur in areas where the model produces ice cloud but not liquid cloud, which is checked by the second criterion. The third makes sure that a non-trivial amount of hydrometeors are present. The screening is far from perfect, and to catch all the areas where the bias occurs, we have had to relax the conditions to the point where many unaffected areas are are also caught.

The use of model quantities makes this an asymmetric condition, and hence this could cause a sampling bias. However, we believe the risk from a sampling bias is smaller than the danger of assimilating large uncorrected biases into the system. Also, the 'cold-sector' screening normally selects big contiguous regions with extents of 500 to 2000 km (panel b). This is much larger than any conceivable displacement error, so sampling error should be relatively insignificant.

The second problem occurs principally in tropical convection with large amounts of precipitating snow or ice. Figure 11 shows histograms of observed and FG channel 37v brightness temperatures, classified by the maximum 'cloud' amount observed in the sample,

$$\mathrm{C37}_{\mathrm{MAX}} = \max(\mathrm{C37}_{\mathrm{OBS}}, \mathrm{C37}_{\mathrm{FG}}) \tag{15}$$

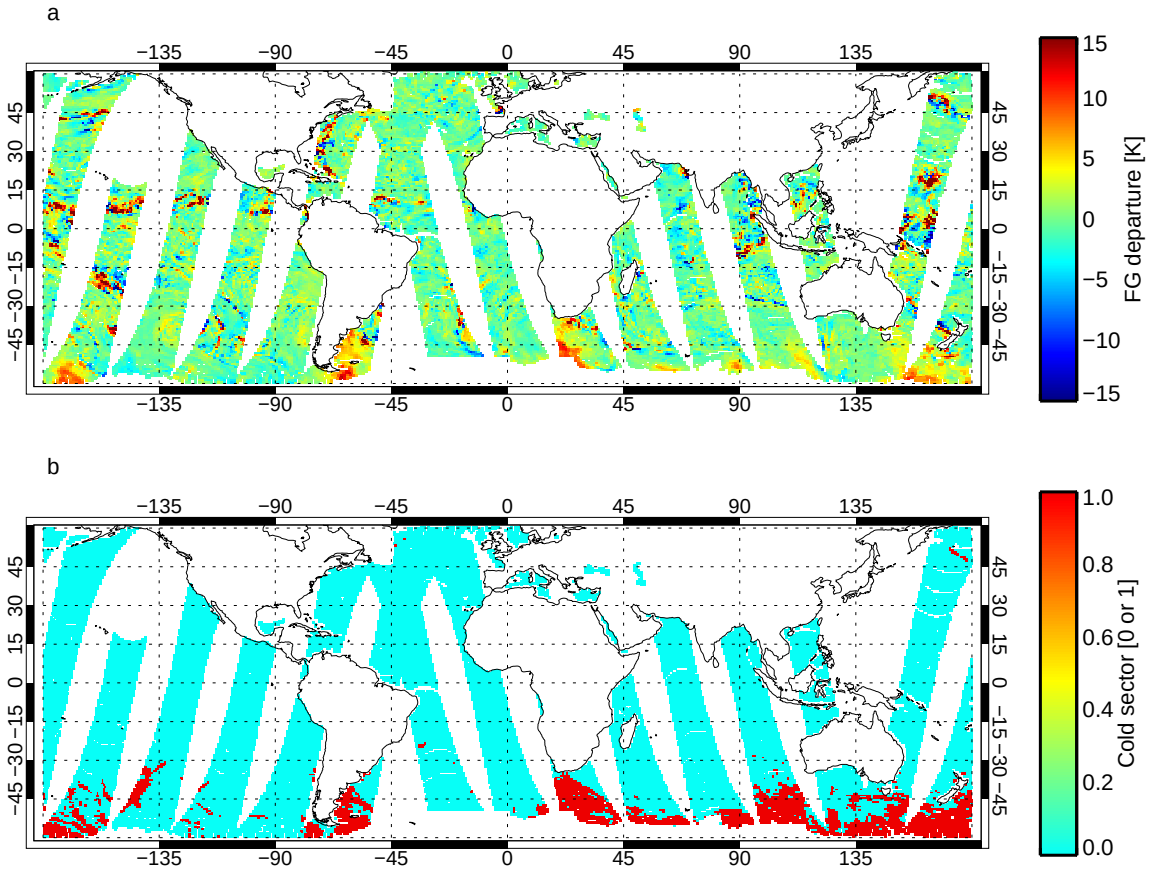


Figure 10: ***(a)*** *DMSP F-13 SSM/I channel 19v FG departures at 12Z on 22 July 2009* ***(b)*** *Cold-sector screening flag, with red indicating areas that are not assimilated.*

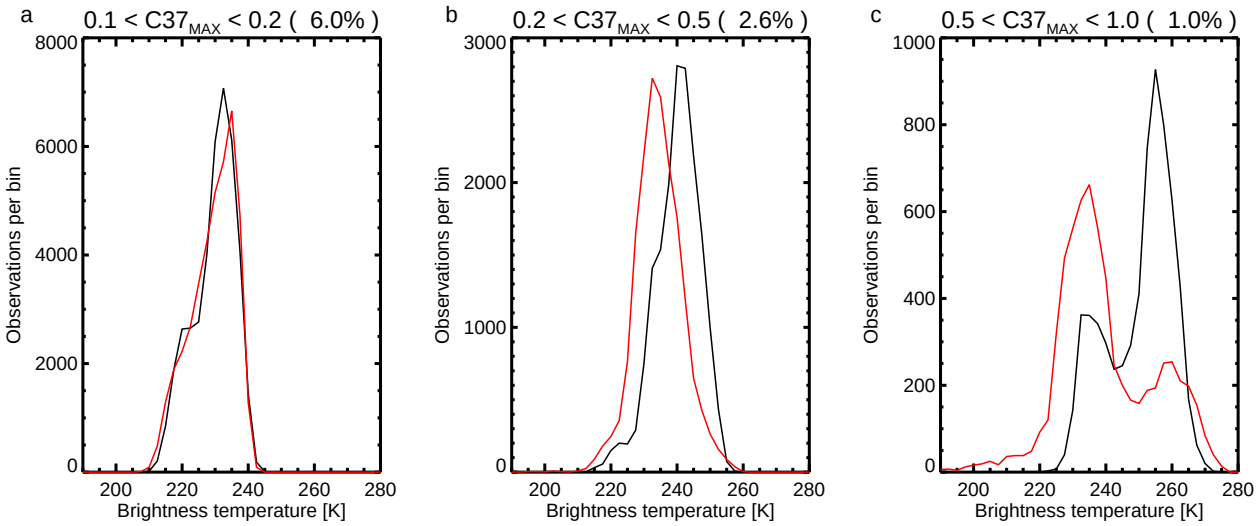


Figure 11: Histograms of observed (black) and bias-corrected FG (red) AMSR-E channel 37v brightness temperatures, classified according to hydrometeor amount, using $C37_{MAX}$ = max($C37_{OBS}$, $C37_{FG}$). Based on a sample from 1 to 10 October 2009, restricted to the tropics (20S to 20N). The percentage in brackets is the fraction of the sample shown in the histogram.

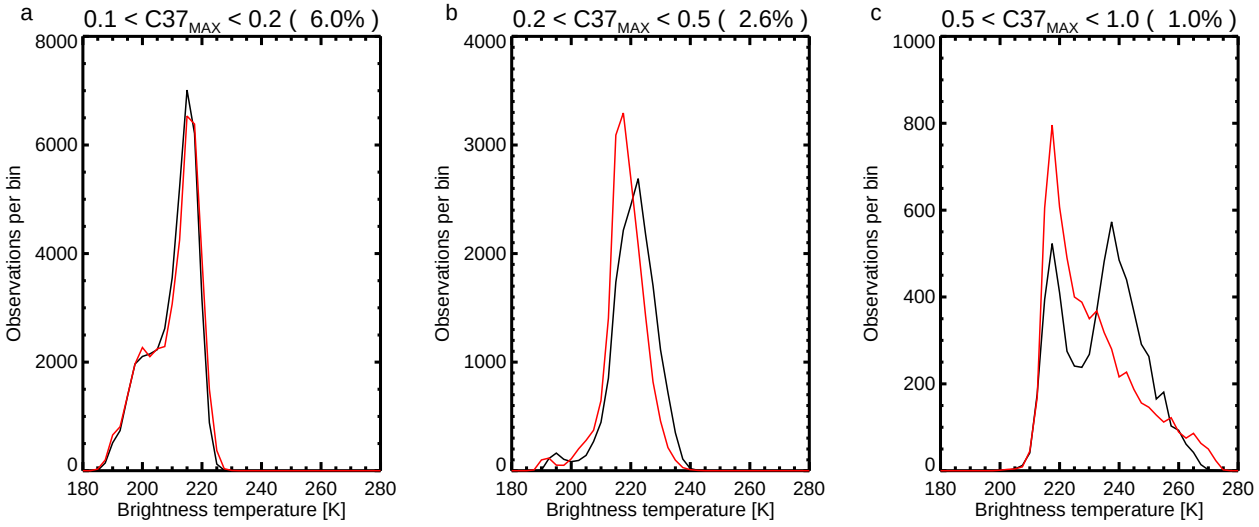


Figure 12: As Fig. 11 but for AMSR-E channel 19v.

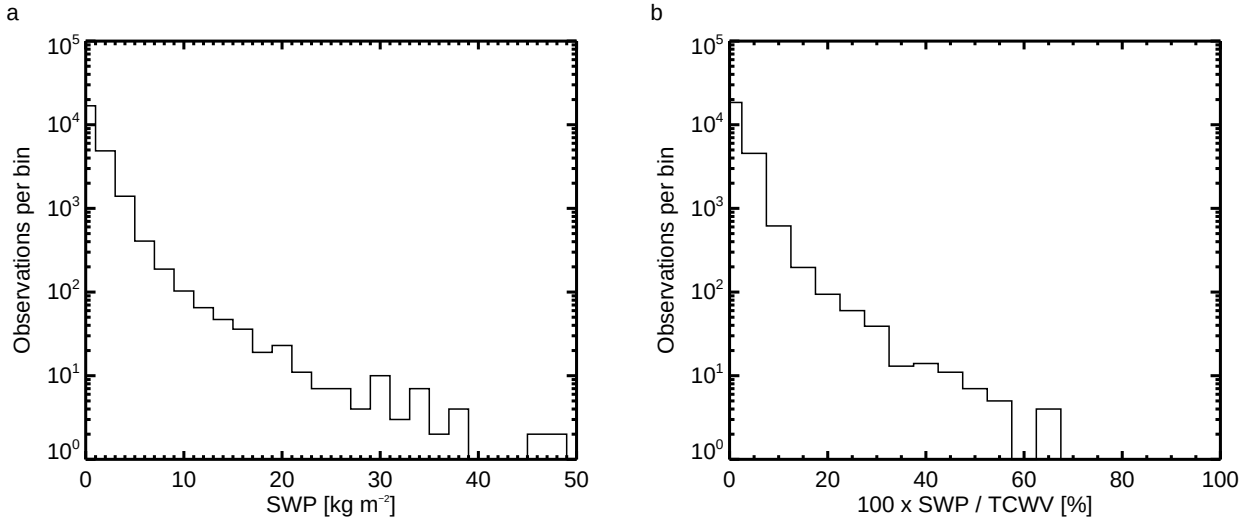


*Figure 13: Histograms of **(a)** SWP in kg m^{-2} and **(b)** SWP as a percentage of TCWV, for observations with $C37_{MAX}$ between 0.2 and 1.0, which corresponds to 3.6% of tropical (20S - 20N) AMSR-E observations between 1 and 10 October 2009.*

The sample has been restricted to tropical areas (30S to 30N). We ignore the 90% of observations with lighter cloud or clear sky. Panel a shows the 6% of situations with $C37_{MAX}$ between 0.1 and 0.2. The agreement between histograms is quite good. However, with the greater amounts of cloud or precipitation in the samples in panels b and c, there is a definite offset in the histograms, with the FG appearing typically 10 K colder than the observation. There are also many FG TBs lower than 220 K in the most precipitation-affected case (panel c), but TBs are never lower than 220 K in the observations. A similar figure for channel 19v (Fig. 12) does not show perfect agreement either, but any shifts are much less obvious. One of the main differences between these channels is that at 37 GHz, there is sensitivity to scattering from snow, whereas 19 GHz is relatively insensitive to scattering. In contrast to the effect of cloud and rain, which typically increases brightness temperatures, scattering typically reduces them. Our hypothesis is that the FG TB is subject to excessive scattering.

The radiative transfer code, RTTOV-SCATT[1] (Bauer et al., 2006) may produce excessive scattering due to the assumption that all snow particles are Mie spheres. The forecast model does not provide any information on microphysical properties of snow, so assumptions have to be made, but they may be inaccurate. However, the FG also seems to contain a large amount of snow. Figure 13a shows a histogram of FG SWP in the sample $0.2 < C37_{MAX} < 1.0$, for which the modelled 37 GHz TBs are too low. Panel b shows the SWP as a percentage of the TCWV at the same points. 4% of the sample show a SWP of larger than 5 kg m^{-2}; 3% show a SWP larger than 10% of the TCWV. These are surprisingly large amounts of snow, considering they represent an average over the 25 km by 25 km grid box associated with a T799 resolution. However, it would be very difficult to investigate further, because we would need measurements from an independent instrument type with global sampling, and none currently exists that can accurately measure the snow column in deep convection. Hence, we cannot yet say if this is a problem of the model or of the observation operator.

For cycle 36r3, the 'excess snow' areas are identified on the basis of the FG SWP being larger than 0.5 kg m^{-2}. Only channels at frequencies greater than 30 GHz are screened out; lower frequencies continue to be assimilated. About 12% of the high frequency observations are removed by this criterion. Again, the use of FG SWP is asymmetric and liable to lead to undesirable sampling effects. For cycle 36r4, we hope to replace it with a

[1]Radiative Transfer model for Television infrared Observation satellite operational Vertical sounder Scattering package

Table 2: Averaging of channel 37v observations in Figure 14.

Number	Approximate resolution (km x km)
1	16 x 27
5	25 x 25
10	35 x 35
20	50 x 50
50	80 x 80
100	110 x 110
200	160 x 160
400	225 x 225
800	320 x 320

criterion based on $C37_{MAX}$, as presented in the analysis here.

2.7 Spatial sampling

There was no time to look at spatial sampling in detail when implementing the initial all-sky approach, although we did find that AMSR-E needed to be superobbed to avoid bringing unwanted small-scale noise into the analysis (Bauer et al., 2010). Here we revisit the issue in more detail.

The dynamical part of the model works in spectral space, at a particular spectral resolution, e.g. T255 or T799. However, the model physics (and the all-sky observation operator) run in normal space on a corresponding Gaussian grid. There is a defined mapping between the spectral and Gaussian grids, so T799 uses a grid with boxes of 25 km by 25 km at the equator, and T255 uses grid-boxes of roughly 78 by 78 km. Model cloud is defined by the grid-box mean hydrometeor amount and the cloud fraction.

The microwave field of view depends on frequency and the size of the instrument's antenna. Antenna sizes have increased in recent years and AMSR-E has a 1.6 m diameter antenna, which is the largest of any of the instruments we use. This results in a high spatial resolution, with a 16 km by 27 km field of view at 19 GHz and 8 km by 14 km at 37 GHz (Kawanishi et al., 2003).

Figure 14 illustrates the effect of spatial resolution on microwave radiances. AMSR-E channel 19v FG and observations have been binned according to brightness temperature. The lower end of the scale corresponds to dry, cold airmasses near the poles, and the upper end represents tropical convection. The FG has been generated from a T255 model. AMSR-E observations have been averaged together to achieve different resolutions. Table 2 lists the number of observations and the approximate resolution this corresponds to. Note that averaging five observations does not broaden the resolution by much. This is because the observations are oversampled to 10 km along track and 9 km across-track at 19 GHz. Another issue is that the conical scan pattern gives a higher observation density on the edges of a swath than in the middle, so when we average a particular number of observations, the effective resolution can vary. Also, for very small numbers of observations, it would be necessary to consider the shape of the antenna gain pattern to get a true idea of the effective resolution. None of these inaccuracies detract from the main aim of this figure, which is simply to demonstrate the impact of changing resolution.

Figure 14b focuses on the upper end of brightness temperatures and uses a log scale. The maximum observed TB drops from about 275 K in raw data to 265 K when the observations are averaged to 110 km by 110 km (broader resolutions do not produce enough observations for reliable statistics in bins above 250 K). As mentioned, the very highest TBs are associated with localised tropical convection. As the spatial scale broadens,

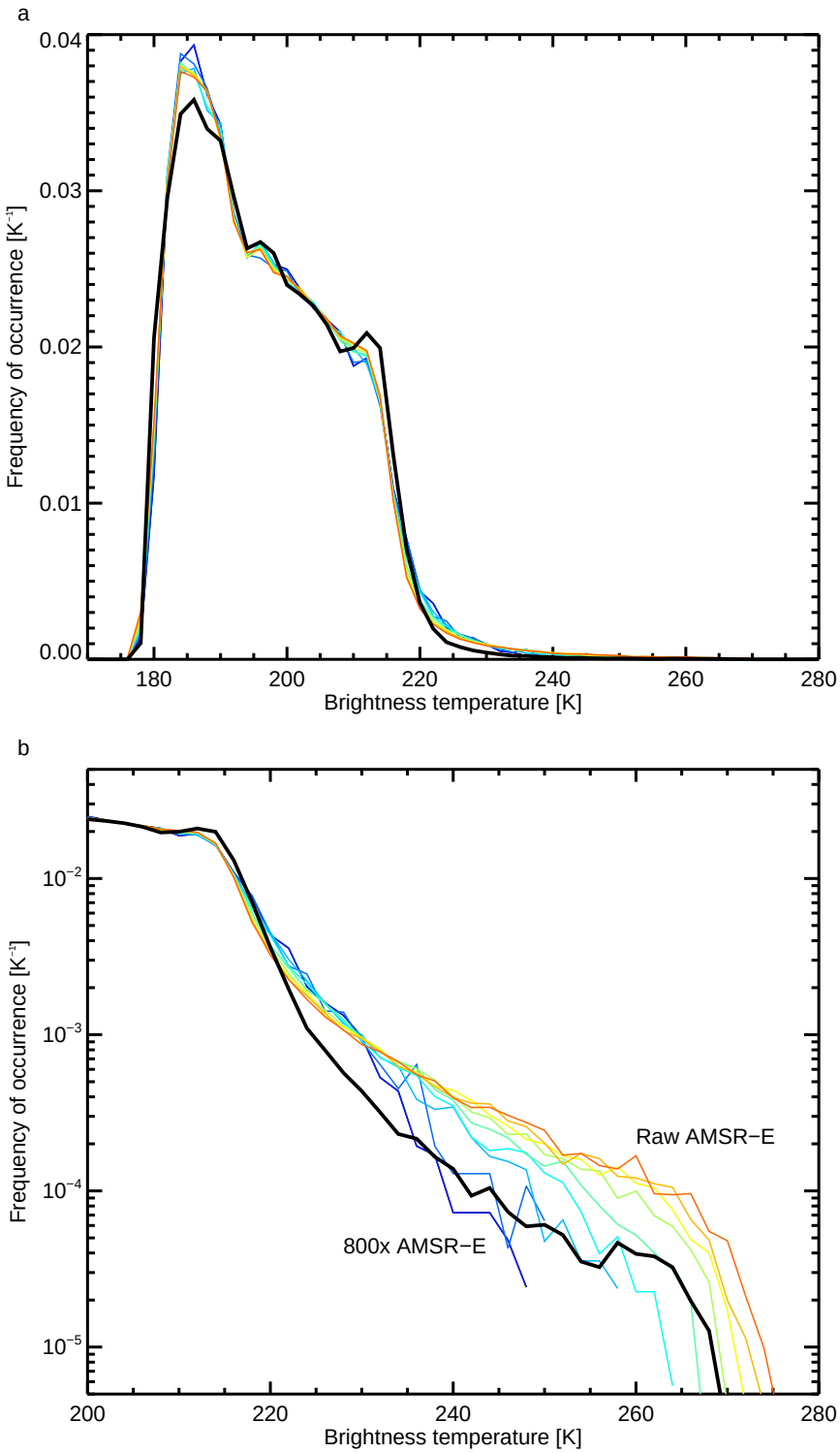


Figure 14: Histogram of AMSR-E channel 19v observations (coloured) and FG (black), based on the period 12 - 16 June 2009: ***(a)*** *linear scale, all observations;* ***(b)*** *log scale zoom on region between 200 K and 280 K. FG has been generated using a T255 forecast model. Different numbers of observations have been binned and averaged in Gaussian grid-boxes to give the rainbow-coloured lines. The numbers of observations binned are: 1 (red), 5, 10, 20, 50, 100, 200, 400, 800 (dark blue). See Table 2.*

these features are averaged out. However, there is a corresponding increase in the number of observations with TBs of around 225 K. It is clear that we must be careful to match the spatial scale of model and observations, or we risk producing artificial biases.

The agreement between the observed and FG histogram is relatively poor. This is likely explained by a bias between model and observations: there appear to be too few intense convective events (220 K to 280 K) but too many general cloud or rainfall events (around 215 K). This is true even when the observations are averaged to 320 km by 320 km. Possibly the 'effective' resolution of the model in cloud and rain areas is even broader than this. This result is not limited to the low resolution of T255 we used to generate Fig. 14. The same effect is seen in Figs. 12b and c, where the FG comes from a T799 model.

It was originally hoped that comparing histograms, as done in Figure 14, would help decide the appropriate averaging resolution for the observations. Unfortunately, the presence of bias means this is not a useful guide. Instead, we have to fall back on theoretical considerations. The outer loop resolution (e.g. T799 or T1279) is the true resolution of the analysis and we should ideally use this. However, averaging to this scale would result in a vast number of observations, which would badly affect computational performance. However, the inner loops (and hence the increments) never have a finer scale than T255. Hence, for performance reasons, we chose T255 as the averaging resolution for the new all-sky system. Nevertheless, this means that there is a mismatch between the T799 resolution of the outer-loop FG and the T255 resolution of the observations.

Figure 15 helps to justify our choice. It shows a single swath of AMSR-E data which encompasses tropical convection (around 5S), trade cumulus (around 20S) and a midlatitude front (around 45S). Observations have been binned on either the T799 Gaussian grid (panel a) or the T255 Gaussian grid (panels b and c, which are identical). As expected, the convective points are smeared out at the lower resolution. The FG has been generated at T799. When we ignore the resolution mismatch by computing the departures between T799 FG and T255 observations, we are effectively sub-sampling the T799 FG to T255 resolution (panel e). It would be more correct to average the T799 FG to T255 resolution (panel f), just as we average the observations. Sub-sampling retains some of the high-resolution variability, which is undesirable, while averaging removes it.

Panels g to i show the FG departures computed using the different approaches. The choice between subsampling and averaging to T255 makes very little real difference. This is because the cloud and rain features in the T799 model are quite smooth, i.e. they have an effective resolution which is much broader than that of the T799 Gaussian grid. Also the spatial scale of the FG departures is quite broad, so there appear to be strong correlations between errors in neighbouring pixels at T255. The error caused by subsampling appears to be much smaller than the amplitude of the FG departures. In summary, using T255 observations in a T799 model appears to cause few real problems, but it has the benefit of substantially reducing the amount of computer processing required. A future refinement would be to include an averaging from T799 to T255 as part of the observation operator, prior to running the radiative transfer.

3 Method

The all-sky upgrade for cycle 36r3 combines these changes:

1. We no longer think about matching a single observation to a single model grid point, but instead we average the observations together on a suitable model grid. In practice, and despite theoretical objections, we use the resolution of the final inner loop, which is T255. There is no longer any need for thinning, so the IFS thinning (e.g.'new_thinn') has been turned off.

2. There is no need to increase observation error as a function of distance from grid point. This has been

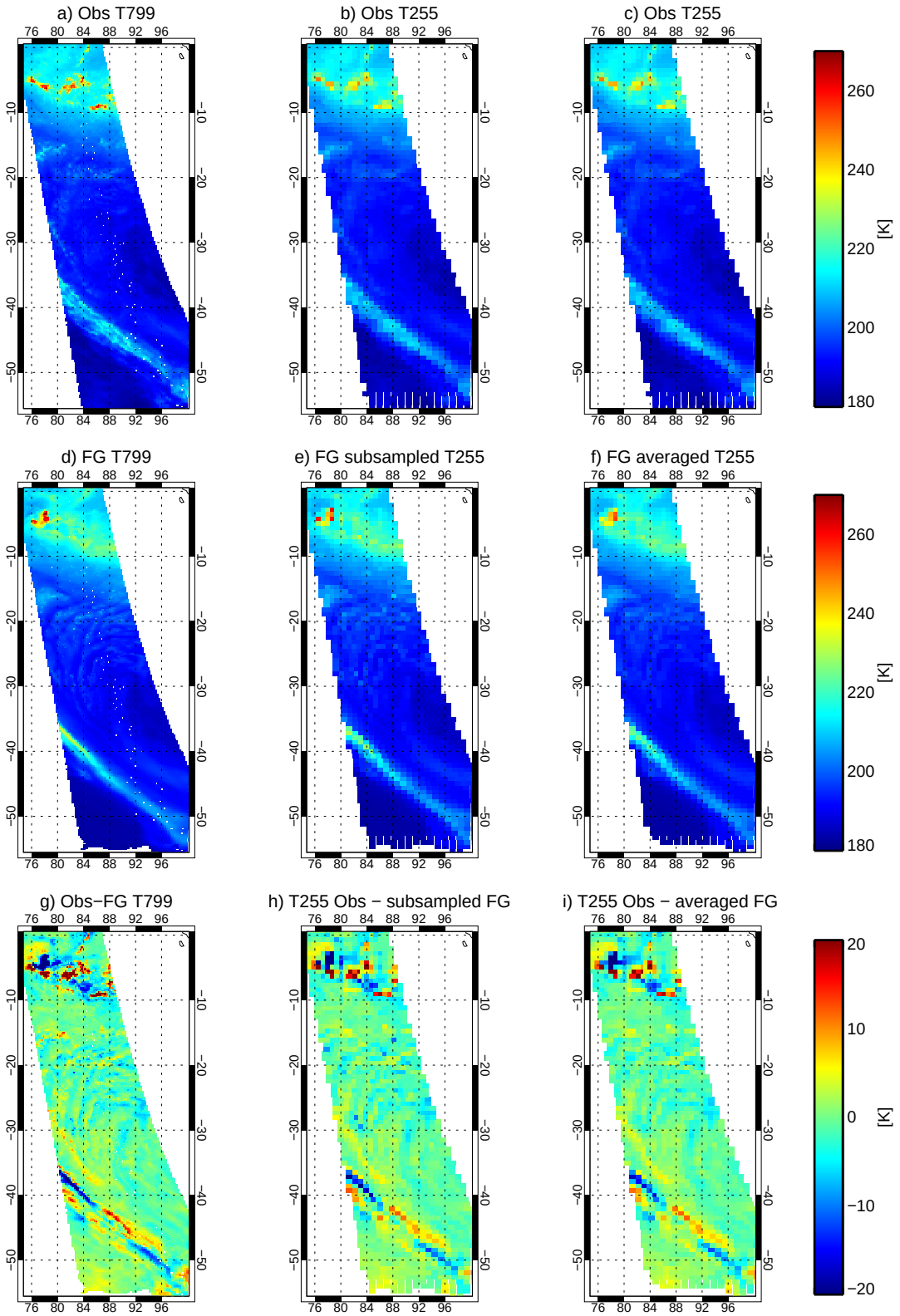


Figure 15: Effect of resolution on AMSR-E channel 19v observations, first guess and departures. White spots in panels a, d and g occur where there are not enough observations per grid box to satisfy our superobbing requirements.

Table 3: Experiments.

Name	Description	ECMWF ID
All-Sky-New	Revised all-sky approach	f8zs
Control	36r1 control	f8g8
Control-Off	As Control, but with microwave imagers switched off	f8g9

abandoned, so observation errors are now much smaller in the inner loops, and the error is constant through all of 4D-Var.

3. BgQC uses the symmetric error model to normalise FG departures. Hence, BgQC allows a lot more observations into the system, particularly in cloudy areas, but is still capable of eliminating bad data.

4. Observation errors are derived from the symmetric model.

5. Observations affected by cold-sector bias are no longer assimilated.

6. Observations at frequencies over 30 GHz with heavy snowfall in the FG are no longer assimilated.

There is also a blacklist change to remove inland seas and lakes, because our surface emissivity model assumes ocean salinity levels. However, the number of observations affected is very small. Overall, the effect of these changes is to give a much greater weight to the all-sky observations in the analysis, and to use more cloud and rain-affected data.

Experiments are based on cycle 36r1 and have a T799 resolution. They have been run for 5 months, starting from the same initial conditions, which are taken from operations on 1 June 2009. A month is allowed for spinup, so the comparison period is 1 July to 31 October 2009. Table 3 summarises the experiments. The all-sky observations come from AMSR-E and the SSM/I instruments on the Defense Meteorological Satellite Program (DMSP) F-13 and F-15 satellites.

4 Results

4.1 Observation error tuning

This section examines the choice of tuning parameter α (Eq. 8) in the observation error formulation. As an a-posteriori justification of our approach, we examined FG fits to other assimilated observations (this has been done previously for advanced infrared sounders). Going from Control-Off (i.e. no microwave imager observations) through Control to All-Sky-New shows the effect of increasing the weight of all-sky observations to the point where the observation errors are roughly the size of the FG departure standard deviation. To increase the weight further, experiments were run with the All-Sky-New configuration but with values of α = 0.67, 0.33 and 0.0. Initial conditions came from All-Sky-New on 1 October 2009 and the experiments were run for 10 days.

Figures 16, 17 and 18 show the standard deviation of Advanced Microwave Sounding Unit B (AMSU-B) and Microwave Humidity Sounder (MHS), AMSU-A, and High Resolution Infrared Radiation Sounder (HIRS) FG departures as a function of experiment, with observation weight increasing towards the right. “New 1.0” corresponds to the All-Sky-New experiment, and “New 0.67” through “New 0.0” are the experiments where α is varied. Ignore the dashed lines for the moment. AMSU-B and MHS channels 3 to 5 are sensitive to humidity in the upper, mid and lower troposphere. AMSU-A channel 5 is sensitive to temperature over the

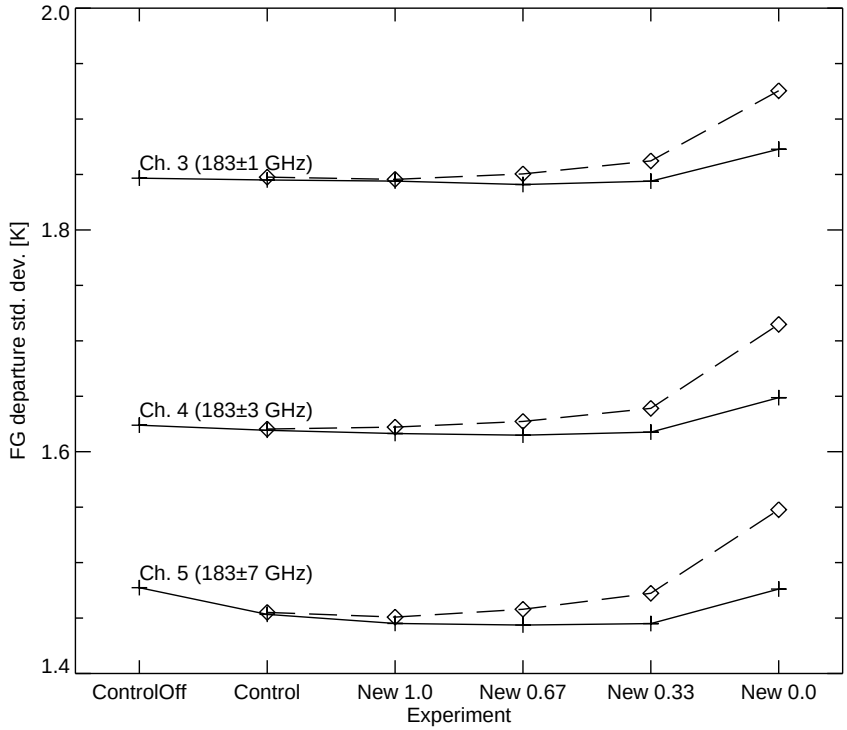


Figure 16: Standard deviation of FG departures, in K, for all assimilated AMSU-B and MHS observations globally from 1 - 10 October 2009, as a function of experiment, ordered so that the weight of All-Sky observations increases towards the right. VarQC is either on (solid line with crosses) or off (dashed line with diamonds) for the all-sky observations.

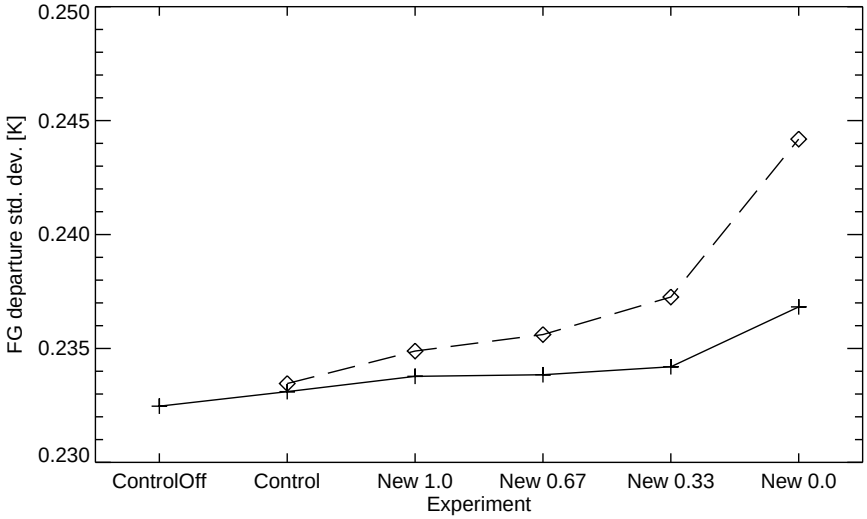


Figure 17: As Fig. 16 but for AMSU-A channel 5.

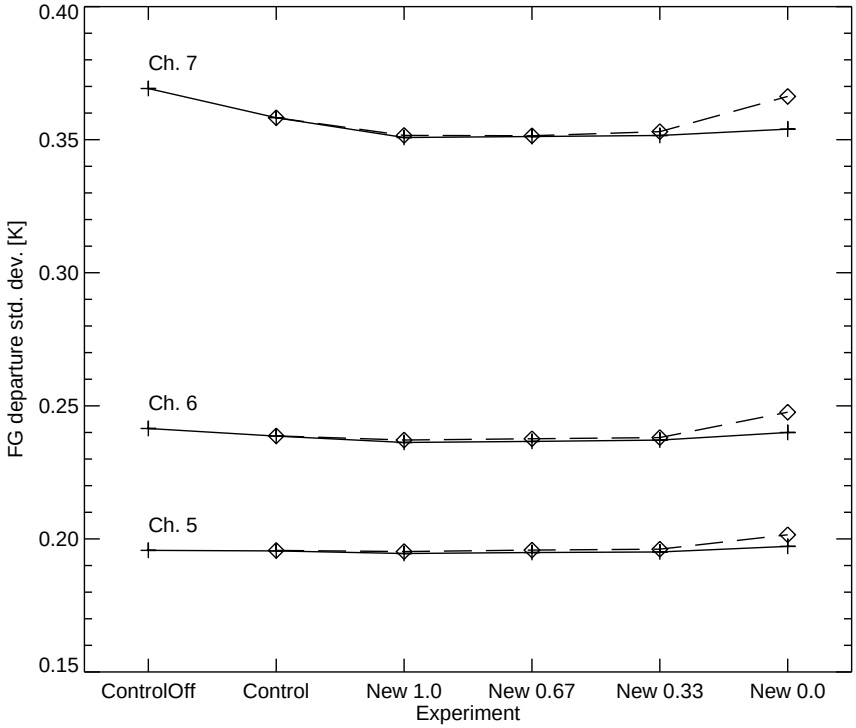


Figure 18: As Fig. 16 but for HIRS channels 5 to 7.

whole troposphere, yet is the lowest-sounding of the AMSU-A channels that we assimilate. HIRS channels 5, 6 and 7 sense tropospheric temperature with weighting functions peaking at 500 hPa, 800 hPa and the surface. Over the 10 day period of these experiments, radiosondes and other in-situ data are not numerous enough to provide reliable statistics.

Overall, All-Sky-New improves 12 h forecasts compared to Control or Control-Off, with a better fit to AMSU-B / MHS and HIRS. Though there is a slight degradation of fit to AMSU-A channel 5, this is only of order 0.001 K. Changing α to 0.67 or 0.33 makes little difference, but going to 0.0 degrades the fit. Here, we are clearly over-constraining the analysis.

Another set of experiments were run with variational quality control (VarQC, Andersson and Järvinen, 1998) switched off for the all-sky observations. VarQC downweights the influence of observations that disagree with the analysis, which masks any problems coming from an excessive fit to the all-sky observations. With VarQC off, it is clear that even values of $\alpha = 0.67$ start to degrade the fit compared to $\alpha = 1.0$. Overall, the All-Sky-New configuration, with $\alpha = 1.0$, looks to be best.

It is strange that results from AMSU-A channel 5 seem to disagree with HIRS channels 5 to 7. AMSU-A shows a very slight increase in FG departure standard deviation between Control-Off and All-Sky-New, while HIRS shows a substantial decrease. There are three possibilities that could explain this: (i) any improvement in FG temperature is limited to the lower troposphere, to which only the HIRS channels have strong sensitivity; (ii) because these HIRS channels have a slight sensitivity to water vapour, it is improvements in lower tropospheric moisture which cause the reduction in standard deviation; (iii) this is a sampling effect, caused by a 1.5% increase in AMSU-A channel 5 usage between Control-Off and All-Sky-New and a 1.4% increase in HIRS channel 7 usage. Future work will attempt to determine which of these effects is most important, but any of

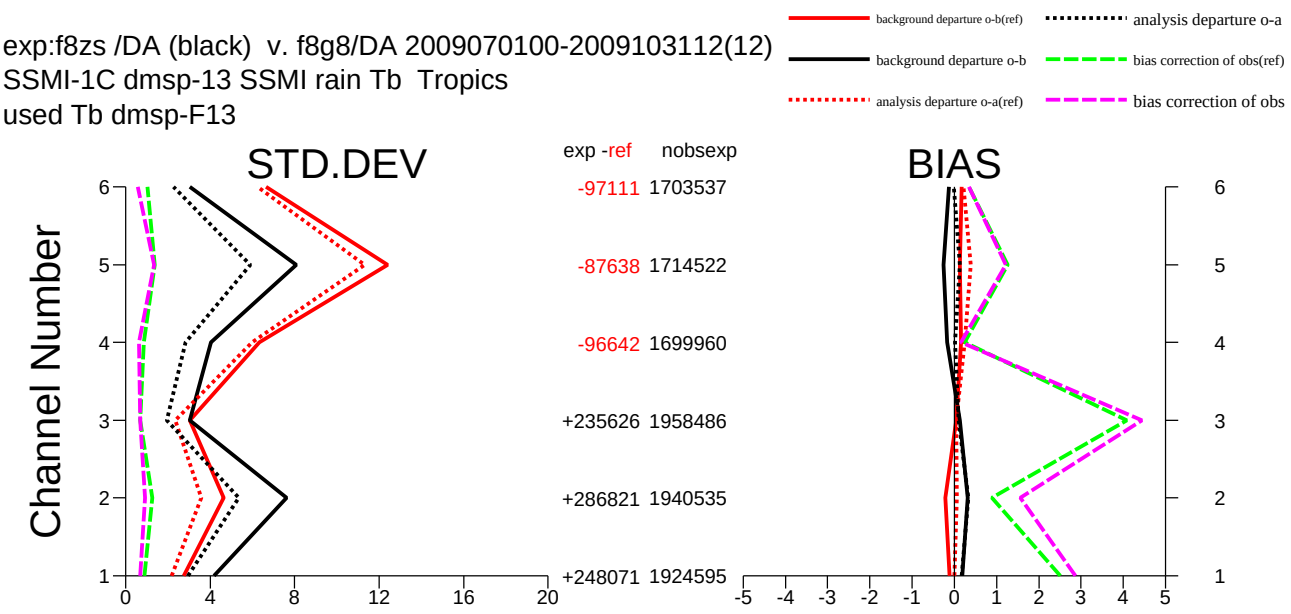


Figure 19: Statistics of tropical SSM/I FG departures, in K, for 1 July to 31 October 2009. All-Sky-New (f8zs) is in black and Control (f8g8) is in red.

these explanations would indicate some benefit from the new approach.

4.2 Observation-based diagnostics

4.2.1 All-sky observations

Figure 19 shows the observation statistics (obstat plot) for SSM/I in the tropics. 12% more observations are assimilated in channels 1 to 3 (19v, 19h and 22v) thanks to the changes in BgQC, which allow much more cloud and rain-affected data into the system. As a result, the standard deviation of FG departures becomes greater. However, the increased weight given to the all-sky observations means that the gap between FG and analysis departure standard deviations is larger than before. At higher frequencies (channels 4-6, i.e. 37v, 37h and 85v) the observations still have more weight than before, but the excess-snow screening also has an effect. In the old approach, observation errors were so large in these channels that most cloud and rain affected data was allowed through BgQC, although the large errors meant the observations had little impact on the analysis. Now we screen out observations where the FG has heavy snowfall, which reduces the FG departure standard deviation, and reduces the number of observations in these channels.

To learn more, we need to look at a consistent sample of observations. A special set of passive monitoring runs was created, in which only the first trajectory was run, and the initial conditions came from a parent,which was either Control or Control-Off. This allowed us to calculate FG departures using the same SSM/I processing and sampling as in All-Sky-New, but without affecting the forecasts. It was not possible to compute analysis departures with this technique. These special experiments were run for 10 days.

Samples have been divided according to the presence of cloud, using a symmetric brightness-temperature based approach. The cloudy sample contains any case where cloud is observed in either FG or observation in any experiment, and the clear sample contains only data where both FG and observation are clear in all experiments.

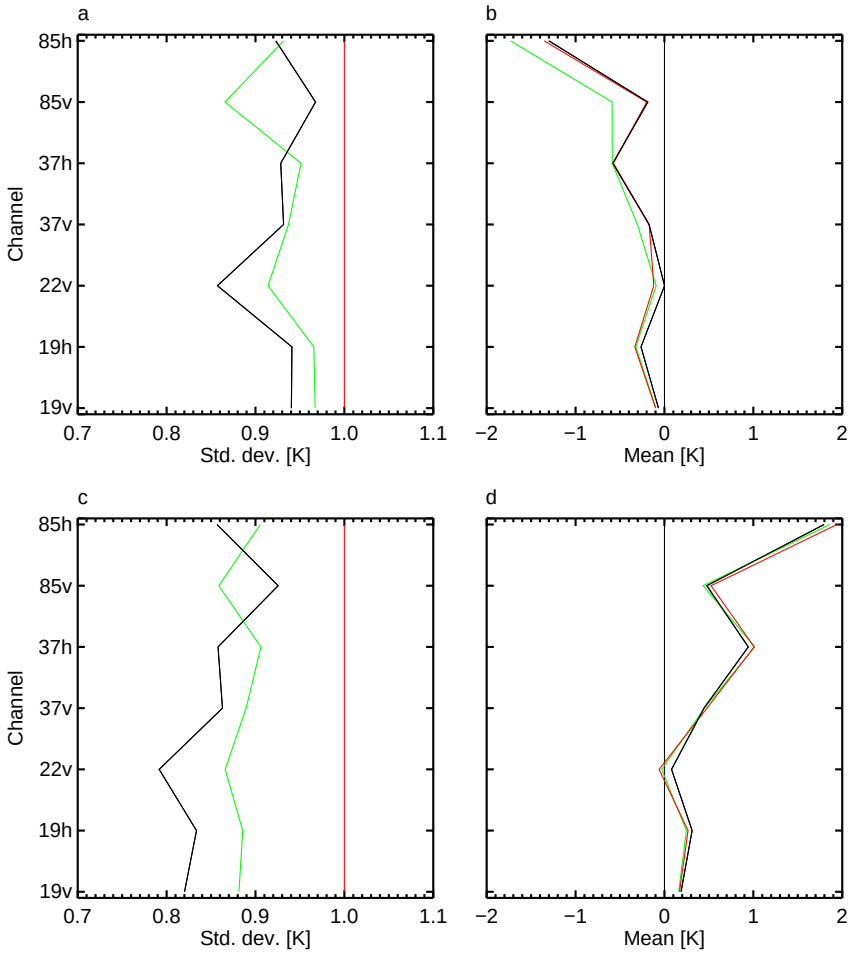


Figure 20: Statistics of SSM/I FG departures, in K, for 1 - 10 October 2009: ***(a,c)*** *standard deviation;* ***(b,d)*** *mean; Control is in green, Control-Off red, and All-Sky-New black. Top row is for observations with cloud observed or in the FG (sample of 629639 in channel 19v); bottom row is for clear observations and first guesses (sample of 338279 in channel 19v). A special set of passive monitoring runs is used to give an identical SSM/I processing and sampling in the calculation of these departures. Standard deviations have been normalised by the Control-Off values.*

This avoids sampling effects that would otherwise confuse the results. (For example, deciding the clear sample on the basis of only one experiment being clear means the other experiments may sometimes still be cloudy and hence will appear, erroneously, to have larger standard deviations.)

Figures 20 and 21 show the results for SSM/I and AMSR-E. The mean FG departure hardly changes between any of the experiments. Standard deviations become much smaller as the observation weight increases. In the lower frequency channels in clear sky situations, the impact of going to the new approach (e.g. All-Sky-New - Control) is roughly half that of adding the all-sky observations in the first place (e.g. Control - Control-Off).

4.2.2 Other assimilated data

Departure statistics for all microwave humidity sounders and all latitudes have been aggregated in Fig. 22. The main impact comes at 183 ± 7 GHz, which of the three AMSU-B / MHS channels has the closest weighting

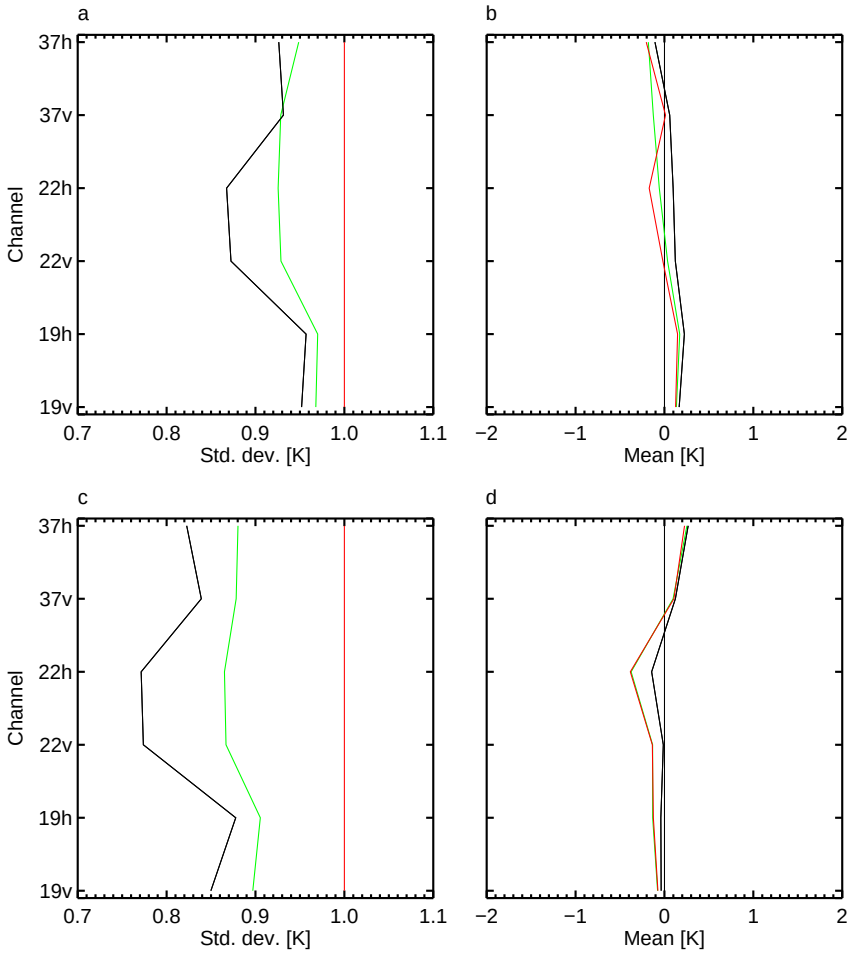


Figure 21: As Fig. 20 but for AMSR-E. Cloud affected sample is 423366; clear sample is 190456

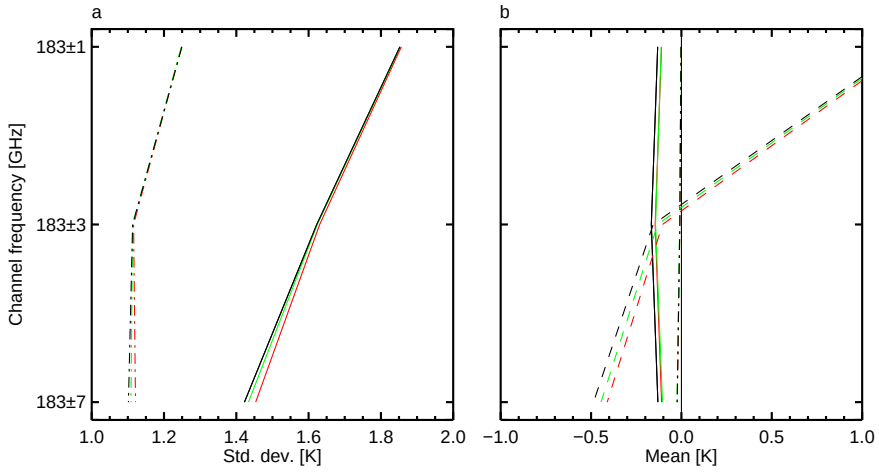


*Figure 22: Statistics of FG (solid) and analysis (dot-dash) departures, in K, for all AMSU-B and MHS instruments assimilated between 1 Jul 2009 and 31 Oct 2009: **(a)** standard deviation; **(b)** mean; Control is in green, Control-Off red, and All-Sky-New black. The dashed line on panel b shows the mean bias correction.*

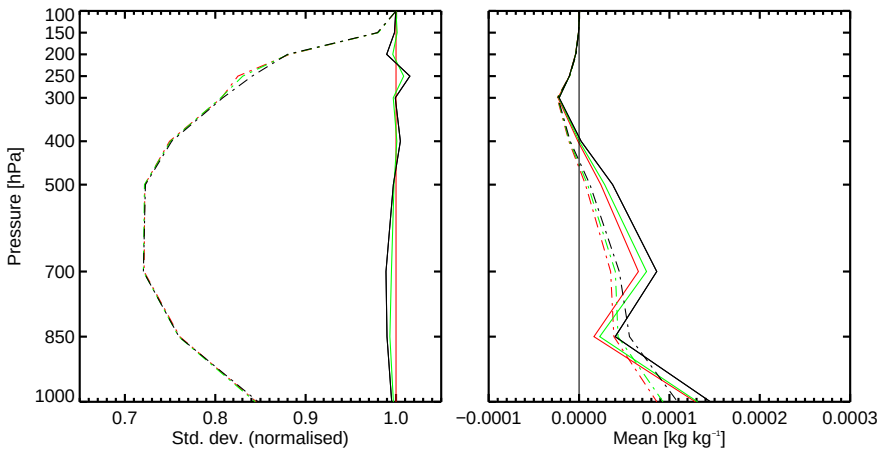


Figure 23: As Fig. 22 but for radiosonde specific humidities in kg kg^{-1}. To cope with several orders of magnitude variation in humidity through the troposphere, standard deviations have been normalised by the Control-Off values.

function to the surface. Standard deviations are lower at FG and analysis in both All-Sky-New and Control, compared to Control-Off, as we have already seen in the observation-tuning section. All-Sky-New has marginally smaller standard deviations, which suggests that the new approach constrains lower tropospheric humidities slightly better than the control. Biases are not affected much. Examining similar plots in latitude bands (not shown) we see that by far the largest effect from microwave imagers comes in the tropics, but the improvements between Control and All-Sky-New are mostly in the southern and northern hemispheres (SH and NH), rather than the tropics.

Figure 23 shows fits to radiosonde humidities. The sample is dominated by NH land observations, so it is not expected to show a very big signal from our changes to the use of ocean-only microwave imager data. Nevertheless, All-Sky-New reduces standard deviations compared to Control and Control-Off, showing again that the new approach is slightly better at constraining lower tropospheric humidities. The main impact is between 700 and 850 hPa. By comparing this to the very similar figure shown in Bauer et al. (2010), we can see that All-Sky-New has roughly regained the quality of radiosonde humidity FG fit that was seen in the approach used prior to cycle 35r2 (e.g. direct radiance assimilation in clear skies and 1D+4D-Var in cloud).

The length of our experiments gives, for the first time, enough dropsonde profiles to produce believable statistics (Fig. 24). Dropsondes are launched on-demand, nearly always over tropical or sub-tropical oceans, either for specific campaigns, or more usually hurricane tracking. Hence, dropsonde profiles are found in just the sort of conditions we would hope to make an impact, and are about the only source of in-situ moisture data over the oceans. Here, the All-Sky-New experiment typically has smaller standard deviations than in Control or Control-Off, particularly in the tropics below 500 hPa. We should be slightly cautious, given that even over 4 months the sample is quite geographically limited, but this is still a promising result.

Finally, we examine the results for HIRS and AMSU-A (Figs. 25 and 26). Fits to HIRS channels 6 and 7 are substantially improved by the all-sky observations, as we saw in the section on observation error tuning. Increasing the weight of all-sky observations degrades the fit to AMSU-A channels 5 and 6, but only by 1%, and only in the FG and not in the analysis. Further investigation shows the effect is limited to the tropics, and the midlatitudes show no degradation in AMSU-A fit. As mentioned in Sec. 4.1, there is a 1.5% increase in the number of AMSU-A observations assimilated in All-Sky-New compared to Control-Off, so we cannot rule out sampling issues. However, a possible explanation is that moisture changes affect the drift of temperature with

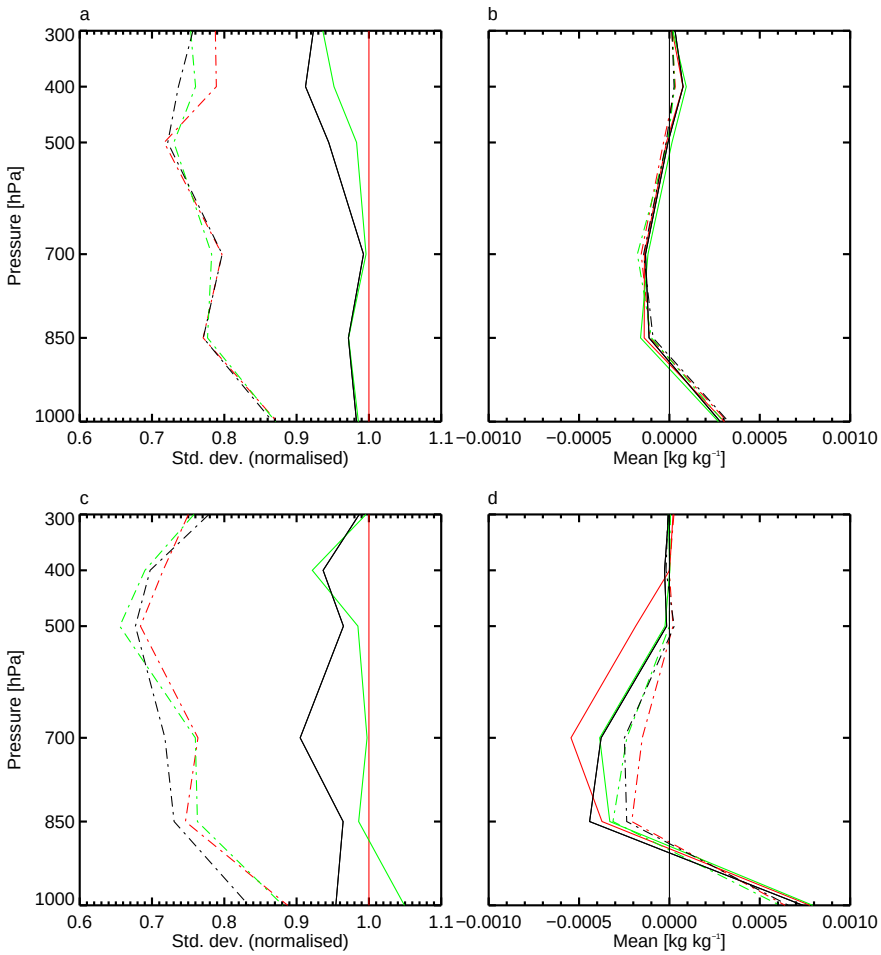


*Figure 24: As Fig. 23 but for dropsondes in **(a,b)** NH only **(c,d)** tropics only*

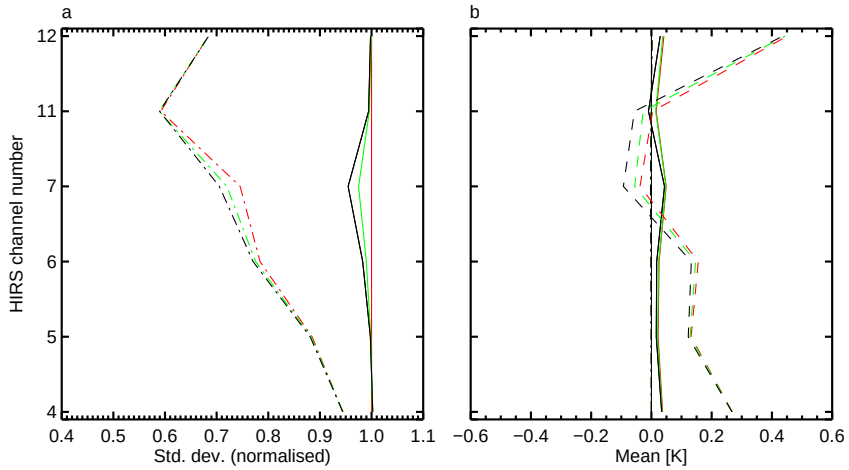


Figure 25: As Fig. 23 but for global HIRS observations

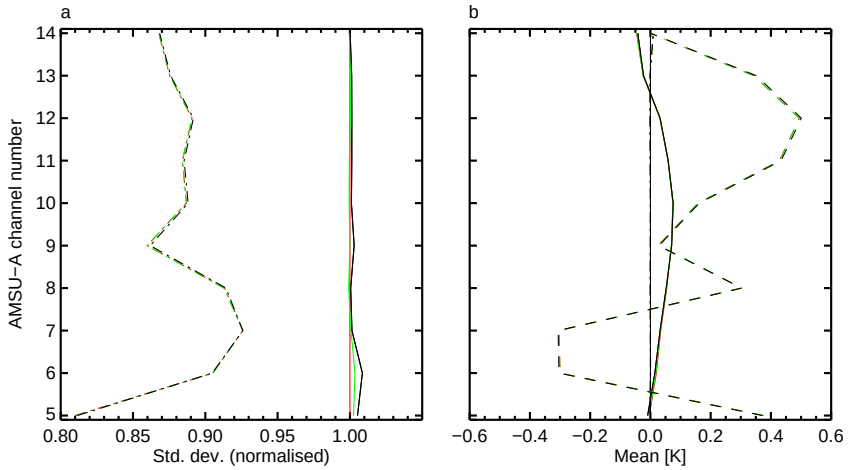


Figure 26: As Fig. 23 but for global AMSU-A observations

forecast time in parts of the tropics.

In summary, going to All-Sky-New improves the fit to all observation types examined here, except for a slight degradation in tropical tropospheric AMSU-A channels. Overall the effect is beneficial.

4.3 Forecast scores

Forecast scores for vector wind (VW) and relative humidity (R) are presented in Figs. 27 and 28. The figures show the normalised difference in RMS error between All-Sky-New and Control. Negative values indicate a reduction in RMS error and hence an improvement compared to Control. Each experiment's own analyses is used as the reference. The methods used for calculating RMS error and the statistical significance of the difference are described in Geer et al. (2010)

There is no significant change to wind forecasts in the midlatitudes between forecast days 2 and 8 (T+48 to T+192). However, the tropics show a significant increase in RMS error that is largest at the beginning of the forecast, reaching locally 0.05, i.e. a 5% increase in RMS error, but then declines with time. In relative humidity there is an increase in RMS error at all latitudes between 60S and 60N in the lower and mid troposphere, which reaches 10 to 20% at T+24.

Maps of the normalised RMS difference are shown in Figs. 29 and 30. In both wind and humidity, the increases are found over oceans only, and predominantly in the subtropics. The locations remain the same throughout the forecast range and there is no growth or spread in the error. This suggests that the difference between experiments comes from the analyses rather than the forecasts, and comes from the all-sky observations, which are assimilated over oceans only.

Figure 31 shows a case study of relative humidity in a small part of the subtropical South Atlantic, on a single day. It is representative of the behaviour throughout the regions showing an increase in RMS error. Panels a and b show that T+12 forecasts are similar in both experiments, and it is the analyses (panels c and d) where the biggest changes have occurred. The increased weight of all-sky observations in All-Sky-New has further dried the two dry features at 10S and 20S, and has moistened an area centred at 35S, 20W. The difference between forecast and analysis (panels e and f) is larger in All-Sky-New than in Control. In fact, at T+12, forecast minus analysis is simply the reverse of the increment (analysis minus forecast). In other words, All-Sky-New makes

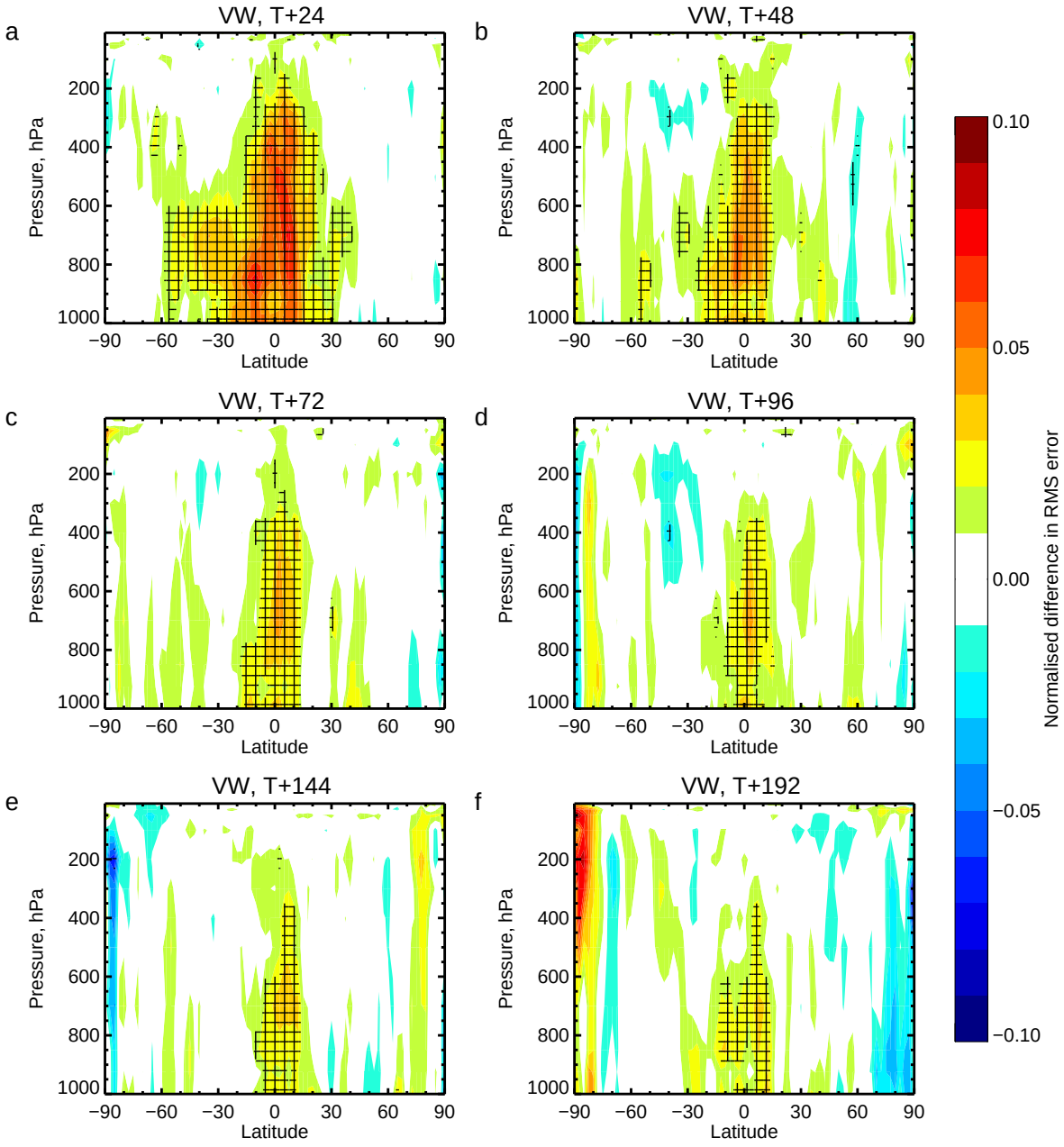


Figure 27: Normalised difference in RMS own-analysis forecast scores in vector wind (VW) between All-Sky-New and Control. Red colours indicate All-Sky-New has larger RMS errors than Control; blue colours that it has lower errors. Scores are calculated at 00Z for the period 1 July to 31 October 2009, giving a sample of 115 to 123 forecasts. Cross-hatching indicates point differences significant at the 99.8% level; this equates to a global significance level of about 95%.

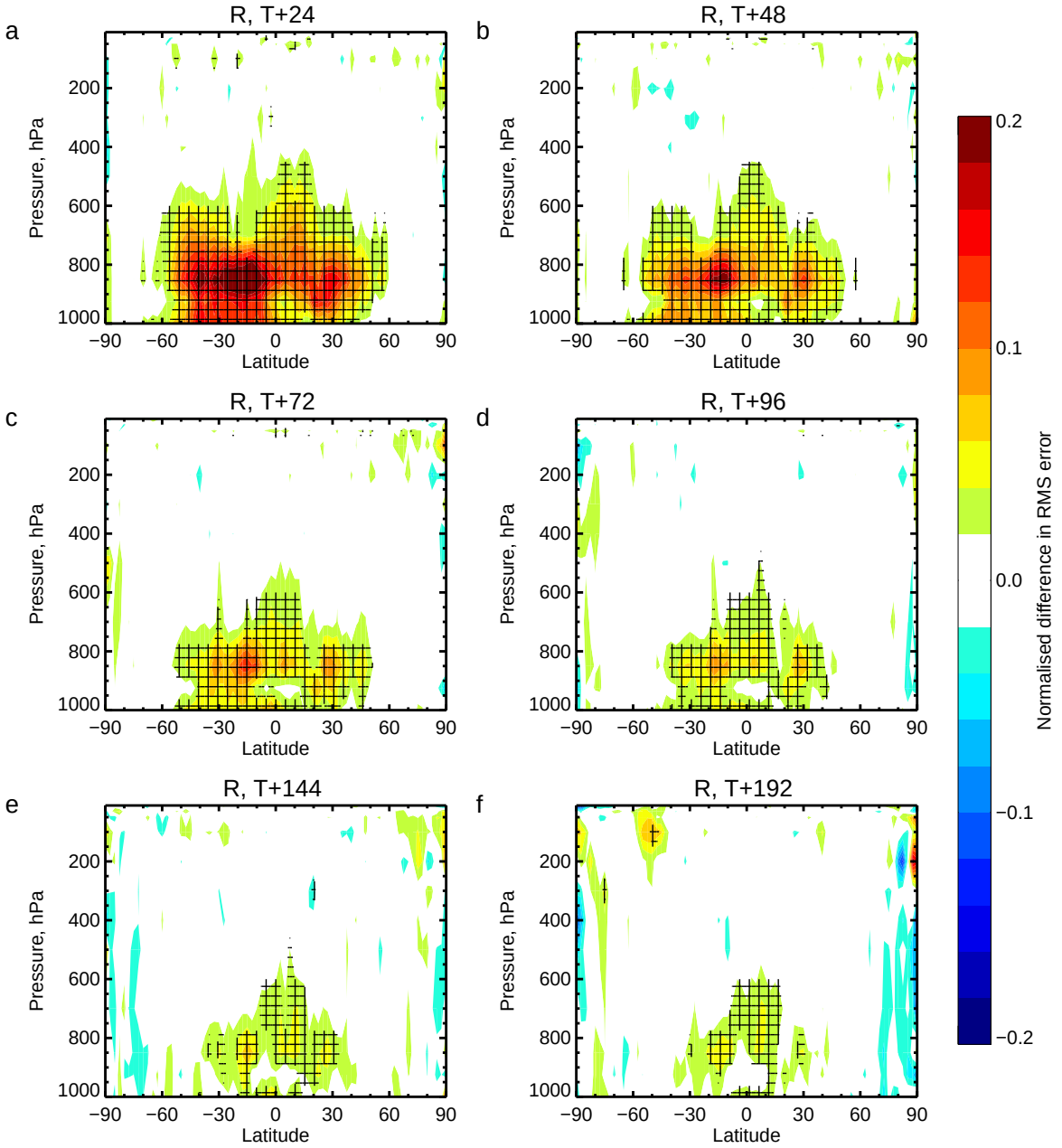


Figure 28: As Fig. 27 but for relative humidity (R).

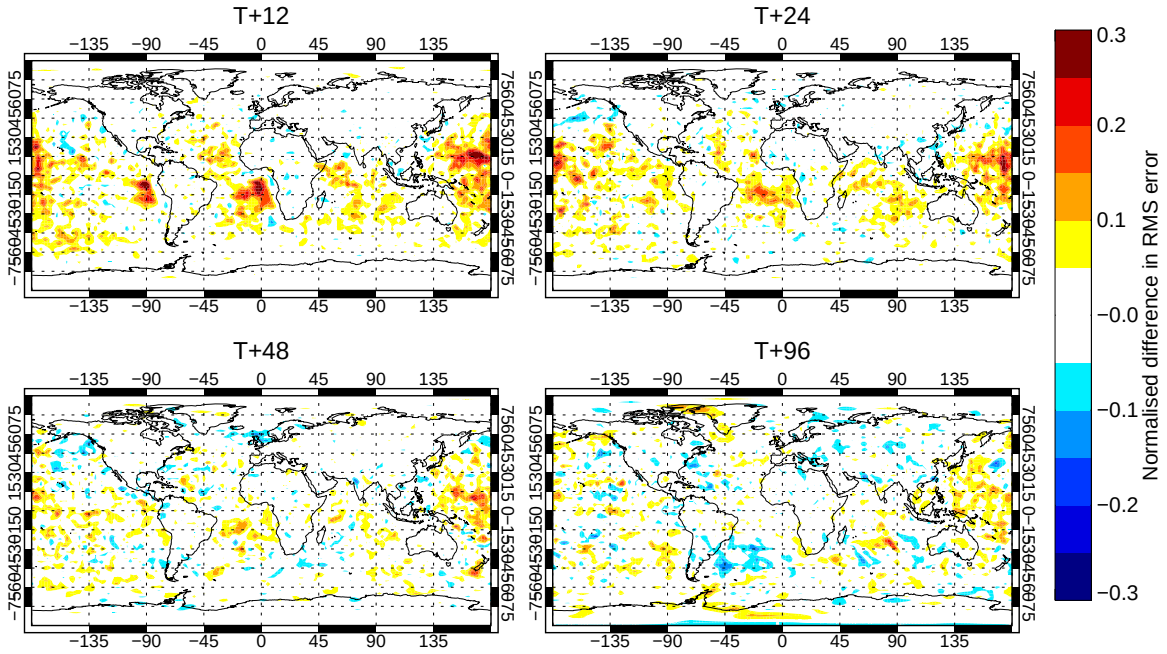


Figure 29: As Fig. 27 but for vector wind at 1000 hPa, in map form. Statistical significances have not been calculated here.

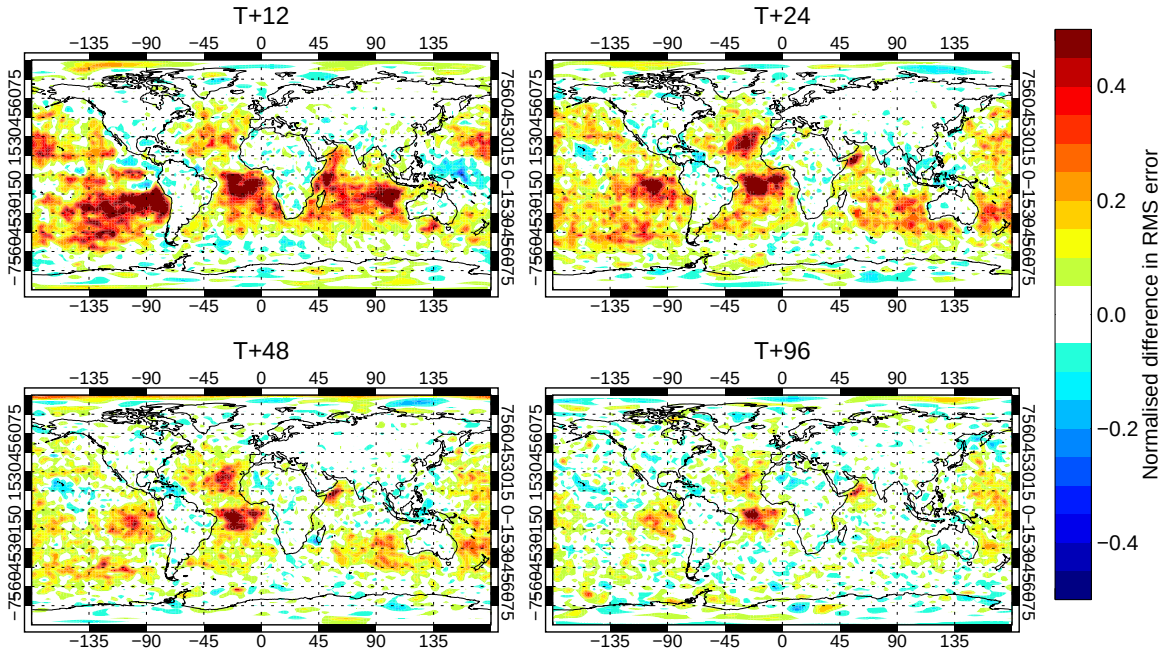


Figure 30: As Fig. 30 but for relative humidity.

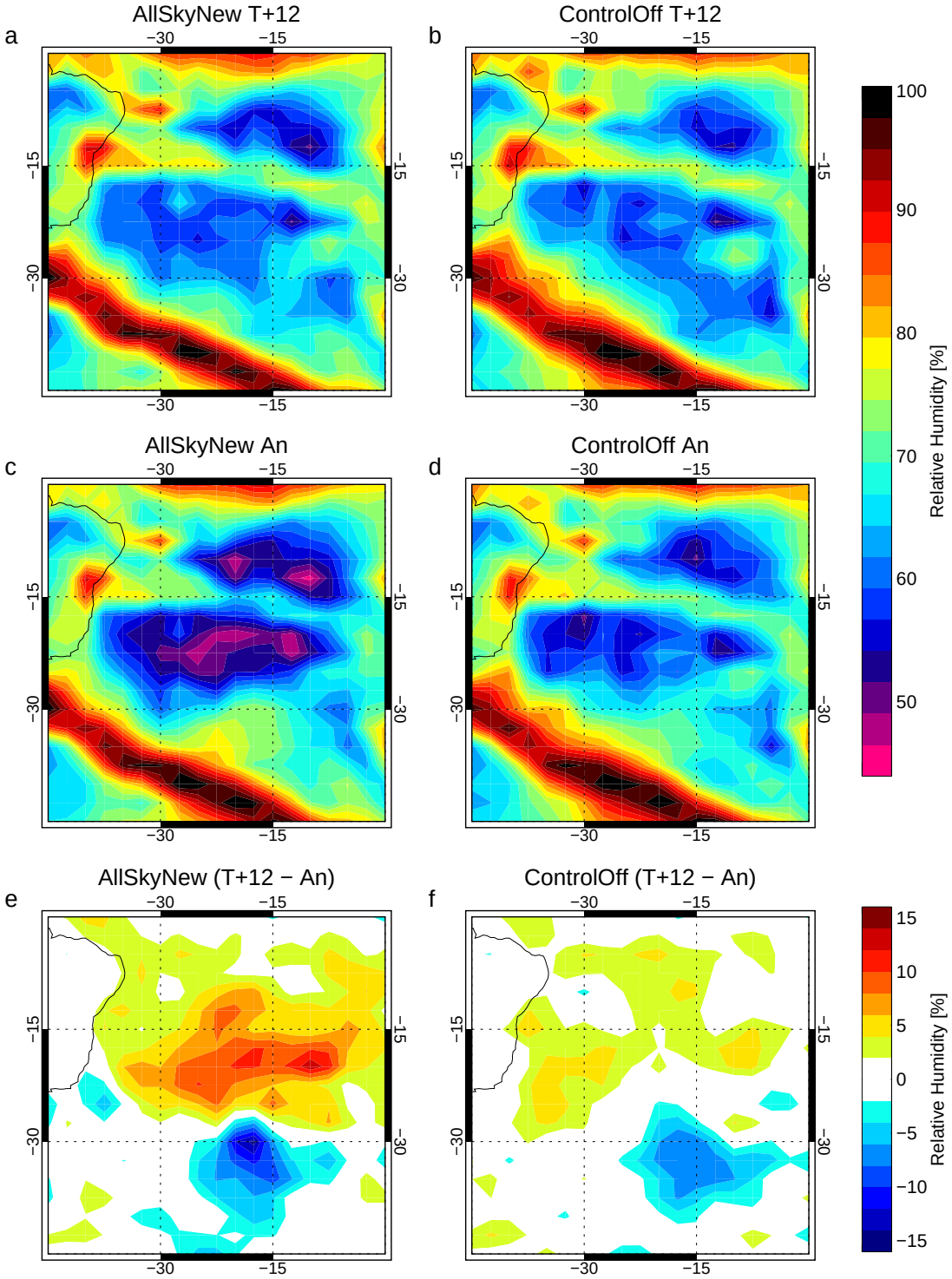


Figure 31: Forecasts (T+12) and analyses of relative humidity at 1000 hPa, valid at 12Z, 9 Jul 2009. ***Left:*** *experiment All-Sky-New,* ***Right:*** *experiment Control-Off;* ***Top:*** *T+12 forecast;* ***Middle:*** *Analysis;* ***Bottom:*** *forecast - analysis*

much larger increments than Control. However, this is what we would expect given the stronger observational constraint.

This effect is largest in the subtropical oceanic regions (Figs. 29 and 30), so it is mainly an effect of trade cumulus or stratocumulus in the maritime boundary layer, and of the humidity structures that drive this. The all-sky observations make large corrections to the model forecast, but these corrections are quickly lost and the model returns to a state that is very similar in both forecasts. This suggests that the subtropical lower tropospheric humidity and cloud are driven by the large-scale circulation and surface conditions. In these situations, which have mainly water cloud or clear sky, we would expect the all-sky observation operator to be quite accurate, because cloud radiative transfer in the microwave becomes mainly an absorption, rather than a scattering problem. Hence, the most likely explanation for the large increments would be shortcomings in the model's boundary layer and cloud parametrisations.

Despite the difficulties we have seen in interpreting own-analysis scores, there is no better way of using model fields and analyses for verification. All other possible approaches are subject to worse problems of interpretation (Geer et al., 2010). In contrast, observation-based scores are not subject to these effects.

Radiosonde-based scores for vector wind in the NH, tropics and SH are shown in Fig. 32. The sense of the figures remains the same as before: negative values indicate that All-Sky-New has smaller errors than Control, and vice-versa. There are no significant degradations in the early forecast range, confirming that the degradation in model-based scores is a result of larger increments, and does not indicate a real problem in the forecasts. An insignificant improvement is seen in the NH around day 8, which agrees with the model-based scores (panel f of Fig. 27). In the SH, however, radiosonde-based vector wind scores beyond day 6 are worse in All-Sky-New than Control. The number of radiosondes available to calculate these scores in the SH is small and concentrated in the land areas of South America and Australia. Thus, these scores are not representative of the whole hemisphere. We believe this potential shortcoming of All-Sky-New should be noted, but it should not form significant grounds for rejecting the new approach.

Scores for geopotential in the NH are shown in Fig. 33, using radiosondes as the verification reference. There is no significant change in geopotential scores against radiosonde, although the minor improvement seen around day 8 again agrees with a small improvement seen in the model-based scores (not shown). Overall, we would conclude that the all-sky observations affect the analyses and short-range forecasts of oceanic moisture and vector wind, and that longer-range forecasts are not affected.

5 Conclusion

The original version of the all-sky assimilation of microwave imager radiances went operational in March 2009 in cycle 35r2. A number of new developments will substantially increase the weight of these observations in the analysis, both in clear skies and cloud and rain-affected areas. This new approach will be made operational with cycle 36r4, and has been evaluated using T799 experiments of four months duration. Analysis and FG departure fits to radiosonde and dropsonde humidities, AMSU-B / MHS humidity channels and HIRS channels 6 and 7 all improve a little (e.g. HIRS departure standard deviations improve from 0.36 K to 0.35 K in channel 7). No fits show any real degradation, though there is a very small decrease in the FG fit to AMSU-A channels 5 and 6 (e.g. departure standard deviations go from 0.233 K to 0.234 K in channel 5). This slight degradation occurs almost entirely in the tropics.

Short-range forecast scores show increases in RMS humidity and vector wind errors, mostly in the tropics. Vector wind RMS errors against radiosondes do not show the same effect. The effect comes from the tighter fit to all-sky observations, which results in much larger humidity increments over the ocean. However, the new

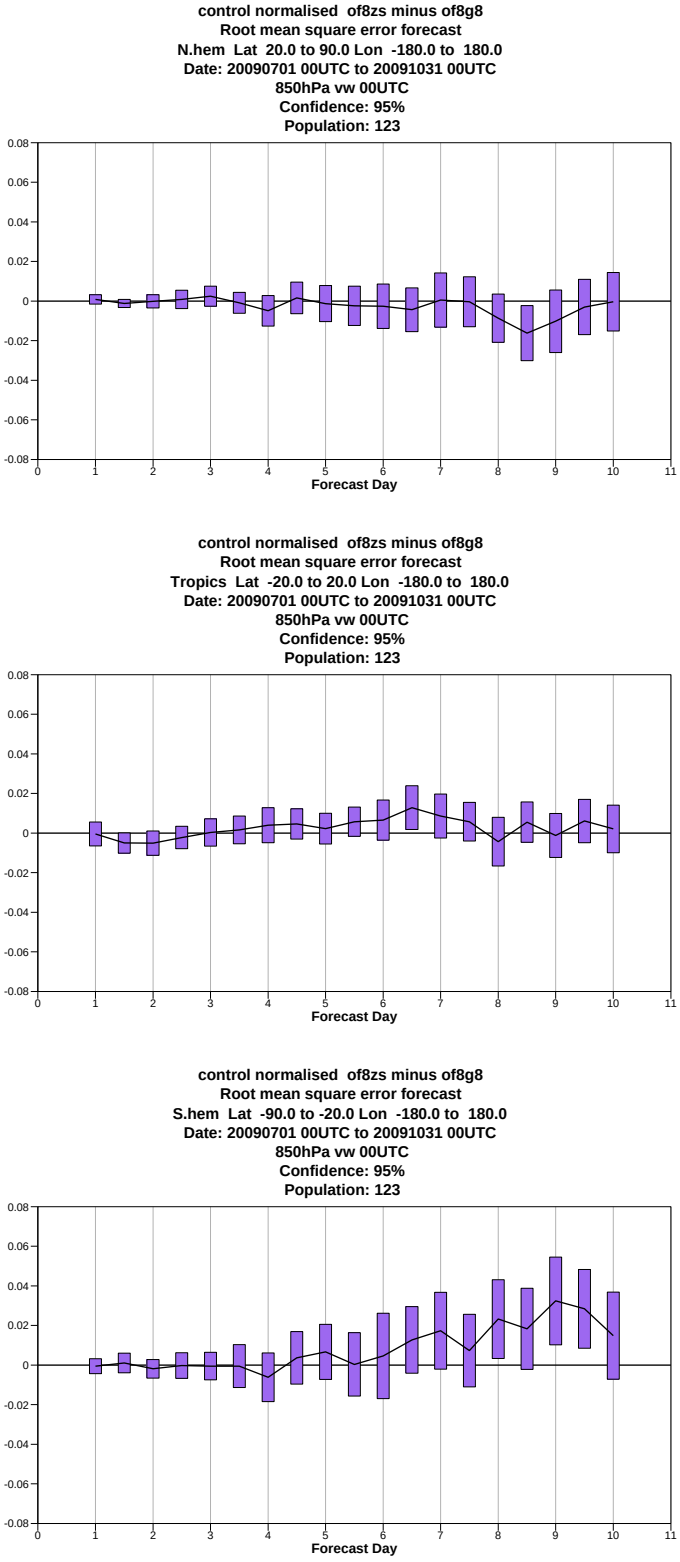


Figure 32: Normalised difference in vector wind forecast scores, with verification against radiosondes, for the NH (top), tropics (middle) and SH (bottom). Negative values indicate that All-Sky-New ('of8zs' here) has smaller RMS errors than Control ('of8g8')..

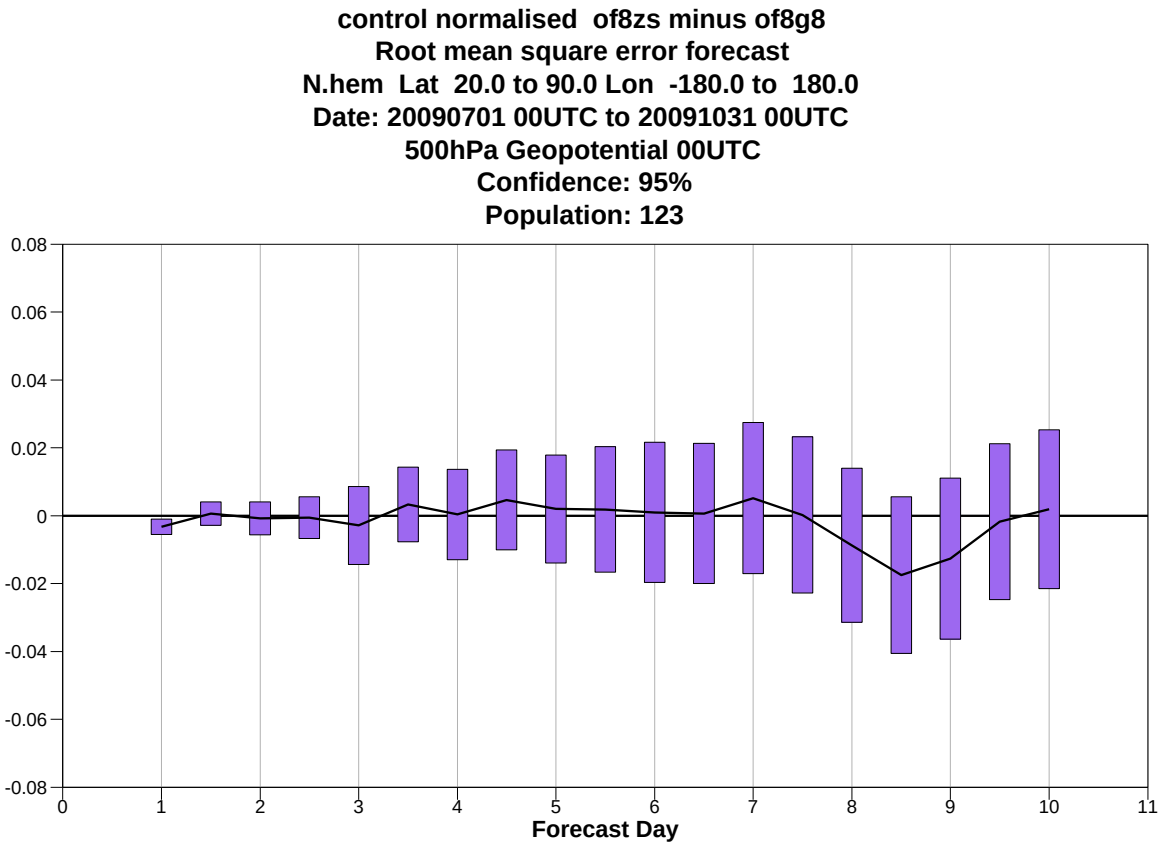


Figure 33: As Fig. 32 but for geopotential against radiosondes in the NH.

humidity information appears to be lost over the first few days of the forecast. Longer-range scores against radiosonde observations (500 hPa geopotential, 850hPa vector wind, from day 3 onwards) are not significantly changed. A degradation for days 7 to 10 at 850 hPa in SH scores is arguably not significant, given the limited geographical coverage of sondes in the SH. In summary, the new approach substantially increases the constraint of lower-tropospheric humidity over oceans in the analysis and first few days of forecast, but has little impact at longer ranges.

The new approach is based on a better understanding of the symmetric nature of FG departures in cloud and rain-affected observations. A conceptual model of displacement error with binary clouds is used to show that binning FG departures by FG cloud amount (or by observed cloud) leads to a sampling bias. These binning strategies are labelled 'asymmetric' and cause several problems. The first occurs when FG cloud (or rain) amount is used as a bias predictor, but the 'bias' as a function of FG cloud will simply be an artefact of the sampling, and should not be removed from the observations. The second problem occurs if observation error is made a function of observed cloud or rain. This will lead to a biased data assimilation system, where if errors are larger in cloudy areas than in clear skies, the increments will always on average dry the system. The solution to both these problems is to find an unbiased or 'symmetric' binning strategy. The average of FG and observed cloud amount is symmetric, as is the maximum cloud amount.

All-sky microwave imager FG departures exhibit similar behaviour to the conceptual model, and the standard deviation of FG departures can be well predicted by the mean cloud amount. This behaviour is used to create a model for the partitioning of observation and background error. The model can be used to normalise the all-sky FG departures, which makes their distribution appear Gaussian. For the first time, this allows an effective quality control for cloud and rain-affected observations. The symmetric model is also used to provide observation errors for the new all-sky approach. The parameter α in the symmetric observation error model controls the partitioning of cloudy FG departure error between background and observation. Note that the background error predicted by this model is only used in QC; it is not yet used to control the background errors of the 4D-Var system. In practice, α is set to 1, meaning that all cloudy FG error is assumed to come from the observations. Sensitivity tests show that decreasing α (giving greater weight to the all-sky observations in cloud and rain) starts to degrade the FG departure fit to other assimilated observations. The new approach is shown give the best fit to other observations in the ECMWF system.

The new approach assigns observation errors of around 20 K in heavy cloud and rain areas. This describes a combination of instrument error (probably by far the smallest contribution), representativity error, radiative transfer operator error, and in practice, model error too. In 4D-Var, the model forms part of the observation operator, and hence a main contribution to these 'observation' errors likely comes from inaccuracies in the treatment of cloud and rain in the forecast model. Weak constraint 4D-Var (e.g. Trémolet, 2006) may in future allow us to handle this kind of error more correctly.

We have not considered observation (or background) error correlations in this work, yet it is clear that FG departures are correlated, and because the model forms part of the observation operator, we may see some benefits from treating these correlations as part of the observation error. Future developments at ECMWF will allow the treatment of correlated observation error. For the future it would also be very interesting (though quite challenging) to apply the diagnostics of Desroziers et al. (2005) to further investigate the error correlations of all-sky observations.

In the previous approach, observation error was inflated as a function of distance from grid point. This was done because the all-sky system does not perform interpolation, but simply matches observations to the nearest grid-point. However, over the maximum distance between observation and grid-point in the T255 resolution of the final inner loop, which is 55 km, the standard deviation of FG departures increases by only 6% in SSM/I channel 19v, going from 4.1 K to 4.3 K. Moreover, in the previous approach, the rate of error inflation with

distance was too fast, resulting in an observation error for all-sky observations that was unnecessarily large. Hence, error inflation has been abandoned in the new system.

We have also considered the spatial scale of the observations. Instead of taking the nearest single all-sky observation to a grid point, we calculate an average or 'superob' of all observations falling into the grid-box, prior to assimilation. This should make the observations more representative of the spatial scales in the model. Particularly in heavy cloud and rain, the spatial scale of the observations substantially affects the histogram of TBs, so a mismatch in the scale of observations and model could cause an artificial bias. However, in the ECMWF model, cloud and rain (and the associated model errors) vary with scales that are much larger than the nominal grid resolution. Theoretically, we should superob to the outer-loop resolution (T1279 as of January 2010). However, the broad scales of model cloud and rain suggest we can get away with using a coarser resolution for the observations. To save computational costs, we chose to superob at the resolution of the final inner-loop, which is T255.

Finally, we have reconsidered our strategy for bias correction in cloudy and rainy areas. In the initial all-sky system, no cloud or rain bias correction was applied. Informed by the symmetric error model, it would now be possible to use mean cloud or rain as a bias predictor in VarBC. However, the bias as a function of mean cloud is actually extremely small. We do see regional and situation-dependent biases but these are very hard to describe with a simple predictor. For these and other reasons, we still cannot use VarBC to correct such biases. Instead, the new approach applies better screening for areas where the model shows significant biases. These are (a) the 'cold sector' bias, where in winter, in dry, cold, polar airmasses, the model shows too little liquid water cloud, and there are mean biases of 5 to 10 K against all-sky observations; (b) 'heavy snowfall bias', where simulated TBs (for the higher frequency channels, i.e. above 30 GHz) are too low by 10 to 20 K in tropical convection, due to excessive scattering. It is not yet clear whether the latter is a problem of the radiative-transfer model, RTTOV-SCATT, or of excessive snow in the forecast model's convection scheme. There also appears to be another bias associated with fronts, as exemplified by Fig. 15. The model's fronts are typically too intense and too narrow compared to the microwave imager observations. This results in a very typical pattern of FG departures, with negative departures in the middle of a front, and positive ones towards the outside. For the moment this last bias remains untreated.

Further work is needed to investigate these model biases, and to fix them in the model moist physics parametrizations if this is where the problems lie. Also, a better understanding of background errors in cloudy and precipitating areas is required.

Acknowledgements

Alan Geer was funded by the EUMETSAT fellowship programme. Philippe Lopez, Bill Bell, Anne Fouilloux, Deborah Salmond, Niels Bormann, Tony McNally, Sabatino DiMichele, Lars Isaksen and Saleh Abdallah are thanked for their help in this work. Jean-Nöel Thépaut is thanked for reviewing the manuscript.

References

Andersson, E. and H. Järvinen (1998). Variational quality control. *ECMWF Tech. Memo., 250, available from http://www.ecmwf.int*.

Auligné, T. and A. P. McNally (2007). Interaction between bias correction and quality control. *Quart. J. Roy. Meteorol. Soc. 133*, 643–653.

Bauer, P., A. J. Geer, P. Lopez, and D. Salmond (2010). Direct 4D-Var assimilation of all-sky radiances: Part I. implementation. *Quart. J. Roy. Meteorol. Soc., submitted*.

Bauer, P., P. Lopez, A. Benedetti, D. Salmond, and E. Moreau (2006a). Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. I: 1D-Var. *Quart. J. Roy. Meteorol. Soc. 132*, 2277–2306.

Bauer, P., P. Lopez, D. Salmond, A. Benedetti, S. Saarinen, and E. Moreau (2006b). Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. II: 4D-Var. *Quart. J. Roy. Meteorol. Soc. 132*, 2307–2332.

Bauer, P., E. Moreau, F. Chevallier, and U. O'Keeffe (2006). Multiple-scattering microwave radiative transfer for data assimilation applications. *Quart. J. Roy. Meteorol. Soc. 132*, 1259–1281.

Benedetti, A. and M. Janisková (2008). Assimilation of MODIS cloud optical depths in the ECMWF model. *Mon. Weath. Rev. 136*, 1727–1746.

Dee, D. (2004). Variational bias correction of radiance data in the ECMWF system. In *ECMWF workshop proceedings: Assimilation of high spectral resolution sounders in NWP, 28 June – 1 July, 2004*, pp. 97–112. Eur. Cent. for Med. Range Weather Forecasts, Reading, UK, available from http://www.ecmwf.int.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteorol. Soc. 131*, 3385–3396.

Errico, R. M., P. Bauer, and J.-F. Mahfouf (2007). Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci. 64*, 3785 – 3798.

Geer, A. J., P. Bauer, and P. Lopez (2007). Lessons learnt from the 1D+4D-Var assimilation of rain and cloud affected SSM/I observations at ECMWF. *Published simultaneously as ECMWF Technical Memoranda 535 and ECMWF/EUMETSAT fellowship reports 17*.

Geer, A. J., P. Bauer, and P. Lopez (2008). Lessons learnt from the operational 1D+4D-Var assimilation of rain- and cloud-affected SSM/I observations at ECMWF. *Quart. J. Roy. Meteorol. Soc. 134*, 1513–1525.

Geer, A. J., P. Bauer, and P. Lopez (2010). Direct 4D-Var assimilation of all-sky radiances: Part II. Assessment. *Quart. J. Roy. Meteorol. Soc., submitted*.

Geer, A. J., P. Bauer, and C. W. O'Dell (2009). A revised cloud overlap scheme for fast microwave radiative transfer. *J. App. Meteor. Clim. 48*, 2257–2270.

Hollinger, J., J. Peirce, and G. Poe (1990). SSM/I instrument evaluation. *IEEE Trans. Geosci. Remote Sensing 28*, 781–790.

Järvinen, H. and P. Unden (1997). Observation screening and background quality control in the ECMWF 3D-Var data assimilation system. *ECMWF Tech. Memo., 236, available from http://www.ecmwf.int*.

Kawanishi, T., T. Sezai, Y. Ito, K. Imaoka, T. Takeshima, Y. Ishido, A. Shibata, M. Miura, H. Inahata, and R. Spencer (2003). The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies. *IEEE Trans. Geosci. Remote Sensing 41*, 184–194.

Petty, G. (1994). Physical retrievals of over-ocean rain rate from multichannel microwave imagery. Part I: Theoretical characteristics of the normalised polarisation and scattering indices. *Meteorol. Atmos. Phys. 54*, 79–99.

Petty, G. and K. Katsaros (1990). Precipitation observed over the south China sea by the Nimbus-7 scanning multichannel microwave radiometer during winter MONEX. *J. App. Meteorol. 29*, 273–287.

Trémolet, Y. (2006). Accounting for an imperfect model in 4D-Var. *Quart. J. Roy. Meteorol. Soc. 132*, 2483–2504.