TECHNICAL MEMORANDUM

619

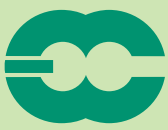
Direct 4D-Var assimilation of all-sky radiances. Part II: Assessment

Alan J. Geer, Peter Bauer and Philippe Lopez

Research Department

Submitted to Quart. J. Roy. Meteorol. Soc.

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

This paper assesses the performance of an all-sky (i.e. clear, cloudy and precipitating) direct assimilation of microwave imagers that was introduced operationally in the four-dimensional variational (4D-Var) assimilation system of the European Centre for Medium-Range Weather Forecasts (ECMWF) in March 2009. Results are based on data-denial experiments and are quantified using forecast verification scores and comparisons to observations. The reliability of forecast verification scores is examined. Humidity-based scores can be misleading in the initial three days of forecasts, but scores based on pressure and wind fields are more reliable, because they are less sensitive to the choice of verifying analysis. The new approach is compared with the previous rain and cloud-affected assimilation, which used a 1D+4D-Var technique. One downside of the old technique is that it 'recycles' the first guess as new information, which causes misleading 'improvements' in forecast scores. The new approach avoids this problem. It also does not require any cloud detection, so unlike the old approach used in clear-skies, it is not susceptible to biases caused by undetected cloud. However the new system is implemented with a cautious quality control and large observation errors. Hence, the analysed and forecast tropical humidity fields are constrained only about half as much in the new system as in the old. However, the new approach is comparable in terms of its positive impact on dynamical fields (pressure and vector wind) in the tropical lower troposphere.

1 Introduction

Part I of this work (Bauer *et al.*, 2010, henceforth B10) introduces a new 'all-sky' four-dimensional variational (4D-Var) assimilation of microwave radiances, which has been operational at the European Centre for Medium-Range Weather Forecasts (ECMWF) since March 2009. This paper, part II, evaluates the impact of these observations on global analyses and forecasts.

In the all-sky approach, microwave imager 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. Channels between 19 GHz and 85 GHz are used, over oceans only. At the frequencies chosen, the atmosphere is semi-transparent except in heavy cloud and precipitation. In clear skies, the 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, and cloud and precipitation. The intention is to use all of this information to improve analyses and forecasts.

Bengtsson and Hodges (2005) questioned whether moisture-sensitive observations have any impact on global forecast skill. However, they were working with a 3D-Var system, where moisture observations can only impact wind fields by changing latent heating in the atmosphere or through background error correlations between moisture and dynamical variables. By contrast, 4D-Var allows tracer-like observations to directly impact the wind fields (e.g. Peubey and McNally, 2009). Just as importantly for cloud and rain observations, the use of adjoint moist physics operators means that temperature and wind fields must be modified in order to fit observed cloud and precipitation structures (B10). In the ECMWF 4D-Var system, clear-sky humidity observations do have a positive impact on moisture and wind fields (Andersson *et al.*, 2007), as did the forerunner of the all-sky technique, the 1D+4D-Var assimilation of cloud- and rain-affected microwave imager observations (Bauer *et al.*, 2006a,b; Kelly *et al.*, 2008). Microwave imager observations also have the ability to improve midlatitude geopotential forecast scores through their sensitivity to wind speed at the surface (Kazumori *et al.*, 2008). However, that study used the low-frequency channels of AMSR-E (at 6 and 10 GHz), which have the greatest sensitivity to wind speed, but these are not yet assimilated in the ECMWF system.

Table 1: Experiment configuration.

	Control	Control-Off	All-Sky	All-Sky-Off
Rainy 1D+4D-Var	On	-	-	-
Clear 4D-Var	On	-	-	-
All-sky 4D-Var	-	-	On	-
Physics in 1st minimisation	No	No	Yes	Yes
ECMWF ID	f4u1	f5rc	f2q6	f3aw

2 Method

2.1 Assimilation system

Experiments are based on cycle 35r1 of ECMWF's operational assimilation and forecasting system. This uses an atmospheric model with a semi-Lagrangian, spectral formulation. The model has 91 levels from the surface to an altitude of 80km, and, in the configuration tested here, a T511 horizontal resolution, corresponding to about 40km. Global analyses and forecasts of wind, temperature, surface pressure, humidity and ozone are produced twice-daily using a 4D-Var assimilation system (Rabier *et al.*, 2000) with a 12-hour time window. The analysis includes in-situ conventional data, satellite radiances from infrared sounders (both polar-orbiting and geostationary), satellite-borne radio-occultation measurements sensitive to temperature, satellite-derived atmospheric motion vectors and surface winds from scatterometers. Microwave imagers are assimilated in clear, cloudy and rainy skies; this paper examines two different methods for doing so.

2.2 Microwave imager assimilation

As explained in B10, between June 2005 and March 2009, ECMWF used a two-stream approach for the assimilation of microwave imager observations. This approach is used in the 'Control' experiment here. Rain and cloud-affected observations are assimilated, between 45°N and 45°S, using a 1D+4D-Var technique (Bauer *et al.*, 2006a,b; Geer *et al.*, 2008). Clear-sky observations are assimilated directly as radiances in 4D-Var, and there is no latitude restriction.

From March 2009, ECMWF have used the new all-sky approach, tested in the 'All-Sky' experiment here. As described in B10, all microwave imager radiances are assimilated directly in 4D-Var. The observation operator uses a scattering radiative transfer model to simulate the effect of cloud and rain. In order to fit the analysis to the observations, modelled wind, temperature, humidity, cloud and precipitation must all be adjusted within the 4D-Var assimilation. Use of the full 4D forecast model ensures consistency between increments in the dynamical and moist physics parameters. Observations are assimilated between 60°N and 60°S.

2.3 Experiments

Table 1 lists the main experiments used in this paper. As well as evaluating the new approach in comparison to the old, we want to evaluate the overall impact of the microwave imager observations, so we perform data-denial experiments by removing all microwave imager observations, producing the experiments 'All-Sky-Off' and 'Control-Off'. These are not identical, because in order to assimilate the all-sky observations, linearised radiation and moist physics schemes needed to be made active in the first inner loop minimisation, which was not previously the case. However, as will be shown later, this change did not have a large forecast impact.

Table 2: Microwave imager observation usage (numbers per day).

Experiment	Instrument	Rainy 1D+4D-Var	Clear direct 4D-Var	All-sky direct 4D-Var
Control	SSM/I	16240	19347	-
	AMSR-E	39231	14483	-
	TMI	21992	12700	-
	SSMIS	-	4654	-
	Total	77463	51184	-
All-Sky	SSM/I	-	-	100123
	AMSR-E	-	-	69502
	TMI	-	-	-
	SSMIS	-	-	-
	Total	-	-	169625

All experiments start from the same initial conditions on 10th August 2007. A twelve-day period is set aside for spin-up and to allow the Variational Bias Correction (VarBC Dee, 2004; Auligné *et al.*, 2007) to stabilise. The period 22nd August until 30th September 2007 is used for study.

3 Results in observation space

3.1 All-sky observations

3.1.1 Observation usage

The number of microwave imager observations used in the All-Sky and Control experiments is shown in Tab. 2. The table lists the maximum number of observations that were used, independent of the number of channels that were available, since this varies. As discussed in B10, the old system used observations from Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) and Special Sensor Microwave Imager Sounder (SSMIS), as well as from AMSR-E and SSM/I. However, SSMIS and TMI are not included in the all-sky system, in order to save computer resources and to avoid some anomalies specific to these instruments (Bell *et al.*, 2008; Geer *et al.*, 2010). We hope to re-introduce TMI and SSMIS in the future.

Despite using fewer instruments and not assimilating microwave imager observations poleward of 60°N, the new system assimilates more observations overall because of changes in the thinning. However, fewer observations are used in areas affected by heavy cloud or rain in the lower frequency channels (19 GHz, 22 GHz) because they are more often removed by quality control in All-Sky than in 1D+4D-Var (see B10).

Table 3 shows mean observation errors used in All-Sky and Control and also serves to indicate which channels have been assimilated. As described in B10, all-sky observation errors are inflated as a function of distance from the grid-point centre. Only observations less than 10 km from the outer-loop (T511) grid points are assimilated. However, the grid resolution and the location of grid-points changes through the three inner-loops of 4D-Var (Fig. 5, B10). Thus the distances to grid points and the observation errors change through the minimisation. The table shows the errors used in the final inner loop, which has the highest resolution (this is T255 or approximately 80 km). As explained in B10, it is the final inner-loop which generates the analysis increment, so it is the observation errors used here which control the weight of all-sky observations in the analysis. In All-Sky, the mean observation error is around 12 K to 17 K in the 19 GHz and 22 GHz channels, roughly 3 to 4 times larger than in Control.

Table 3: Mean observation error, in K, of observations used in the final inner-loop of the 00Z analysis on 22 Aug 2007.

	All-Sky		Control			
			1D+4D-Var		Clear	
Channel	SSM/I	AMSR-E	SSM/I	AMSR-E	SSM/I	AMSR-E
19v	12.5	13.7	3.0	4.0	3.0	3.0
19h	14.2	16.8	6.0	8.5	4.5	4.5
22v	12.5	-	3.0	-	4.0	-
24v	-	13.7	-	3.5	-	3.5
24h	-	15.2	-	-	-	4.5
37v	37.1	36.5	-	-	3.5	3.5
37h	142	140	-	-	4.0	4.0
85v	111	-	-	-	4.0	-
85h	240	-	-	-	6.0	-

Additionally in All-Sky channels 37v and 85v, observation error is inflated as a function of the hydrometeor amount (see B10). This results in a large average observation error, but be aware that the clear-sky observations retain an observation error of order 5 K, similar to that used in clear-sky assimilation in Control. Channels 37h and 85h are not intended to be used in the All-Sky approach, and this is achieved by setting a minimum 99 K observation error, rather than by eliminating these channels completely.

Overall, observation errors are substantially larger in All-Sky than in Control. This would be expected to reduce the impact of the observations However, this is partly countered by the larger number of microwave imager observations assimilated (170625 in All-Sky compared to 128647 in Control, Tab. 2). As will be shown later, the net effect is a reduction of the observational constraint in All-Sky compared to Control.

3.1.2 Mean departures

We now investigate the first guess and analysis departures of the microwave imager observations in the All-Sky experiment. These are calculated in brightness temperature space as

$$\mathbf{d} = \mathbf{y}^{o} - \mathbf{b} - H[M[\mathbf{x}(t_0)]], \tag{1}$$

with observations $\mathbf{y}^{o}$ and bias correction $\mathbf{b}$. To get the model equivalents of the all-sky observations, a non-linear observation operator H and forecast model M are applied to the forecast or analysis $\mathbf{x}$ at time t_0. See B10 for further explanation.

Although the assimilation of microwave imager observations in clear, cloudy and rainy areas is unified in the all-sky approach, it is helpful to divide them up again for the investigations that follow. As in Geer *et al.* (2008) we use simple, brightness temperature (TB) based tests to categorise the observations and their model equivalents.

Cloudy observations are defined as having greater than 0.05 kg m^{-2} in cloud liquid water path (LWP) retrieved using the Karstens *et al.* (1994) regression:

$$\begin{aligned} \mathrm{LWP} &= 4.2993 + 0.3996\, ln(280 - \mathrm{TB}_{22\mathrm{v}}) \\ &\quad -1.4069\, ln(280 - \mathrm{TB}_{37\mathrm{v}}) \end{aligned} \tag{2}$$

Here, LWP is given in kg m^{-2}, and TB$_x$ indicates the brightness temperature (in Kelvin) in channel x. Separately, rain (or heavy cloud) is indicated by a polarisation difference of less than 40 K in the 37 GHz channels (e.g. Petty, 1994).

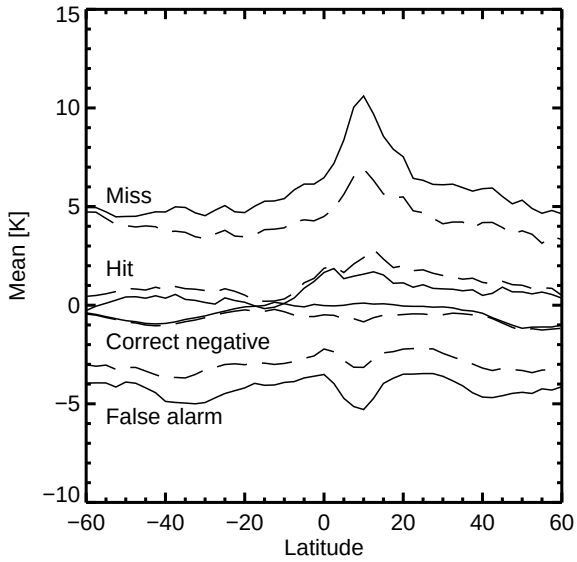


Figure 1: Zonal mean of FG (solid) and analysis (dashed) departures for channel 37v on SSM/I in the All-Sky experiment, divided into contingency table categories for cloud, defined on the basis of observed cloud versus FG cloud (even for the analysis departures). Cloud is defined by the criterion LWP $> 0.05\,\mathrm{kg\,m^{-2}}$.

These tests were originally fitted to observations, but as is typical in data assimilation, the simulations have biases compared to the observations. These biases are large enough to significantly change the results of the Karstens *et al.* (1994) regression. To avoid this we apply the bias correction in the opposite sense to usual, so we use TBs either from uncorrected observations $\mathbf{y}^o$ or from the bias-corrected model equivalent, $H[M[\mathbf{x}(t_0)]]+\mathbf{b}$.

We will also make a contingency-table style analysis of the data. Based on the detection of cloud (or alternatively rain) in the first-guess (FG) or observed TBs, these categories are:

- **Hit:** Observations and FG show cloud;
- **Miss:** Observations show cloud; FG does not;
- **False alarm:** Observations are clear; FG show cloud;
- **Correct negative:** Observations and FG are clear.

Mean FG and analysis departure biases from All-Sky are shown in Fig. 1 as a function of latitude for the four cloud contingency categories. The example shows SSM/I channel 37v but is fairly representative of behaviour in all channels and for AMSR-E too. In the miss category, the lack of cloud or moisture in the model naturally causes positive FG departures, but the analysis correctly reduces their size, showing that moisture or cloud has been created. Similarly, the false alarm category has negative departures due to excess cloud or moisture in the model, and this is successfully reduced in the analyses. Biases are largest in the inter-tropical convergence zone (ITCZ), at around 10°N, associated with the heavy cloud and precipitation there.

In the hit and correct negative categories, the bias gets slightly worse between FG and analysis, but as we will see, this is a problem of the 4D-Var analyses irrespective of the assimilation of all-sky observations. Overall, we see an analysis that is working well on a basic level, and adjusting moisture, cloud and rain fields to fit the observations.

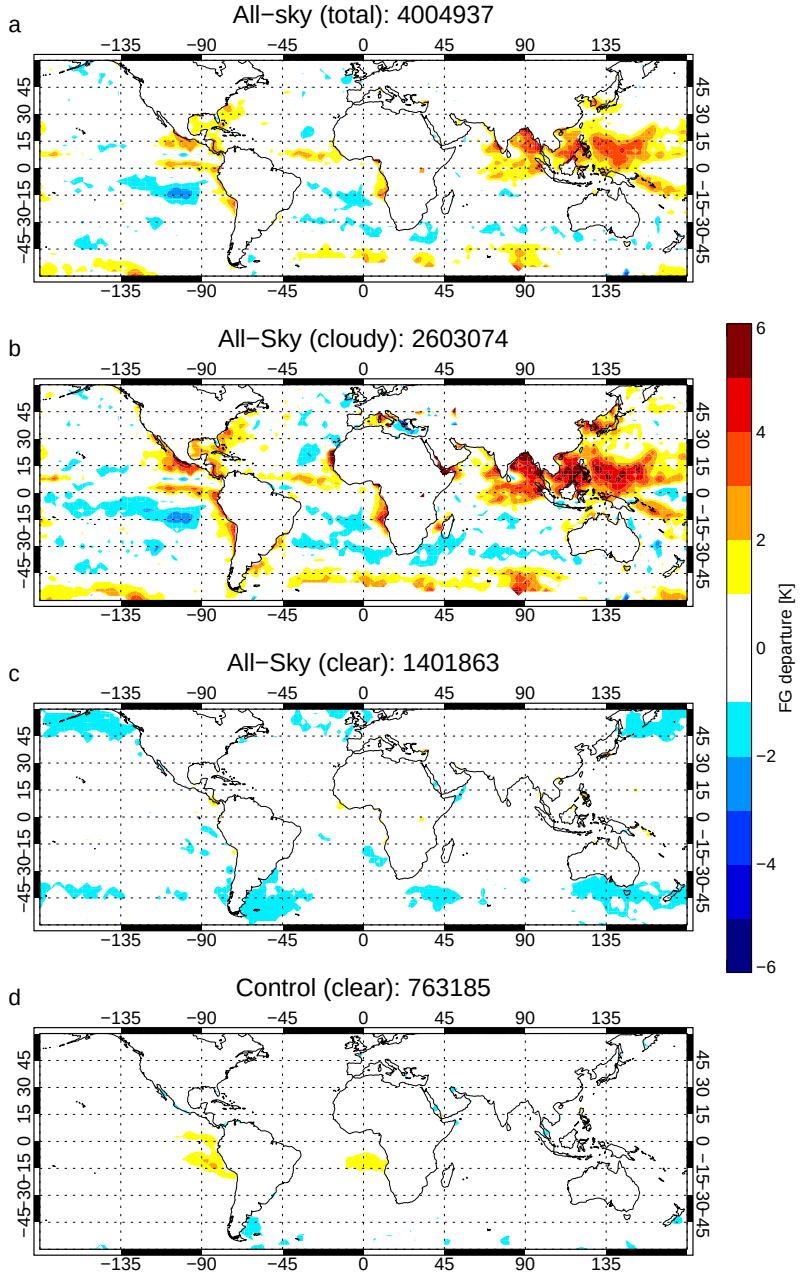


Figure 2: Mean departures over the period 22nd August to 30th September 2007, for SSM/I channel 37v, (a-c) for the All-Sky experiment and (d) for Control: ***a)*** *All assimilated observations* ***b)*** *Cloud-affected cases (includes all rain-affected)* ***c)*** *Clear-sky cases.* ***d)*** *All observations assimilated in the clear-sky route in Control*

Figure 2 shows time-mean FG departures for SSM/I channel 37v. In the All-Sky average there is a broad pattern of positive biases in the ITCZ and warm pool and negative biases in the subtropics (panel a). The pattern is similar for other channels and for AMSR-E (not shown). However, the behaviour at high latitudes, and of some of the smaller features, varies depending on a particular channel's information content. We do not intend to investigate these biases in detail here; Geer *et al.* (2009) give more insight. However, data assimilation only works properly with unbiased observations, so ideally we need these biases to be very small.

In panels b and c we split the sample into a 'cloudy' category combining the hits, misses or false alarms, i.e. wherever cloud is detected in either FG or analysis, and a 'clear' category (the correct negatives). This shows that the positive biases in the ITCZ and warm pool are associated with cloudy situations. Clear observations show little bias in the tropics. At high latitudes, there is often a negative bias of around 1 K in the clear observations. In contrast, the cloudy observations show positive biases in the Southern Ocean.

It is interesting to compare the clear biases in the All-Sky experiment (panel c) with those in the clear-sky path in the Control experiment (panel d). The Control departures are different in two ways. First, the sample contains the false alarm category as well as the correct negatives, i.e. it contains cases where the observations are clear but the first guess is cloudy. Second, to compensate for this, only clear-sky radiative transfer is done, so the radiative effect of any first guess cloud or precipitation is completely ignored. The Control clear-sky biases are generally small, but there are biases of up to +1.5 K in the eastern coastal areas of the Pacific and South Atlantic, where maritime stratocumulus is often found. No such bias is present in the All-Sky clear sample. This could be explained if the observation-based cloud screening in Control was failing to detect some of the lighter stratocumulus, and cloud-affected observations were being assimilated as clear. In the all-sky assimilation, if light cloud is present in the model, the effects are still included in the radiative transfer simulations, and on average there should be no bias. This illustrates one of the big advantages of the all-sky approach: there is no need to do a possibly unreliable cloud-detection.

However, it is noticeable that the All-Sky clear departures (panel c) have larger negative biases at high latitudes than do the Control clear-sky departures (panel d). This is essentially a problem of the VarBC bias correction in the All-Sky experiment, which cannot distinguish between clear and cloudy situations (see B10). VarBC ensures that global mean biases are very small, but unless it has the right predictors, it may still allow regional or situation-dependent biases. Hence, in the All-Sky departures in panel a, positive biases in the ITCZ and at high latitudes balance a negative bias in the subtropics. Moreover, in the southern high latitudes the positive bias in cloudy sky situations (panel b) balances the negative bias in clear situations (panel c). As a result, clear sky biases at high latitudes are larger in All-Sky than they were in Control. A future solution to these problems may be to introduce cloud-dependent bias predictors, though we have found it difficult in practice. For the moment, we must live with these biases, and the system works adequately well, as will be shown.

3.1.3 Fits to all-sky observations

Figure 3 shows the standard deviations and means of first guess and analysis departures (Eq. 1) as a function of SSM/I channel. Data are grouped according to the presence of cloud or rain in the FG or observations. Panel a contains those data where either the FG or observations have rain ('rainy'). Panel b contains those data where either the FG or observations have cloud ('cloudy'), so this is a superset of the data in panel a. Panel c shows those data where neither FG or analysis has any cloud ('clear'), so it contains the complement of the data in Panel b.

The figure includes the experiments All-Sky and All-Sky-Off. The sample of observations is very nearly equal in each case. This has been achieved by including all observations passing the initial quality checks, rather than the smaller set which get past thinning and quality control (QC, B10). It has not been possible to match

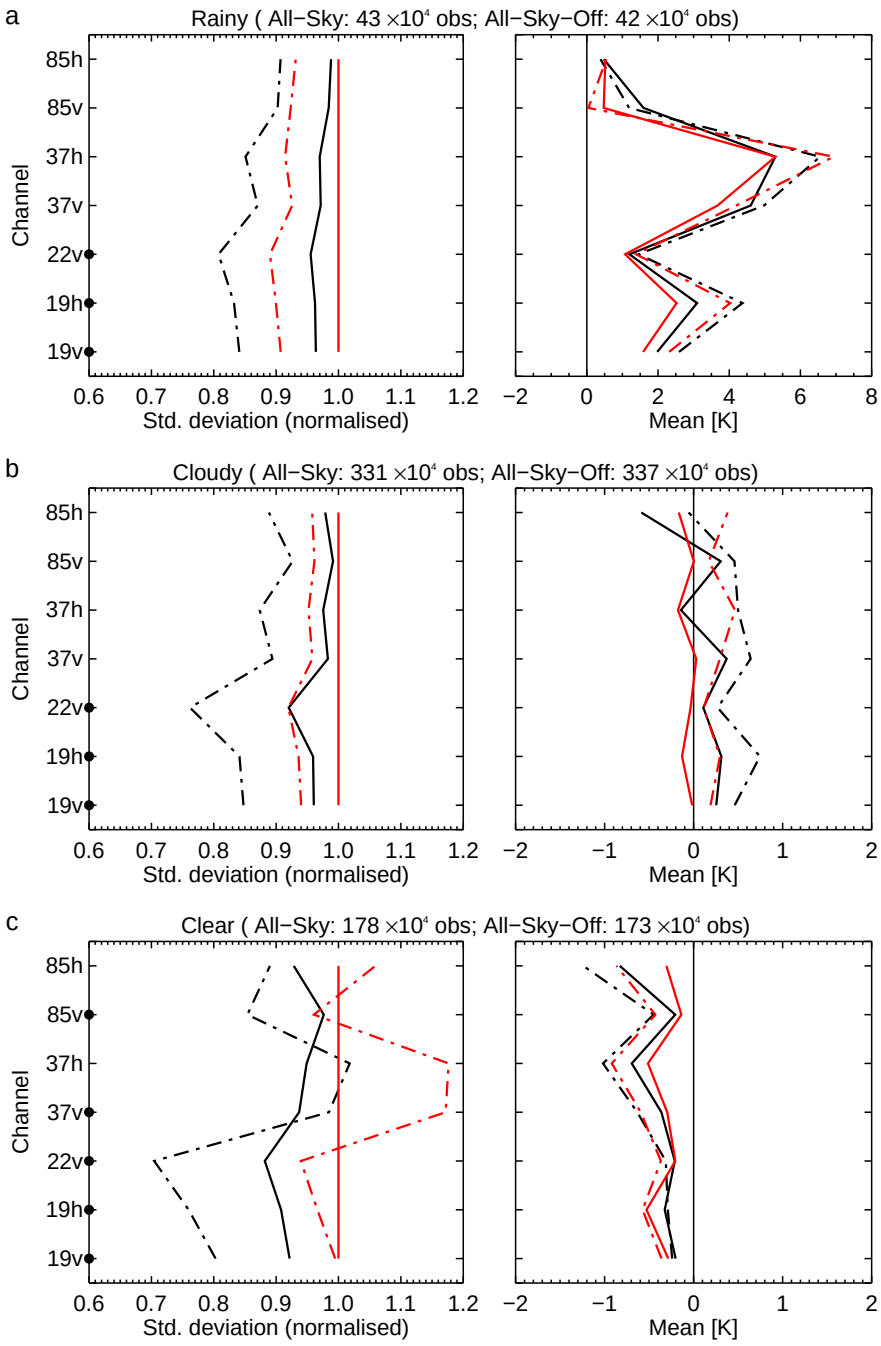


Figure 3: Standard deviation and mean departures over the period 22nd August to 30th September 2007, for all SSM/I channels: ***a)*** *Rain affected cases;* ***b)*** *Cloud-affected cases (includes all rain-affected)* ***c)*** *Clear-sky cases. All-Sky-Off is in red; All-Sky is in black. Analysis departures are shown as dot-dash lines and FG departures as solid lines. Standard deviations are normalised by the All-Sky-Off FG values. Black dots on the y-axis indicate channels which can be considered 'assimilated', i.e. with errors typically smaller than 20 K in that sample in the All-Sky experiment.*

a similar sample of observations in the Control experiment because there have been substantial changes to the pre-thinning. Hence, this figure can only illustrate the impact of microwave imagers in the all-sky system. Later, using other humidity data, we will see how this compares to the Control experiment.

Starting with the rainy sample standard deviations, Fig. 3a shows that analysis departures are smaller than first guess departures, as would be expected in a working assimilation system. In the case of All-Sky-Off, this shows that the other observations assimilated in the system bring a substantial amount of information that can be used to predict rain and cloud structures better. However, turning on the microwave imager assimilation in All-Sky reduces standard deviations still further, improving both the analyses (by about 3% to 10%, depending on channel) and the first guess (by about 2% to 4%). This shows that short-term forecasts in rainy areas are drawing closer to the microwave observations. The situation is similar in the cloudy sample (Fig. 3b). The addition of microwave imager observations in All-Sky improves the first guess forecast to almost the quality of the All-Sky-Off analysis, and the All-Sky analyses are 7% to 15% closer to cloudy observations than in All-Sky-Off.

The 'clear' sample standard deviations (Fig. 3c) show a similar picture for all channels except 37v and 37h, where the analysis in All-Sky-Off is 18% worse than the first guess. The 37 GHz channels have the greatest sensitivity to cloud liquid water of any of the microwave imager channels, and examining maps of the standard deviations (not shown), it is clear that this effect is mainly associated with the ITCZ and warm pool regions, so despite this being a 'clear' sample the effect is related to cloud. In fact, it is an artefact of our sampling. In order to compare the same samples for both analysis and FG, our contingency analysis is made on the basis of FG cloud only. However, it is quite possible that cloud may appear in the analysis where there was none in the FG. Because standard deviations in the correct negative category are very small, if some cloud appears in the analysis, standard deviations can easily be larger than in the FG. Turning on microwave imager observations in All-Sky mostly suppresses this behaviour, showing that clear-sky microwave imager observations constrain the analysis and help to prevent it from producing cloud where the observations show clear skies.

The right hand side of Fig. 3 shows the mean FG and analysis departures. The bias is generally larger in the analysis than in the FG. This is seen in both All-Sky and All-Sky-Off so it is independent of the all-sky observations. Other observations in the system are causing the analyses to move away from both the model (as represented by the FG) and the SSM/I observations. This suggests that in cloudy and rainy areas, the analysis is becoming either drier or less cloudy than the first guess, and in clear areas, it is being moistened. Turning on microwave imager observations in the All-Sky experiment does not suppress this behaviour.

Figure 4 shows that results for AMSR-E are similar to those for SSM/I: the analyses draw to the observations in standard deviation terms, but biases tend to become worse in the analysis. The analysis does not draw quite so close to AMSR-E as it does to SSM/I observations in similar channels. A likely explanation is that we have intentionally set the AMSR-E observation errors a little larger than those for SSM/I (Tab. 3).

3.2 Fits to other assimilated data

Figure 5 shows FG and analysis departure statistics for the microwave humidity sounders assimilated in the ECMWF system (Advanced Microwave Sounding Unit B, AMSU-B, on NOAA-16, 17 and 18; Microwave Humidity Sounder, MHS, on Metop-A). All-Sky-Off shows the highest standard deviations: the assimilation of microwave imager data in All-Sky and Control improves the fit, principally in the 183 ± 7 GHz channel. Of the three channels, this one senses lowest in the atmosphere, with a weighting function that peaks at around 800 hPa. However, the All-Sky standard deviations are marginally larger than in Control. This shows that All-Sky has a slightly weaker constraint on moisture than does Control. The changes to bias correction and mean departures are essentially negligible. The number of observations assimilated in the different experiments is

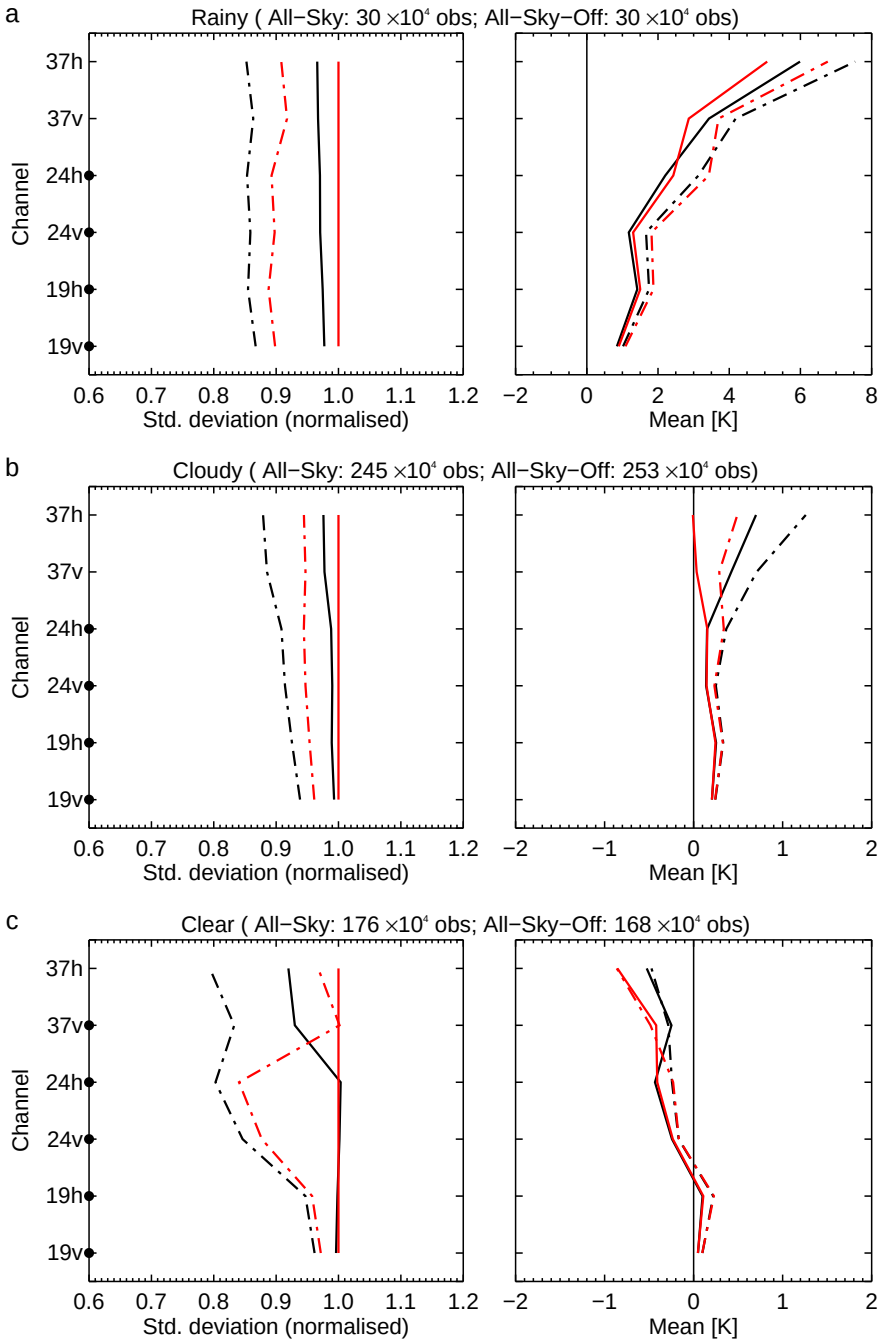


Figure 4: As for Fig. 3, but for AMSR-E.

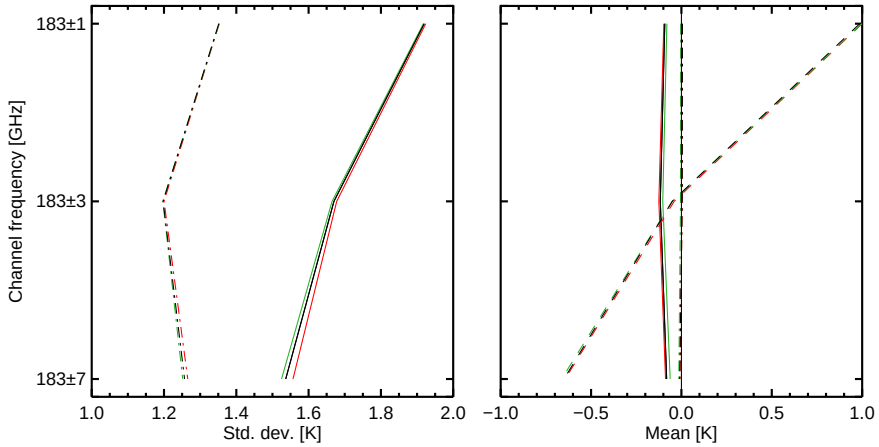


Figure 5: Standard deviation and mean of departures over the period 22nd August to 30th September 2007, for all assimilated AMSU-B and MHS radiances [in K]. All-Sky-Off is in red; All-Sky is in black; Control is in green. Analysis departures are shown as dot-dash lines and FG departures as solid lines. Dashed lines (in the mean plot only) show the bias correction.

Table 4: Pearson correlation coefficients between model RWP (at SSM/I locations) and colocated PR observations.

	log(RWP)		RWP		
Experiment	First Guess	Analysis	First Guess	Analysis	Sample
All-Sky	0.38	0.42	0.24	0.41	2795
All-Sky-Off	0.32	0.30	0.36	0.40	3002
Control (4D-Var)	0.45	0.44	0.31	0.23	406
Control (1D-Var)	0.43	0.62	0.30	0.76	412

also nearly constant. For example, there are roughly 2.5 million AMSU-B / MHS observations in Control in the 183±7 GHz channel, 0.3% fewer in All-Sky and 0.5% fewer in All-Sky-Off.

Radiosonde humidity observations have a constant bias correction (i.e. not evolved by VarBC) so they act as an anchor point for the system. Compared to these (Fig. 6), the standard deviations confirm what was seen from the microwave sounder fits: All-Sky FG standard deviations are slightly larger than Control in the 700 to 850 hPa region, but both are smaller than All-Sky-Off. However, Control has substantially larger lower-tropospheric biases compared to the All-Sky experiments, suggesting that the old two-stream technique for assimilating microwave imagers worsened the bias compared to radiosondes by drying the system. All-Sky and All-Sky-Off have very similar biases compared to radiosonde, showing that the additional humidity information included in All-Sky is compatible with and does not substantially affect the rest of the system.

3.3 Fits to independent rain retrievals

Geer *et al.* (2008) compared the ECMWF 1D-Var rain retrievals to independent '2A25' observations from the Precipitation Radar (PR) on the Tropical Rainfall Measuring Mission (TRMM, Kummerow *et al.*, 1998; Iguchi *et al.*, 2000). They found that the vertically-integrated Rain Water Path (RWP, measured in kg m^{-2}), rather than the surface rain rate, was most representative of the information content of the microwave imager retrievals. Correlation coefficients between SSM/I RWP and PR observations increased from 0.49 in the FG to 0.83 in the retrievals, clearly demonstrating the quality of the 1D-Var retrievals.

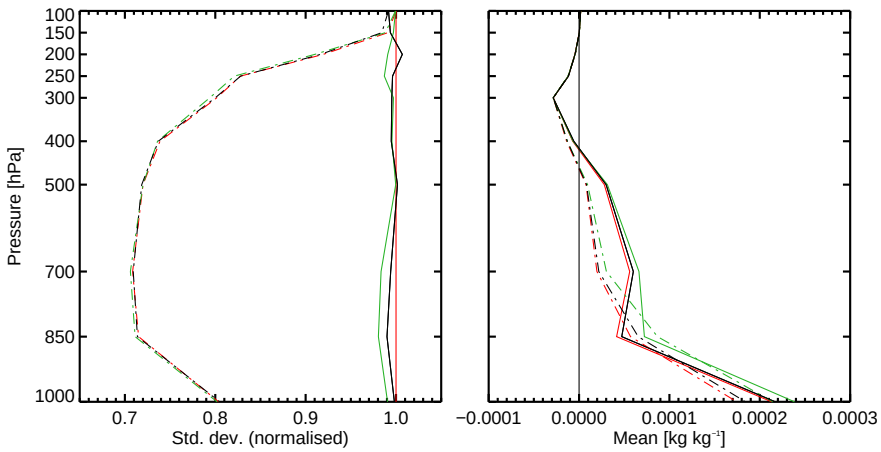


Figure 6: Standard deviation and mean of departures over the period 22nd August to 30th September 2007, for all assimilated radiosonde specific humidities. All-Sky-Off is in red; All-Sky is in black; Control is in green. Analysis departures are shown as dot-dash lines and FG departures as solid lines. The radiosonde bias correction is constant, so it is not shown. Standard deviations are normalised by the All-Sky-Off FG values.

Here, we repeat these comparisons for the All-Sky, All-Sky-Off and Control experiments (Tab. 4). Comparisons are made between PR and the 4D-Var model RWP at locations with successfully assimilated SSM/I observations. Colocations are made almost exactly in time (within ± 7.5 minutes) and exactly in space. PR has a horizontal resolution of 5 km which is matched to the 70 km by 50 km resolution of the 19 GHz channels on SSM/I by averaging all PR observations found within 25 km of the SSM/I location. The comparison is insensitive to the exact choice of radius; similar results are obtained with a 50 km radius. Correlation coefficients are calculated from the set of all points with RWP $> 10^{-4}$ kg m^{-2} in both PR and model. Calculations are done both in log(RWP), to give weight to smaller rain rates, and in RWP directly, to concentrate on the higher rain rates. See Geer *et al.* for further details.

Correlations between PR and the ECMWF model are almost the same, whether we consider All-Sky, All-Sky-Off or Control, FG or analysis. Correlations are of order 0.3 to 0.4, indicating very little agreement with the independent PR data. Numbers of colocations are quite small, as would be expected given how different TRMM's orbit is from that of the satellites which carry SSM/I. Numbers for the Control experiment are particularly low because (a) a combined pre-thinning actually removes some SSM/I observations where those from TMI (also on TRMM) are available and (b) only cloudy observations are considered, i.e. those successfully assimilated in the 1D+4D-Var path. However, similar calculations can be made at AMSR-E locations and, in Control only, TMI locations, and these back up the results presented here (not shown).

In the Control experiment we can repeat the comparisons of Geer *et al.* for the 1D-Var retrievals. Correlations between 1D-Var FG and PR are of order 0.3 to 0.4, similar to those in the full model, but the 1D-Var retrievals are substantially better, with RWP correlations reaching as high as 0.76. As shown by Geer *et al.* (2008), 1D-Var retrievals are reasonably accurate and there is useful rain information content in the microwave imager observations. Rain information from the 1D-Var retrievals does not get into the 4D-Var analysis in the Control experiment because only a total column water vapour (TCWV) pseudo-observation is assimilated (Geer *et al.*, 2008). However, in the All-Sky experiment, we hoped that direct 4D-Var assimilation would produce a better transfer of rain information into the analyses; these results show that it does not. A partial explanation is the low weight in the analysis given to the all-sky observations. However, we should also note that much work still needs to be done to allow 4D-Var to produce a credible rain analysis at the very local time and space scales of this comparison.

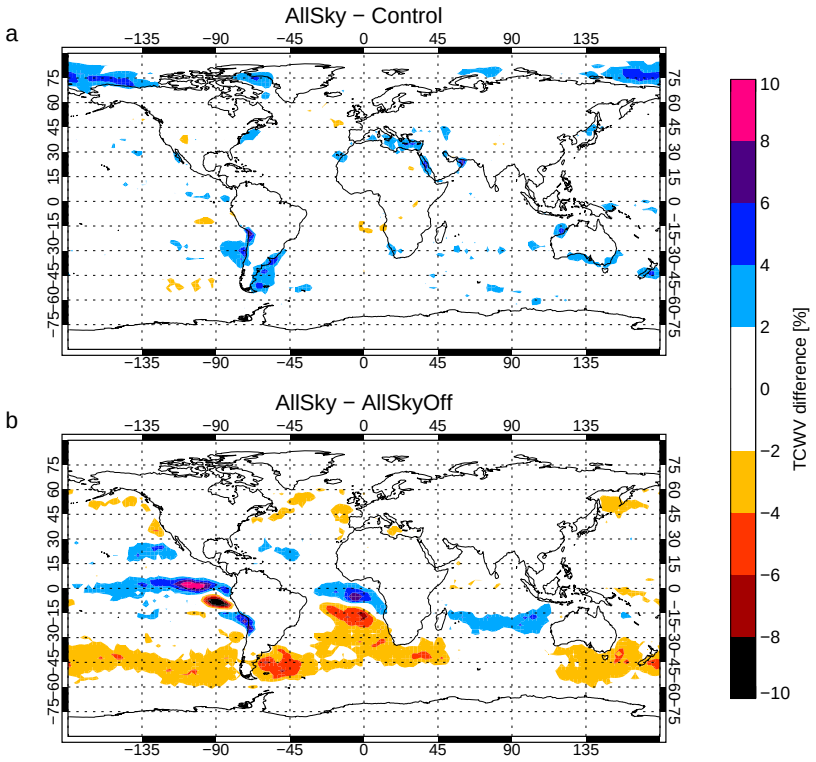


*Figure 7: Percentage change in TCWV analyses between **a)** All-Sky and Control and **b)** All-Sky and All-Sky-Off. Means calculated over the period 22 August 2007 to 30 September 2007, using analyses at 00Z and 12Z.*

4 Results in model space

4.1 Mean impacts

Moving to the all-sky approach has not made a big impact on the analysed TCWV fields. The difference between All-Sky and Control (Fig. 7a) is in most areas much smaller than the impact of adding the microwave imagers to the system in the first place (Fig. 7b). The patterns of moistening and drying caused by adding microwave observations are, in general, common to both the All-Sky and the Control sets of experiments. These patterns decay over the first 48 hours of the forecast (not shown). The main difference between All-Sky and Control is in the Arctic Ocean, where All-Sky is moister than Control. In Control, clear-sky microwave imager observations were drying the analysis there (not shown). No microwave imager observations are assimilated in this region in All-Sky, and as shown later, removing these observations is beneficial in terms of forecast scores.

Figure 8 shows the mean tropical precipitation against forecast range. In all four experiments, the 3 h forecast is about 20% too low compared to the model's equilibrium value, which it attains after about 120 h (5 days) of forecast. Precipitation forecasts at 12 h and 18 h overshoot the equilibrium value by about 12%. All experiments show similar behaviour except Control, where the rain overshoot at 12 h and 18 h is reduced. All-Sky precipitation is consistent with All-Sky-Off and Control-Off: in other words, the current All-Sky approach is in better balance with the rest of the system, and is not able to constrain the erroneous precipitation spinup, unlike the old system.

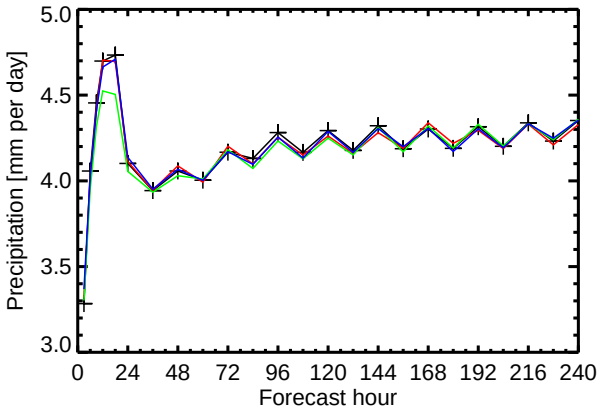


Figure 8: Precipitation from All-Sky (black with crosses), Control (green), All-Sky-Off (red) and Control-Off (blue), versus forecast time. Mean is calculated over 20°N to 20°S and 22 August to 30th September 2007.This is based on the accumulated rain in the last 3 hours (up to T+12), in the last 6 hours (T+18 and T+24), or the last 12 hours (T+36 onwards), converted into an equivalent rain rate in mm per day.

4.2 Forecast scores

Here we examine the statistics of forecast error, e_i, with index i indicating the position on a 2.5° latitude / longitude grid. The error for an experiment, X, is calculated as:

$$e_i^{\mathrm{X}} = \mathrm{forecast}_i^{\mathrm{X}} - \mathrm{analysis}_i^{\mathrm{X}}, \tag{3}$$

An analysis valid at the same time as the forecast is taken to represent 'truth'. By construction, data assimilation is supposed to produce an analysis that is a better estimate of the true atmospheric state than either model or observations alone. In this work, we use each experiment's 'own analysis' as the reference, to avoid making a prior assumption about which analysis (e.g. Control or All-Sky) is best. However, we still have to be cautious with these kind of verification measures as they are in no way independent of the system being tested. Some of the problems of forecast verification are examined in depth later on; for the moment we must be aware that caveats apply.

To compare experiment X to experiment Y for a particular pressure level and latitude range, on day j, we calculate the normalised difference in root mean squared (RMS) error:

$$D_j = \frac{\mathrm{RMS}(e^{\mathrm{X}}) - \mathrm{RMS}(e^{\mathrm{Y}})}{\mathrm{RMS}(e^{\mathrm{Y}})}. \tag{4}$$

The RMS is calculated over all i for that level falling and latitude range. For display we then calculate the mean of D_j over all n available days. The forecasts have been started at 00Z each day. Hence, for 12 h forecasts ('T+12'), there are 40 days' scores available, for there is a matching 12Z analysis available on every day of the experiment period. This reduces to 32 days for the T+192 (8-day) forecasts, since analyses are not available to verify the forecasts at the end of the experiment period.

Statistical significance is calculated using the test statistic, t, defined by

$$\Delta_j = \mathrm{RMS}(e^{\mathrm{X}}) - \mathrm{RMS}(e^{\mathrm{Y}}) \tag{5}$$

$$t = \frac{\mathrm{mean}(\Delta)}{\mathrm{Std.Dev.}(\Delta)/n^{1/2}}. \tag{6}$$

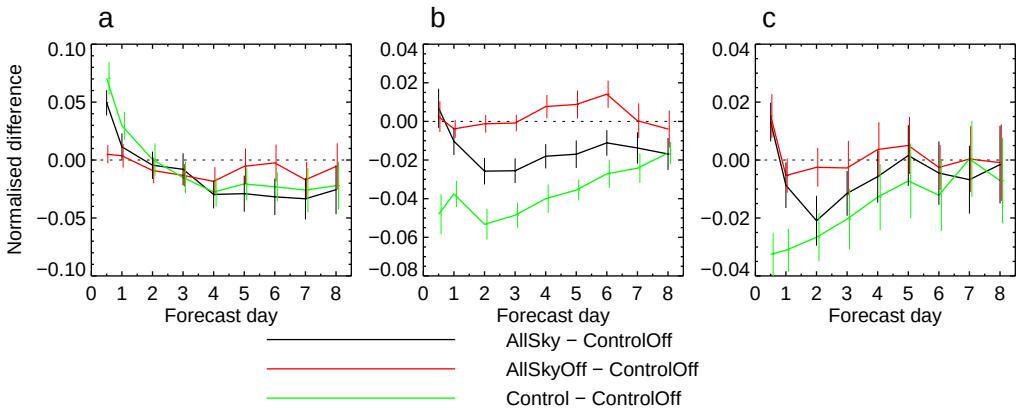


Figure 9: Normalised difference in forecast score i.e. $\frac{RMS(e^X)-RMS(e^Y)}{RMS(e^Y)}$*, against forecast time in days, for total column water vapour (TCWV):* ***(a)*** *the Southern Hemisphere (90°S to 20°S);* ***(b)*** *Tropics (20°S to 20°N),* ***(c)*** *Northern Hemisphere (20°N to 90°N). Error bars indicate the 90% confidence range. The key lists three experiments, which are all compared to Control-Off. Forecasts are verified against 'own analyses', i.e. analyses from the experiment that is being verified.*

This test statistic follows a t distribution (Sec. 5.2.3, Wilks, 2006). $\mathrm{RMS}(e^{\mathrm{X}})$ and $\mathrm{RMS}(e^{\mathrm{Y}})$ vary with the synoptic situation and hence are correlated with each other and from day to day. However, the difference between them is mostly free from serial correlation, meaning we can treat each day's Δ_j as independent.

The null hypothesis is that differences in RMS error between the two experiments are explained by random variablity. The test is two-tailed, because we want to identify both positive and negative differences. A 90% significance level is required in our regional comparisons; higher significance levels are required in latitude-pressure plots to account for multiplicity (see Sec. 4.2.3).

It is important to be clear that even when we attain statistical significance according to these tests, we might reach different conclusions if we performed our experiments in another time period (e.g. Andersson *et al.*, 1998; Bouttier and Kelly, 2001). Thus, we will concentrate on the largest and most obvious features, assuming that these have the broadest 'significance', which is backed up by noting that many of these features have been seen in our other experiments with microwave imager assimilation (e.g. Bauer *et al.*, 2006b; Kelly *et al.*, 2008).

4.2.1 Total column water vapour

Figure 9 shows the normalised difference in total column water vapour (TCWV) RMS errors for All-Sky, All-Sky-Off and Control compared to Control-Off. The most significant differences are in the tropics, where Control has lower RMS errors than Control-Off for forecast times out to 8 days. This confirms that the old method of assimilating microwave imager data worked well in the tropics. As found by Andersson *et al.* (2007), microwave imagers are the main observational constraint on tropical ocean TCWV, and as shown by Kelly *et al.* (2008), this translates into significant improvements in TCWV forecast scores. These scores are not affected by mean changes in the moisture fields; RMS errors are almost identical to the standard deviations (not shown).

Comparing All-Sky-Off to Control-Off (red line, Fig. 9) isolates the impact of turning on model physics in the first minimisation. There is very little difference between the two, so the differences between All-Sky and Control must come purely from changes in the usage of microwave imager observations.

The new system, All-Sky (black line, Fig. 9), improves the scores (reduces the RMS error) in the tropics, but

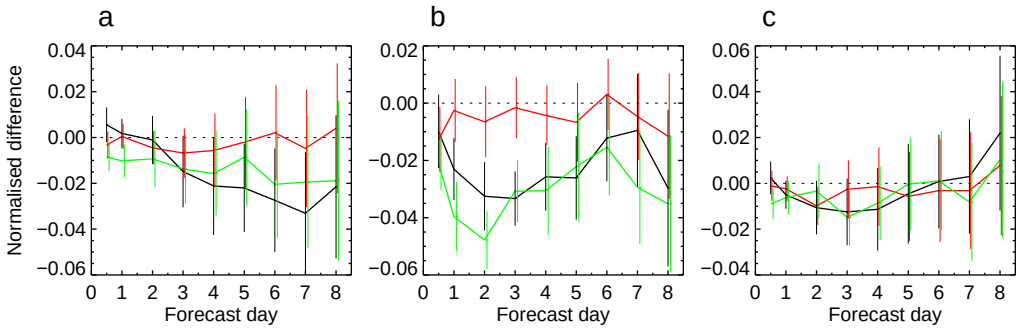


Figure 10: As Fig. 9 but for mean-sea-level pressure (MSL)

only by half as much as Control. This is roughly consistent with the FG fits to humidity observations in Figs. 5 and 6, which are slightly better for Control than for All-Sky. From B10 and Tab. 4 we know that the typical observation error is four times larger in All-Sky than in Control, and that QC removes many more cloud- and rain-affected observations. Opposed to this is a slight increase in the numbers assimilated, but overall, it is clear that the new approach does not constrain the water vapour field as well as the old system. As shown in the single observation tests of B10, it should eventually be possible to reduce the observation error assigned to All-Sky observations. However, the current observation errors and QC represent a cautious approach to including a new type of assimilation into an operational system.

In general in the northern hemisphere (NH) and southern hemisphere (SH), the RMS errors in TCWV are either unaffected or slightly reduced by microwave imager assimilation, whether in Control or in All-Sky (Fig. 9a and c). The main exception is in the first day, particularly in the SH, where microwave imager observations cause a small increase in RMS forecast error in both Control and All-Sky. The SH high latitudes in winter have always posed a challenge for the rain and cloud affected microwave assimilation, due to possible errors in surface emissivity or modelled cloud (e.g. Geer *et al.*, 2009). Also, the bias correction does not work perfectly here (Fig. 2). However, since forecasts at day 2 and beyond are unaffected, this increase in RMS error is most likely due to the natural effect of adding observations in data-sparse areas, combined with the loss of the 'recycling' effect of 1D+4D-Var. This is examined in Sec. 4.2.4.

4.2.2 Dynamical fields

Figure 10 shows the scores for mean-sea-level pressure (MSL). The main impact is in the tropics, where both Control and All-Sky significantly reduce MSL errors between days 1 and 5. Both experiments are comparable. Figure 11 shows vector wind (VW) forecast scores at 925 hPa and 700 hPa, the levels where microwave imagers have their largest impact. Again, the impact of microwave imager observations in the tropics is significant and positive in both All-Sky and Control. All-Sky appears to have slightly less impact than Control, though this difference is barely statistically significant. In both MSL and VW, the consistency between All-Sky-Off and Control-Off forecast scores (red line) confirms that the addition of model moist physics in the first minimisation did not have much impact on dynamical scores.

In the SH in the first two days, RMS errors in MSL and VW are significantly worse in All-Sky than Control, but no worse than All-Sky-Off. As for TCWV, the most likely explanation is the effects described in Sec. 4.2.4. Overall, despite the reduced impact on moisture fields compared to Control, All-Sky maintains a similar positive impact on wind and pressure scores.

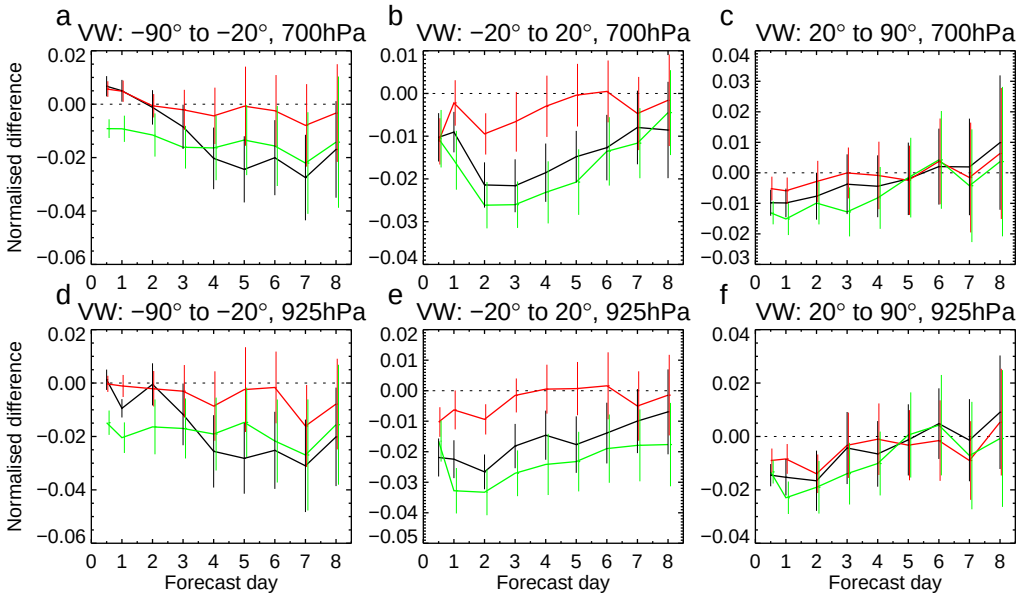


Figure 11: Normalised difference in forecast score, against forecast time in days, for vector wind (VW): ***(a - c)*** *at 700 hPa;* ***(d - f)*** *at 925 hPa, for (left to right) the Southern Hemisphere, Tropics, and Northern Hemisphere. Other details and key as for Fig. 9.*

4.2.3 Relative humidity scores

Figure 12 shows the relative humidity scores. Comparing Control to Control-Off (the green line) shows that in the old approach, microwave imager assimilation had its main positive impact on humidity forecast scores (i.e. it reduced RMS forecast errors) in the tropics and NH at 700 hPa. In contrast, forecasts at 925 hPa were apparently being degraded by microwave imager assimilation for about the first 2 days of the forecast in the SH and tropics, and over all forecast times in the NH. All-Sky (the black line) shows neither the large positive nor the large negative impacts on forecast scores, suggesting either a more balanced assimilation system, or as we might expect from the TCWV scores, a weaker observational constraint on humidity.

Figure 13 shows the difference in forecast scores between All-Sky and Control as a function of latitude and pressure level. To account for the statistical multiplicity inherent in such a figure (e.g. Livezey and Chen, 1983) we have simply increased the required confidence level to 99.5% for each individual point when determining its significance. There are 876 points in each panel of this figure, but there is often strong correlation between them, so we could think of each panel as representing a much smaller, unknown number of independent tests. Our approach would then be equivalent to making a Šidák correction for multiplicity (e.g. Abdi, 2007), assuming 21 independent tests in order to obtain a family-wide significance of 90%. Hence, in the context of the whole figure, the cross-hatching indicates significance to a much lower level than 99.5%, perhaps 90%, though we do not know exactly.

Relative humidity forecasts have lower RMS errors in All-Sky than Control at most latitudes in the lower troposphere (up to about 800 hPa), where the apparent degrading effect of imagers has been removed. Scores are degraded in the tropics at around 700 hPa, where the apparent improving effect of imagers has also been removed. These changes are largest in the first 24 h and decline thereafter, except for at very high northern latitudes.

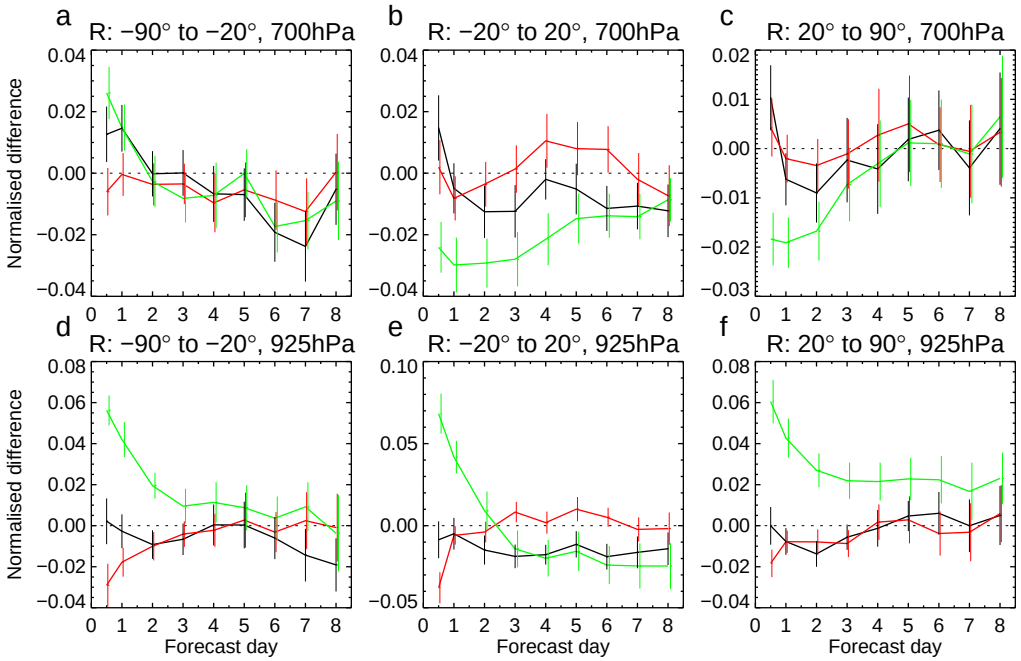


Figure 12: As Fig. 11 but for relative humidity (R) .

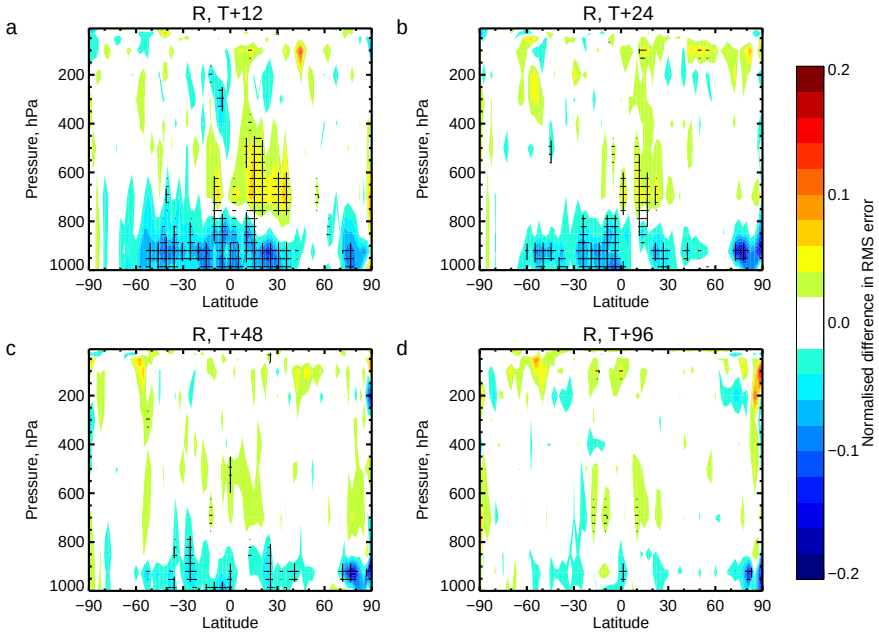


Figure 13: Normalised difference between All-Sky and Control in RMS relative humidity forecast error. Blue colours indicate where the All-Sky experiment has smaller forecast errors than Control. Red colours indicate the opposite. Cross-hatching indicates point differences significant at the 99.5% level; in the text we argue that this equates to a global significance level of roughly 90%.

Table 5: Special experiments to investigate forecast score impacts, by comparison to the Control and Control-Off experiments which were also described in Tab. 1.

Experiment name	Control	Control-No-Rain	Control-Fg-Rain	Control-Off
Rainy 1D+4D-Var	On	-	Special	-
Clear 4D-Var	On	On	On	-
ECMWF ID	f4u1	f659	f9d7	f5rc

Over the Arctic sea, we have already seen that All-Sky analyses are moister than Control (Fig. 7). In Control, clear-sky observations are assimilated over the Arctic sea and they cause a dry bias in the analyses (compared to a system without imager assimilation, i.e. Control-Off). In All-Sky, by contrast, no microwave imager observations are assimilated polewards of 60° so there is no drying effect. This effect shows up in the forecast scores poleward of 80°N in Fig. 13 and in the NH generally in Fig. 12f, where Control has higher RMS errors than All-Sky and Control-Off all the way out to 8 days. However, this is not due to the forecasts being different, because the moisture field quickly drifts back to the model's preferred state. The increased RMS error in Control comes from the clear-sky observations pulling the analyses away from this preferred state.

Comparing All-Sky-Off to Control-Off (red line, Fig. 12) again shows no significant impact on forecast scores from turning on moist physics in the first inner loop, except a minor change at T+12.

4.2.4 Forecast scores re-examined

In the Control experiment, it appears that microwave imager assimilation degrades relative humidity forecast scores at 925 hPa and improves them at 700 hPa. Table 5 introduces two new experiments which help to investigate these effects. We will mainly concentrate on relative humidity scores, but we will check the reliability of other parameters at the end of this section.

Experiment 'Control-No-Rain' is identical to Control except that 1D+4D-Var rain assimilation has been switched off. Figure 14 shows the effect on relative humidity forecast scores in the tropics. We see that 1D+4D-Var assimilation must have been responsible for the apparent improvement in forecast scores at 700 hPa in the first 24 h but the apparent degradation at 925 hPa came from the clear-sky microwave imager assimilation.

This 'degradation' shows more about the subtleties of forecast verification at short ranges than about the observations themselves. We have chosen to use own analyses as the reference. However, as noted by Bouttier and Kelly (2001), adding observations in data-poor areas can perturb the analyses relative to the forecasts. The result is that the forecasts can look worse, when verified against their own analyses, than if the observations had not been assimilated at all.

It would be tempting to choose one or other analysis as the reference for all experiments. Figure 15 shows RMS errors for the four possible combinations of forecast and verifying analysis for Control and Control-Off. As shown in Fig. 14, verifying against own analysis results in larger RMS errors for Control than Control-Off in the first 2 days. However, using the alternative analysis (e.g. Control forecasts verified against Control-Off analyses or vice-versa) results in very different errors. This is because, by construction, own-analysis RMS errors will tend to zero at the beginning of the forecast range. Using an alternative analysis for verification, the RMS error at short forecast ranges tends towards that of the difference between the two analyses, which in this case is 4.3% in relative humidity.

Rather than verifying against own analysis, if we were to take just one of the experiments as the reference, we would have an even bigger problem. If we took Control as the reference, then we would be comparing Control forecasts to Control analyses, so short range RMS forecast errors would be very small. However, we would

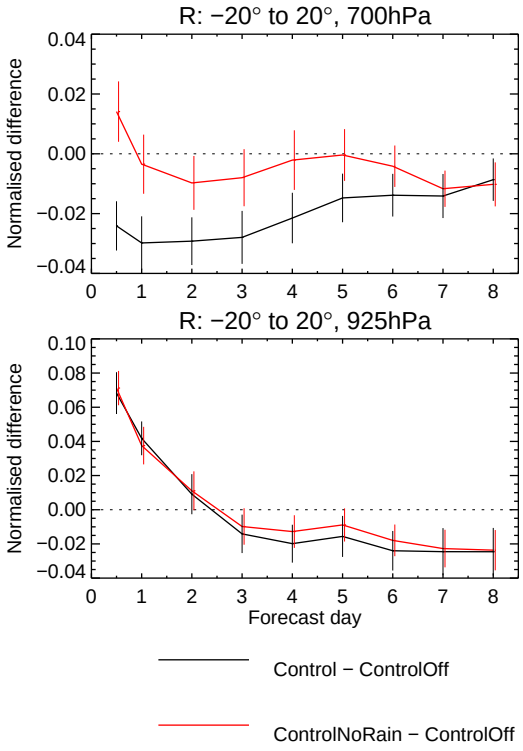


Figure 14: As Fig. 12, but concentrating on the tropics, and comparing Control and Control-No-Rain forecast scores to those of Control-Off.

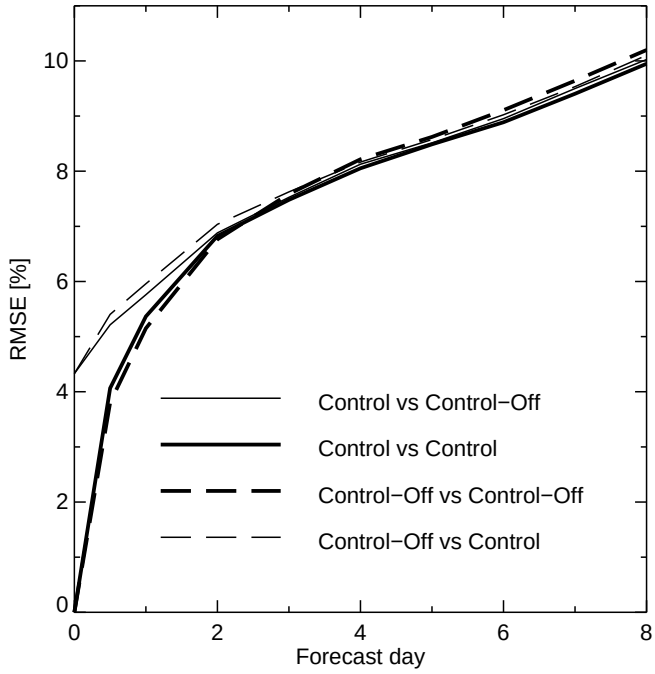


Figure 15: RMS error of forecasts verified against own or different analyses, for relative humidity in the tropics at 925 hPa.

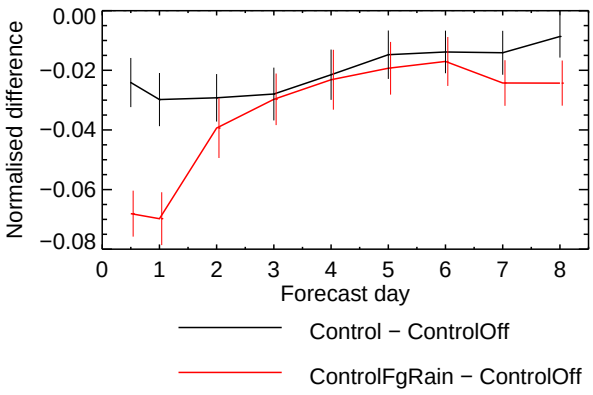


Figure 16: As Fig. 12, but concentrating on the tropics at 700 hPa, and comparing Control and Control-Fg-Rain forecast scores to those of Control-Off.

be comparing Control-Off forecasts to Control analyses, and the errors would be large and dominated by the random error of the analyses themselves. Thus, Control would appear enormously better than Control-Off, but this conclusion would be wrong.

A final possibility would be to use a third analysis as the reference, so that the random error component was similar in all the scores computed. However (though without going into details here) the presence or absence of similar observations would also make a difference to the results. For example, an experiment using microwave imagers would still show smaller RMS errors if compared to an experiment that also contained that data type than it would if compared to one that did not. Hence, it appears that verifying against 'own-analysis' is the least worst option. The apparent short-range degradation when observations are added in data-poor areas is just something we have to be aware of when looking at these statistics.

From Fig. 14 we have seen that 1D+4D-Var cloud- and rain-affected microwave assimilation appears to improve forecast scores at 700 hPa in the tropics. This has been noted in a number of studies of 1D+4D-Var over the years (e.g. Bauer *et al.*, 2006b; Kelly *et al.*, 2008). However, there must be some inconsistency if 1D+4D-Var reduces RMS errors at short ranges while clear-sky assimilation increases them.

One of the known disadvantages of 1D+4D-Var (e.g. Lopez and Bauer, 2007; Geer *et al.*, 2008) is that the 1D-Var retrieval is a weighted combination of FG and observation. When the retrieval is assimilated in 4D-Var, some part of the FG is assimilated as new data. Simply because of this, 1D+4D-Var may be making the analyses more consistent with the forecasts, and 'improving' forecast scores. To test this, we ran another experiment, Control-Fg-Rain (Tab. 5), where the FG TCWV was assimilated instead of the 1D-Var retrieval. Everything else (for example, 1D+4D-Var observation usage, quality control and observation error) was the same as in Control. Figure 16 shows the results. As expected, Control-Fg-Rain reduces RMS errors even more than Control, particularly in the first day. We can conclude that the apparent improvement in short-range relative humidity forecast scores with 1D+4D-Var actually came from the 'recycling' of the FG.

The TCWV scores (Control minus Control-Off) also suffer from these effects, but because TCWV is a vertically integrated quantity there is competition between the short-range 'degradation' coming from adding the clear-sky data, and enhancement coming from 'recycling' effect of 1D+4D-Var (not shown). However, the effects are smaller than for relative humidity scores. The FG departures shown earlier (e.g. Figs. 5 and 6) can be considered an observation-based verification of the 12 h specific humidity forecast, and these show that the humidity field is improved by microwave imager assimilation.

The 'degradation' effect is not seen in MSL and vector wind scores, apart from in the first day in the SH. We examined figures similar to Fig. 15 and in general there were consistently smaller RMS errors in the Control experiment than in Control-Off, so long as we avoided the mistake of trying to compare 'own-analysis' scores to 'other-analysis' scores.

Why are the relative humidity scores so sensitive? Fields of relative humidity, particularly in the boundary layer over ocean at 925 hPa, are dominated by fluctuations with very small spatial scales. In comparison, TCWV, MSL and wind fields are smooth and variability is dominated by synoptic or global structures. Geopotential scores outside the tropics, not shown here but widely used, are similar. For all these fields, random errors between different analyses are smaller than the typical size of forecast errors. Additionally, these fields are strongly constrained in modern analysis systems by temperature-related observations, even in the SH. Hence, there are now few 'data-sparse regions' like those encountered by Bouttier and Kelly (2001). However, the relative humidity field can still be considered data-sparse, particularly in the lower troposphere over ocean where microwave imagers remain virtually the only direct source of data (e.g. Andersson *et al.*, 2007). So, relative humidity forecast scores should be treated with caution, particularly for forecasts shorter than 3 days. TCWV is less affected than relative humidity. MSL, wind and (outside the tropics) geopotential forecast scores are more generally reliable. However, relative humidity is interesting to us in this work because it is more related to cloud formation.

In summary, the apparent degradation in relative humidity forecast errors with clear-sky microwave imager assimilation in the Control experiment is completely to be expected when introducing observations in data-poor areas, particularly when using 'own-analysis' scores (Bouttier and Kelly, 2001). The same effect results in the absurd situation that when we experimentally assimilate FG TCWV information as new observations we can substantially 'improve' forecast scores. This effect is responsible for the apparent improvements in relative humidity scores with 1D+4D-Var assimilation in the Control experiment and in the original 1D+4D-Var studies (Bauer *et al.*, 2006b).

However, we have seen in the observation-based part of this paper that both Control and All-Sky improve the specific humidity information in the 12 h forecasts. At long forecast ranges (e.g. greater than 3 days) the choice of verifying analysis is much less important. Beyond day 6 there is little difference in the relative humidity forecast scores between any of our experiments, except where the RMS error is dominated by mean effects, such as in the Arctic ocean.

5 Information transfer

Geer *et al.* (2008) investigated the transfer of information into the 4D-Var analyses by comparing FG departures and increments, and hypothesised that an all-sky direct 4D-Var would produce better results than 1D+4D-Var, particularly for the transfer of cloud and precipitation information. In the context of the old 1D+4D-Var system it was quite easy to study information transfer in terms of TCWV, LWP and RWP (their fig. 2) since the 1D-Var retrieval produces water vapour and hydrometeor profiles which can be compared to their equivalents in the 4D-Var forecast model. In the all-sky system, information is presented to 4D-Var in terms of TB, so we must work with this. However, it is possible to use simple approximate relationships to interpret TB changes in terms of TCWV and LWP. From earlier we have the Karstens *et al.* (1994) regression for LWP (Eq. 2). These authors also observed that $ln(280 - TB_{22v})$ was 97% correlated with TCWV, but did not provide regression coefficients. From our own simulations we derived the following:

$$TCWV = 161 - 34.78\, ln(280 - TB_{22v}) \tag{7}$$

It is important to point out that we are still working in observation space and we are just looking at transformed

TBs. These are not accurate retrievals of LWP or TCWV and in reality it is very hard to completely separate the effects of LWP and TCWV on TBs. However, this approach still gives us more geophysical insight than looking at TBs directly. In the rest of this section we will refer to our 'retrievals' as LWP* and TCWV* to make this obvious.

These relationships have been derived for SSM/I observations that have not been bias corrected. Biases between observations and simulations can be of order 1K, which substantially affects the results of these regressions. To avoid these biases being carried into the LWP and TCWV estimates, model TBs are corrected using the VarBC bias corrections, as done for the cloud and rain tests in Sec. 3.1.2.

To understand the transfer of information from all-sky observations into the analyses, we consider histograms of the FG departure (observation - FG) compared to the increment (analysis - FG), at observation locations. In a hypothetical assimilation system with no background constraint or any other competing observations, a scatter plot of increment vs. FG departure would lie on the 1:1 line, showing that all information in the observations had been transferred into the analyses. Figure 17 shows histogram plots for TCWV* and LWP*, based on a sample of SSM/I observations from the All-Sky experiment. The TCWV* histogram lies relatively close to the 1:1 line, showing a reasonable transfer of moisture information into the analyses, though with a less steep gradient, as would be expected given that the model background constrains the analyses.

LWP* shows a similar quality of agreement between negative (e.g. cloud-removing) departures and negative increments. However, positive (e.g. cloud-creating) departures create little new cloud in the analyses. This problem also affected the 1D+4D-Var approach (Geer *et al.*, 2008).

The TCWV* and LWP* increments in the All-Sky experiment come not just from all-sky observations, but also from everything else going into the analyses, particularly those observations that influence winds or humidities directly. To make sure that our conclusions were valid for the all-sky observations alone, we ran an experiment that assimilated nothing but all-sky observations. To avoid the analyses degrading over time, this experiment's FG was re-initialised from the All-Sky experiment before each assimilation window. Histograms like Fig. 17 were extremely similar to those shown here, indicating that our conclusions are valid.

To understand the difficulty of creating cloud in the analyses, we divided the observations according to their time in the assimilation window (Fig. 18). In the first few hours of the assimilation window (Fig. 18a), it was also quite difficult either to create or remove cloud. In roughly the last hour (Fig. 18b), it was possible to reduce cloud amounts but still difficult to increase them. It is also noticeable that a greater range of increments occurs. This suggests several problems in the current 4D-Var system. First, cloud is not part of the analysis control variable, so it is difficult either to create or destroy at the beginning of the time window. Here, cloud can only be changed by modifying wind, temperature or humidity fields. This points to the need for cloud to be included as a control variable. Later in the assimilation window, the model is more able to respond to the observations through changes in the dynamics and in the moist physics, but it is still very hard to create cloud. This may come either from the restriction on supersaturated humidity increments implied by the normalised humidity control variable (Hólm *et al.*, 2002), or because the cloud created in the analysis 'rains out' before the observation time is reached (Lopez *et al.*, 2006).

6 Conclusion

We have performed two pairs of data-denial experiments based around the ECMWF operational forecasting system. The first pair (Control and Control-Off) examine the impact of microwave imager observations with the old approach, which used a 1D+4D-Var technique for cloud and rain affected observations, and direct radiance assimilation for clear sky data. The second pair (All-Sky and All-Sky-Off) examine the performance

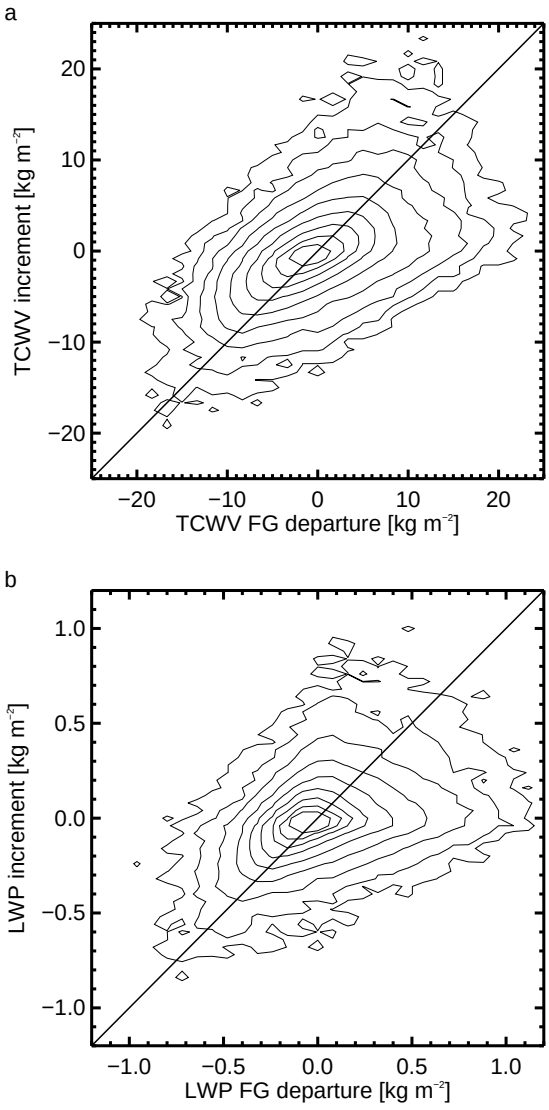


Figure 17: "Information transfer" into the 4D-Var analyses, based on 678,388 SSM/I observations from 23rd - 29th August 2007 in the All-Sky experiment. 2D histograms show FG departure against increment, for: (a) TCWV* and (b) LWP*, as derived from observed or simulated TBs. The 1:1 line is overplotted. Contours are in logarithmic steps, starting from the outermost contour: 3, 10, 32, 100, 316 etc.

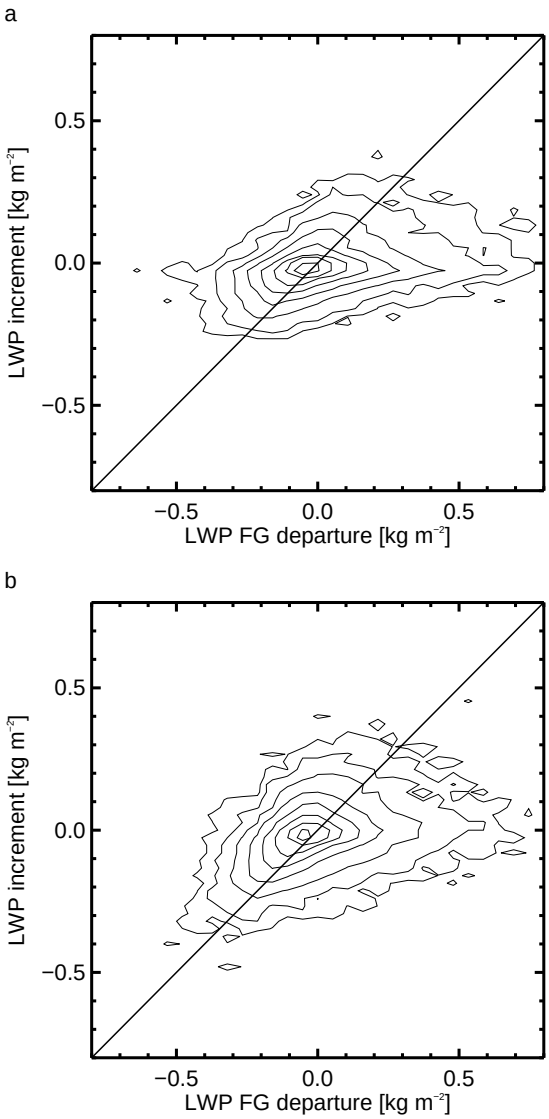


Figure 18: As for Fig. 17, but looking at LWP only and dividing the sample according to time through the assimilation window: (a) from the first 1 h 50 min (84,221 observations); (b) the last 1 h 20 min (82,649 observations).*

of the new all-sky direct assimilation that is described in Bauer *et al.* (2010) and was introduced operationally in March 2009. Our assessment is based on forecast verification scores and comparisons to observations, both assimilated and independent.

We tested the reliability of forecast scores for the validation of these experiments. Relative humidity scores show misleading effects in the initial three days of forecasts, being strongly affected by the choice of verifying analysis. Microwave imager assimilation appears to increase RMS errors in some circumstances. As pointed out by Bouttier and Kelly (2001), this can occur when including new observations in data-sparse areas. Scores based on TCWV, MSL, wind and (outside of the tropics) geopotential are more reliable, and the choice of verifying analysis is less important, unless the mistake of comparing own-analysis and other-analysis scores is made. These fields tend to be dominated by synoptic and global scale patterns. By contrast, relative humidity fields, particularly in the lowest levels, are dominated by small-scale variability. Hence, the largest component of 'forecast' error in the first few days comes from the large random errors in the analyses themselves, which are associated with these small-scale fluctuations. Because of the link between relative humidity and cloud formation, it is likely that cloud verification scores based on model and analysis fields will suffer similar problems. Hence, although relative humidity scores (and cloud scores in the future) are worth investigating, they should be treated with caution.

We took the opportunity to re-assess the old approach. There were two minor issues with the clear-sky direct assimilation. Assimilation over the Arctic Ocean dried the analyses relative to the forecasts, suggesting that better bias correction may have been required. Also, in the subtropical stratocumulus regions there were positive biases of 1 to 2 K. These disappear with the all-sky approach, which suggests that cloud-detection was failing in these areas and cloud-affected observations were being treated as clear-sky. Cloud-detection is unnecessary in an all-sky assimilation system.

As shown by Kelly *et al.* (2008), both clear-sky and 1D+4D-Var assimilation of microwave imagers are able to improve humidity and wind scores in the tropics. However, relative humidity forecast score 'improvements' at short ranges with 1D+4D-Var (e.g. Bauer *et al.*, 2006b; Kelly *et al.*, 2008) actually come from a flaw in the 1D+4D-Var technique. The 'recycling' of the humidity first guess has the effect of reducing the differences between analyses and short-range forecasts, hence artificially improving the forecast scores.

One area where the 1D-Var technique remains superior to all-sky direct 4D-Var is in the quality of the 1D-Var retrievals. The rain amount is retrieved by 1D-Var but not assimilated. However, it shows reasonable instantaneous agreement with retrievals from PR onboard TRMM (Geer *et al.*, 2008). For this purpose only, 1D-Var has an advantage in not being constrained by a 4D model or by any other observational information. In contrast, because all-sky observations are just one part of a comprehensive 4D-Var analysis, they have far less control over the model's humidity, cloud and rain fields.

Despite the limitations described above, the old approach to microwave assimilation in clear-skies, cloud and precipitation worked reasonably well in improving wind and moisture forecasts, particularly in the tropics. However, the future aim is to improve cloud and rain analyses and forecasts, and an all-sky direct radiance assimilation should be more flexible, simpler, and more accurate. To help reduce the risks of introducing such a new technique, it has been implemented with a very cautious approach to quality control, and by applying larger observation errors than used in the old system.

The result of taking this cautious approach is that, compared to a system without microwave imagers, the analysed and forecast tropical humidity fields are improved only about half as much as with the old system. This can be seen in fits to other assimilated humidity observations, such as AMSU-B mid-tropospheric humidity channels and radiosonde measurements. It is also seen in the TCWV forecast scores. However, the all-sky system is comparable with the control in terms of its positive impact on dynamical fields in the tropics, e.g. in MSL and vector wind in the lower troposphere.

Our results point to some areas requiring further work:

- Only SSM/I and AMSR-E are assimilated operationally in the new system. We would like to reintroduce SSMIS and TMI.
- Quality control excludes a lot of cloud and precipitation-affected situations, and the rather large observation errors (e.g. 12 K in the 19v channel) reduce the impact of the data. We are currently experimenting with a more complete use of the data.
- Creating cloud in the 4D-Var analysis is far more difficult that removing it, which may come from the normalised relative humidity control variable, or perhaps a 'rain-out' problem (Lopez *et al.*, 2006).
- It is easier to create or destroy cloud at the end of the 4D-Var analysis window than it is at the beginning. This might be remedied by the use of a cloud-related control variable, though background errors may also be constraining the creation of cloud.

Finally, we have not considered the issue of background errors in the presence of cloud and precipitation. It is likely that errors and correlation lengths may be very different in these areas, compared to in clear sky regions, and more work is needed there too.

Acknowledgements

Alan Geer was funded by the EUMETSAT fellowship programme. Jean-Noël Thépaut and Erland Källén are thanked for reviewing the manuscript.

References

Abdi H. 2007. The Bonferroni and Šidák corrections for multiple comparisons. In: *Encylopedia of Measurement and Statistics*, Salkind N (ed), Sage, pp. 103–107.

Andersson E, Haseler J, Undén P, Courtier P, Kelly G, Vasiljevič D, Brankovi C, Cardinali C, Gaffard C, Hollingsworth A, Jakob C, Janssen P, Klinker E, Lanzinger A, Miller M, Rabier F, Simmons A, Strauss B, Thépaut JN, Viterbo P. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). III: Experimental results. *Quart. J. Roy. Meteorol. Soc.* **124**: 1831–1860.

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

Auligné T, McNally AP, Dee DP. 2007. Adaptive bias correction for satellite data in a numerical weather prediction system. *Quart. J. Roy. Meteorol. Soc.* **133**: 631–642.

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

Bauer P, Lopez P, Benedetti A, Salmond D, Moreau E. 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, Lopez P, Salmond D, Benedetti A, Saarinen S, Moreau E. 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.

Bell W, Candy B, Atkinson N, Hilton F, Baker N, Bormann N, Kelly G, Kazumori M, Campbell W, Swadley S. 2008. The assimilation of SSMIS radiances in numerical weather prediction models. *IEEE Trans. Geosci. Remote Sensing* **46**: 884–900.

Bengtsson L, Hodges K. 2005. On the impact of humidity observations in numerical weather prediction. *Tellus* **57A**: 701–708.

Bouttier F, Kelly G. 2001. Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Quart. J. Roy. Meteorol. Soc.* **127**: 1469–1488.

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*. Eur. Cent. for Med. Range Weather Forecasts, Reading, UK, available from http://www.ecmwf.int, pp. 97–112.

Geer AJ, Bauer P, Bormann N. 2010. Solar biases in microwave imager observations assimilated at ECMWF. *IEEE Trans. Geosci. Remote Sens.* : in press.

Geer AJ, Bauer P, Lopez P. 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 AJ, Bauer P, O'Dell CW. 2009. A revised cloud overlap scheme for fast microwave radiative transfer. *J. App. Meteor. Clim.* **48**: 2257–2270.

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

Hólm E, Andersson E, Beljaars A, Lopez P, Mahfouf JF, Simmons A, Thepaut JN. 2002. Assimilation and modelling of the hydrological cycle: ECMWF's status and plans. *ECMWF Tech. Memo., 383, available from http://www.ecmwf.int* .

Iguchi T, Kozu T, Meneghini R, Akawa J, Okamoto K. 2000. Rain-profiling algorithm for the TRMM precipitation radar. *J. App. Met.* **75**: 181–189.

Karstens U, Simmer C, Ruprecht E. 1994. Remote sensing of cloud liquid water. *Meteorol. Atmos. Phys.* **54**: 157–171.

Kawanishi T, Sezai T, Ito Y, Imaoka K, Takeshima T, Ishido Y, Shibata A, Miura M, Inahata H, Spencer R. 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.

Kazumori M, Liu Q, Treadon R, Derber J. 2008. Impact study of AMSR-E radiances in the NCEP global data assimilation system. *Mon. Weather Rev.* **136**: 541–559.

Kelly GA, Bauer P, Geer AJ, Lopez P, Thépaut JN. 2008. Impact of SSM/I observations related to moisture, clouds and precipitation on global NWP forecast skill. *Mon. Weather Rev.* **136**: 2713–2726.

Kummerow C, Barnes W, Kozu T, Shiue J, Simpson J. 1998. The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Ocean. Tech.* **15**: 809–817.

Livezey RE, Chen WY. 1983. Statistical field significance and its determination by Monte Carlo techniques. *Mon. Weath. Rev.* **111**: 46–59.

Lopez P, Bauer P. 2007. "1D+4D-Var" assimilation of NCEP stage IV radar and gauge hourly precipitation data at ECMWF. *Mon. Weather Rev.* **135**: 2506–2524.

Lopez P, Benedetti A, Bauer P, Janisková M, Köhler M. 2006. Experimental 2D-Var assimilation of ARM cloud and precipitation observations. *Quart. J. Roy. Meteorol. Soc.* **132**: 1325–1347.

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.

Peubey C, McNally A. 2009. Characterization of the impact of geostationary clear-sky radiances on wind analyses in a 4D-Var context. *Quart. J. Roy. Meteorol. Soc.* **135**: 1863 – 1876.

Rabier F, Järvinen H, Klinker E, Mahfouf JF, Simmons A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteorol. Soc.* **126**: 1148–1170.

Wilks D. 2006. *Statistical methods in the atmospheric sciences*. Academic Press, 2 edn.