TECHNICAL MEMORANDUM

802

All-sky assimilation of Megha-Tropiques/SAPHIR radiances in the ECMWF numerical weather prediction system

Philippe Chambon and Alan J. Geer

Research Department

June 2017

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/en/research/publications

Contact: library@ecmwf.int

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

Abstract

The Megha-Tropiques mission is an Indo-French satellite, on a low-inclination orbit, which is dedicated to the study of tropical water and energy cycles. This mission carries a suite of payloads among which is the SAPHIR microwave humidity sounder. This instrument sounds the atmosphere with six channels around the 183.31GHz water vapour absorption band. The present study assesses the impact of SAPHIR radiance assimilation within the IFS 4D-Var data assimilation system, and it is performed within the all-sky framework which offers the unique capability of directly assimilating observations affected by clouds and precipitation. A first part of the study consists of specifying observation errors for this instrument, which provides a different set of channels compared to other microwave humidity sounders assimilated in the all-sky system. The impact of SAPHIR on ECMWF forecasts is evaluated over several periods of time and for two different IFS cycles, and is also compared to the impact of the four all-sky MHS sounders (which is evaluated by removing these observations in the band 30 °N to 30 °S). Analysis-based verification is difficult to interpret in the tropics, so the evaluation is based on fits to observations, including precipitation observations, for which a confidence test was derived. By most measures, SAPHIR produced similar benefit to activating the four MHS sounders in the tropics. Neither instrument type could significantly improve precipitation forecasts on its own, but assimilated together they improve short-range tropical precipitation skill by around 1%. When considering the activation of SAPHIR on top of the already very dense observing system including MHS and many other all-sky instruments, there was still beneficial impact on humidity fields. This shows that the impact of adding new all-sky microwave humidity sounding information has not yet saturated.

Contents

1 Introduction

A few years ago, humidity observations in cloudy and precipitating areas were thought to provide little benefit to Numerical Weather Prediction (NWP); they are now becoming one of the most important sources of data in the space-based observing system (Geer et al., 2017). Pioneering work has been performed at ECMWF to assimilate microwave humidity observations in the so-called 'all-sky' framework (Bauer et al., 2010; Geer et al., 2010). This system offers the unique capability of directly assimilating observations affected by clouds and precipitation within a 4D-Var system. Following work allowing the all-sky assimilation of SSMIS and MHS observations (Geer and Baordo, 2014; Geer et al., 2014; Baordo and Geer, 2016) the present paper describes studies preparing for the operational use of the microwave sounder SAPHIR on-board the Megha-Tropiques satellite within the ECMWF forecasting system. Operational assimilation of SAPHIR is expected to start at ECMWF with the implementation of Cycle 43r3 in mid 2017.

The Megha-Tropiques satellite (MT) is an Indo-French satellite mission built by the Indian Space Research Organization (ISRO) and the Centre National d'Etudes Spatiales (CNES) that was launched in 2011 (Roca et al., 2015). MT flies a suite of payloads to monitor the water and energy exchanges in Tropical regions, with a low-inclination orbit offering an enhanced sampling (Capderou, 2006). In particular, the SAPHIR microwave sounder (Sounder for Atmospheric Profiling of Humidity in the Intertropics by Radiometry) provides observations in six channels around the 183.31GHz water vapour absorption band which are of potential interest for all-sky assimilation. SAPHIR is a well calibrated sounder (Clain et al., 2015) which is currently used as reference for the inter-calibration effort of 183.31 GHz observations led by the GPM Intercalibration (X-CAL) Working Group (Wilheit et al., 2013).

The SAPHIR Level 1 brightness temperatures have been disseminated via Eumetsat to NWP centres since May 2014 (Level 1A2 products corresponding to the brightness temperature samples suppressing overlaps between pixels). These data are now assimilated operationally in a number of NWP centres (e.g. Meteo-France since April 2015, Chambon et al. 2015; the Japanese Meteorological Agency since June 2015; the UK Met Office since May 2016, Indira Rani et al. 2016); some centres are currently doing experiments which should lead to operational assimilation of SAPHIR data as well (e.g. the National Oceanic and Atmospheric Administration, Jones et al. 2017). All the latter centres assimilate SAPHIR clear-sky observations only; hence the ECMWF all-sky framework offers a unique opportunity to reassess SAPHIR impact for NWP taking benefit from 183 GHz observations' sensitivity to hydrometeors.

The first section of this paper describes the methodology used for SAPHIR data assimilation. In particular, the adaptation of the 'symmetric error model' which prescribes observation errors as function of identified cloudy and precipitating regions, is presented, as well as some specific aspects of the SAPHIR data quality control. The second section then describes the results of data assimilation experiments, which can be split into two categories, respectively denial experiments in the tropics and full observing system experiments.

2 Methodology

2.1 SAPHIR observations

The Megha-Tropiques orbit has a low inclination of 20 ° with respect to the equator which leads to an enhanced sampling of tropical regions as can be seen on Figure 1. While microwave sounders onboard sun-synchronous satellites typically provide two overpasses per day at the same location in the tropics,

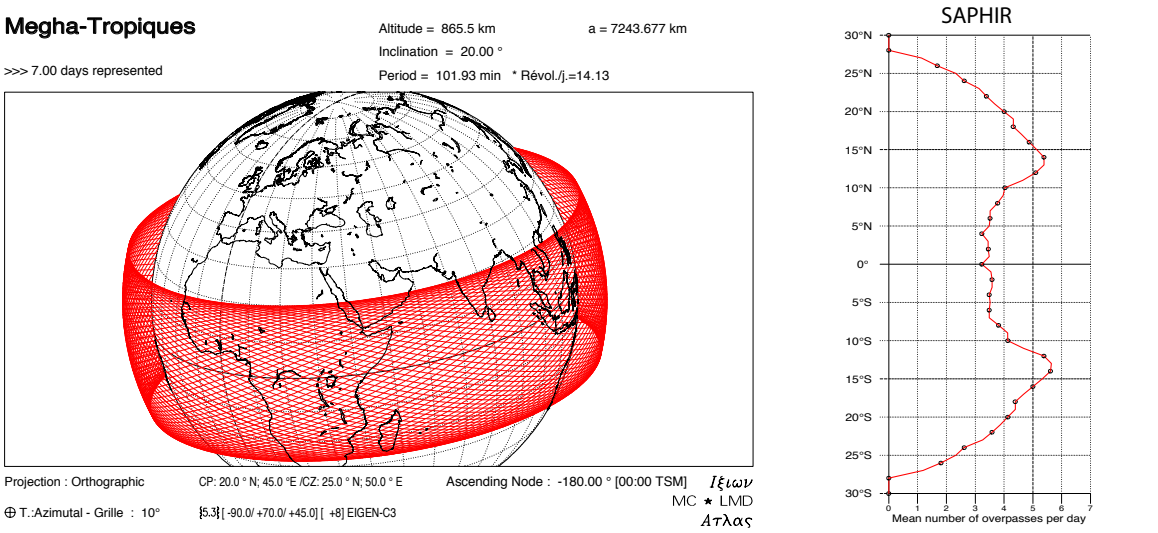


Figure 1: Left figure: Representation of 7 consecutive days of Megha-Tropiques orbit, simulated using the IXION software (Capderou, 2006). Right figure: Number of observation per 1 ° box as function of the latitude. (Figures adapted from Capderou, 2009)

the SAPHIR/Megha-Tropiques observing system provides up to five observations per day depending on the latitude (Figure 1, right). As can be seen on Figure 2, successive Megha-Tropiques observations drift westward in time, offering the opportunity to observe the same meteorological scene several times consecutively. One example is Typhoon Maysak which can be seen in the Pacific Ocean between 08h00 UTC and 12h00 UTC on March 31st, 2015 on Figure 2. This figure also illustrates the data thinning applied before assimilation in the all-sky system, which (in timeslots of 60 minutes) takes the nearest observation to every second point on a reduced Gaussian model grid, resulting in a diamond sampling pattern and a spacing of about 100 km between observations.

The SAPHIR instrument sounds the atmosphere with six channels around the 183.31 GHz water vapour absorption band, listed in Table 1. Hence, one of the main difference with an MHS-like sounder is that it observes the atmosphere with twice the number of channels, at the cost of a lack of window channels like the 89 GHz or the 157 GHz channels available from MHS. These latter window channels would have been available from the Megha-Tropiques platform with the MADRAS conical scanning radiometer but this failed and was declared unoperational by ISRO and CNES after January 2013 (Roca et al., 2015).

SAPHIR channels are sensitive to both atmospheric humidity and hydrometeors. Figure 3 illustrates typical sensitivities of these channels for a clear atmosphere. While similar frequencies are available from other platforms for channels 2 to 5, channels 1 and 6 are unique and extend the sounding range up to around 200 hPa at the top and 900 hPa at the bottom. Figure 4 gives an example of SAPHIR channels in a precipitating environment for which one can see, as expected, that SAPHIR brightness temperatures are mainly sensitive to snow profiles. As reported in the literature (e.g. Hong et al., 2005), sensitivities of 183.31GHz observations to other hydrometeors like cloud ice are of one order of magnitude smaller.

2.2 Setting of observation errors

Within the ECMWF all-sky framework, a brightness temperature-based cloud predictor is used to prescribe observation errors used in the 4D-Var data assimilation system. This flow-dependency of obser-

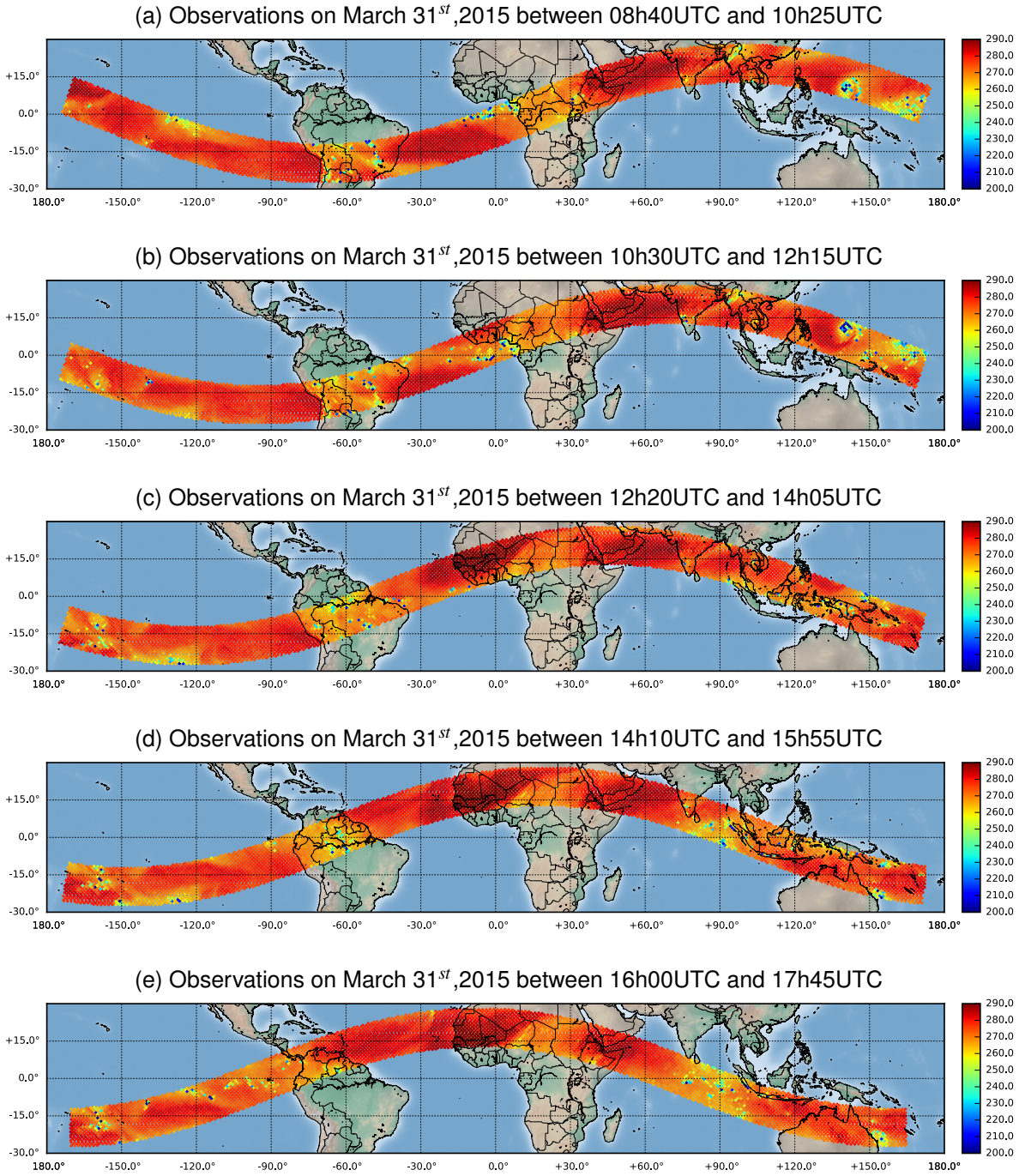


Figure 2: SAPHIR observations at 183.31±6.8GHz (channel 5) derived from five consecutive orbits of the Megha-Tropiques satellite.

Channel Number	Central Frequencies	Channel bandwidth
Channel 1	183.31 ± 0.2GHz	200MHz
Channel 2	183.31 ± 1.1GHz	350MHz
Channel 3	183.31 ± 2.8GHz	500MHz
Channel 4	183.31 ± 4.2GHz	700MHz
Channel 5	183.31 ± 6.8GHz	1200MHz
Channel 6	183.31 ± 11.GHz	2000MHz

Table 1: List of SAPHIR channels.

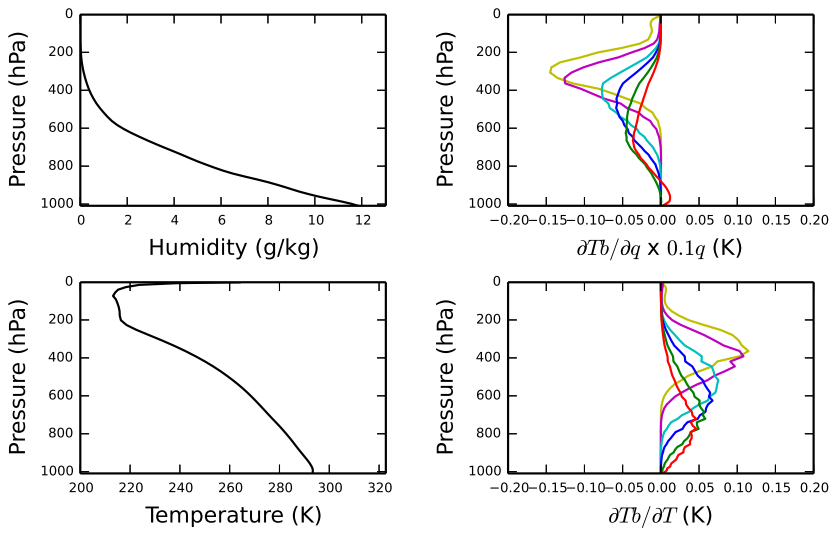


Figure 3: Example of Jacobians for Specific Humidity and Temperature for SAPHIR channels for a clear atmosphere. The mean profiles used are shown in the first column. Sensitivities of the SAPHIR channels to a change of 10% of the humidity profile (first row), to a change of 1 Kelvin (second row) are shown in the second column. Yellow: 183.31±0.2GHz, Magenta: 183.31±1.1GHz, Cyan: 183.31±2.8GHz, Blue: 183.31±4.2GHz, Green: 183.31±6.8GHz, Red: 183.31±11GHz.

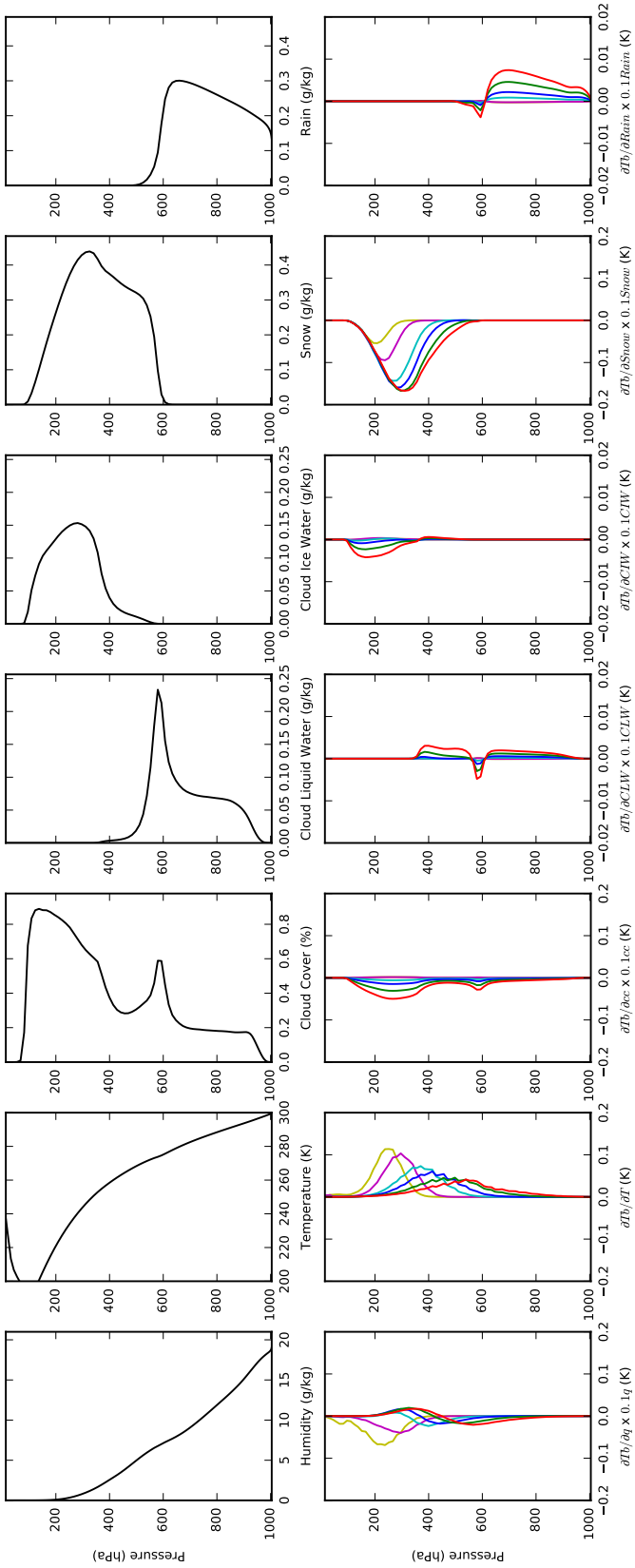


Figure 4: Example of Jacobians for Specific Humidity, Temperature, Cloud cover, Rain and Snow content for SAPHIR channels for a rainy atmosphere. The mean profiles used are shown in the first row. Sensitivities of the SAPHIR channels to a change of 10% of the humidity profile (first column), to a change of 1 Kelvin (second column), and to a change of 10% of cloud cover (third column), rain (fourth column) and snow (last column) are shown in the right column. Yellow: 183.31±0.2GHz, Magenta: 183.31±1.1GHz, Cyan: 183.31±2.8GHz, Blue: 183.31±4.2GHz, Green: 183.31±6.8GHz, Red: 183.31±11GHz.

vation errors takes into account the larger first guess departures that are encountered when at a given location clouds are observed but not predicted or vice versa.

The symmetric cloud predictor currently used over oceans for MHS all sky assimilation is based on the Scattering Index defined as follows (Geer et al., 2014):

$$\text{Scattering Index for MHS over Oceans (SI)} = (\text{TB}_{90GHz} - \text{TB}_{150GHz}) - (\text{TB}_{90GHz}{}^{clr} - \text{TB}_{150GHz}{}^{clr})$$

$$\text{Symmetric Cloud Predictor for MHS over Oceans} = \frac{1}{2} \times (\text{SI}_{FG} + \text{SI}_{OBS})$$

This index approximately measures the amount of scattering from frozen hydrometeors using the difference between 90 GHz and 150 GHz brightness temperatures; this difference is corrected from the water vapour absorption by subtracting the simulated difference without the hydrometeors effect (brightness temperatures with 'clr'). It can be computed for both observations and model first guess and then averaged out to derive the so-called 'symmetric cloud amount' (Geer and Bauer, 2011).

As mentioned above, no 90 GHz nor 150 GHz channels are available from the Megha-Tropiques platform, hence, the approach used for MHS observation errors need to be adapted. The scattering index for MHS over Oceans can be computed either purely from simulations or from a mixture of observed and simulated brightness temperatures as it needs clear sky radiative transfer calculations even in cloudy situations. Therefore, two methods would be possible to investigate which cloud predictor would be suitable for SAPHIR: (a) colocate MHS, SAPHIR data and IFS profiles and then compare the MHS predictor and potential ones for SAPHIR, (b) work purely on simulated data and compute the MHS predictor from radiative transfer simulations as well as potential predictors for SAPHIR with the same dataset of atmospheric profiles. The second method was chosen for simplicity and a study was performed to select a scattering index for SAPHIR which would behave similarly to the MHS one: (i) Forward RTTOV V11 simulations of both MHS and SAPHIR channels have been performed on the same sample of 3.8×10^6 atmospheric model profiles (ii) the Scattering Index for MHS was computed as well as several potential cloud predictors for SAPHIR made of combinations of channels; (iii) correlations were computed between SI MHS and the potential cloud predictors for SAPHIR. A high correlation is an indicator that the scattering index for MHS and the one for SAPHIR behave consistently.

Table 2 indicates that a 0.9 correlation can be reached by simply considering an index based on SAPHIR channel 6, which is the difference between the all-sky brightness temperature (simulated here, but it could also be the observed value) and the simulated clear-sky brightness temperature. This simple difference hence explains more than 80% of the MHS cloud predictor variance; it was therefore selected as new cloud predictor for both SAPHIR observations and SAPHIR model first guess and then symmetrised by averaging the latter two as for other sensors within the all-sky system. Note that this predictor is used equally over land and ocean, unlike the equivalent MHS predictor. This is because in the tropics, the high water vapour burden means 183 GHz channels are less sensitive to the surface (and for SAPHIR a screening has been implemented for rare situations when this is not true) so there is no need to construct separate indices for land and ocean surfaces.

First guess departure statistics have been collected for a 10-day period in January 2015, and categorized as a function of the symmetric cloud predictor for SAPHIR. Figure 5 shows the standard deviation of first guess departures for each of the six SAPHIR channels. For each of the figures, one can see the increase in standard deviation of first guess departures with the symmetric cloud amount. The blue curves correspond to simple modeling of the empirical values, with a piece-wise approach as described in Geer and Bauer (2011): (i) a constant error g_{Clr} prescribes clear sky errors up to a given cloud amount C_{Clr}, (ii) a function prescribes increasing errors as a function of the symmetric cloud predictor, with

a linear, quadratic or mixture of linear and quadratic dependence, up to a given cloud amount C_{Cld}, (iii) a maximum constant error g_{Cld} prescribes errors when the symmetric cloud amount is greater than C_{Cld}. The numerical values of these parameters, corresponding to the blue curves on Figure 5, are given in Table 3. One can note that the linear model fits better the behaviour of channels 5 and 6 first guess departures while the quadratic model fits better that of channels 1 and 2. Channels 3 and 4 show intermediate characteristics which led to the choice of a mixture of linear and quadratic model.

The lowest peaking channels' first guess departure distributions are characterized by long positive and negative tails (left columns of Figures 6 and 7), corresponding to mis-located clouds in the first guess with respect to observations. Also visible is an asymmetry between positive and negative first guess departures, with an excess of positive values between ∼10 K and ∼20 K particularly visible for channel 5 on Figure 7. This latter excess of positive first guess departures can also be observed with the MHS sounder (see Figure 2c of Geer et al., 2014) ; it may correspond to a model bias in the tropics related to convective activity in the Inter-Tropical Convergence Zone with small cloud fractions. The necessary investigation to fully understand this bias is not in the scope of this paper and left for future work.

Applying varying observation errors improves the Gaussianity of normalized first guess departures as can be seen on Figures 6 and 7 (right column). The 3^{rd} and 4^{th} moment of each distribution is also significantly reduced and closer to that of a Gaussian distribution (see Table 4).

2.3 Quality control

As for all observations in the IFS, a first guess check is applied, based on the size of the normalised FG departures. For SAPHIR, the cutoff is at 2.5 times the normalised departure, clearly visible in Figure 6. Also, as for most observations, VarQC (Andersson and Järvinen, 1998) is applied to downweight outliers but this affects only a few observations; see e.g. Bauer et al. (2010) or Geer and Bauer (2011) for more details on how the FG check and VarQC affect all-sky observations. There are also SAPHIR-specific screening decisions. The first comes because the cloud predictor used for prescribing SAPHIR observation errors is based on channel 6 as explained above. Following the choices made for the assimilation of MHS all-sky observations for which channels 1 and 2 are discarded from the minimization, it was chosen to not assimilate SAPHIR channel 6 and rely on the information content of channels 1 to 5 to improve IFS analyses and forecasts.

The second main quality control which has been performed for the SAPHIR instrument is related to surface transmittance. As SAPHIR has no window channels, the dynamical estimation technique of surface emissivity used for other microwave sounders like AMSU-B or MHS (Karbou et al., 2010; Baordo and Geer, 2016) cannot be applied and it has been replaced by the use of static atlases. In the tropics where the water vapour leads to a quite opaque atmosphere at 183.31 GHz, this is a reasonable compromise. Nonetheless, in some regions of the tropics where the atmosphere can be very dry like over deserts, the surface transmittance can go up to 0.6 (see Figure 8). Based on the fact that surface emissivity atlases may not give accurate enough estimations for radiative transfer simulations in these regions, the conservative choice was made to not assimilate SAPHIR observations for which the computed surface transmittance is higher than 0.25. An example of resulting coverage is provided on Figure 9.

Correlation with	SI MHS	Correlation with	SI MHS	Correlation with	SI MHS	Correlation with	SI MHS	Correlation with	SI MHS	Correlation with	SI MHS
$\mathbf{TB_1}$ - $\mathbf{TB_1}^{clr}$	-0.62										
$\mathbf{TB_2}$ - $\mathbf{TB_2}^{clr}$	-0.71	$(\mathbf{TB_1}\text{-}\mathbf{TB_2})$ - $(\mathbf{TB_1}^{clr}\text{-}\mathbf{TB_2}^{clr})$	0.75								
$\mathbf{TB_3}$ - $\mathbf{TB_3}^{clr}$	-0.81	$(\mathbf{TB_1}\text{-}\mathbf{TB_3})$ - $(\mathbf{TB_1}^{clr}\text{-}\mathbf{TB_3}^{clr})$	0.83	$(\mathbf{TB_2}\text{-}\mathbf{TB_3})$ - $(\mathbf{TB_2}^{clr}\text{-}\mathbf{TB_3}^{clr})$	0.85						
$\mathbf{TB_4}$ - $\mathbf{TB_4}^{clr}$	-0.85	$(\mathbf{TB_1}\text{-}\mathbf{TB_4})$ - $(\mathbf{TB_1}^{clr}\text{-}\mathbf{TB_4}^{clr})$	0.86	$(\mathbf{TB_2}\text{-}\mathbf{TB_4})$ - $(\mathbf{TB_2}^{clr}\text{-}\mathbf{TB_4}^{clr})$	0.87	$(\mathbf{TB_3}\text{-}\mathbf{TB_4})$ - $(\mathbf{TB_3}^{clr}\text{-}\mathbf{TB_4}^{clr})$	0.87				
$\mathbf{TB_5}$ - $\mathbf{TB_5}^{clr}$	-0.88	$(\mathbf{TB_1}\text{-}\mathbf{TB_5})$ - $(\mathbf{TB_1}^{clr}\text{-}\mathbf{TB_5}^{clr})$	0.88	$(\mathbf{TB_2}\text{-}\mathbf{TB_5})$ - $(\mathbf{TB_2}^{clr}\text{-}\mathbf{TB_5}^{clr})$	0.89	$(\mathbf{TB_3}\text{-}\mathbf{TB_5})$ - $(\mathbf{TB_3}^{clr}\text{-}\mathbf{TB_5}^{clr})$	0.88	$(\mathbf{TB_4}\text{-}\mathbf{TB_5})$ - $(\mathbf{TB_4}^{clr}\text{-}\mathbf{TB_5}^{clr})$	0.88		
$\mathbf{TB_6}$ - $\mathbf{TB_6}^{clr}$	**-0.90**	$(\mathbf{TB_1}\text{-}\mathbf{TB_6})$ - $(\mathbf{TB_1}^{clr}\text{-}\mathbf{TB_6}^{clr})$	0.90	$(\mathbf{TB_2}\text{-}\mathbf{TB_6})$ - $(\mathbf{TB_2}^{clr}\text{-}\mathbf{TB_6}^{clr})$	0.90	$(\mathbf{TB_3}\text{-}\mathbf{TB_6})$ - $(\mathbf{TB_3}^{clr}\text{-}\mathbf{TB_6}^{clr})$	0.89	$(\mathbf{TB_4}\text{-}\mathbf{TB_6})$ - $(\mathbf{TB_4}^{clr}\text{-}\mathbf{TB_6}^{clr})$	0.88	$(\mathbf{TB_4}\text{-}\mathbf{TB_6})$ - $(\mathbf{TB_5}^{clr}\text{-}\mathbf{TB_6}^{clr})$	0.87

Table 2: Correlation coefficients between several potential cloud predictors for the SAPHIR sounder and the Scattering Index (SI) used as cloud predictor for the MHS sounder over oceans. These correlations are computed from 3.8×10^6 samples. Each cloud predictor value has been computed from the Meteo France AROME model running over the Indian Ocean and the RTTOV-SCATT radiative transfer model over the January and February 2015 period.

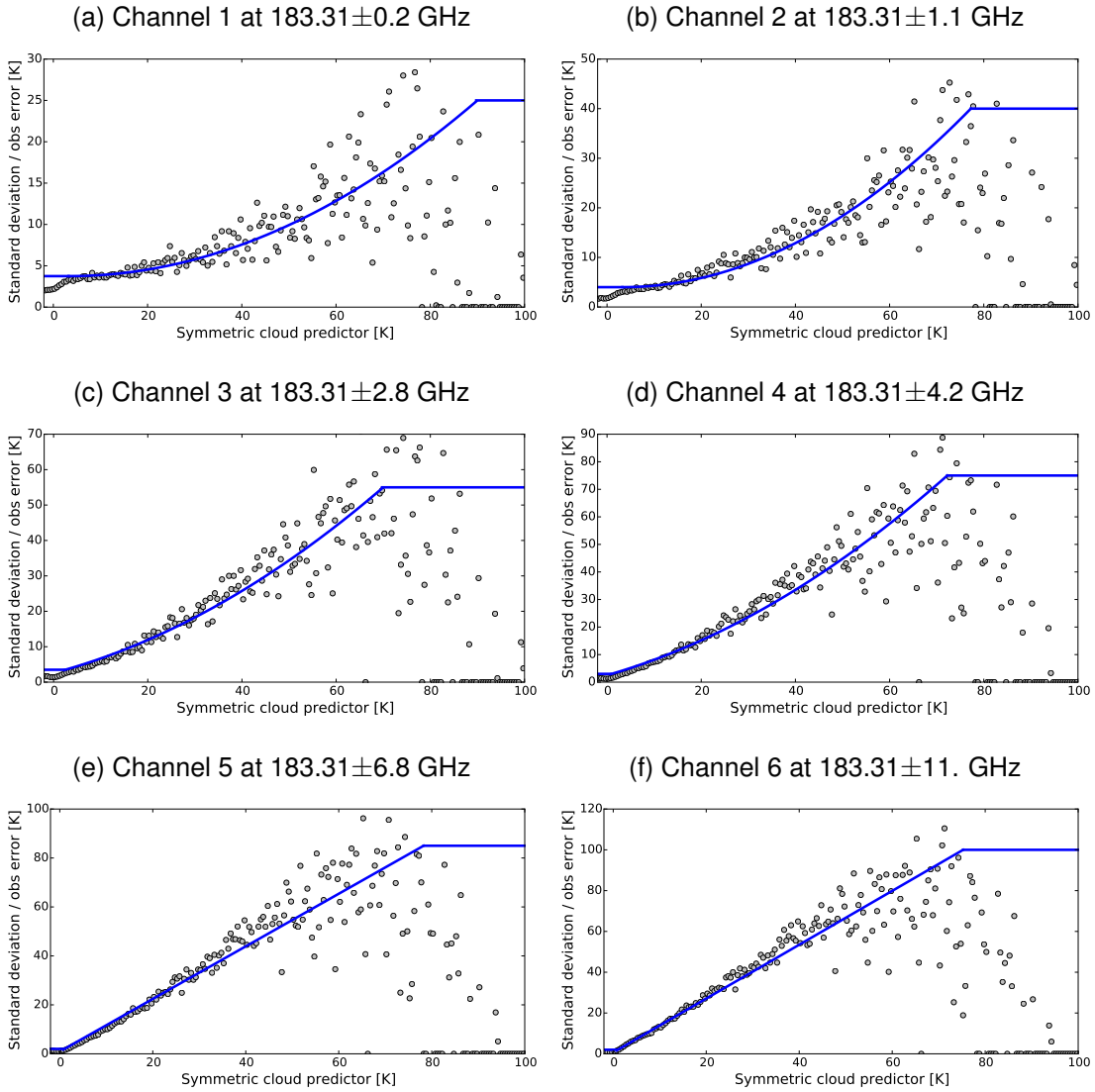


Figure 5: Standard deviation of SAPHIR first guess departures over the 10-day period from January 1^{st}, 2015 to January 10^{st}, 2015, categorized as function of the selected symmetric cloud predictor (grey dots). Modeling of the observation errors (blue lines) for each channel

	Model Type	g_{Clr}	g_{Cld}	C_{Clr}	C_{Cld}
Channel 1	Quadratic	3.8	25.	3.2	90.
Channel 2	Quadratic	4.0	40.	3.2	77.2
Channel 3	0.5 Linear and 0.5 Quadratic	3.5	55.	2.5	70.
Channel 4	0.5 Linear and 0.5 Quadratic	3.0	75.	1.0	72.2
Channel 5	Linear	2.0	85.	1.0	78.2
Channel 6	Linear	2.0	100.	1.0	75.2

Table 3: Parameters for the observation error models. All units are in Kelvin

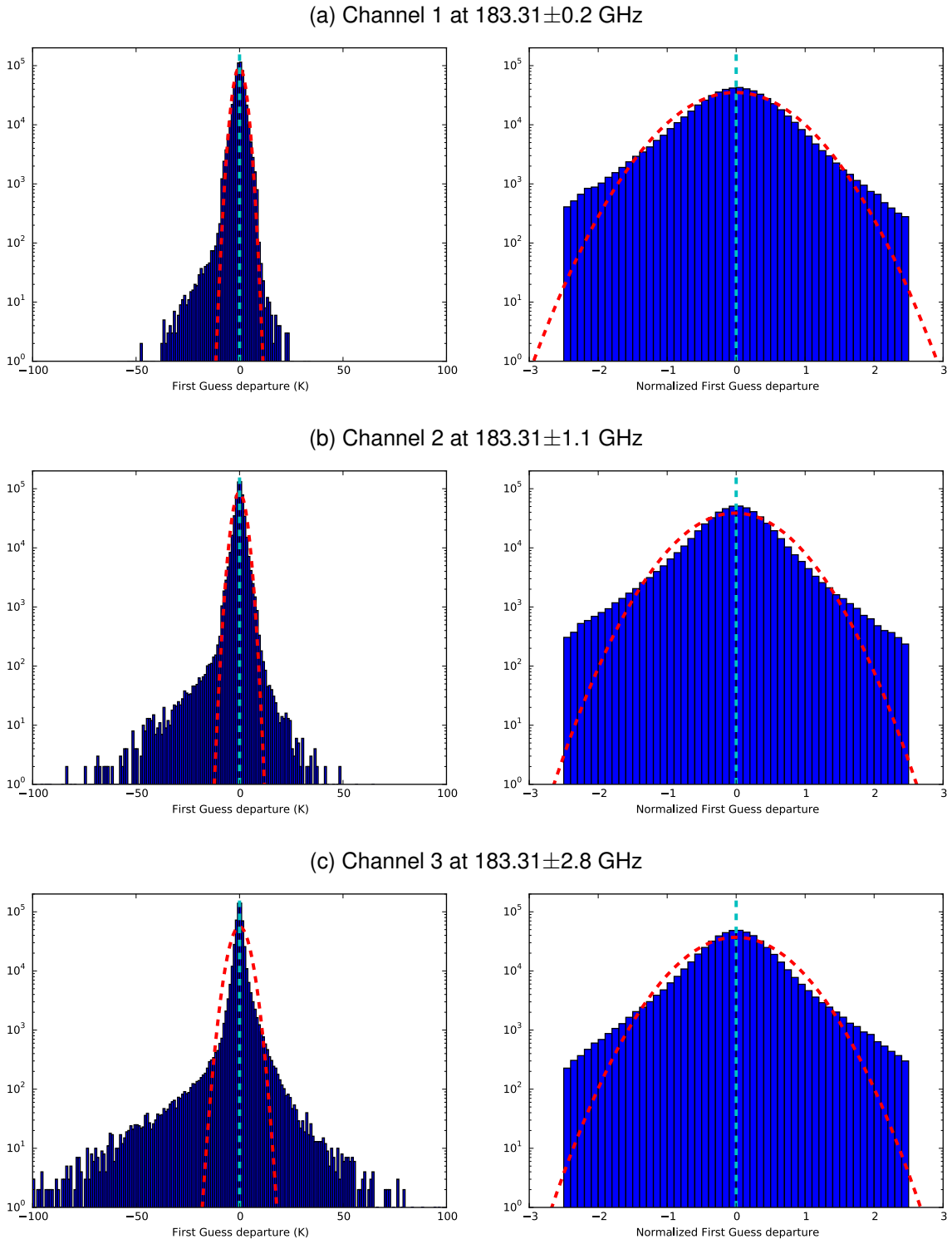
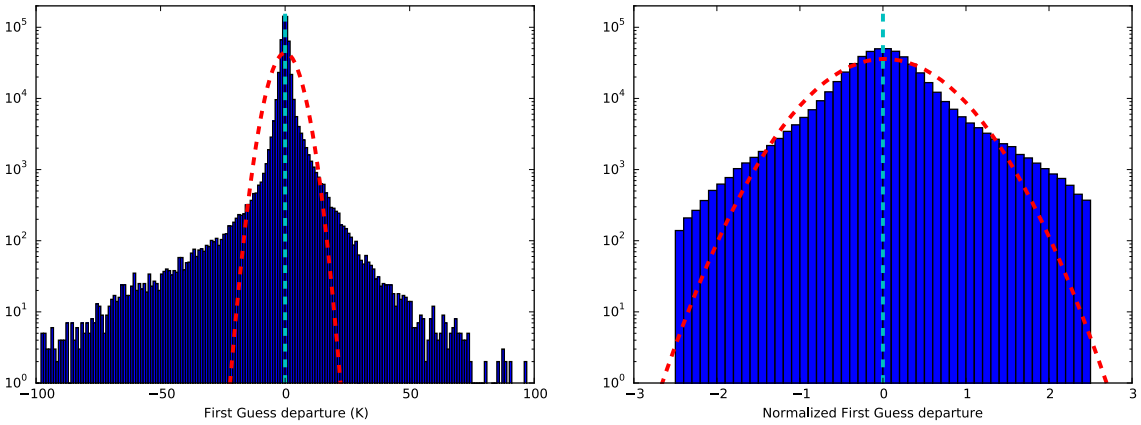
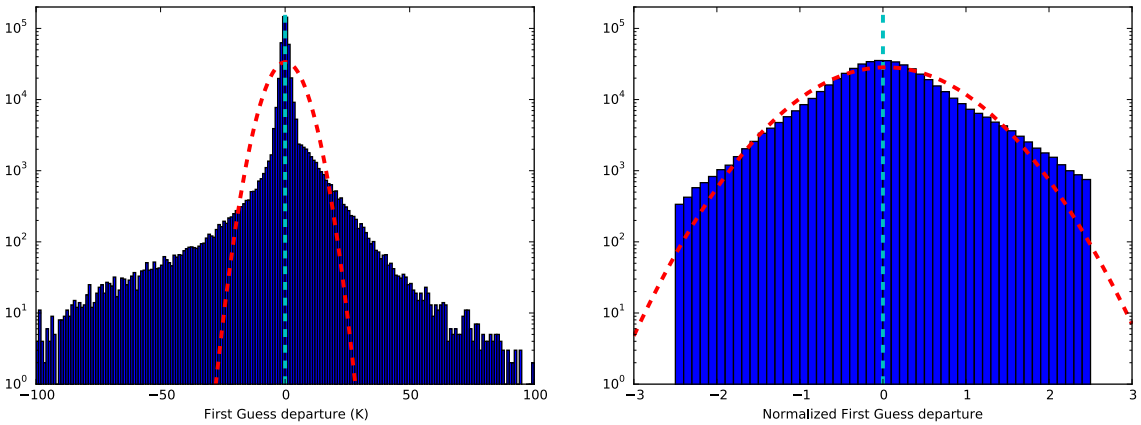


Figure 6: Histograms of first guess departures (left column) and normalized first guess departures (right column) for SAPHIR channels 1 to 3 over the 10-day period from January 1^{st}, 2015 to January 10^{st}, 2015. Observations have been screened as for assimilation (e.g. the FG check has been applied.) The red dashed lines correspond to Gaussian distributions of mean and standard deviation taken from the first two moments of the first guess departures and normalized first guess departure populations.

(a) Channel 4 at 183.31±4.2 GHz



(b) Channel 5 at 183.31±6.8 GHz



(c) Channel 6 at 183.31±11. GHz

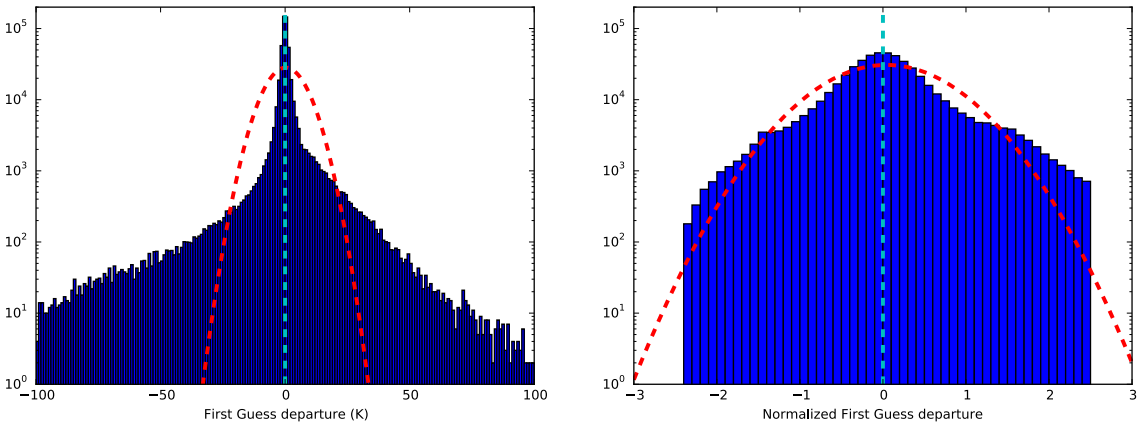


Figure 7: Same as Figure 6 for SAPHIR channels 4 to 6.

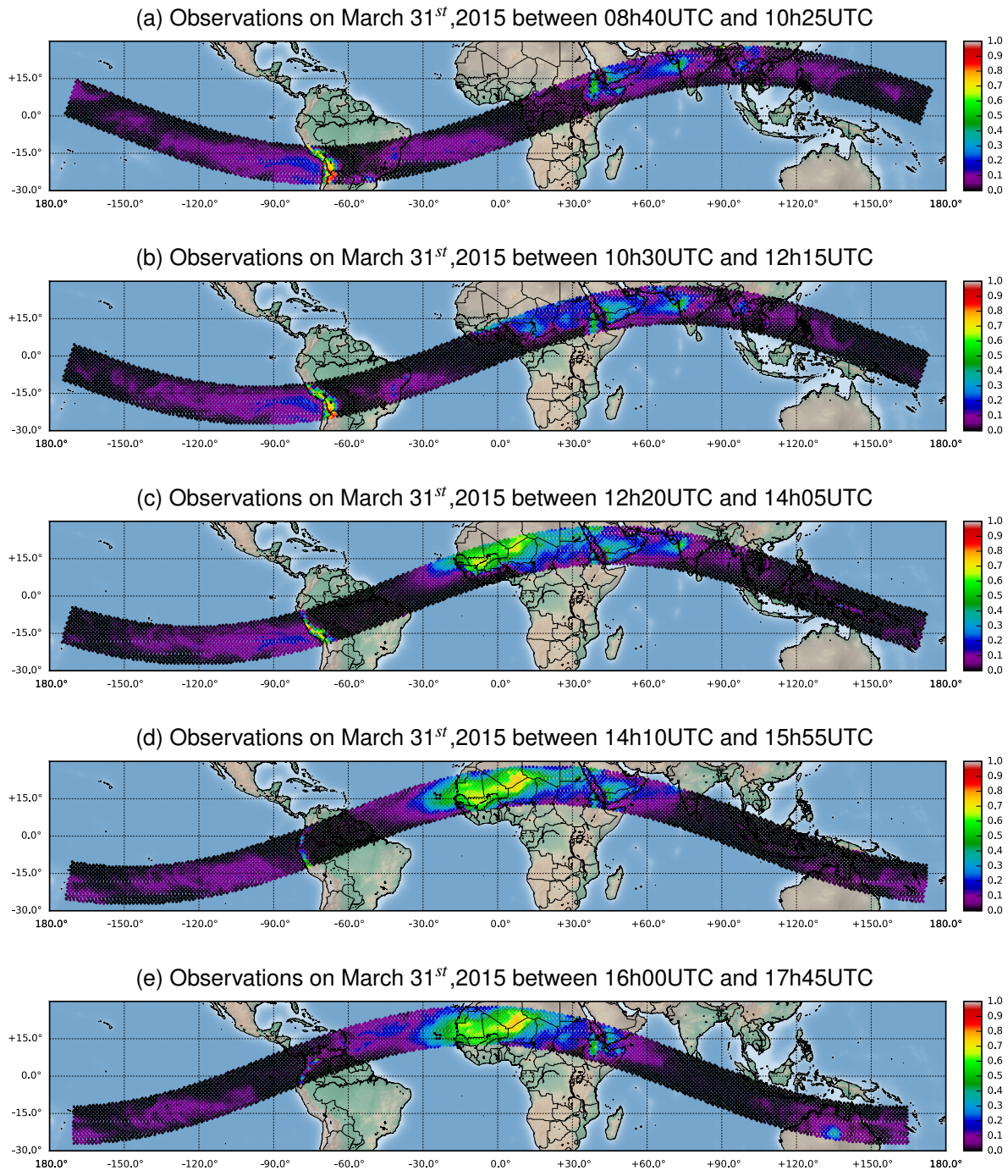


Figure 8: Surface transmittance for SAPHIR observations at 183.31±11GHz (channel 6) for five consecutive orbits of the Megha-Tropiques satellite.

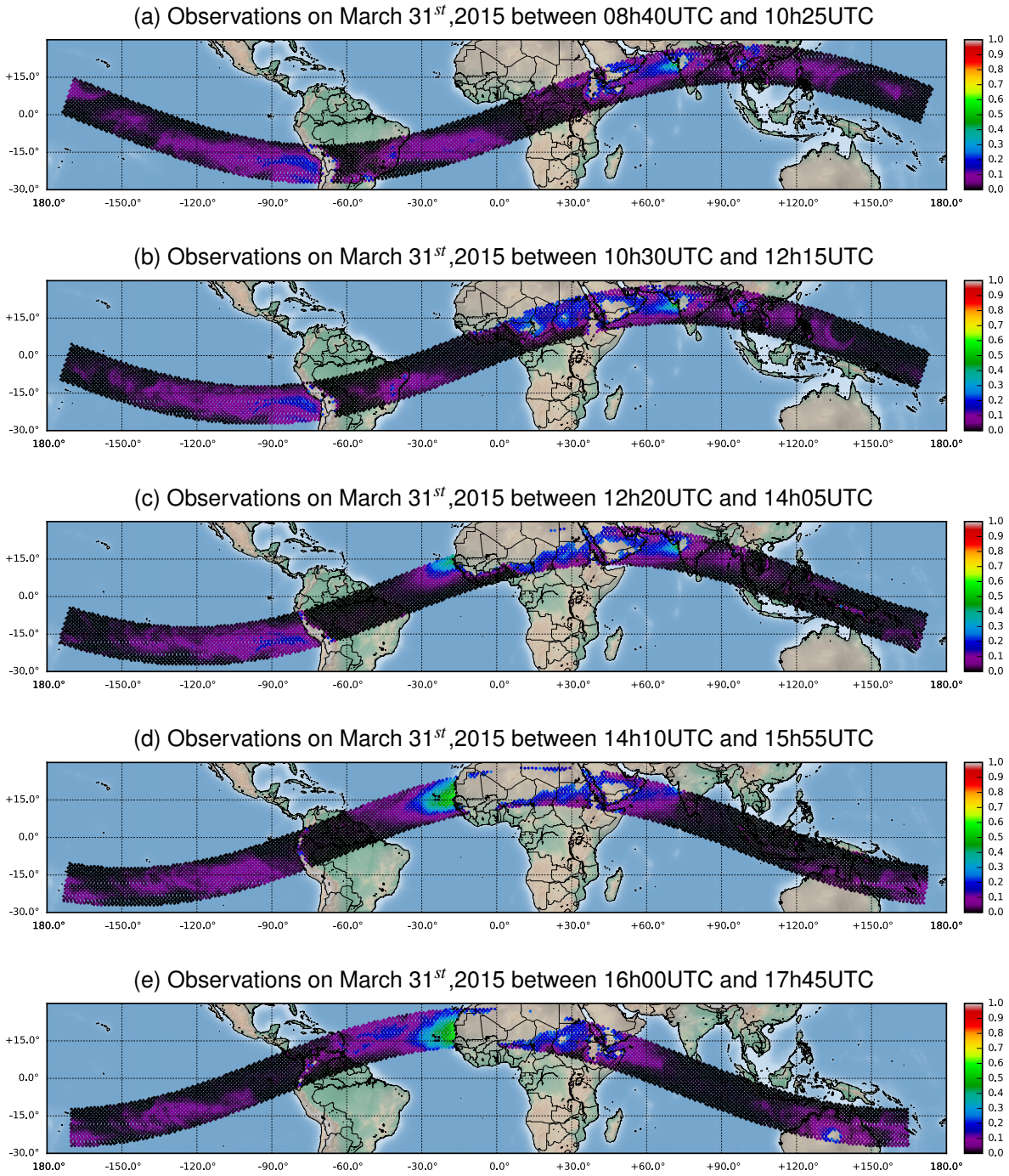


Figure 9: Same as Figure 8 with screened locations when the transmittance is higher than 0.25.

	Kurtosis without normalization by error model	**Kurtosis with normalization by error model**	Skewness without normalization by error model	**Skewness with normalization by error model**
Channel 1	9.42	**1.26**	-0.78	**-0.14**
Channel 2	89.20	**2.17**	-3.76	**-0.12**
Channel 3	123.50	**2.04**	-4.54	**0.04**
Channel 4	100.35	**2.27**	-3.61	**0.23**
Channel 5	75.25	**0.83**	-2.63	**0.17**
Channel 6	61.13	**1.51**	-2.20	**0.31**

Table 4: Third and fourth moments of the first guess departure distributions displayed on Figures 6 and 7. The kurtosis values are normalized (0 indicates Gaussian behaviour.)

3 Data assimilation experiments

3.1 Experimental set up

Several experiments have been performed over three different periods to assess the impact of SAPHIR data within the IFS data assimilation system. Experiments using cy42r1 have been performed over the first period which covers six months in 2015 from January to June. Experiments using cy43r1 have been performed over the two periods to cover one winter from November 2015 to January 2016, and one covering one summer from May 2016 to August 2016.

Over the three periods, the real-time dissemination of SAPHIR data suffered from discontinuities as can be seen on the time series shown on Figure 10. The data gaps led to a use of $\sim$81% of SAPHIR data during the January to June 2015 period, a use of $\sim$65% for the winter 2015/2016 experiments, and a use of $\sim$70% for the summer 2016 experiments. From Figure 10, one can see that the SAPHIR/Megha-Tropiques observing system brings a number of observations similar to 2 to 3 AMSU-B/MHS instruments onboard sun-synchronous satellites. This number may increase if the quality of the real-time dissemination improves in the next years.

The results from nine of the experiments are summarized in the following section. Each of them corresponds to a specific assimilated observing system which is described in Table 5. The series of experiments performed in cy42r1 helps to understand and compare the impact of assimilating either SAPHIR or MHS in the tropics. The series of experiments performed in cy43r1 assesses the impact of the instrument within an observing system increasingly full of microwave humidity data. In particular, the cy43r1 experiments reflect the addition of AMSR2, GMI and MWHS-2 to the ECMWF operational observing system during 2015 and 2016 (see e.g. Geer et al., 2017).

Experiment Name	Covered Period	Res.	IFS cycle	Observing system assimilated
$CTRL_{2015}$	01/02/2015 - 06/30/2015	TL511	cy42r1	As in cy42R1 + passive monitoring of SAPHIR
$CTRLdenial_{2015}$	01/02/2015 - 06/30/2015	TL511	cy42r1	As $CTRL_{2015}$ - 4 MHS between 30 °S and 30 °N
$SAPHIR_{2015}$	01/02/2015 - 06/30/2015	TL511	cy42r1	As $CTRLdenial_{2015}$ + assimilation of chan. 1 to 5 of SAPHIR
$SAPHIR3ch_{2015}$	01/02/2015 - 06/30/2015	TL511	cy42r1	As $CTRLdenial_{2015}$ + assimilation of chan. 2,3 and 5 of SAPHIR
$SAPHIR+MHS_{2015}$	01/02/2015 - 06/30/2015	TL511	cy42r1	As $CTRL_{2015}$ + assimilation of chan. 1 to 5 of SAPHIR
$CTRL_{2015/2016}$	11/01/2015 - 01/31/2016	TCo399	cy43r1	As $CTRL_{2015}$ + AMSR2, GMI and MWHS-2
$SAPHIR+MHS_{2015/2016}$	11/01/2015 - 01/31/2016	TCo399	cy43r1	As $CTRL_{2015/2016}$ + assimilation of chan. 1 to 5 of SAPHIR
$CTRL_{2016}$	05/01/2016 - 08/31/2016	TCo399	cy43r1	As $CTRL_{2015}$ + AMSR2, GMI and MWHS-2
$SAPHIR+MHS_{2016}$	05/01/2016 - 08/31/2016	TCo399	cy43r1	As $CTRL_{2016}$ + assimilation of chan. 1 to 5 of SAPHIR

Table 5: List of IFS-4DVar experiments discussed in the present paper.

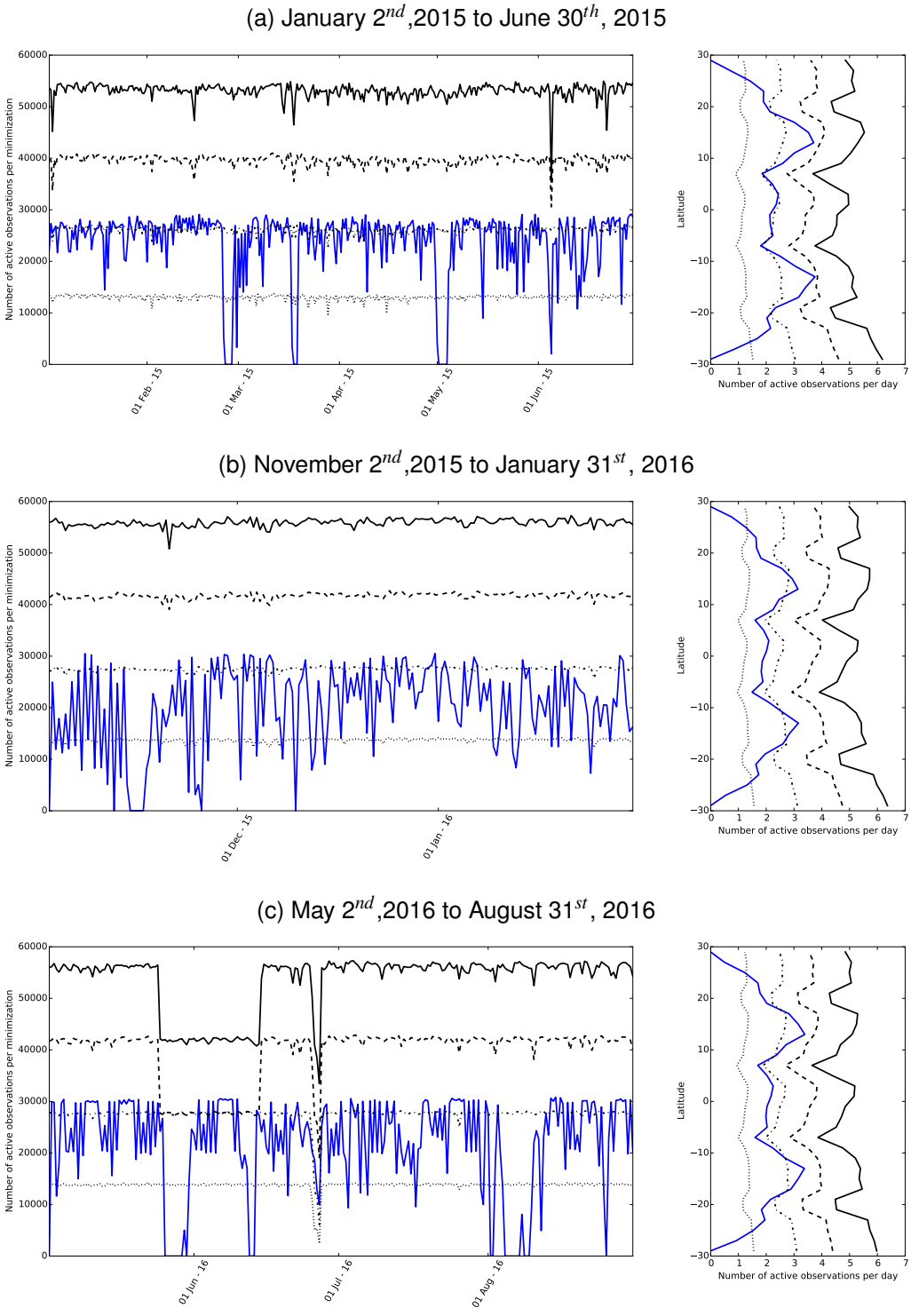


Figure 10: Timeseries of number of used observations (left column) and number of used observations per day as function of latitude (right column) over the three periods of test at 183.31±7GHz. Totals are given for SAPHIR (blue line), MHS/MetOp-A (dotted black line), and then cumulatively adding to MHS/MetOp-A the data from MHS/MetOp-B (dashed dotted black line), MHS/NOAA-18 (dashed black line) and MHS/NOAA-19 (full black line).

3.2 Rationale on Tropical forecast verification

To evaluate observing system experiments, forecast verification is required, usually based on both independent observations and NWP analyses as the reference. One advantage of using NWP analyses as reference in a verification framework is that forecasts can be compared over the full domain of investigation instead of a subsample of it where independent observations are available. This is particularly useful in the tropics where observing networks are overall less dense than in the midlatitudes. For these reasons, 'own analysis verification' is a widely used method; nonetheless the forecast errors derived from such comparisons can be contaminated by the error covariance between analyses and forecasts themselves, and the differences between them can be affected by changes in the standard deviation of the analysis. Error covariances are likely to be stronger for short term than longer term forecasts, but because error growth in the tropics is often smaller than in midlatitudes, these effects can persist for days into the forecast for some atmospheric variables and corrupt forecast impact evaluations.

One example is provided on Figures 11, 12 and 13 where one can see the impact of assimilating 4 MHS instruments in the tropics (left series of plots) and the SAPHIR instrument in the tropics (right series of plots) ; these two groups of figures respectively correspond to an own analysis verification of the CTRL_{2015} versus CTRLdenial_{2015} and of the SAPHIR_{2015} versus CTRLdenial_{2015}. Regarding the addition of MHS observations in the tropics, independent verifications shows that they improved humidity forecasts (see next section). The apparent degradations are hence an artefact of the verification method itself. This artefact in the tropics is consistently seen in the own analysis verification of observing system experiments performed to assess the impact of the all-sky instruments, as described in Geer et al. (2017, in the current study, see e.g. the relative humidity scores in Fig. 11 which is impacted to T+216h). It is likely that this effect is always present in own-analysis results from global OSEs, but it is more clearly seen in experiments that only assimilate additional data in the Tropics.

Own analysis verification being likely less reliable in the tropics than elsewhere, an alternative solution to verify the experiments performed would have been to use the ECMWF operational analysis as reference. This method would suffer less from the problem of correlated errors in analyses and forecasts. One drawback of this solution is that the ECMWF operational analysis does not include SAPHIR data yet, which would be in favor of the control experiments presented in this study. Therefore the following evaluations focus on comparisons with independent datasets like satellite and conventional observations as well as with satellite-derived precipitation products. The latter comparisons are performed using a specific metric adapted to precipitation forecast verification, known as Fractions Skill Score (FSS, Roberts and Lean, 2008; Ebert, 2008) for which the same statistical significance tests as for classical 'obstat' comparisons are appropriate (see Appendix A).

3.3 Results from denial experiments in the tropics

Comparisons have been performed between first guess departures statistics of the CTRLdenial_{2015} with the ones of the CTRL_{2015}, SAPHIR_{2015}, SAPHIR3ch_{2015} and SAPHIR+MHS_{2015} experiments for several observing systems. The experiments compare the impact of SAPHIR and MHS in the tropics, in the context of the rest of the observing system available in 2015. This already includes SSMIS (from DMSP F-17 and F-18), MWHS and ATMS microwave humidity sounding observations, although only the first of these instruments is assimilated in all-sky conditions. Figures 14, 15 and 16 show first guess departure statistics for instruments relating respectively to humidity, temperature and winds.

- Regarding humidity forecasts (see Figure 14), it is clear that assimilating either 4 MHS instru-

(a) $CTRL_{2015}$ versus $CTRLdenial_{2015}$

(b) $SAPHIR_{2015}$ versus $CTRLdenial_{2015}$

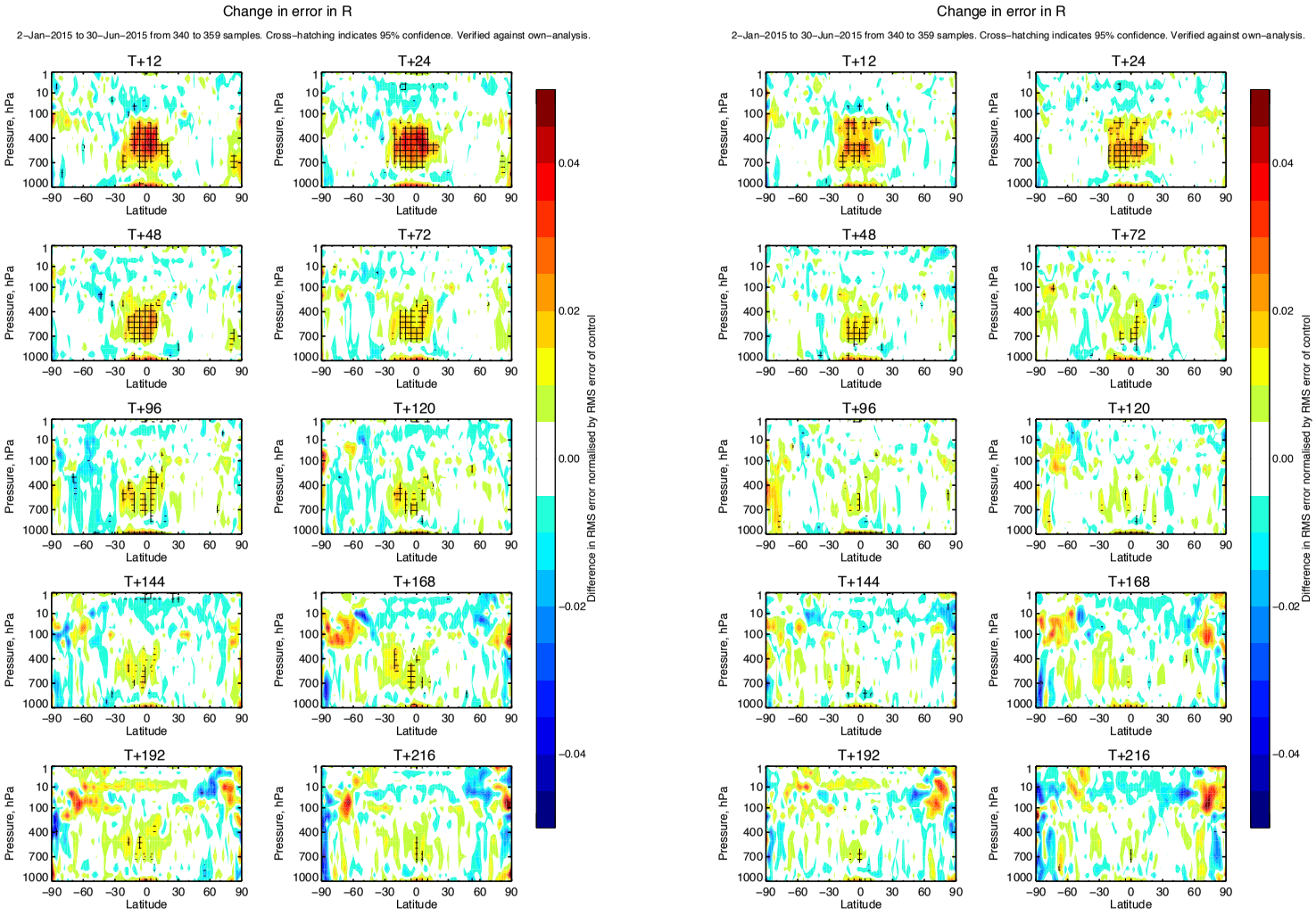


Figure 11: Zonal Differences of Root Mean Square Forecast Errors on Relative Humidity between the $CTRL_{2015}$ and the $CTRLdenial_{2015}$ experiments (left series of figures) and between the $SAPHIR_{2015}$ and the $CTRLdenial_{2015}$ experiments (right series of figures). Each Root Mean Square Forecast Error is computed using the experiment's own analysis as the reference.

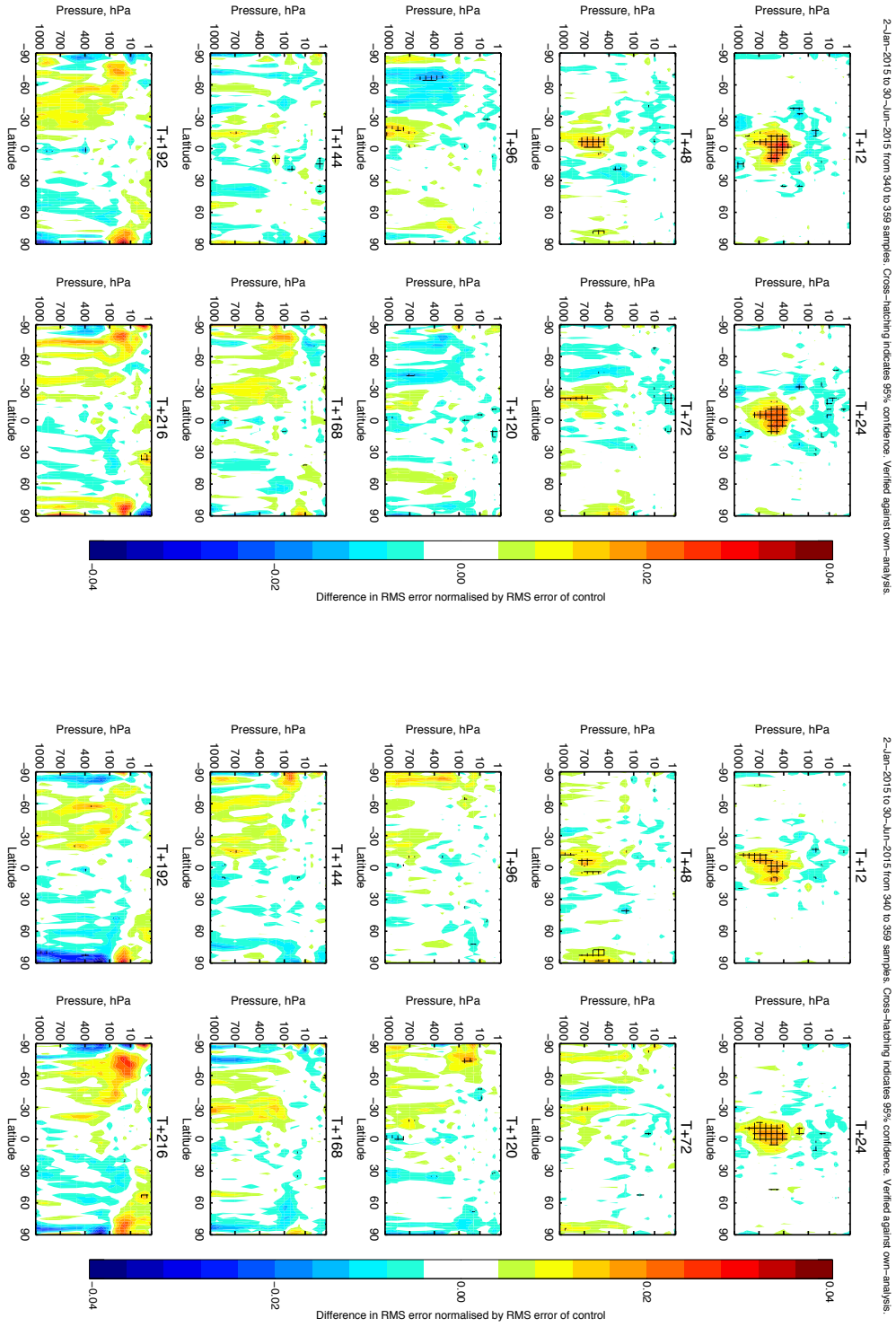


Figure 12: Zonal differences in Root Mean Square Forecast Errors, with details as for Fig. 11, but for Horizontal Winds.

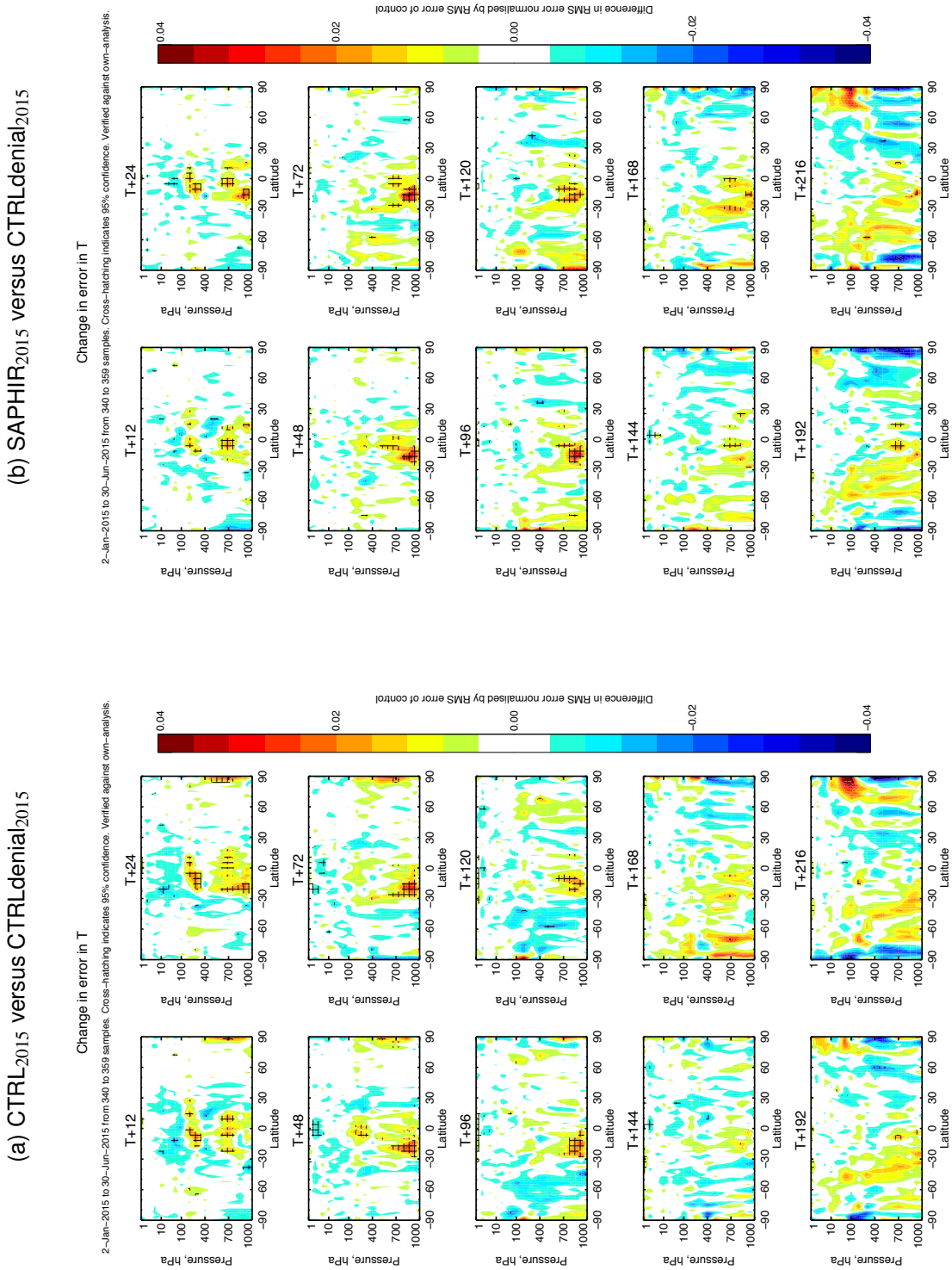


Figure 13: Zonal differences in Root Mean Square Forecast Errors, with details as for Fig. 11, but for tempearture.

ments, SAPHIR, only 3 channels of SAPHIR, or 4MHS+SAPHIR, all improve humidity forecasts. This is visible on HIRS channels 6, 7, 11 and 12; ATMS channels 18 to 22; SSMIS channels 9 to 11 as well as humidity channels from the hyperspectral instruments AIRS and IASI. Assimilating the full set of SAPHIR channels (1 to 5) compared to only assimilating channels 2, 3 and 5 (MHS-like configuration) improved the first guess fit to some ATMS and SSMIS channels (compare green and red curves). SAPHIR channels 1 and 4 hence seem to bring some added value with respect to an AMSU-B/MHS instrument. From ATMS first guess departure improvements, it can be seen that assimilating 4 MHS instruments improves humidity forecasts roughly twice as much compared to the SAPHIR instrument which is consistent with the factor of two in terms of number of observations (compare black and green curves). Another interesting feature of this comparison is that assimilating SAPHIR data on the top of 4 MHS instruments still seems to bring some additional information to improve humidity forecasts (compare blue and black curves); this seems to be statistically significant for ATMS water vapour channels, several HIRS channels and SSMIS channels 9 to 11.

- Regarding temperature forecasts (see Figure 15), AMSU-A first guess departure statistics show some statistically significant improvements which are also visible on radiosonde statistics. While only assimilating three channels of the SAPHIR instrument does not lead to statistically significant improvements, except for AMSU-A channel 8, the three experiments with 4 MHS, the full SAPHIR channel set, or 4MHS+SAPHIR, seem to improve temperature fields similarly, except again for AMSU-A channel 8 where the full configuration 4MHS+SAPHIR provides the best results.

- Regarding horizontal wind forecasts (see Figure 16), SATOB first guess departure statistics indicate that SAPHIR observations are beneficial to the IFS, at the same vertical levels as the MHS observations: close to the surface and above 400hPa. This is confirmed with conventional observations where an improvement is detected between 400hPa and 150hPa. Nonetheless, it is less clear that adding SAPHIR observations on the top of 4 MHS instruments brings some added value for horizontal winds. Another feature which characterizes both the assimilation of the 4MHS or SAPHIR in the tropics is the impact on mean meridional winds which can be seen on Figure 17: a dipole of positive and negative increments propagates from the analysis time and almost up to +12h forecast range, both close to the surface and around 100hPa. The signs of these dipoles correspond to a decrease in intensity of the large scale circulation (e.g. "Hadley cell" like circulation) and may show a possible correction of a model bias by the 4D-Var system inferring large scale dynamical information from these humidity observations.

As mentioned above, IFS precipitation forecasts have also been evaluated, over the whole tropics, using the satellite rainfall product TRMM 3B42 V7 (Huffman et al., 2007) as reference. This dataset combines observations from microwave imagery acquired from low earth orbiting satellites (e.g. DMSP/SSMIS, GCOM-W1/AMSR2, GPM/GMI etc) and infrared imagery from geostationary satellites. In such products, microwave observations are used to detect precipitation and derive instantaneous rain rates which are then integrated in time to provide rain accumulations with a data fusion technique using geostationary infrared imagery. Rain accumulations are then bias-adjusted at the monthly scale with respect to rain gauge data. The TRMM 3B42 V7 product has been extensively validated in the tropics, in particular at the daily scale at which it was found to be characterized by low bias and high correlation with rain gauge data (e.g. over Western Africa: Roca et al., 2010; Gosset et al., 2013); it is therefore at this daily scale that the comparisons are performed.

Hence, for the model forecasts, 24h rain accumulations are computed, leaving out a 12h initial spin up

period (+36h forecast minus +12h forecast for the first 24h accumulation, +60h forecast minus +36h forecast for the second 24h accumulation, etc).

Fuzzy verification have then been used in order to compare the satellite products and the rainfall forecasts from the different experiments. This kind of verification metric does take into account the fact that numerical weather prediction models do not necessarily accurately forecast the position of precipitating events by introducing a tolerance in space in the comparisons with one reference (Ebert, 2008). Several tolerance distances can be used and various conclusions can be drawn as function of the results. For instance, an improved score at large tolerance distance, together with a neutral score at small tolerance distance are indicators of improved predictions in terms of rain intensity but not in terms of rain events locations.

Among existing fuzzy verification scores, the FSS (Roberts and Lean, 2008) was selected for its 'neighbourhood observation - neighbourhood forecast' matching strategy (Ebert, 2008). Figure 18 shows the FSS values for the CTRLdenial$_{2015}$ experiment computed for the first 24h accumulations over the whole tropics and for the full six months period, for eight different rain intensities and four different neighborhood sizes. The highest (and best score) corresponds to the 3mm per day threshold and a window size of 200km. The lowest scores correspond to the highest precipitation accumulations evaluated which basically mean that the highest rain accumulations estimated from space rarely match in space and time with IFS precipitation forecasts, even with a 200km tolerance.

The differences between the FSS of two different experiments are evaluated with the same methodology as first guess departure differences, using normalized mean differences from which a statistical significance testing can be applied (see Appendix A for details on the validity of statistical significance testing for FSS). Figure 19 shows the comparison of FSS from three experiments with respect to the CTRLdenial$_{2015}$ experiment. The main conclusions are:

- the improvements detected with this particular metric on Tropical daily rainfall forecasts at the +36h range from either the 4 MHS or SAPHIR are not statistically significant at the 99% confidence level.
- with the assimilation of 4 MHS and SAPHIR data all together, the improvements become large enough to pass the statistical significance test at the 99% level for daily rain accumulations ranging from 5mm to 20mm per day.
- the largest improvement detected corresponds to a 1% increase of the FSS at the 20mm per day threshold and for a 25km neighborhood distance. This is an indicator that the locations of the daily rain accumulations greater than 20mm, which were already in the CTRLdenial$_{2015}$ experiment, have been improved.
- the improvements shown at large neighborhood distances are not statistically significant which may mean the rain events which are not predicted in the CTRLdenial$_{2015}$ experiment with a 200km uncertainty are not predicted either in the other experiments.
- no improvement on Tropical daily rainfall forecasts at longer forecast range have been detected for any of the tested configurations (not shown).

To summarise, both the observation fits and the precipitation verification indicate similar behaviour of SAPHIR and MHS data assimilation in the tropics; it is also clear that the combination of the two sensors is more effective than one on its own. This is demonstrated by, for example, ATMS first guess departures, and FSS for rain intensities ranging from 5mm to 20mm per day.

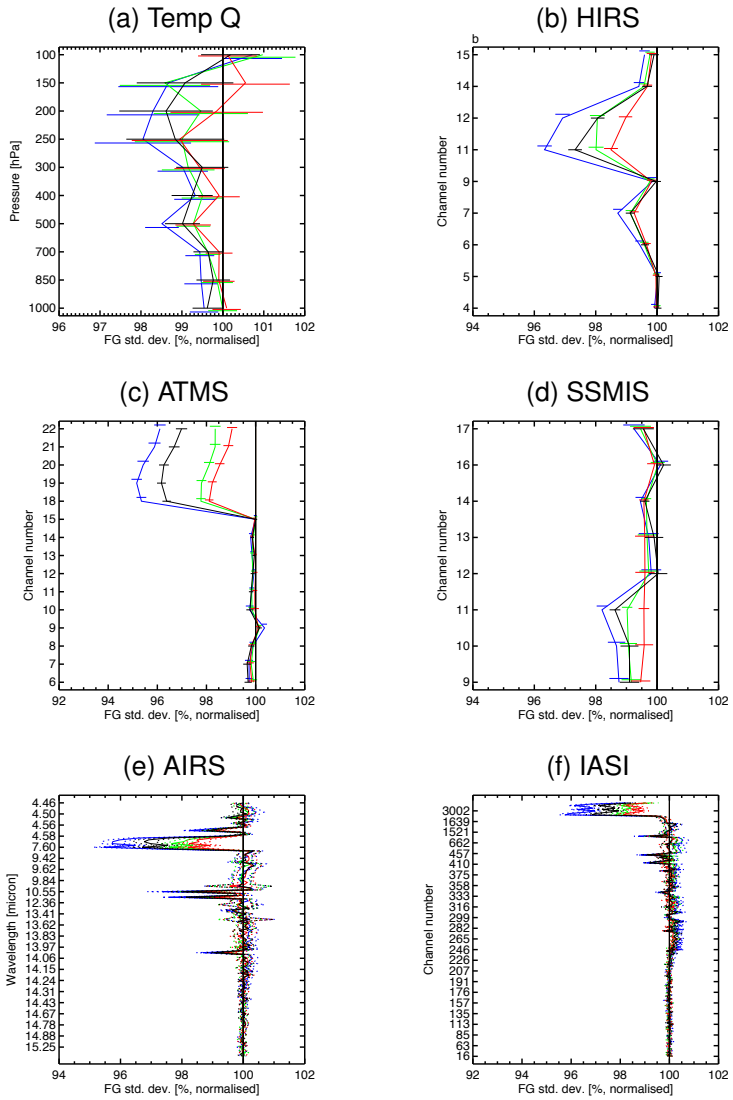


Figure 14: First guess departure standard deviations in the tropics for humidity-related observations, normalised with respect to the **CTRLdenial**$_{2015}$ experiment: **SAPHIR3ch**$_{2015}$, **SAPHIR**$_{2015}$, **CTRL**$_{2015}$, **SAPHIR+MHS**$_{2015}$**. Hence, CTRLdenial**$_{2015}$ **corresponds to 100% on the horizontal axis.**.

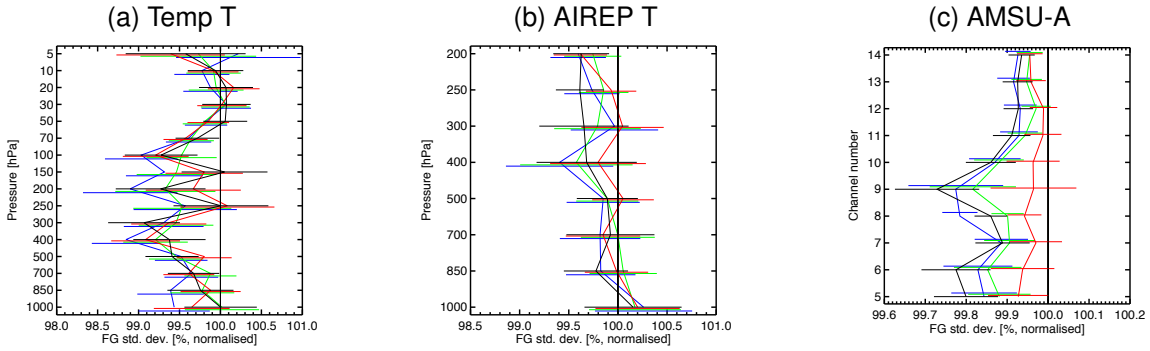


Figure 15: First guess departure standard deviations in the tropics for temperature-related observations, normalised with respect to the **CTRLdenial**$_{2015}$ experiment: **SAPHIR3ch**$_{2015}$, **SAPHIR**$_{2015}$, **CTRL**$_{2015}$, **SAPHIR+MHS**$_{2015}$.

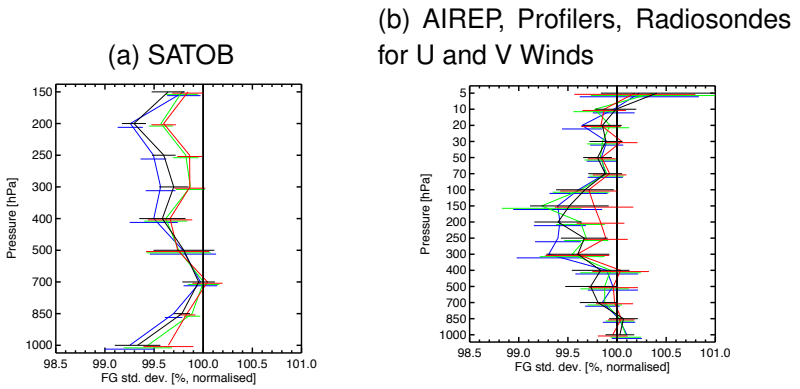


Figure 16: First guess departure standard deviations in the tropics for wind observations, normalised with respect to the **CTRLdenial**$_{2015}$ experiment: **SAPHIR3ch**$_{2015}$, **SAPHIR**$_{2015}$, **CTRL**$_{2015}$, **SAPHIR+MHS**$_{2015}$.

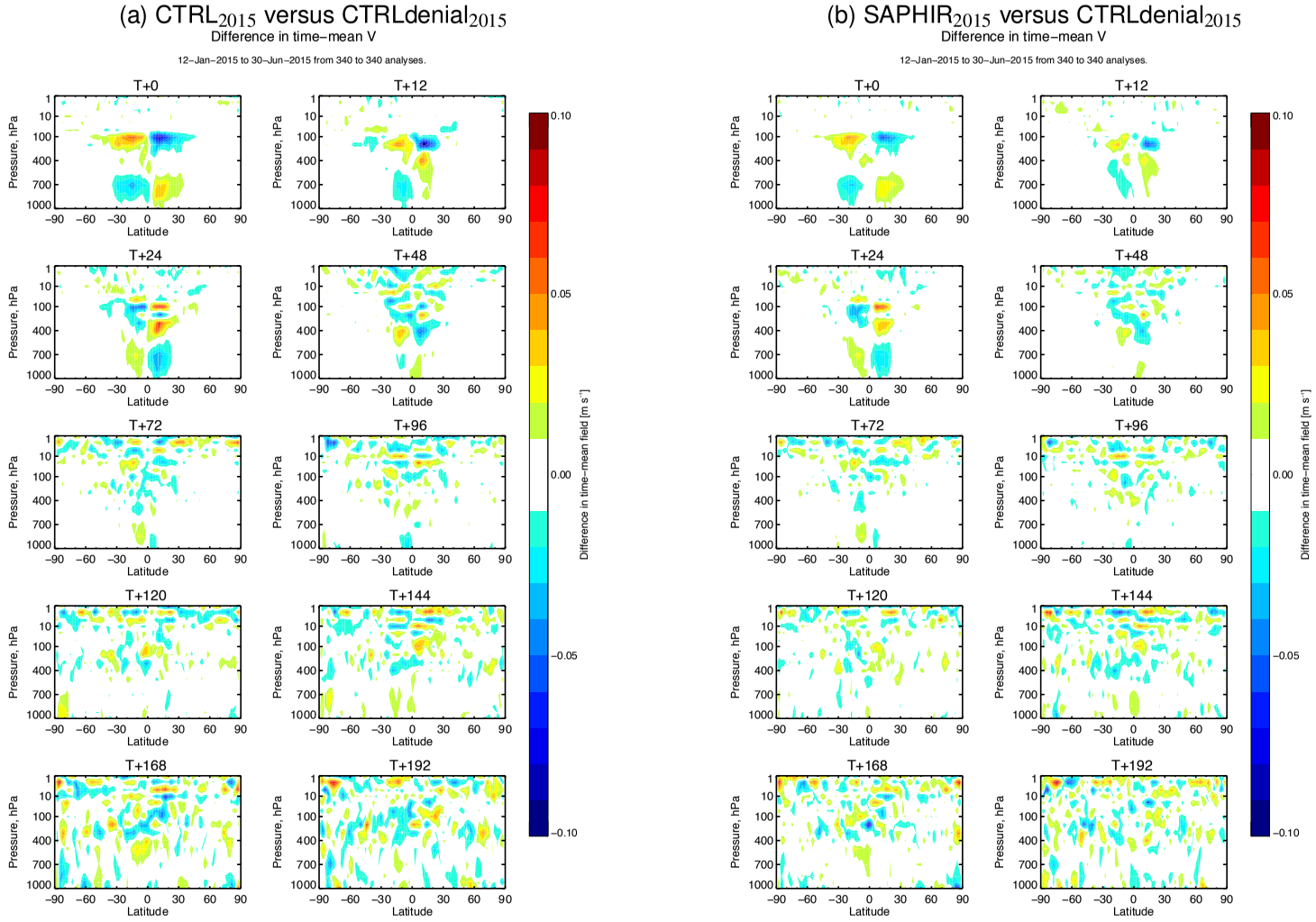


Figure 17: Zonal Differences of Mean Meridian Winds between the $CTRL_{2015}$ and the $CTRLdenial_{2015}$ experiments (left series of figures) and between the $SAPHIR_{2015}$ and the $CTRLdenial_{2015}$ experiments (right series of figures).

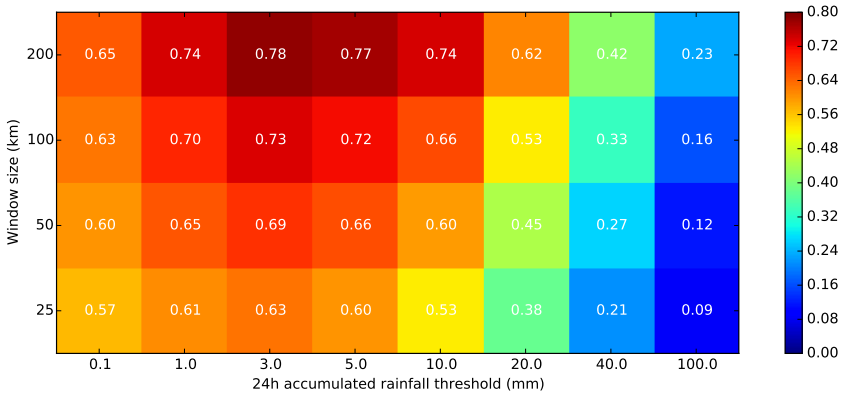


Figure 18: Fraction Skill Scores of the CTRLdenial$_{2015}$ experiment with respect to the TRMM 3B42 V7 precipitation estimates in the tropics, for rainfall threshold from 0.1mm to 100m and neighborhood sizes from 25km to 200km over the full six-month period. The calculations have been performed at the 0.25 $^\circ$ $\times$ 0.25 $^\circ$ resolution and for daily rain accumulations. The first 12h of precipitation have not been verified, to avoid model spin-up in precipitation fields in the very short range. Hence, the model precipitation fields compared to TRMM 3B42 V7 hence correspond to 24h precipitation accumulation computed as +36h forecast minus the +12h forecast.

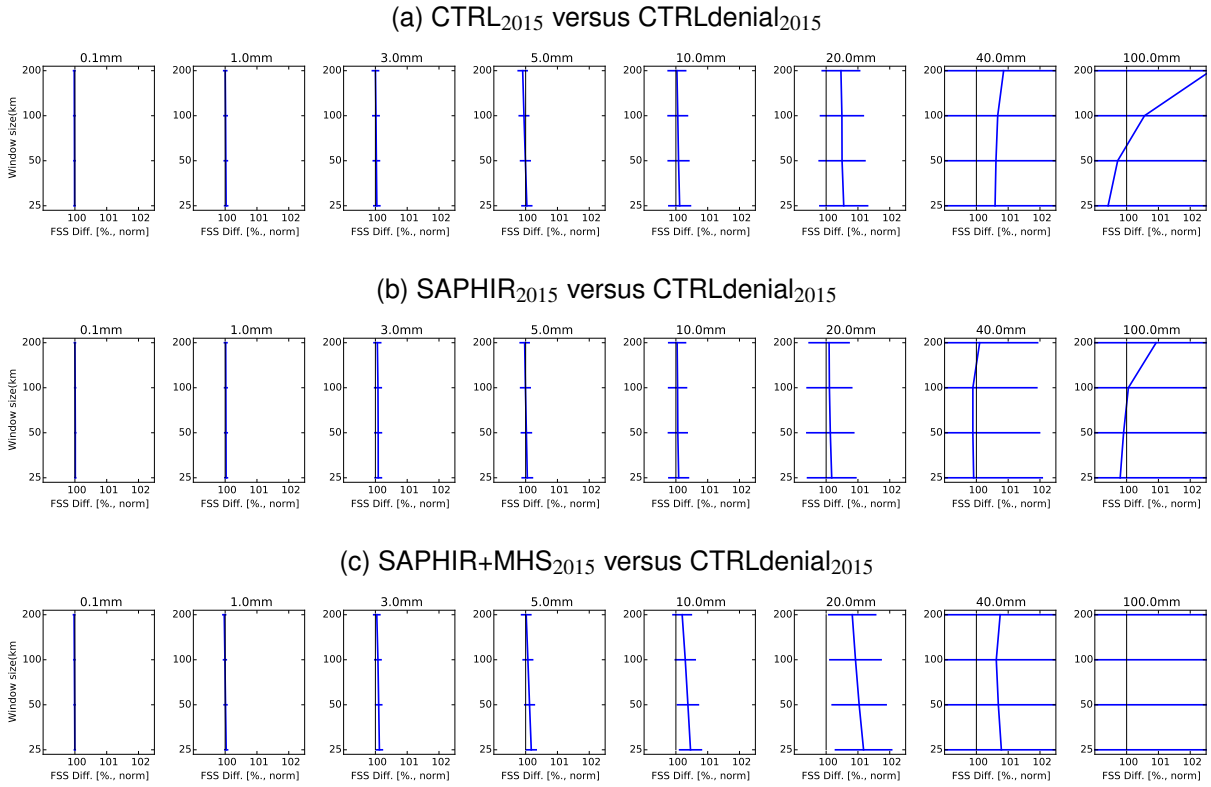


Figure 19: Normalized differences of mean Fraction Skill Scores for three comparisons of 24h precipitation forecasts with respect to the CTRLdenial$_{2015}$ experiment, over the whole tropics and for the 01/02/2015 to 06/30/2015 period. The error bars correspond to confidence intervals on the mean Fraction Skill Scores, at the 99% level.

3.4 Results with a full observing system

Because the use of all-sky observations has been evolving rapidly (Geer et al., 2017), additional comparisons have been performed in order to quantify the value of adding SAPHIR assimilation within the IFS system for more recent dates and with a newer cycle of the IFS, giving an even denser observing system. Table 5 gives details of the experimentation. The series of experiments conducted with cy43r1 of the IFS include the following additional microwave observations compared to those performed with cy42r1: AMSR2 (Kazumori et al., 2016), GMI, MWHS-2. Figures 20, 22 and 24 (resp. Figures 21, 23 and 25) show first guess departures statistics for several observing systems providing information on humidity, temperature and winds, within the context of cy42r1 (respectively cy43r1). The results presented in this section demonstrate the impact of adding SAPHIR into the otherwise full observing system.

- Regarding humidity forecasts (see Figures 20 and 21), one can see improvements on HIRS channels 11 and 12 and ATMS channels 18 to 22 by assimilating SAPHIR, of 1% in the cy42r1 experiment, and 0.75% in the cy43r1 experiments. Similar conclusions can be drawn for some water vapour channels of the AIRS and the IASI instruments. Conclusions are less clear for the SSMIS sensors for which channels 11 and 12 are characterized by improved first guess departures in the cy42r1 experiment. It is only the case for channel 11 in the cy43r1 experiments.

- Regarding temperature forecasts (see Figures 22 and 23), the main signal, indicating slightly improved temperature forecasts is in channel 8 of the AMSU-A sensors with a 0.07% improvement of first guess departures, common to cy42r1 and cy43r1 experiments. At some pressure levels, small signals are visible as well for temperature radiosoundings but their statistical significance is less clear. There are significant improvements in both experiments, but on different vertical levels in the cy42r1 and cy43r1 experiments. This is consistent with the large size of the confidence range for this data, which is a consequence of small data numbers.

- Regarding horizontal winds forecasts (see Figures 24 and 25), only a few signals are statistically significant. Some improvements and degradations can be seen in the SATOB first guess departures for cy43R1 experiments, respectively at 500hPa and 700hPa. The Aircraft reports first guess departures, combined with wind profilers and radiosoundings indicates small improvements at some levels.

Overall, one can conclude that the assimilation of SAPHIR data brings consistent and positive improvements, mainly to water vapour, over the three tested periods and the two different IFS cycles. The magnitude of these improvements is slightly reduced for the most recent periods and cy43r1 which make sense for the two following reasons: (i) less SAPHIR data were available in the winter 2015/2016 and summer 2016 experiments (resp. 70% and 65% of the full dataset) compared to the experiments conducted from January to June 2015 (81% of the full dataset), (ii) as mentioned above more microwave sensors are assimilated within the IFS cy43r1, including AMSR-2, GMI, MWHS and MWHS-2 data.

Precipitation forecasts of the experiments have been evaluated as well; with the FSS metric computed over the whole tropics and for daily accumulations, no statistically significant impact has been detected assimilating SAPHIR only on the top of the considered observing system. As an example, Figure 26 shows the normalized difference of FSS between the SAPHIR+MHS_{2015} versus $CTRL_{2015}$ experiments. This finding is consistent with the previous section where it was shown that the assimilation of 4 MHS sensors and SAPHIR together is needed to detect an improvement with this particular metric and statistical significance testing method.

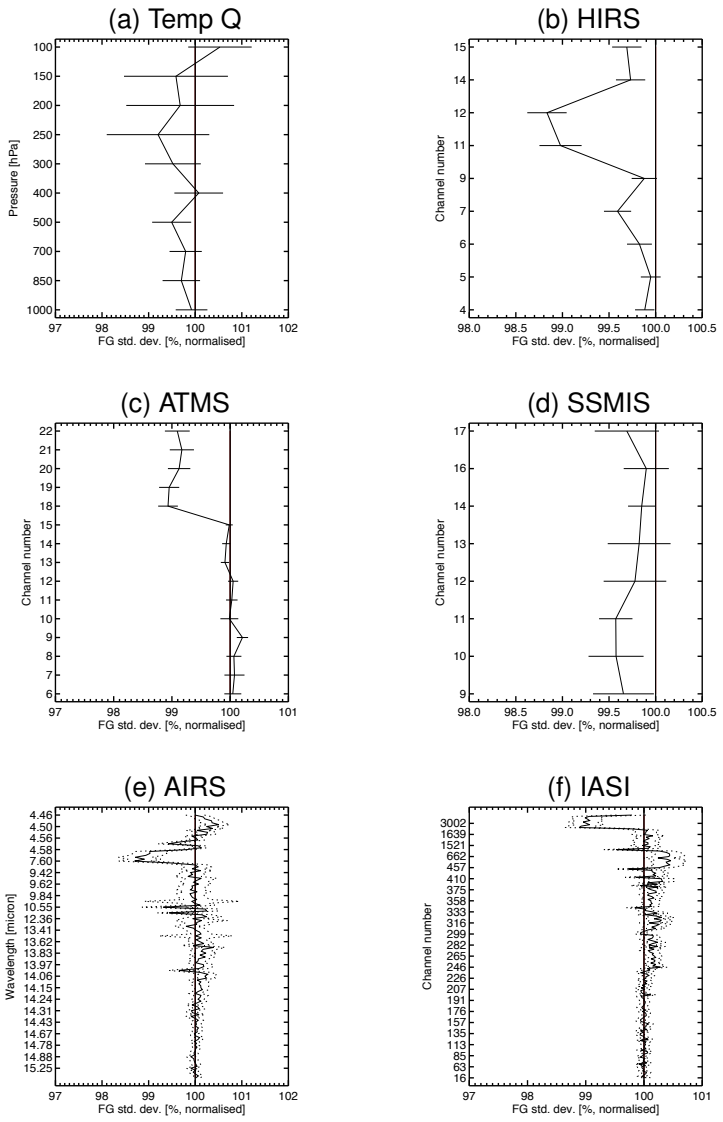


Figure 20: First guess departure standard deviations in the tropics, for mostly humidity-related observations, in the SAPHIR+MHS$_{2015}$ and the CTRL$_{2015}$ experiments, normalised by CTRL$_{2015}$. These figures measure the impact of adding SAPHIR in the cy42r1 configuration with all other observations active.

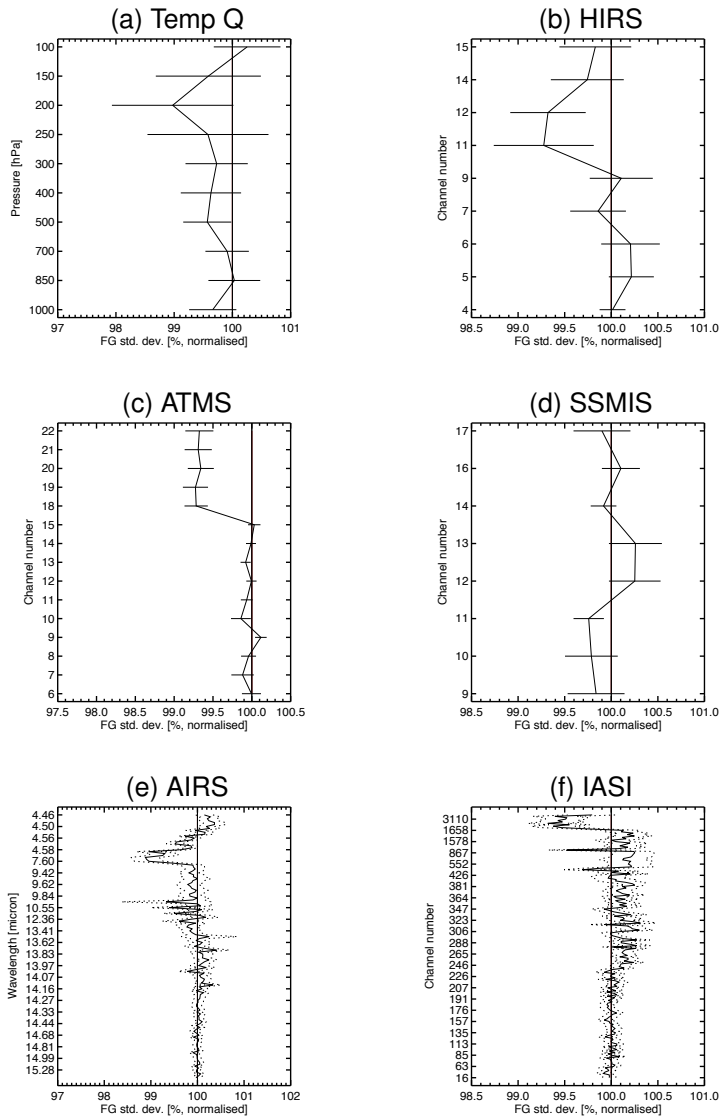


Figure 21: First guess departure standard deviations in the tropics, for mostly humidity-related observations, for the SAPHIR+MHS$_{2015/2016}$ and the CTRL$_{2015/2016}$ experiments, together with the ones between the SAPHIR+MHS$_{2016}$ and the CTRL$_{2016}$ experiments, normalised by the combined control. These figures are hence built up from a 7-month period in total, and measure the impact of adding SAPHIR in the cy43r1 configuration with all other observations active.

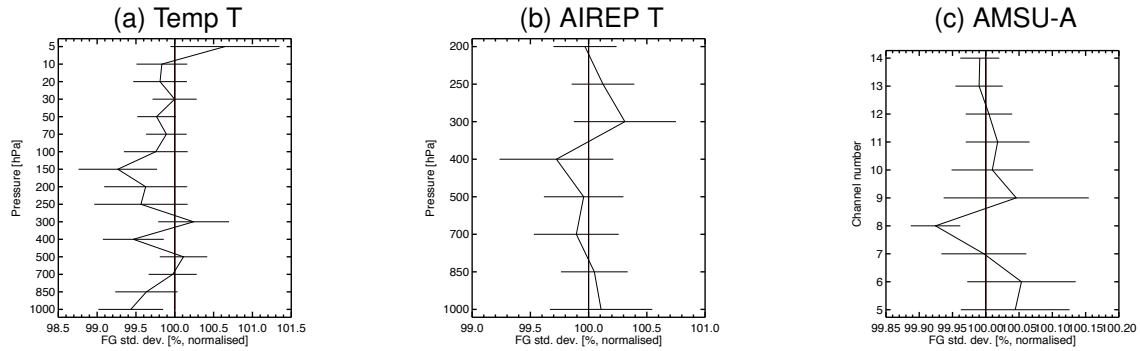


Figure 22: First guess departure standard deviations in the tropics, normalised by control, for temperature-related observations, in the cy42r1-based experiments (SAPHIR+MHS$_{2015}$ and CTRL$_{2015}$). Other details as Fig. 20.

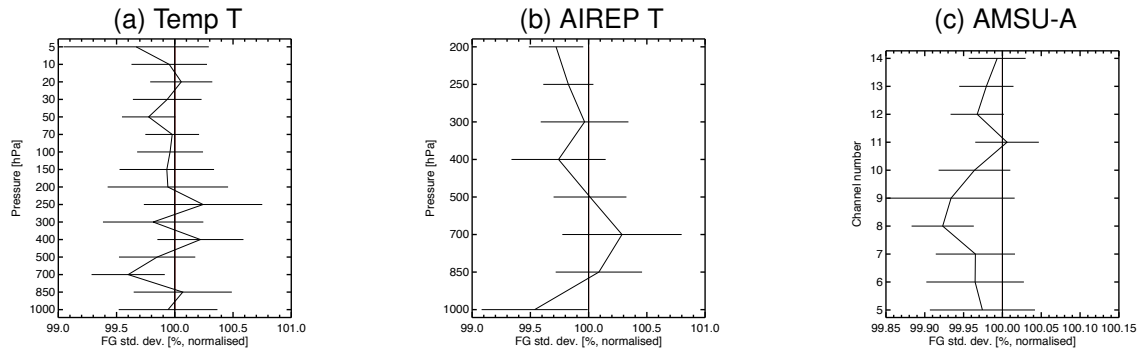


Figure 23: First guess departure standard deviations in the tropics, normalised by control, for temperature-related observations, for the cy43r1-based experiments (SAPHIR+MHS$_{2015/2016}$ combined with SAPHIR+MHS$_{2016}$ and CTRL$_{2015/2016}$ combined with CTRL$_{2016}$). Other details as Fig. 21.

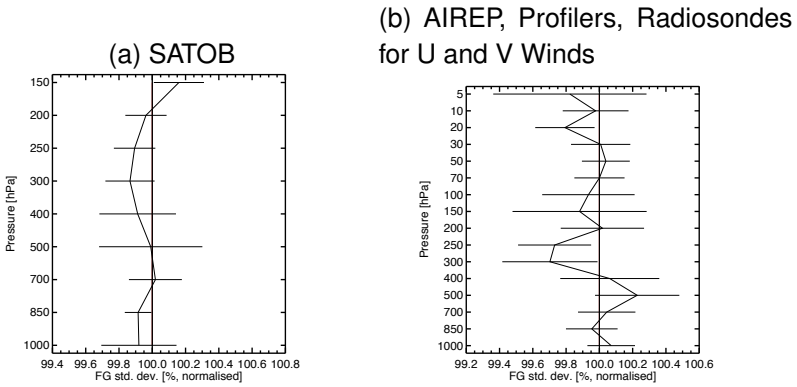


Figure 24: First guess departure standard deviations in the tropics, normalised by control, for wind-related observations, for the cy42r1-based experiments (SAPHIR+MHS$_{2015}$ and the CTRL$_{2015}$). Other details as Fig. 20.

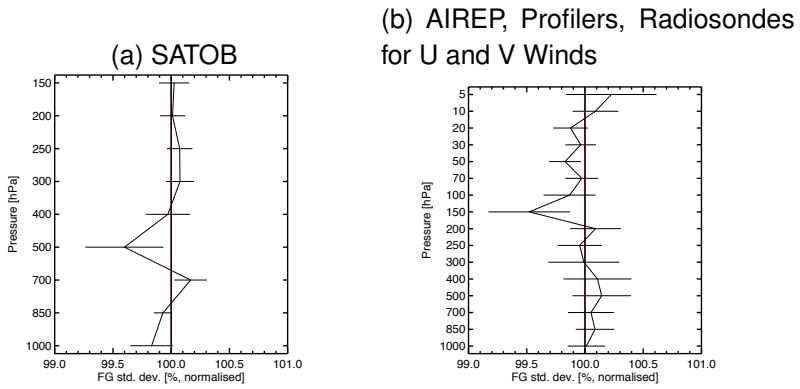


Figure 25: First guess departure standard deviations in the tropics, normalised by control, for wind-related observations, for the cy43r1-based experiments (SAPHIR+MHS$_{2015/2016}$ combined with SAPHIR+MHS$_{2016}$ and CTRL$_{2015/2016}$ combined with CTRL$_{2016}$). Other details as Fig. 21

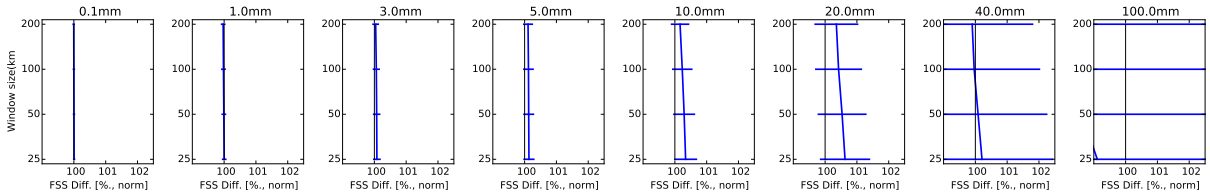


Figure 26: Same as Figure 19 but for the SAPHIR+MHS$_{2015}$ versus CTRL$_{2015}$

3.5 A non sun-synchronous orbit for diurnal cycle studies

One specific of the Megha-Tropiques' orbit is a 51-day precession cycle which means that after this period of time, all local times have been observed at least once for a given location within the Tropical band. Hence, MT is an interesting tool for making diurnal cycle studies (e.g. Chambon et al., 2015) where one can avoid using multiple instruments that are not necessarily inter-calibrated (e.g. Chung et al., 2013). As an illustration, observed and simulated SAPHIR brightness temperatures of the SAPHIR+MHS$_{2015/2016}$ experiment have been categorized by local times of observations, both for oceanic and land surface over the whole tropics (Figures 27 and 28) and for the Sahelian band (Figure 29).

Over tropical oceans (Figure 27 a), the distributions shown on Figure 27 illustrate the continuous convective activity observed through day and night. This is true even for the coldest brightness temperatures of SAPHIR channel 6, the number of which corresponds roughly to the number of deep convective scenes. This continuous convective activity is reproduced by the model, at a lower intensity though as illustrated by the negative relative difference shown on Figure 27 d) for all local times. Note that this negative bias is not necessarily due to a lack of convective activity in the model but could also be due to a lack of scattering simulated by the observation operator at 183.31 ± 11 GHz.

Over tropical land surfaces (Figure 28 a), the population of cold brightness temperatures of SAPHIR channel 6 is characterized by a minimum around noon and multiple maxima overnight. This is a well known feature which is related to the life cycle of convective systems in the tropics and has been characterized in the literature in many ways (e.g. with ground rainfall observing networks or satellite-based rainfall products Sane et al., 2012; Roca et al., 2010).

Both numerical weather prediction and climate prediction models suffer from deficiencies in the physical

parametrizations which aim at simulating deep convection; among others, one of the well documented issue is a too-early diurnal peak, so that modelled convective activity is maximum around local noon (Kidd et al., 2013). While some recent parametrizations clearly improved the representation of diurnal processes in the models (e.g. Grandpeix and Lafore, 2010; Bechtold et al., 2014; Lopez, 2014; Piriou and Guérémy, 2017), significant differences remain between observations and simulations. This can be seen on Figure 28 (b) and (d) where there are an excess of first guess SAPHIR brightness temperatures colder than 260K: this excess lies between 10am to 5pm and ranges up to 400% with respect to the observations. On Figures 28 (c) and (e) corresponding to simulated brightness temperatures from the ECMWF analyses, one sees that the excess of cold brightness temperatures around noon have been reduced by the 4D-Var system, and this decrease in convective activity seems to persist, but only as far as roughly midnight. The features seen over the whole tropics for continental surfaces can also be observed locally over the Sahelian band (Figure 29). Further understanding of the differences between observed, first guess and analysis diurnal cycles would require a dedicated study with several seasons of data and is out of the scope of the present paper. Nonetheless, it is an illustration of what Megha-Tropiques and SAPHIR data could be used for, after a few years of assimilation within the ECMWF system.

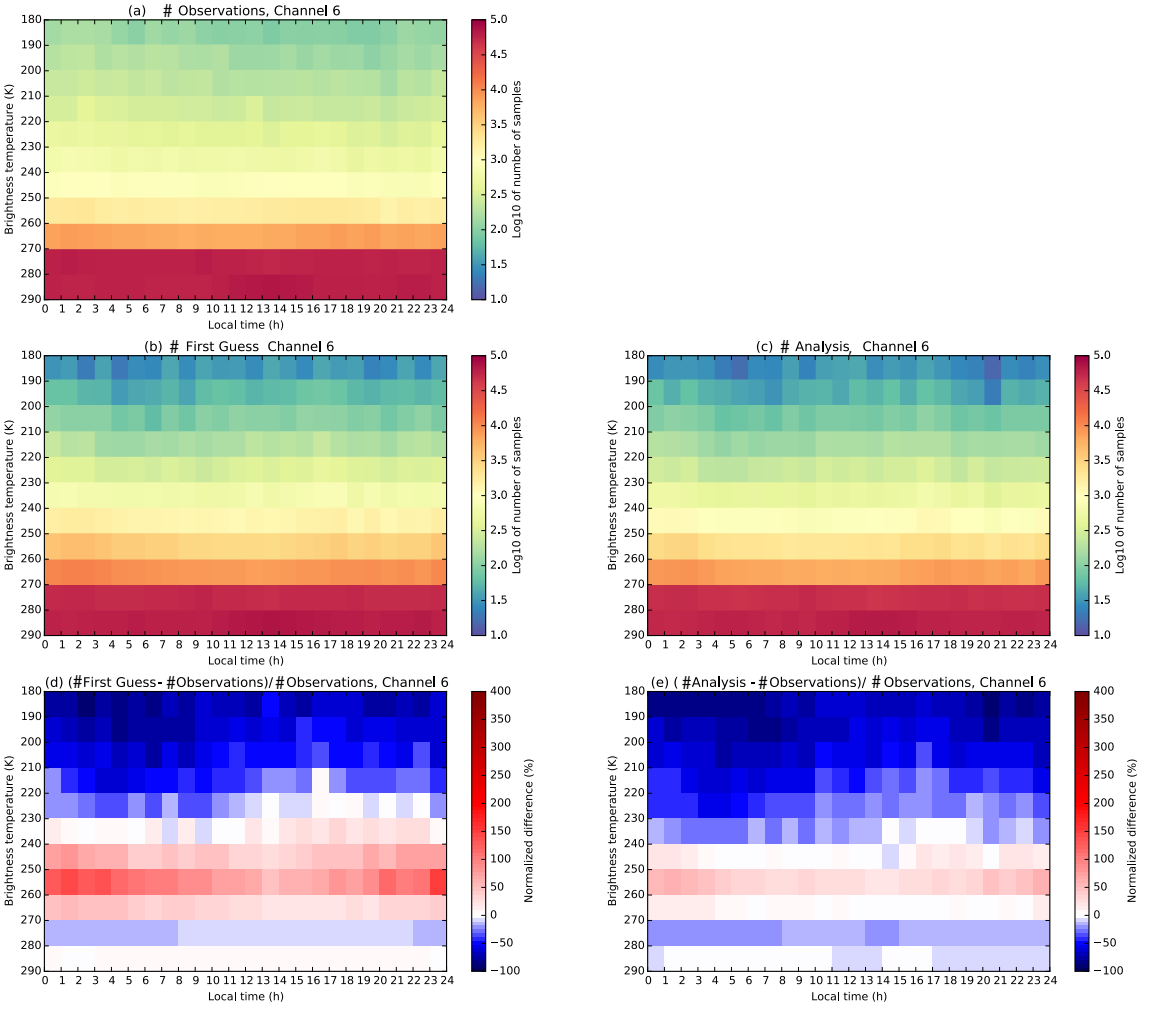


Figure 27: Two dimensional histograms of observed and simulated brightness temperatures over Oceans for the SAPHIR+MHS$_{2015/2016}$ experiment. The brightness temperatures are categorized by local time bins and brightness temperature bins over the whole tropics. The colours indicate the number of observed or simulated brightness temperatures in log scale for figures (a), (b) and (c). They indicate the relative difference of counts in percentage for figures (d) and (e).

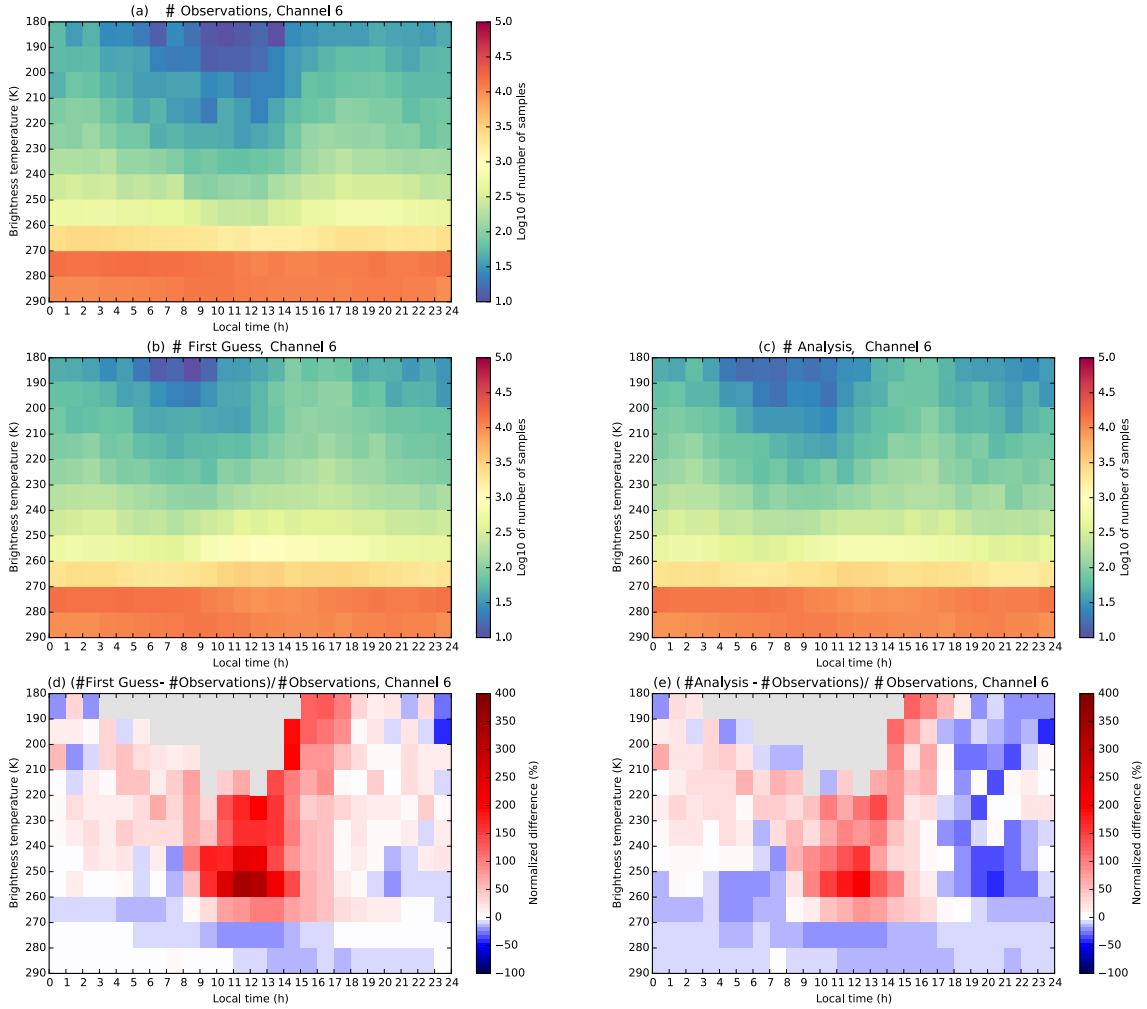


Figure 28: As Fig. 27), showing two dimensional histograms of observed and simulated brightness temperatures, but for land surfaces for the SAPHIR+MHS$_{2015/2016}$ experiment.

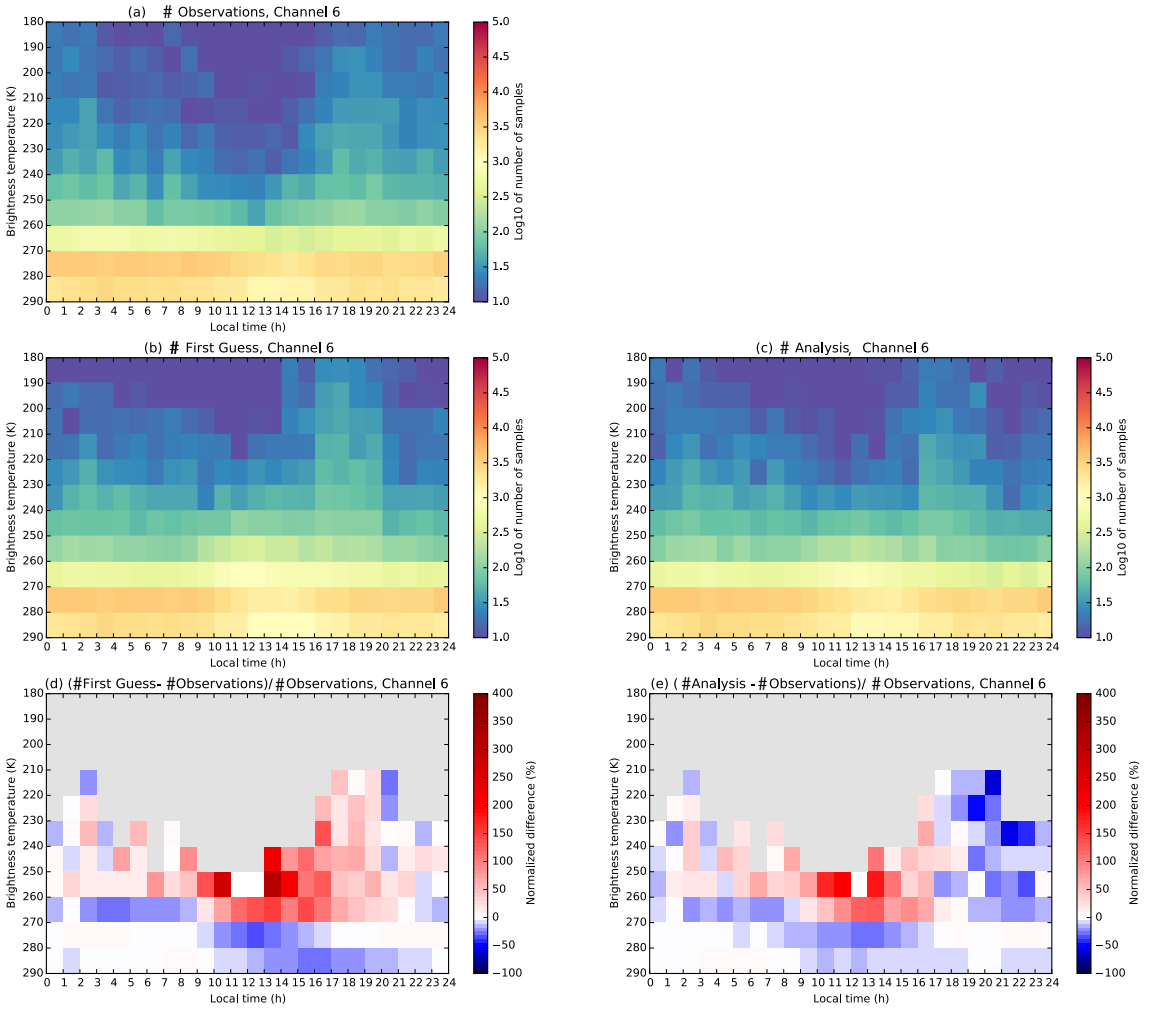


Figure 29: As Fig. 27), showing two dimensional histograms of observed and simulated brightness temperatures for the SAPHIR+MHS$_{2015/2016}$ experiment, but over the Sahelian band (20W40E10N20N).

4 Conclusion

The SAPHIR/Megha-Tropiques observing system has been in orbit since October 2011 and since then has provided well calibrated observations at 183.31GHz over the Tropical belt. These observations are now assimilated operationally in a number of NWP centres in clear sky environments. The present study assesses for the first time the impact of SAPHIR observations within an all-sky framework, close to the operational configuration of the ECMWF system. The main findings of this study are as follows:

- A cloud predictor based on SAPHIR 183.31 $\pm$ 11 GHz channel can be defined and it behaves similarly to the scattering-index based cloud predictor defined for the MHS sounder over oceans.
- Using experiments where the four available MHS instruments from different satellites were removed within the tropical belt (30 °N to 30 °S) it was shown that the assimilation of SAPHIR data provides similar results to those achieved by assimilating the 4 MHS instruments in the tropics.
- SAPHIR data assimilation led to improved humidity forecasts, even on the top of the already very dense observing system used for recent dates at ECMWF with the latest operational cycle (cy43r1 experiments). This observing system already includes mid and upper-tropspheric humidity data from a number of infrared sensors, plus 9 other microwave sensors (4 MHS, 2 SSMIS, MWHS, MWHS-2, ATMS). This shows that the beneficial impact of adding new all-sky microwave humidity sounding information has not yet saturated, although it is now limited to improvements of just several tenths of a percent in the first guess departures of some sensors.
- Neither SAPHIR nor the 4 MHS sensors provide statistically significant impacts on daily rainfall forecasts when evaluated on their own, but the two sensors together act in synergy to significantly improve IFS daily rainfall forecasts, up to a +36h forecast range.
- The assimilation of SAPHIR channels 1 and 4 led to statistically significant positive impacts on the top of SAPHIR channels common with MHS. The assimilation of SAPHIR channel 6 was tested in the IFS framework (not shown) and while improving first guess departures of some humidity observations like MHS ones, degradations were noted as well on some temperature observations which may be due to a too strong constrain on clouds and precipitation. SAPHIR channel 6 was assimilated successfully in clear sky in both the Meteo France and the Met Office global models, therefore this channel may have some potential as well within the IFS.
- Taking benefit from the Megha-Tropiques orbit, SAPHIR is a useful tool for identifying diurnal biases and assessing the progress of improvements in the IFS representation of the convective diurnal cycle. A first investigation suggests that short term forecasts are characterized by an excess of convective activity during daytime over continental surfaces.

The near-real-time dissemination of SAPHIR data suffers from discontinuities which have probably attenuated its impact on the IFS data assimilation system. Hence, increasing the number of observations available for NWP users could benefit its impact in data assimilation. Nonetheless, the positive impacts discussed in this paper are already important and statistically significant, over a total of 13 months of 4D-Var assimilation both in cy42r1 and cy43r1.

The contamination of own-analysis verification in the tropics (likely by error covariances between forecasts and analysis and/or changes in analysis standard deviation) prevented the assessment of the impact of SAPHIR data on ECMWF long range forecasts. Alternative verification methods, like observational verification for forecasts, were not available for the present study but would be useful for the impact assessment of future tropical sensors.

Acknowledgements

Stephen English and Jean-François Mahfouf are thanked for initiating this collaboration, and Fabrizio Baordo for preparatory work on SAPHIR assimilation in the ECMWF system. Stephen English and Niels Bormann are thanked for reviews of the manuscript.

A Significance of changes in medium-range forecast precipitation scores

Rainfall forecast verification is important for assessing the impacts of scientific developments on Numerical Weather Prediction. However, it requires quite speciallised statistical tools. As for other atmospheric variables, improvements or degradations of precipitation fields resulting from one single scientific development are typically small; this means that just as for other forecast verification, it is important to detect if observed modifications to rainfall predictions have any statistical significance. Geer (2016) explains that one of the issues in forecast verification is that modifications in forecast skill that are actually related to scientific or technical changes are mixed with modifications coming from chaotic error growth. These chaotic errors can generate surprisingly large differences in skill between any two forecast systems, even when compared over a long period.

A null population has been explicitly generated by Geer (2016) using a pair of experiments in order to verify the validity of the statistical tests often used in forecast verification. This pair of experiments are two and a half years long IFS experiments, characterized by a single technical difference. The latter change results in tiny numerical differences in the data assimilation but not in any scientific or technical change that should lead to a modification on forecast scores. One of the conclusions of Geer (2016) was that, due to temporal autocorrelaion in the skill of succesive forecasts, confidence intervals typically derived for t-tests at a given percentage require some inflation in order to achieve the required percentage of confidence. The FSS (Roberts and Lean, 2008) used in the present study for rainfall forecasts verification has a non linear behavior by construction as it takes into account thresholds on precipitation fields and neighborhood distances. Therefore, it is interesting to revisit the conclusions of Geer (2016) for this particular metric.

To this aim, 24h precipitation accumulations have been extracted from the pair of IFS "chaos" experiments computed by Geer (2016) over the two and a half year period. Then these daily rainfall forecasts have been compared to the TRMM 3B42 V7 satellite-based rainfall dataset (Huffman et al., 2007), by computing daily FSS for the same rain thresholds and same neighborhood sizes used in the present study. The mean FSS scores of the control experiment of Geer (2016) are shown on Figure 30 and the differences between the two chaos experiments are shown on Figure 31.

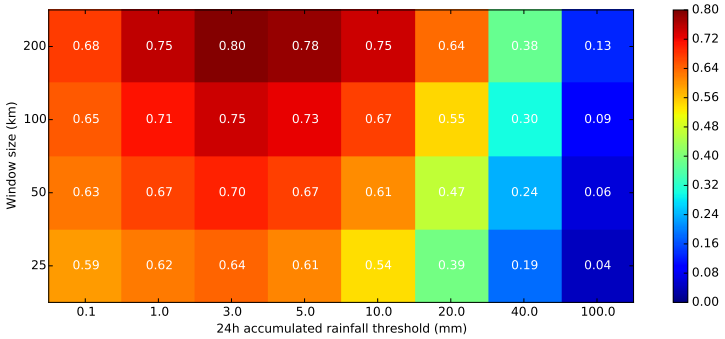


Figure 30: Fraction Skill Scores of the control experiment of Geer (2016) with respect to the TRMM 3B42 V7 precipitation estimates in the tropics, for rainfall thresholds from 0.1mm to 100m and neighborhood sizes from 25km to 200km over the full two and a half year period. The calculations have been performed at the $0.25\,^{\circ} \times 0.25\,^{\circ}$ resolution and for daily rain accumulations.

One can see that the differences between the scores are very tiny and none pass the t-test at 99%, except for the 100mm rain intensity threshold for which one may almost conclude that extreme rainfall events are better predicted in one of the chaos experiments than in the other one.

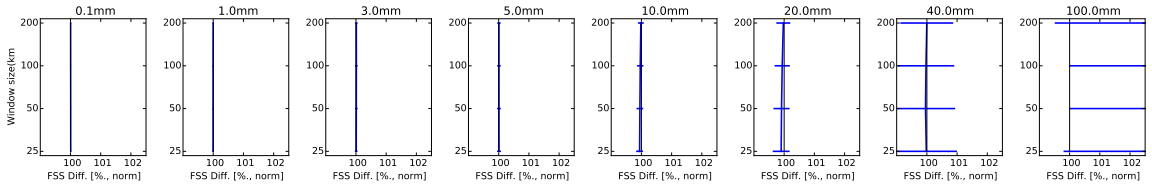


Figure 31: Normalized differences of mean Fraction Skill Scores between the two chaos experiments of Geer (2016), over the whole tropics and for the 01/01/2011 to 06/30/2012 period. The error bars correspond to confidence intervals on the mean Fraction Skill Scores, at the 99% level. Note that the mean FSS differences are larger than 2.5% for the 100mm rain intensity which lead to not see the full lines and error bars for this particular rain threshold with the chosen fixed horizontal range.

Following the same methodology as in Geer (2016), the result of performing t-tests on 10-day blocks of samples from the FSS differences between the two "chaos" experiments are shown on Figures 32 and Figures 33, respectively for neighborhood sizes of 200km and 25km.

A first interesting feature one can note on both Figures is that the cumulated distribution functions of z-tests from FSS differences behaves similarly to the Student's t-distributions. This would not necessarily have been the case as FSS is a quite non-linear score, nevertheless this is likely an example of the central limit theorem in practice. Looking at the details of these distributions, computing the percentage of z-tests samples smaller than a given value help to check whether confidence intervals on mean FSS differences, estimated from standard errors on the mean, would need some inflation or not. Tables 6 and 7 report these percentages for two confidence ranges at 95% and 99%, i.e. z-values respectively smaller than 1.96 and 2.576. One can conclude from these tables that for the two thresholds, fractions of populations are very close to 95% and 99%, except mainly for the smallest rain and highest rain intensities. For example, it seems that computing a confidence interval at 99% for FSS differences for 0.1mm rain thresholds and a 50km neighborhood does in fact correspond to 95% confidence range. Overall, it seems that confidence intervals would need either small inflation or deflation depending on the rain intensity and the neighborhood size. For the present study, the conservative choice was made to apply 99% confidence ranges on all FSS differences calculations, keeping in mind that in some cases, this better match with a 95% confidence range.

200 km	92.2%	95.5%	96.6%	96.6%	95.5%	93.3%	95.5%	95.5%
100 km	90.0%	96.6%	95.4%	96.6%	96.6%	94.4%	96.6%	93.3%
50 km	90.0%	96.6%	94.4%	97.7%	98.8%	93.3%	96.6%	91.1%
25 km	91.1%	95.5%	94.4%	97.7%	98.8%	95.5%	97.7%	92.2%
	0.1mm	1.0mm	3.0mm	5.0mm	10.0mm	20.0mm	40.0mm	100.0mm

Table 6: Percentage of z-tests smaller than 1.96, generated from blocks of 10 paired FSS differences between the Noise and the Control experiments. A value smaller (resp. greater) than 95% indicates that the confidence interval should be inflated (resp. deflated)

200 km	98.8%	98.8%	98.8%	98.8%	97.7%	98.8%	98.8%	96.6%
100 km	96.6%	98.8%	97.7%	98.8%	100.0%	98.8%	100.0%	96.6%
50 km	95.5%	100.0%	98.8%	98.8%	98.8%	98.8%	100.0%	97.7%
25 km	97.7%	98.8%	98.8%	100.0%	98.8%	97.7%	100.0%	97.7%
	0.1mm	1.0mm	3.0mm	5.0mm	10.0mm	20.0mm	40.0mm	100.0mm

Table 7: Percentage of z-tests smaller than 2.576, generated from blocks of 10 paired FSS differences between the Noise and the Control experiments. A value smaller (resp. greater) than 99% indicates that the confidence interval should be inflated (resp. deflated).

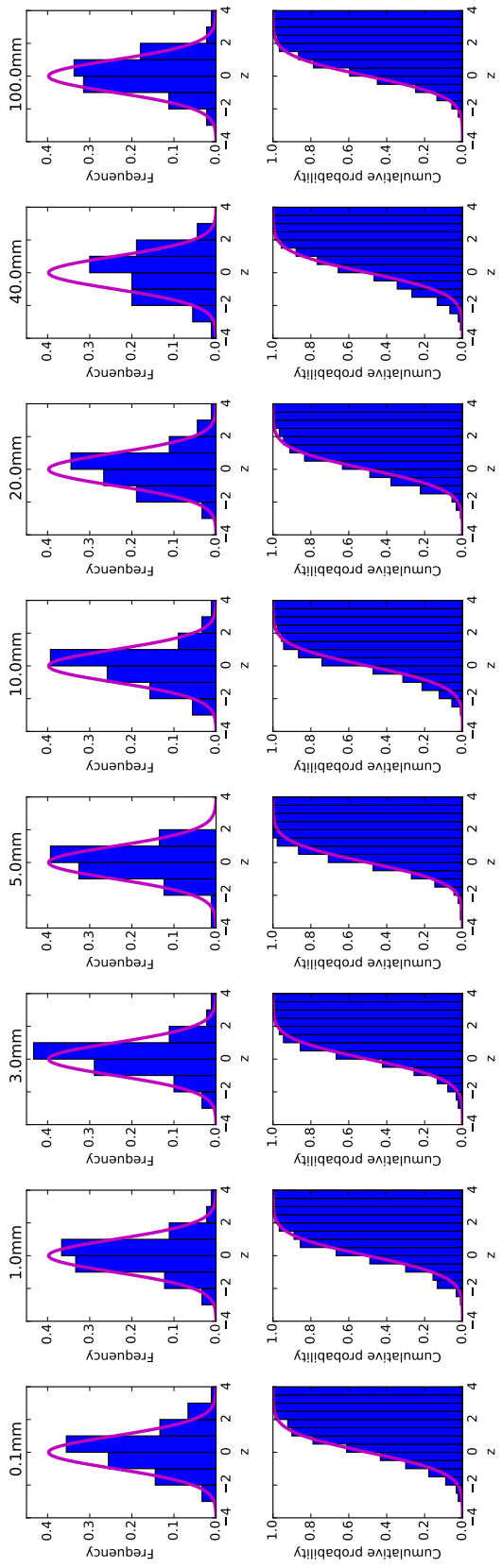


Figure 32: First row: Probability distribution functions of z-tests generated from blocks of 10 paired FSS differences (blue histograms) and the Student's t-distribution (pink line). Second Row: Cumulate distribution of the same. The used FSS scores are computed with a 200km neighborhood.

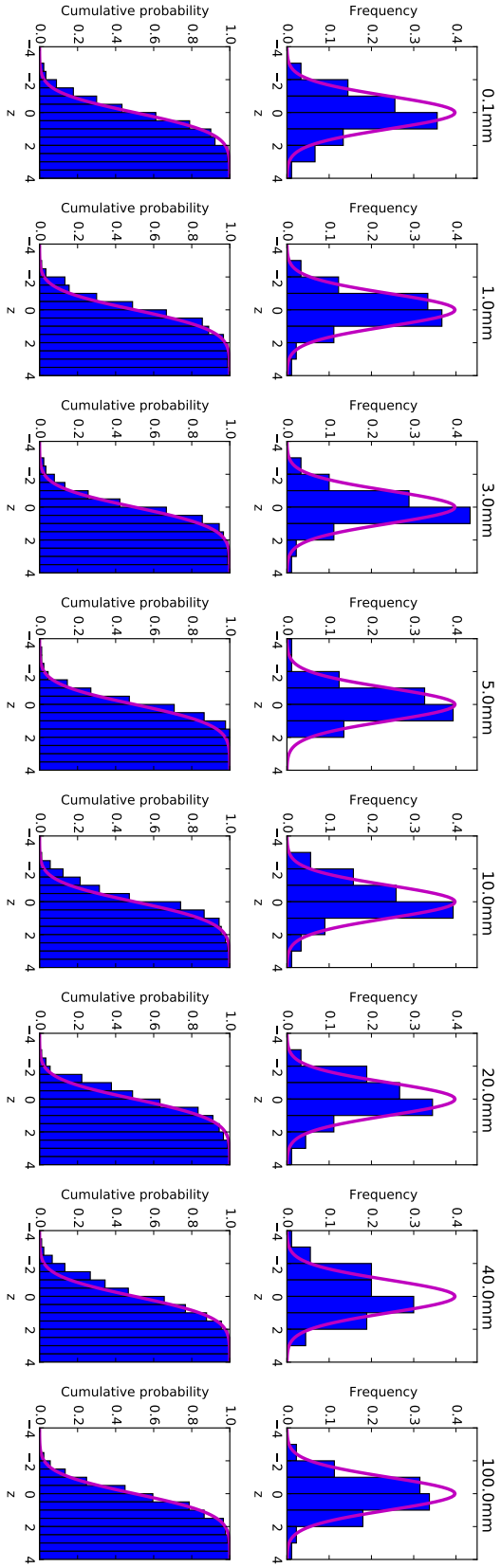


Figure 33: Same as Figure 32 except for the used FSS scores which are computed with a 25km neighborhood.

References

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

Baordo, F. and A. J. Geer (2016). Assimilation of SSMIS humidity-sounding channels in all-sky conditions over land using a dynamic emissivity retrieval. *Quart. J. Roy. Meteorol. Soc. 142*, 2854–2866.

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. 136*, 1868–1885.

Bechtold, P., N. Semane, P. Lopez, J.-P. Chaboureau, A. Beljaars, and N. Bormann (2014). Representing equilibrium and nonequilibrium convection in large-scale models. *J. Atmos. Sci. 71*(2), 734–753.

Capderou, M. (2006). *Satellites: Orbits and missions*. Springer Science & Business Media.

Capderou, M. (2009). Sampling. comparison with other meteorological satellites. *Megha-Tropiques Technical Memorandum*.

Chambon, P., L.-F. Meunier, F. Guillaume, J.-M. Piriou, R. Roca, and J.-F. Mahfouf (2015). Investigating the impact of the water-vapour sounding observations from SAPHIR on board Megha-Tropiques for the ARPEGE global model. *Quarterly Journal of the Royal Meteorological Society 141*(690), 1769–1779.

Chung, E.-S., B. J. Soden, B. J. Sohn, and J. Schmetz (2013). An assessment of the diurnal variation of upper tropospheric humidity in reanalysis data sets. *Journal of Geophysical Research: Atmospheres 118*(9), 3425–3430.

Clain, G., H. Brogniez, V. H. Payne, V. O. John, and M. Luo (2015). An assessment of SAPHIR calibration using quality tropical soundings. *Journal of Atmospheric and Oceanic Technology 32*(1), 61–78.

Ebert, E. E. (2008). Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications 15*(1), 51–64.

Geer, A. J. (2016). Significance of changes in forecast scores. *Tellus A 68*, 30229, http://dx.doi.org/10.3402/tellusa.v68.30229.

Geer, A. J. and F. Baordo (2014). Improved scattering radiative transfer for frozen hydrometeors at microwave frequencies. *Atmos. Meas. Tech. 7*, 1839–1860.

Geer, A. J., F. Baordo, K. Befort, N. Bormann, S. English, M. Kazumori, H. Lawrence, P. Lean, and C. Lupu (2017). The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quart. J. Roy. Meteorol. Soc.*, submitted.

Geer, A. J., F. Baordo, N. Bormann, and S. English (2014). All-sky assimilation of microwave humidity sounders. *ECMWF Tech. Memo., 741, available from http://www.ecmwf.int*.

Geer, A. J. and P. Bauer (2011). Observation errors in all-sky data assimilation. *Quart. J. Roy. Meteorol. Soc. 137*, 2024–2037.

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. 136*, 1886–1905.

Gosset, M., J. Viarre, G. Quantin, and M. Alcoba (2013). Evaluation of several rainfall products used for hydrological applications over West Africa using two high-resolution gauge networks. *Quarterly Journal of the Royal Meteorological Society 139*(673), 923–940.

Grandpeix, J.-Y. and J.-P. Lafore (2010). A density current parameterization coupled with Emanuel's convection scheme. Part I: The models. *Journal of the Atmospheric Sciences 67*(4), 881–897.

Hong, G., G. Heygster, J. Miao, and K. Kunzi (2005). Detection of tropical deep convective clouds from AMSU-B water vapor channels measurements. *J. Geophys. Res. 110*, D05205.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker (2007). The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology 8*(1), 38–55.

Indira Rani, S., A. Doherty, N. Atkinson, W. Bell, S. Newman, R. Renshaw, J. P. George, and E. N. Rajagopal (2016). Assimilation of SAPHIR radiance: impact on hyperspectral radiances in 4D-VAR. Volume 9880, pp. 98800M–98800M–6.

Jones, E., K. Garrett, and S.-A. Boukabara (2017). Assimilation of Megha-Tropiques SAPHIR observations in the NOAA global model. *Submitted to Monthly Weather Review*.

Karbou, F., E. Gérard, and F. Rabier (2010). Global 4DVAR assimilation and forecast experiments using AMSU observations over land. Part I: Impacts of various land surface emissivity parameterizations. *Weather and Forecasting 25*(1), 5–19.

Kazumori, M., A. J. Geer, and S. J. English (2016). Effects of all-sky assimilation of GCOM-W/AMSR2 radiances in the ECMWF numerical weather prediction system. *Quart. J. Roy. Meteorol. Soc. 142*, 721–737.

Kidd, C., E. Dawkins, and G. Huffman (2013). Comparison of precipitation derived from the ECMWF operational forecast model and satellite precipitation datasets. *Journal of Hydrometeorology 14*(5), 1463–1482.

Lopez, P. (2014). Comparison of ncep stage iv precipitation composites with ecmwf model. Technical report, ECMWF Technical Memorandum.

Piriou, J.-M. and J.-F. Guérémy (2017). A subgrid convection scheme, PCMT, for predicting dry, moist and precipitating convection in numerical weather prediction and climate modes. Part 1: equations. *Submitted to Quart. J. Roy. Meteorol. Soc.*.

Roberts, N. M. and H. W. Lean (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev. 136*, 78–96.

Roca, R., H. Brogniez, P. Chambon, O. Chomette, S. Cloché, M. E. Gosset, J.-F. Mahfouf, P. Raberanto, and N. Viltard (2015). The Megha-Tropiques mission: a review after three years in orbit. *Frontiers in Earth Science 3*, 17.

Roca, R., P. Chambon, I. Jobard, P.-E. Kirstetter, M. Gosset, and J. C. Bergès (2010). Comparing satellite and surface rainfall products over west Africa at meteorologically relevant scales during the AMMA campaign using error estimates. *Journal of Applied Meteorology and Climatology 49*(4), 715–731.

Sane, Y., M. Bonazzola, C. Rio, P. Chambon, T. Fiolleau, I. Musat, F. Hourdin, R. Roca, J.-Y. Grandpeix, and A. Diedhiou (2012). An analysis of the diurnal cycle of precipitation over Dakar using local rain-gauge data and a general circulation model. *Quarterly Journal of the Royal Meteorological Society 138*(669), 2182–2195.

Wilheit, T., H. Brogniez, S. Datta, W. Liinwood Jones, V. Payne, E. Stocker, and J. Wang (2013, September). The use of SAPHIR on Megha-Tropiques for intercalibration of polar-orbiting microwave water vapor sounders. In *2013 EUMETSAT Meteorological Satellite Conference 19th American Meteorological Society (AMS) Satellite Meteorology, Oceanography, and Climatology Conference: Understanding the past, observing the present and protecting the future, Vienna 2013*, Vienna, Austria.