ESA CONTRACT REPORT

Contract Report to the European Space Agency

Tech Note - Phase-II - WP1300: SMOS Report on Noise Filtering

Joaquín Muñoz Sabater,
Patricia de Rosnay

Technical Note - Phase-II - WP1300
ESA/ESRIN Contract
4000101703/10/NL/FF/fk

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

ECMWF

Series: ECMWF ESA Project Report Series

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

Contract Report to the European Space Agency

Tech Note - Phase-II - WP1300: SMOS Report on Noise Filtering

Authors: Joaquín Muñoz Sabater,
Patricia de Rosnay
Technical Note - Phase-II - WP1300
ESA/ESRIN Contract 4000101703/10/NL/FF/fk

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, Berkshire, UK

November 2011

	Name	Company
First version prepared by (November 2011)	J. Muñoz Sabater	ECMWF
Quality Visa	E. Källén	ECMWF
Application Authorized by		ESA/ESTEC

Distribution list:

ESA/ESRIN
Luc Govaert
Susanne Mecklenburg
ESA ESRIN Documentation Desk

SERCO
Raffaele Crapolicchio

ESA/ESTEC
Tania Casal
Matthias Drusch
Klaus Scipal

ECMWF
HR
Division & Section Heads

Contents

Abstract

Contracted by the European Space Agency (ESA), the European Centre for Medium-Range Weather Forecasts (ECMWF) is involved in global monitoring and data assimilation of the Soil Moisture and Ocean Salinity (SMOS) mission data. For the first time, a new innovative remote sensing technique based on radiometric aperture synthesis is used in SMOS to observe soil moisture over continental surfaces and ocean salinity over oceans. Monitoring SMOS data (i.e. the comparison between the observed value and the model equivalent of that observation) is therefore of special interest and a requirement prior to assimilation experiments. In this report a simple but effective approach to reduce observational noise from SMOS data is addressed. This report is the technical note, phase II, workpackage number 1300 of the ESA Request for Quotation RfQ 3-13053.

1 Introduction

SMOS provides multi-angular measurements of polarized brightness temperatures, i.e. a region on the Earth's surface is being observed under different viewing angles [3]. Depending on the location of the observed area within the Field Of View (FOV), the number of views can vary from a few units/tens up to 160. In general, the observed areas furthest from the centre of the satellite track are sampled less times than those located near the center. The geometry of the observation is complex, a complete image of the surface emission is produced by inverting the visibilities associated to the interferometric technique. Errors in the reconstructed image should be expected due to innacuracies in the antenna pattern estimation, the algorithm which reconstruct the image and the Noise Injection Radiometers (NIR) brightness temperatures measurements. This is translated into a potential degradation of the radiometric sensitivity in terms of a higher noise. Another source of noise comes from the nature of the radiometers. Any imaging radiometer, in fact, is affected by three types of noise [2]: a) the radiometric resolution (temporal standard deviation of the zero-mean random error due to the finite integration time) [6], b) the radiometric bias (spatial average of all the systematic errors) and c) the radiometric accuracy (spatial standard deviation of the sum of all the systematic errors [10]).
Although the previous sources of noise are linked to the instrumentation and measurement technique used for SMOS, another new potential large source of noise is embedded in the measurements. In SMOS, the same area of the Earth surface can be observed at different viewing geometries, and that can turn into quite different pixel shapes and sizes, specially at large incidence angles. This implies that even at slightly different viewing angles, the surface contribution to the final measurement value can arise from very different areas and land cover types. This is especially important when looking at inhomogeneous targets, such as the Earth's surface. Therefore it is expected a significant angular noise contribution due to the diverse nature of the Earth's surface.
The objective of this workpackage is to develop, test and validate a methodology to reduce the random noise in the observations and the number of observations entering the data assimilation system by sampling the multi-angular measurements.

2 Methodology

There are different possible approaches of diverse nature and complexity aiming at reducing noise from the observations. In order to make it usable in the operational Integrated Forecasting System, the methodology used in our case should be rather simple and efficient.
The natural microwave emission of the soil depends on several surface variables. For electromagnetic waves polarised horizontally, in general, the observed brightness temperatures decrease with increasing the incidence angle, whereas the opposite behaviour is observed for the vertical polarised component. The ratio of increase or decrease with the incidence angle depends critically on the soil state (soil moisture, soil temperature), type of vegetation cover and soil properties (mainly soil roughness). However, as a first approximation, the angular signature of a target at a given time can be fitted to a n-th order polynomial. In this study only polynomals of second and third order were used. The method employed here consists of fitting all the observations of the same DGG (Discrete Global Grid) node (corresponding to spatial averaged values centered on the node) and orbit to 2^{nd} and 3^{rd} order polynomials. If the polymonial regression model is a good representation of the observed brightness temperatures then the coefficient of determination (r^2) will have a high value. r^2 explains what percent of the total T_B variance is explained by the polynomial regression model and varies from 0 to 1. The rest of the variance (1-r^2) is the variability of the observations from the model. The standard deviation of the residues to the fitted curve (STD) provides an approximate indication of the noise associated to the observations. By averaging the observations in angular bins of different size it is expected to reduce the value of the residues to the curve and hence the noise associated to different viewing geometries. On top of that, the number of entries for the soil moisture analysis is reduced, contributing to a practical data thinning scheme as detailed in [8].
The simple methodology applied in this study can be used under the following considerations:

- As brightness temperatures measured at the top of the atmosphere are very sensitive to the soil state, the polynomial fit will be representative of these observations if only measurements acquired during a single orbit pass are considered. Otherwise the natural geo-physical variability of the signal could be embedded in the estimated level of noise of the observations. For example, if over a certain target it rains between two satellite passes, the observed brightness temperatures will be very different, yielding wrong conclusions of the noise affecting the observations. If several orbits are wished to be used, then areas which have demonstrated to be very stable in time (Antarctic, a dessert) may be used.

- In order to avoid mixing up with the Radio Frequency Interference (RFI) contamination on the signal, only Australia, South America and North America were analysed separately. Although these continents are not free from RFI, at least the RFI effect is less serious than over Europe and Asia.

- Averaging over relatively small angular bins should be acceptable to reduce observational noise due to surface heterogeneities, but over large bins (5 degrees or more) this method could mistakenly not only reduce the random nature of the noise affecting the observations, but also the natural variability of the signal, because T_B can quickly change with the incidence angle (specially for large incidence angles).

- The data used in this report are not re-processed data, however as single orbits are considered and not temporal trends are accounted for, the conclusions should be equivalent.

Based on the previous considerations, data acquired during two single days representing two distinguished seasons were selected: 1 December 2010 and 1 June 2011. Ascending and descending orbits were analysed separately, as well as the XX and YY polarisations. Firstly, all the observed brightness temperatures recorded over the same node of the SMOS DGG grid, and for the same polarisation and type of orbit, were fitted to a 2^{nd} and 3^{rd} polynomials. The minimum number of observations per node necessary to compute a fit was set up to 10. As the size of the angular bin increases, then the number of observations available for the fit is lower.

Data was averaged in bins of 0.5, 1, 2 and 3 degrees. Larger bin sizes were avoided because there is a risk to mix with the angular natural variability of the observations, specially for large incidence angles where the angular gradient can be large. In order to consider only significant correlations, the Pearson's r coefficient with 5% significance level was evaluated at each time. In this case, configurations where the p-value was larger than 0.05 were rejected and not accounted for in the statistics, as the correlation value is from the statistical point of view a pure coincidence [9].

2.1 Examples for single-points

To better understand how this method works, some single nodes showing different geophysical characteristics and type of soil cover the 1 December 2010 were selected and analysed. The following figures show the observations fitted to only an 2^{nd}-order polynomial, by using all the observations collected over the node and by averaging them in bins of 0.5, 1, 2 and 3 degrees, respectively. These points are located in South East of Australia, North West of South America and North of North America, as shown in Fig. 1.

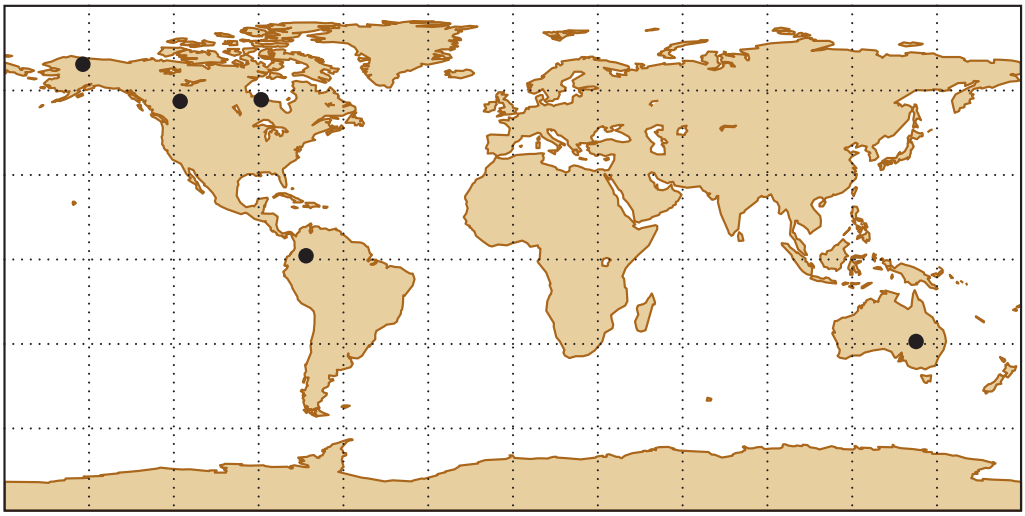

Figure 1: Location of the single points analysed in this section.

2.1.1 Australia, grid point id 8167123

This grid point is located in South Australia [lat=29.017 S, lon=143.001 E], and corresponds to a dessert area. The relation between T_B and incidence angle is very well described by a 2^{nd} order polynomial regression model (see Fig. 2). For this particular case and the particular soil conditions of this day of December 2010, the YY polarisation (right panel) has a wider dynamical range as a function of the incidence angle than the XX polarisation (left panel), but in both cases the polynomial fits well the observations. In both cases, the small incidence angles seem to be more affected by noise, which also seems to be the case for the highest incidence angles of the XX polarisation. The binning method is effective at removing large noise (see figures from top [no binning] to bottom [3-degrees binning]), and at the end of the process a reduced dataset with lower noise is remaining.

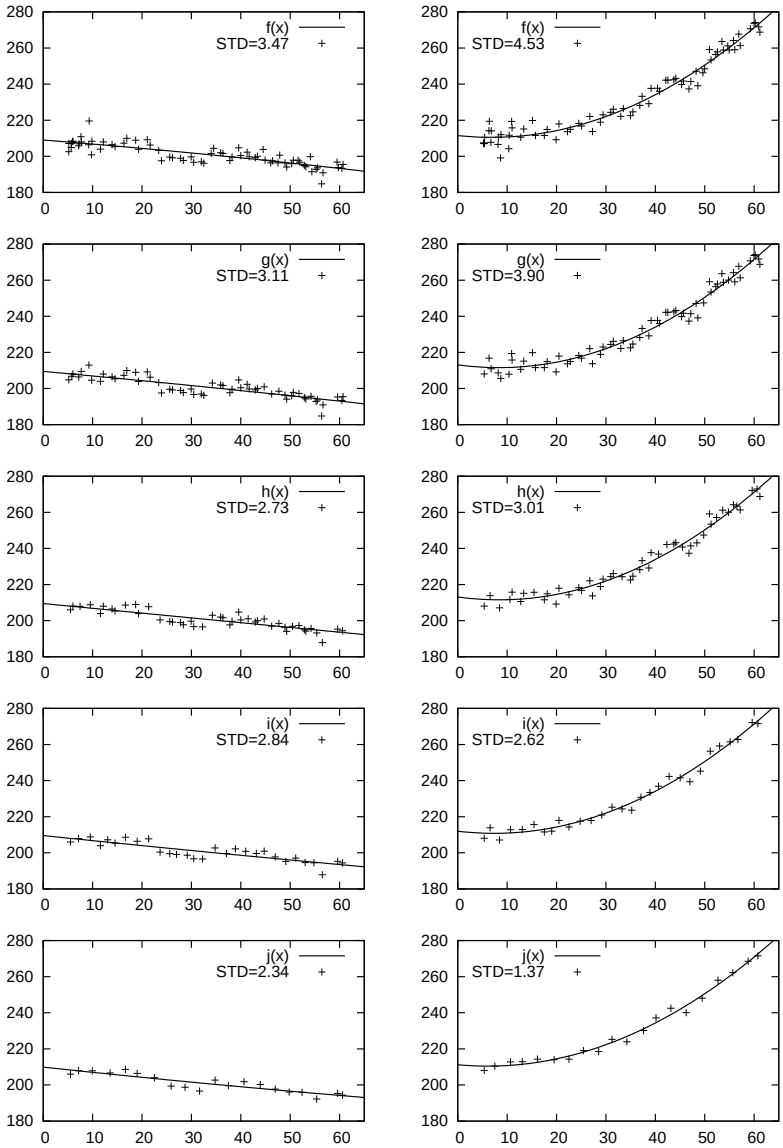

Figure 2: Australia grid id 8167123. Observations (plusses) and the 2^{nd} order polynomial fitted curve (solid line) the 1 December 2010. Left panel is for the XX polarisation and right panel for the YY polarisation. In the top figures all the observations acquired at this grid point are included, then bins of 0.5, 1, 2 and 3 degrees are applied, respectively, to average the observations. STD shows the standard deviation of the residues to the fitted curve.

2.1.2 Dense forest, grid point id 1136747

This grid point [lat=1.279 N, lon=73.565 W] analyses the angular signature of brightness temperatures in an area heavily forested, in particular this point is located in the Amazon forest of South America. As expected the signal is very flat for both polarisations and with relative low noise for the YY polarisation (see Fig. 3). In contrast, the XX polarisation suffers of significant noise from 10 to 35 degrees, which is only partially reduced about 5 K by averaging the observations in angular bins.

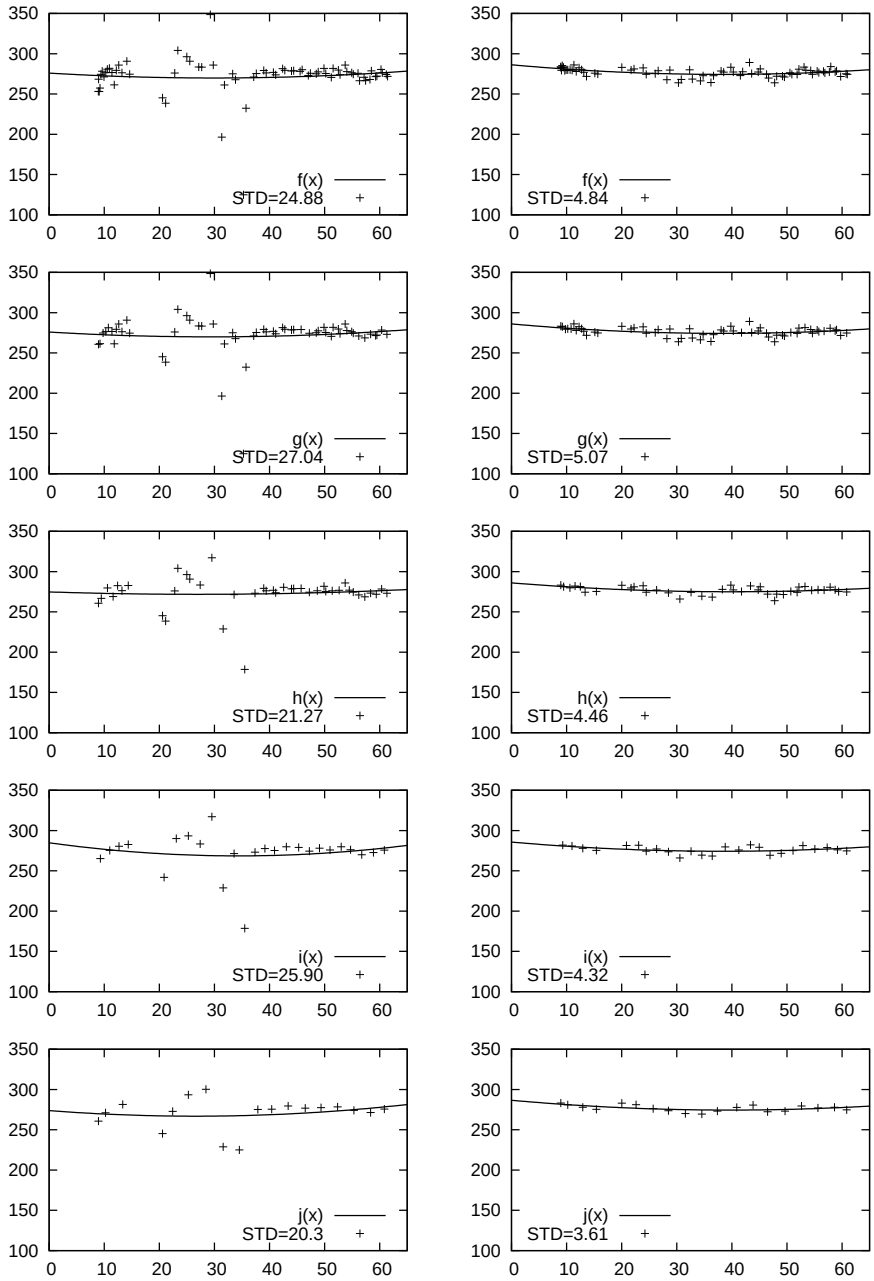

Figure 3: As in Fig. 2, but for the grid point id 1136747 (Amazon forest).

2.1.3 Shallow snow depth, grid point 111836

This grid-point is located in North America [lat=56.319 N, lon=117.879 W], with a shallow forecasted value of snow. The snow depth value forecasted for this grid-point and date was 3 cm. In this case the signal is less flat that for the Amazon case, but the dynamical range of T_B is lower than for a standard case (see Fig. 4). The signal looks very noisy specially for low incidence angles, and the binning method makes a good job by removing the strong noise of the lowest incidence angles.

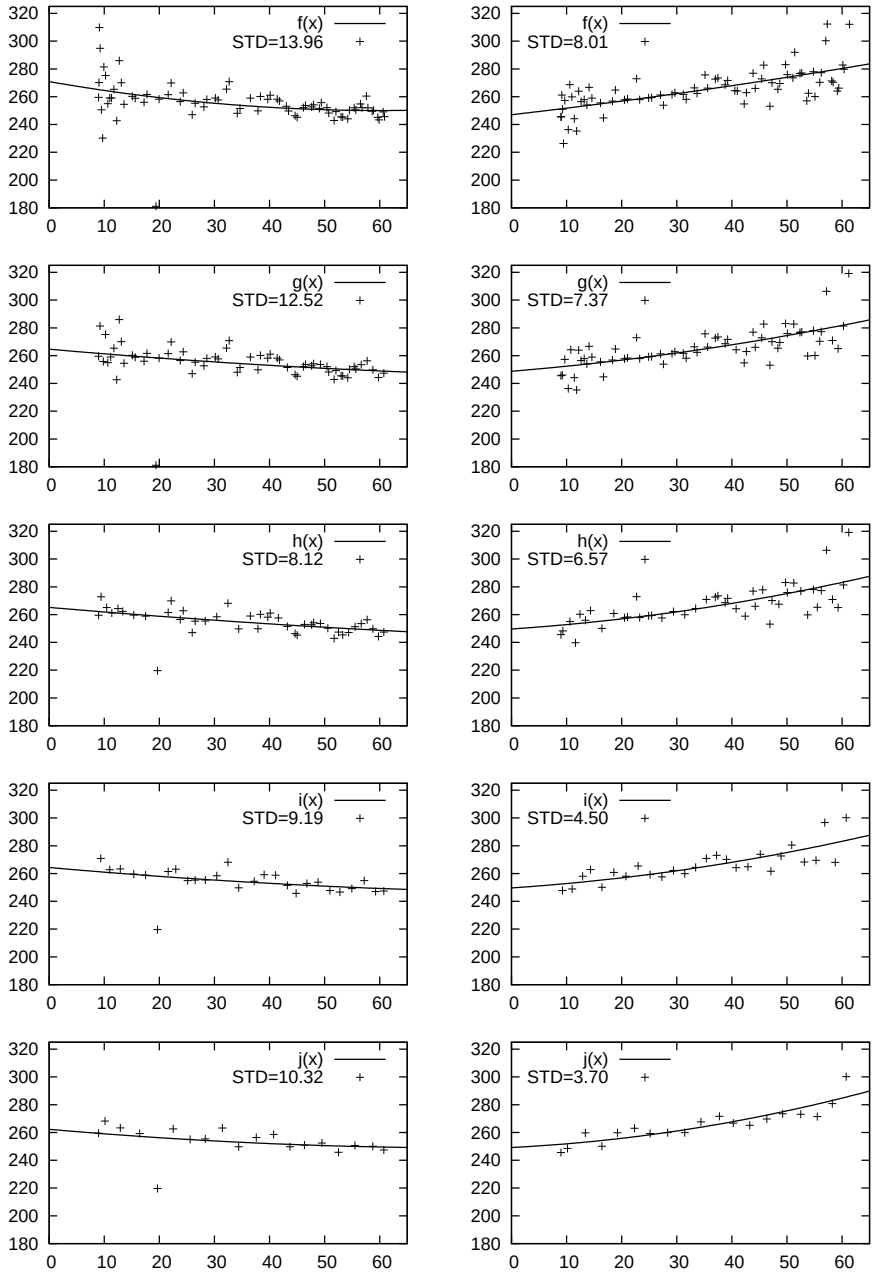

Figure 4: As in Fig. 2, but for the grid point id 111836 (North America point with shallow snow layer)

2.1.4 Large snow depth, grid point id 23165

If the snow depth is increased (in this case the forecasted snow depth value was 0.42 m), then the signal seems to be more stable with lower STD (see Fig. 5), as it happens for this location [lat=69.599 N, lon=151.914 W] the 1 December 2010. This result is in agreement with other angular signatures presented at other locations, as it is the case of the very stable time series observed over the Dome-C site in the Antarctica (see for instance [5]).

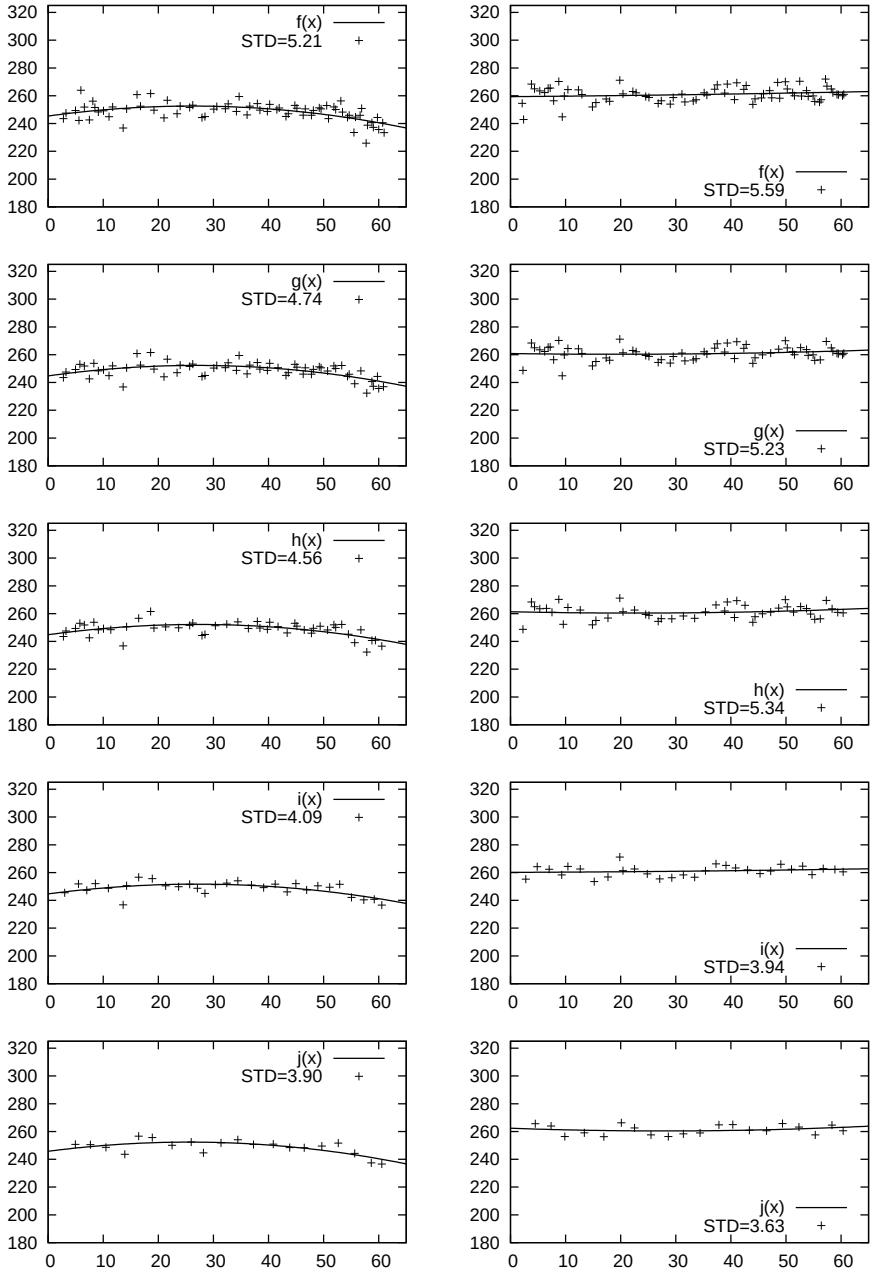

Figure 5: As in Fig. 2, but for the grid point id 23165 (North America point with 0.42 m of forecasted snow).

2.1.5 Coastal point, grid point id 140644

This grid-point [lat=56.765 N, lon=89.018 W] corresponds to a point near the coast of the Hudson Bay in Canada, with some snow (0.12 m according to the IFS first-guess). The coefficient of determination is very good for the XX polarisation and very bad for the YY polarisation, which means the polynomial fit is poor for YY. The same is found in some coastal points in the Caribbean islands. The signal for the XX polarisation is quite stable, whereas it is very noisy for the YY polarisation (see Fig. 6). Binning is effective at removing noise for the YY polarisation, by almost 5 K, whereas a slightly reduction (around 0.5 K) is achieved for the XX polarisation. A possible reason for this dual behaviour is the higher sensitivity of the YY polarisation to the amount of water content in the pixel.

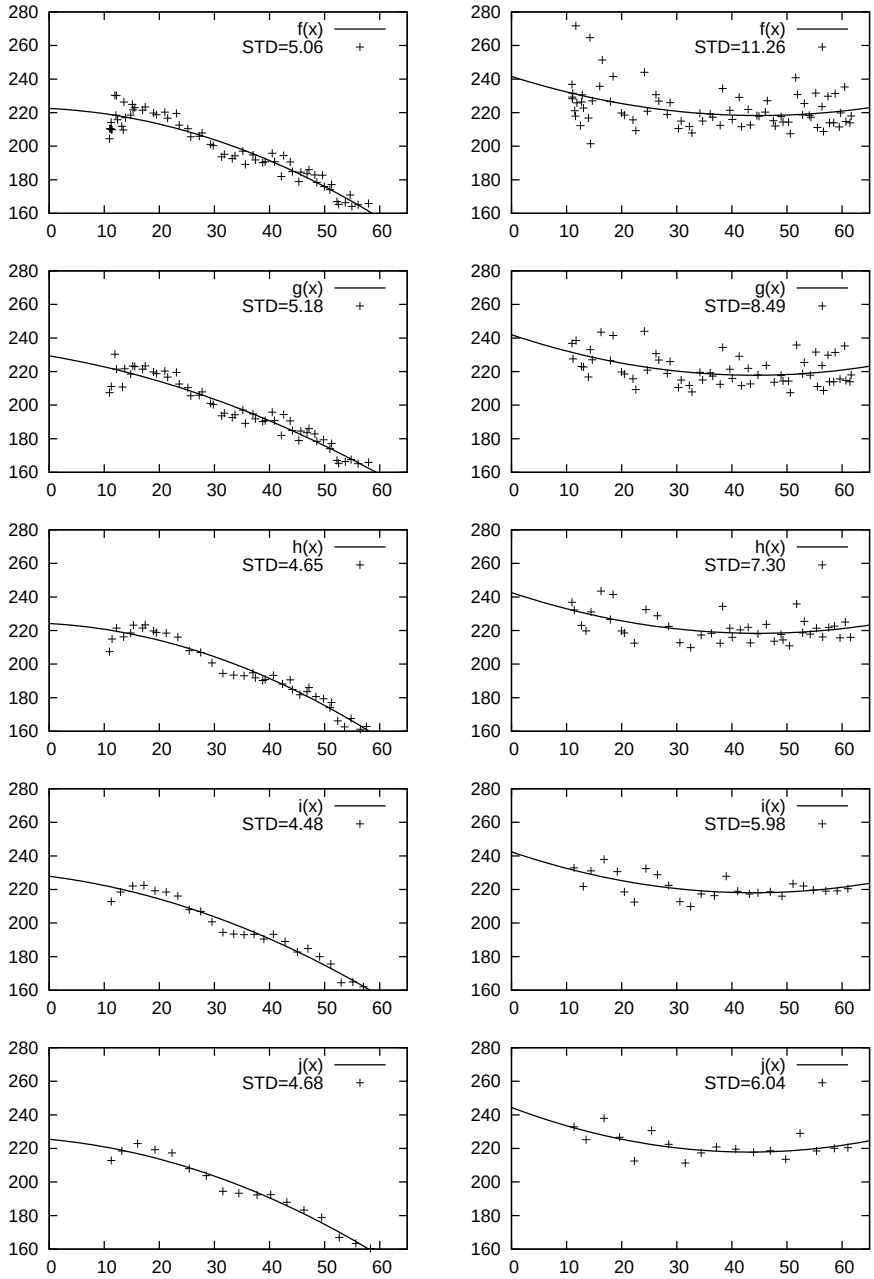

Figure 6: As in Fig. 2, but for the grid point id 140644 (North America point near the coast).

3 Maps of the binning effect

The left panel of Fig. 7 shows the r^2 maps of SMOS brightness temperatures fitted to a 2^{nd} order polynomial model, for the different angular bins used in this study, for ascending orbits over Australia, on 1 December 2010 for the XX polarisation. Each r^2 value is computed individually per node. The right panel of Fig. 7 shows the equivalent maps of the variability of the residues (STD) to the fitted curve. In section 7 the same results shown in this section are presented for the YY polarisation for Australia. Also section 7 shows the results obtained the 1 June 2011, and for North America and South America, respectively. The analysis of these maps show the following:

- As expected, the edges of the satellite track are the most noisy areas of the FOV. This can be clearly observed in the r^2 and STD figures when all the observations are used. These low performance areas correspond to the extended-alias FOV (EA-FOV), mostly containing incidence angles greater than 45 degrees. A 2^{nd}-order polynomial regression model is not a good representation of this area of the FOV, which seems to be very noisy. By increasing the size of the angular bin where the observations are averaged, the most outer parts of the edges of the satellite track move progressively from large STD to very low, in fact zero STD. This effect has no physical meaning, it is an artefact. This is because in the plots the data including grid points with only 2 observations are automatically included too, whereas in the mean statistics these points are filtered out. If only two points are available to compute an order-2 fitting, it is always possible to find a perfect curve matching exactly these two points, and hence a zero STD and $r^2 = 1$. As the size of the angular bin increases, then the number of nodes with only two points increases, thus the STD and r^2 seem to increase, when what it really happens is that there is an insufficient number of points to produce a significant fit. This happens mainly at the edges of the satellite track where less data is available.

- Many pixels near the coast, picking up contribution from sea, show very high correlation whereas the opposite behaviour is found for YY polarisation (see for instance the Australia maps for XX polarisation in Fig. 7 or the South America maps in section 7). This behaviour is common either for ascending or descending orbits. By increasing the angle of the observation larger areas contribute to the signal, and near the coast this means larger contribution of the sea. While the XX polarisation behaves as theoretically expected, the signal of the YY polarisation stays quite noisy and flat instead of growing with the incidence angle. This effect might be explained by the combination of the proximity to oceans (which compensate an increase of brightness temperatures with increasing the incidence angle) and a lower sensitivity to water bodies.

- In general it is found that there is very low correlation and larger noise in areas with snow (see North America figures in winter, sections 7.1 and 8.1) and this is stronger for YY polarisation than for XX polarisation. This is caused by a lower sensitivity of the YY polarisation to the snow water content, which produces flatter signals and therefore the presence of noise quickly reduces the coefficient of determination. The same behaviour occurs for very dense forests as the Amazons (see South America figures in sections 7 and 8), but in this case some areas present stronger noise at the XX polarisation. These areas present flat angular signatures and the presence of noise makes the coefficient of determination reduce quickly.

- In general it is found that a 2^{nd} order polynomial represents better the angular behaviour of the observations in the XX polarisation than in the YY polarisation, because the last one is slightly more noisy. A quantitative analysis can be found in section 4.

- The December and the June cases show equivalent results. Some geographical changes are mainly found in North America, which is likely due to the differences in snow covered areas. A 2^{nd} order polynomial

regression model should not be used as a reference model when snow or dense canopies are present. In any case, for assimilation experiments these types of surface cover will be flagged and rejected, as the sensitivity to soil moisture is lost.

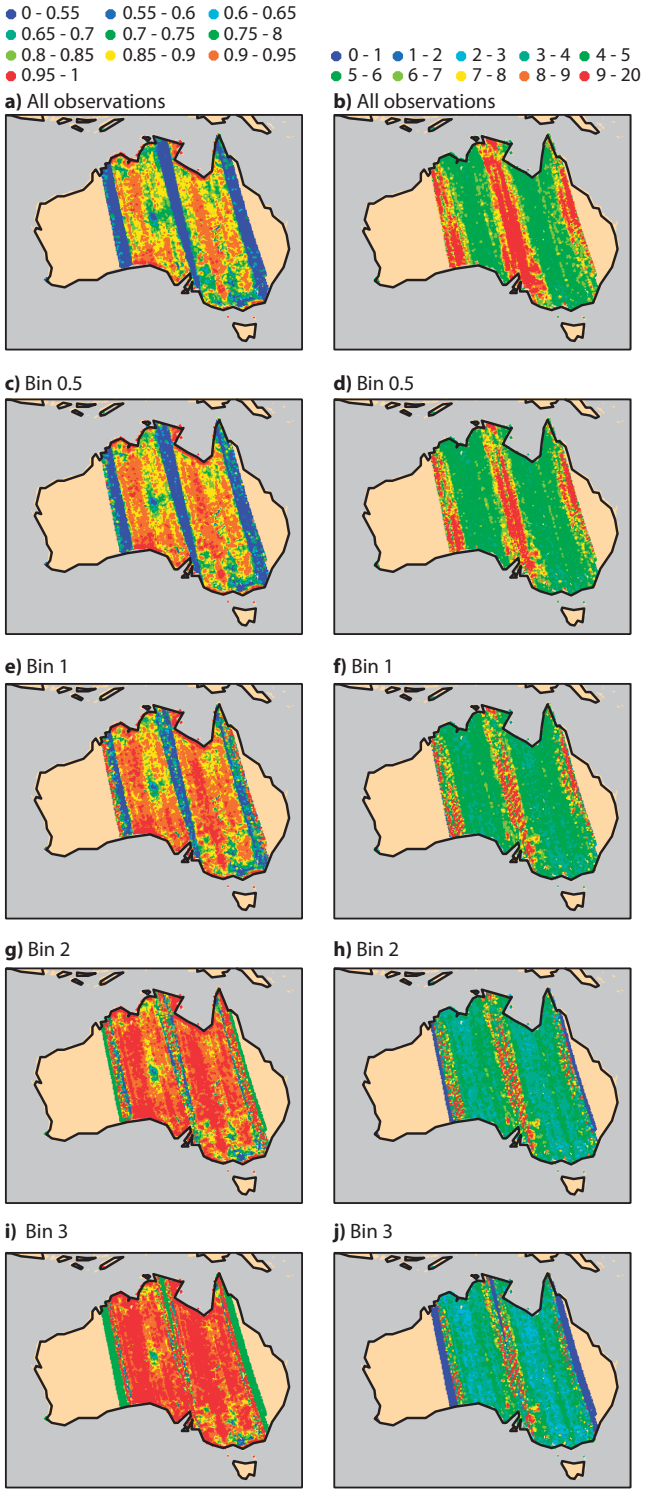

Figure 7: Coefficient of determination (r^2) and STD of the residues (in K) between the SMOS TB angular signature and its 2nd order polynomial regression model for the XX polarisation, the 1 December 2010, ascending orbits.

4 Mean statistical results over continents

In this section the average of the r^2 and the STD statistical variables are computed independently per continent, type of orbit, polarisation, bin size and type of regression model. Only significant correlations are considered (p-value lower than 0.05 [9]). Tables 1, 2 and 3 group the mean statistics, fitting the observations to a 2^{nd} and 3^{rd} order polynomial, on 1 December 2010 for Australia, North and South America, respectively. Tables 4, 5 and 6 show the same averaged values but only for a 2^{nd} order polynomial function and on 1 June 2011. From the values shown in these tables it can be observed that:

- The STD of the residues to the fitted curve consistently decreases with increasing the size of the angular bin. The geographical averaged noise of the observations varies from 5.4 to 6.4 K without any binning to 2.3 to 2.8 K when the maximum bin size of 3 degrees is used for the December case. The June case shows very similar results except for the ascending orbits of North America which show larger STD values. By averaging the observations in bins up to three degrees a potential noise reduction around 3.5 K could be achieved. However, in 86 % of cases (and all of the summer cases), averaging the observations in bins of 3 degrees does not result in better r^2 than for bins of 2 degrees. In most cases 3 degrees averaging is even worse. This is an indication that when an excessive binning is applied, the natural variability of the observations is included, and thus the polynomial fit is no longer a better representation of the observations. Higher order polynomials change the sign of the gradient several times and they have the potential to better fit the noise, which is not the objective of this methodology. Therefore, averaging the observations in bins of 2 degrees should be the maximum acceptable to reduce noise from the observations. It was checked (not shown) that by using larger angular bins (4, 5 and 6 degrees) the r^2 is slowly getting worse. Thereafter, only a 2^{nd} order polynomial regression model will be used for further analysis.

- In most of the cases studied, the polynomial fit represents better the XX polarisation than the YY polarisation, except for some orbits over South America where the behaviour is found to be more alike. Likewise, the STD is, in general, also lower for the XX mode than for the YY mode. Although the XX polarisation is more sensitive to the soil water content and shows greater dynamics, according to these results the larger dynamics does not necessarily means larger noise. After averaging the observations in angular bins of 2 degrees, the noise difference between XX and YY polarisations is lower than 1 K, often less than 0.5 K, except for the ascending case over North America where the difference is 1.16 K.

- While the STD of the residues is larger in ascending than in descending orbits, the trend is that the polynomial regression model explains better the distribution of T_B for ascending orbits. However, these results change for continent, polarisation and period of the year. For Australia and South America, in general, r^2 is better for ascending than for descending orbits, however descending orbits show lower STD. The North America case seems to be more dependent on the period of the year, which it is likely associated to the snow covered area during the winter period. In this case, the r^2 values are significantly higher for the YY polarisation in the June case.

	model: Ax^2+Bx+C								model: Ax^3+Bx^2+Cx+D							
	ASCENDING				DESCENDING				ASCENDING				DESCENDING			
	XX		YY		XX		YY		XX		YY		XX		YY	
bin	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD
no-bin	0.65	5.16	0.56	6.23	0.58	4.65	0.48	5.46	0.68	5.39	0.61	6.39	0.63	4.98	0.55	5.63
0.5	0.66	4.33	0.59	5.21	0.59	3.91	0.51	4.61	0.71	4.56	0.66	5.36	0.65	4.27	0.59	4.82
1	0.70	3.80	0.63	4.45	0.63	3.38	0.55	3.91	0.75	3.95	0.70	4.54	0.70	3.61	0.63	4.02
2	**0.73**	**3.10**	**0.66**	**3.54**	**0.63**	**2.55**	**0.56**	**2.94**	0.77	3.13	0.71	3.52	0.65	2.58	0.58	2.87
3	0.73	2.73	0.66	2.98	0.61	2.14	0.53	2.40	0.73	2.65	0.68	2.87	0.60	2.06	0.55	2.26

Table 1: Mean r^2 and STD between the SMOS observed brightness temperatures angular signature and its 2^{nd} and 3^{rd} order polynomial fitted curve, for Australia, on 1 December 2010.

	model: Ax^2+Bx+C								model: Ax^3+Bx^2+Cx+D							
	ASCENDING				DESCENDING				ASCENDING				DESCENDING			
	XX		YY		XX		YY		XX		YY		XX		YY	
bin	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD
no-bin	0.40	5.89	0.32	8.15	0.42	5.42	0.32	6.07	0.45	6.56	0.37	9.16	0.47	5.85	0.38	6.65
0.5	0.43	4.87	0.35	6.80	0.45	4.46	0.35	4.89	0.49	5.48	0.42	7.67	0.52	4.80	0.43	5.44
1	0.48	4.18	0.40	5.76	0.50	3.81	0.39	4.10	0.55	4.60	0.48	6.40	0.57	4.04	0.48	4.52
2	0.51	3.10	0.45	4.26	0.53	2.89	0.44	3.09	0.56	3.32	0.50	4.59	0.57	2.96	0.50	3.31
3	0.52	2.54	0.46	3.34	0.53	2.41	0.45	2.53	0.54	2.64	0.50	3.60	0.55	2.42	0.50	2.67

Table 2: As in Table 1, but for North America.

	model: Ax^2+Bx+C								model: Ax^3+Bx^2+Cx+D							
	ASCENDING				DESCENDING				ASCENDING				DESCENDING			
	XX		YY		XX		YY		XX		YY		XX		YY	
bin	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD	r^2	STD
no-bin	0.41	7.06	0.42	6.09	0.33	5.21	0.28	4.68	0.47	7.87	0.49	6.40	0.40	5.62	0.37	5.51
0.5	0.43	5.95	0.44	5.12	0.35	4.42	0.29	3.80	0.50	6.75	0.52	5.52	0.44	4.83	0.41	4.68
1	0.47	4.96	0.48	4.30	0.40	3.77	0.33	3.13	0.55	5.61	0.56	4.63	0.48	4.02	0.45	3.83
2	0.48	3.57	0.50	3.22	0.40	2.62	0.32	2.19	0.53	3.96	0.55	3.36	0.44	2.80	0.40	2.66
3	0.47	2.83	0.48	2.60	0.38	2.08	0.30	1.69	0.50	3.15	0.53	2.69	0.41	2.18	0.36	2.04

Table 3: As in Table 1, but for South America.

model: $Ax^2 + Bx + C$

	ASCENDING				DESCENDING			
	XX		YY		XX		YY	
bin	r^2	STD	r^2	STD	r^2	STD	r^2	STD
no-bin	0.63	4.64	0.50	5.55	0.59	4.65	0.52	5.63
0.5	0.64	3.93	0.53	4.72	0.60	3.89	0.54	4.78
1	0.68	3.38	0.57	4.00	0.65	3.32	0.59	3.98
2	0.68	2.60	0.59	3.08	0.64	2.46	0.59	3.01
3	0.66	2.22	0.57	2.51	0.60	2.07	0.55	2.43

Table 4: Mean r^2 and STD between SMOS observed brightness temperatures angular signature and its 2^{nd} order polynomial fitted curve, for Australia, on 1 June 2011.

model: $Ax^2 + Bx + C$

	ASCENDING				DESCENDING			
	XX		YY		XX		YY	
bin	r^2	STD	r^2	STD	r^2	STD	r^2	STD
no-bin	0.30	11.44	0.26	14.11	0.46	5.79	0.45	6.53
0.5	0.33	8.77	0.29	10.94	0.48	4.91	0.48	5.51
1	0.38	6.91	0.34	8.56	0.53	4.23	0.52	4.63
2	0.41	4.81	0.37	5.56	0.56	3.18	0.56	3.52
3	0.41	3.76	0.37	4.33	0.56	2.64	0.56	2.91

Table 5: As in Table 4 but for North America.

model: $Ax^2 + Bx + C$

	ASCENDING				DESCENDING			
	XX		YY		XX		YY	
bin	r^2	STD	r^2	STD	r^2	STD	r^2	STD
no-bin	0.44	5.32	0.36	5.52	0.37	6.49	0.39	6.44
0.5	0.46	4.54	0.38	4.67	0.39	5.54	0.41	5.43
1	0.51	3.94	0.41	3.95	0.44	4.81	0.45	4.60
2	0.53	2.98	0.43	2.92	0.46	3.61	0.48	3.55
3	0.52	2.50	0.43	2.33	0.45	2.98	0.47	2.91

Table 6: As in Table 4 but for South America.

5 Results by soil type, vegetation cover type and incidence angle

In order to study the potential influence of the soil and vegetation cover type on the angular signature of the observations, mean statistics were computed independently for each type of soil texture and vegetation cover type used in H-TESSEL [1]. According to this classification, up to seven different soil textures are allowed, all of them with their own wilting point and field capacity characteristic values: coarse, medium, medium-fine, fine, very fine, organic and tropical organic. For vegetation, H-TESSEL uses the classification of the Global Land Cover Characterization (GLCC) database which has been derived using one year of Advanced Very High Resolution Radiometer (AVHRR) data and ancillary information ([4]; http://edcdaac.usgs.gov/glcc/glcc.html), being the nominal resolution 1 km. The vegetation cover types are split in high and low vegetation type. High vegetation types are evergreen needleleaf trees, deciduous needleleaf trees, deciduous broadleaf trees, evergreen broadleaf trees, mixed forest/woodland and interrupted forests. Low vegetation classes include crops/mixed farming, short grass, tall grass, tundra, irrigated crops, semidesert, bogs and marshes, evergreen shrubs and decidious shrubs.

5.1 Soil type influence

Fig. 8 (left) shows the map of the soil texture for the orbits swath covering Australia 1 December 2010. In Fig. 8 (right) the number of SMOS observations for each soil type are shown. This figure shows that for these orbits the soil is dominated by coarse and medium textures, but also a significant contribution of fine and medium-fine textures is present too. The averaged r^2 and STD mean values of its 2^{nd} order polynomial fit for the XX polarisation are shown in Fig. 9 and for the YY polarisation in Fig. 10. Results are presented separately for ascending and descending orbits and for all angular bins used in this study. These figures show that all types of soil texture show similar statistics, except the very fine soil texture type. However, the very fine texture statistics are not significant as the number of observations collected for this type were very low (see Fig. 8 right). There is a good fit between observations and the polynomial regression model, slightly better for ascending orbits and exceeding $r^2 = 0.7$ when the observations are averaged in angular bins. The best scores are obtained when angular bins of 2 degrees are used. The STD is in agreement with the statistics presented in Table 1. This figure also shows the advantages of averaging the data in bins of up to 2 degrees, reducing in all cases the noise level by 3 K. In section 9 the results obtained for North and South America are shown. For the North America case, the correlation of the observations with the polynomial regression model are lower for all soil types, which is also due to larger soil heterogeneities of North America compared to Australia. However, after binning up to 2 degrees the levels of noise are quite close to those of the Australia case, being slightly lower for the descending orbits. While for North America these results show a slight trend to decrease the noise of the observations with decreasing the size of the soil particles, the contrary trend is observed for South America. Therefore, these results do not provide any evidence of any type of soil texture over which observations are more noisy than over the others. So the soil texture type is not the dominant factor in the production of noise in the observations.

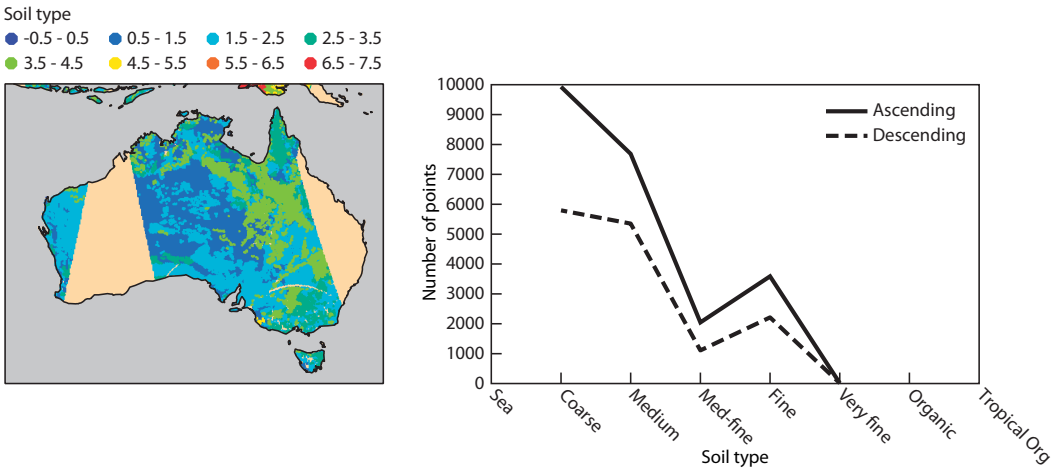

Figure 8: Soil type map and number of observations per soil type the 1 December 2010 in Australia. Legend color texture code: 0=Sea, 1=Coarse, 2=Medium, 3=Medium-fine, 4=Fine, 5=Very fine, 6=Organique, 7=Tropical-organique.

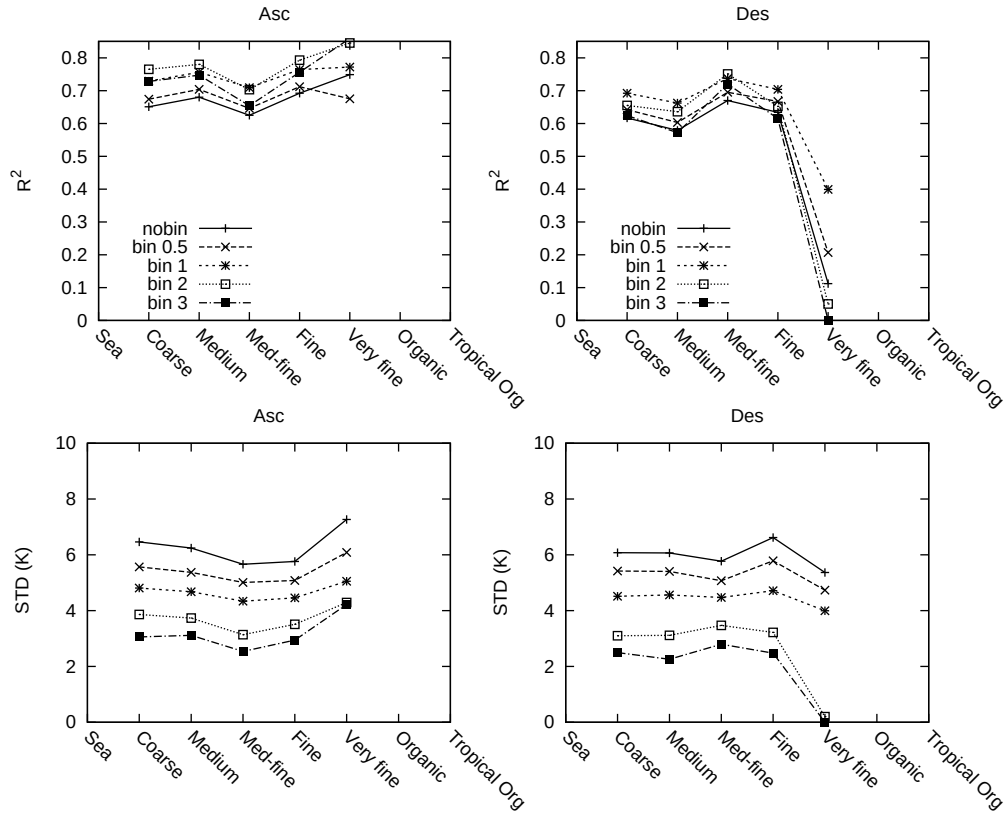

Figure 9: Mean r^2 and STD (in K) per type of soil texture class for Australia, between SMOS T_B and its 2^{nd} order polynomial fit, for the XX polarisation, the 1 December 2010.

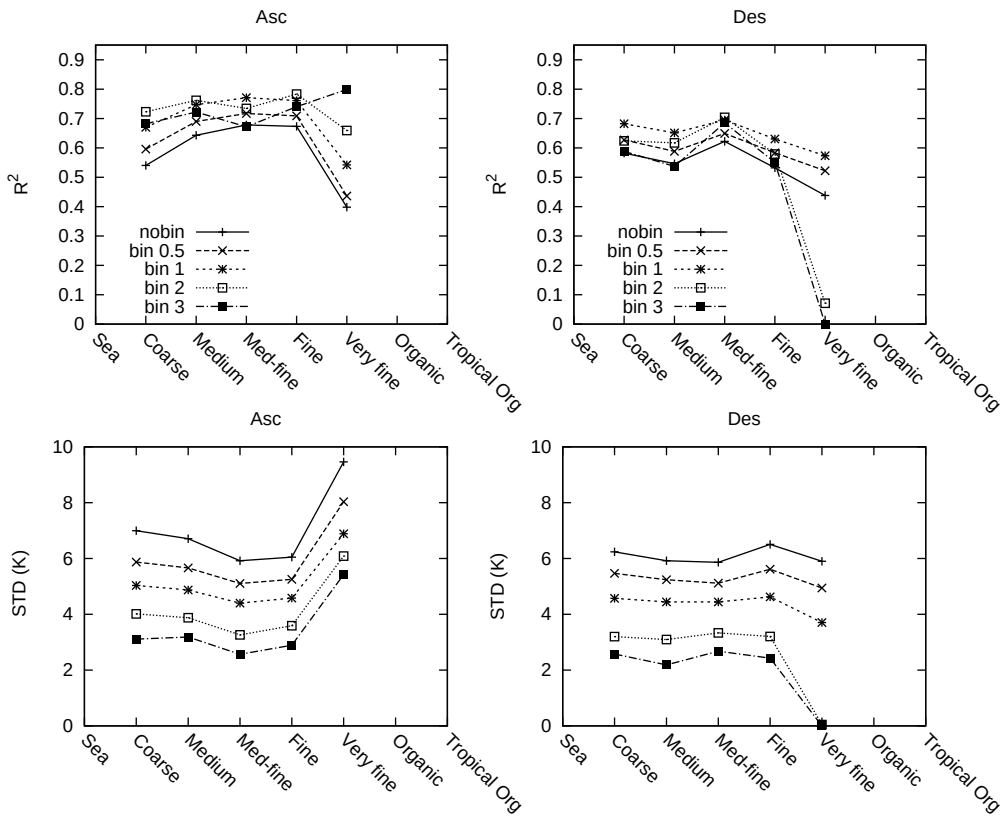

Figure 10: As in Fig. 9 but for the YY polarisation.

5.2 Vegetation cover types

The vegetation cover type was also studied as a possible factor influencing the noise of the observations. Statistics were computed independently for each type of vegetation cover type used in H-TESSEL, as it was done in section 5.1. For Australia, the semi-desert and tall grass types are the most representative for the area covered by the SMOS overpasses on 1 December 2010 (see Fig. 11). Australia obtains the best scores in terms of r^2 as a significant fraction of bare soil is commonly present in a pixel, and the angular signature can be explained quite well by a 2^{nd} order polynomial, being the ascending orbits of the XX polarisation those obtaining the best results (see Figs. 12 and 13). In terms of STD, the results are quite equivalent for both types of orbits and slightly better for the XX polarisation. Both, North and South America have a wider range of vegetation types over which significant correlation values than Australia are available. For South America (see Figs. 44 to 46) quite similar levels of noise are observed for each type of vegetation cover type. No evidence of different behaviour between high or low vegetation types were found. Correlation values are a bit lower for the high vegetation type as the angular signature in this case is flatter and small deviations from the flat behaviour has a significant impact on the r^2 values. It was also found higher noise levels for the XX polarisation and slightly larger for ascending orbits. Anomalous behaviours can be found where very few number of pixels with significant correlation values were present, because from a statistical point of view they are not representative of the whole type. See for example the abnormal high r^2 values for the evergreen needleleaf trees classe of South America in Figs. 45 and 46. For North America (Figs. 47 to 49), a larger representation of evergreen forests is present. In this case slightly larger noise is found for ascending orbits as for the South America case, however the YY polarisation is more noisy than the XX polarisation. The results obtained for the June case are similar for Australia and South America (see Figs. 50 to 52 and Figs. 53 to 55, respectively), with small differences reflecting the different areas covered by the orbits in December and June and the changes on the soil cover characteristics. The North America case seems a bit more complex (Figs. 56 to 58), mainly for the ascending orbits with large values of noise when all the observations are used.

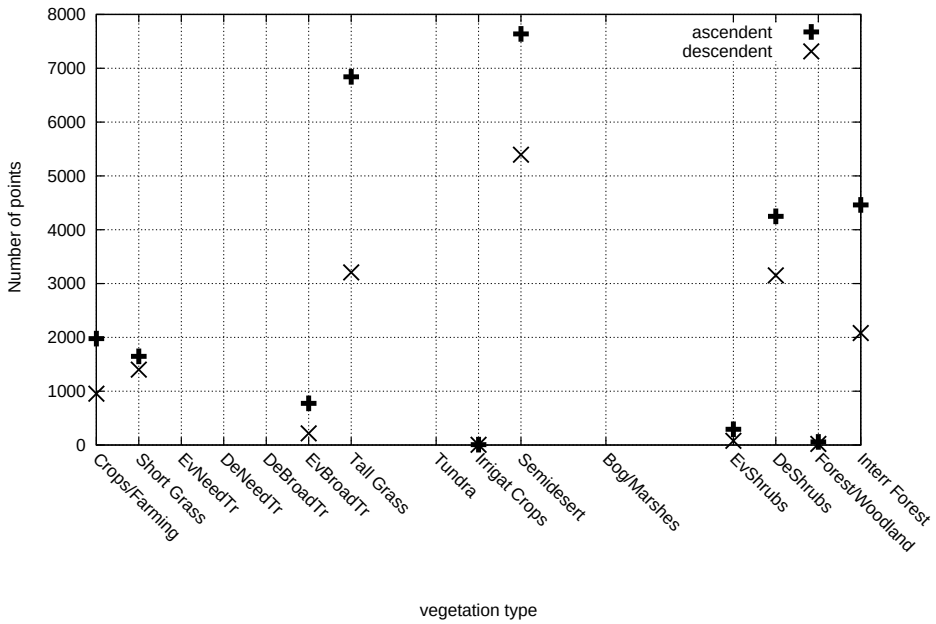

Figure 11: Number of measurements per vegetation cover type the 1 December 2010 in Australia.

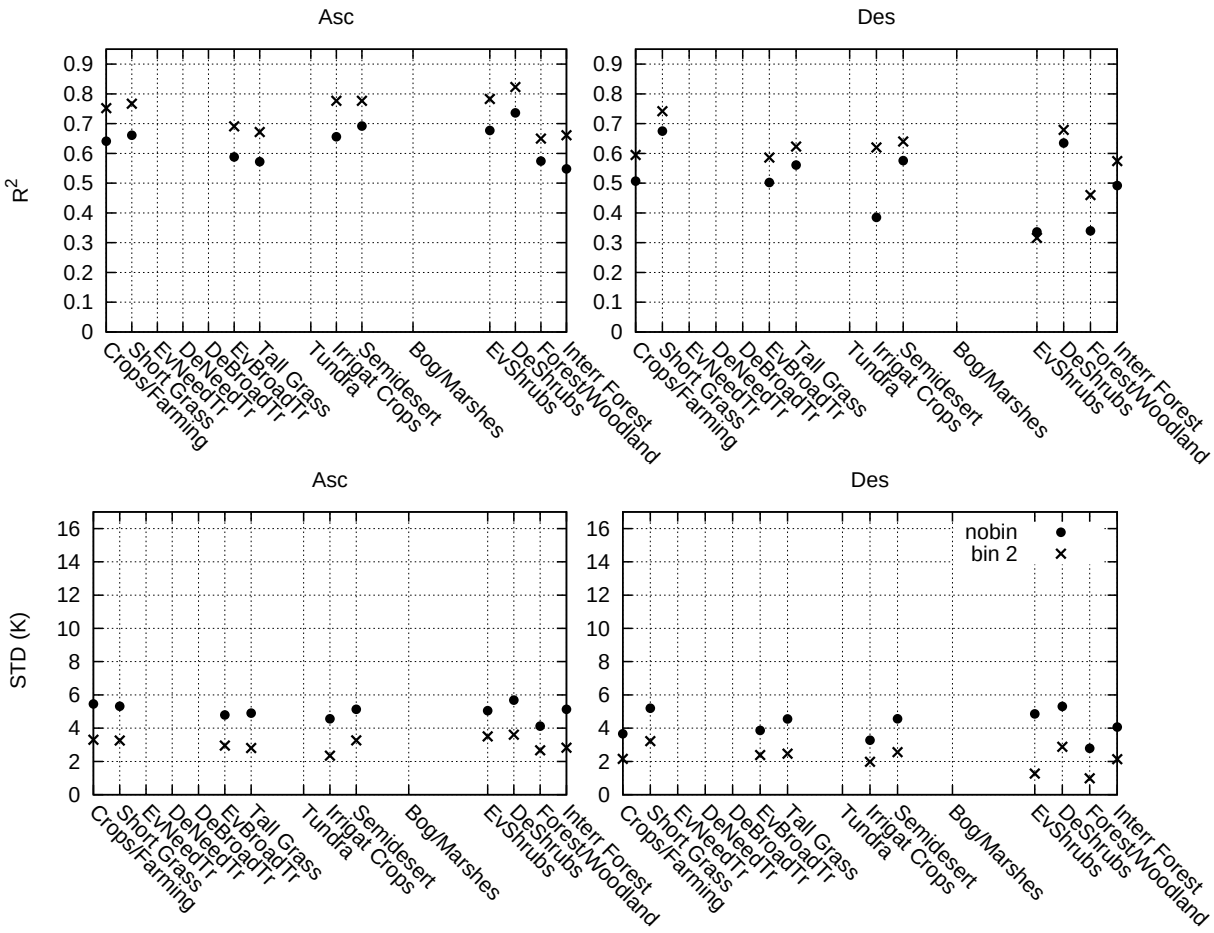

Figure 12: Mean r^2 and STD (in K) per type of vegetation type in Australia, between SMOS T_B and its 2^{nd} order polynomial fit, for the XX polarisation the 1 December 2010.

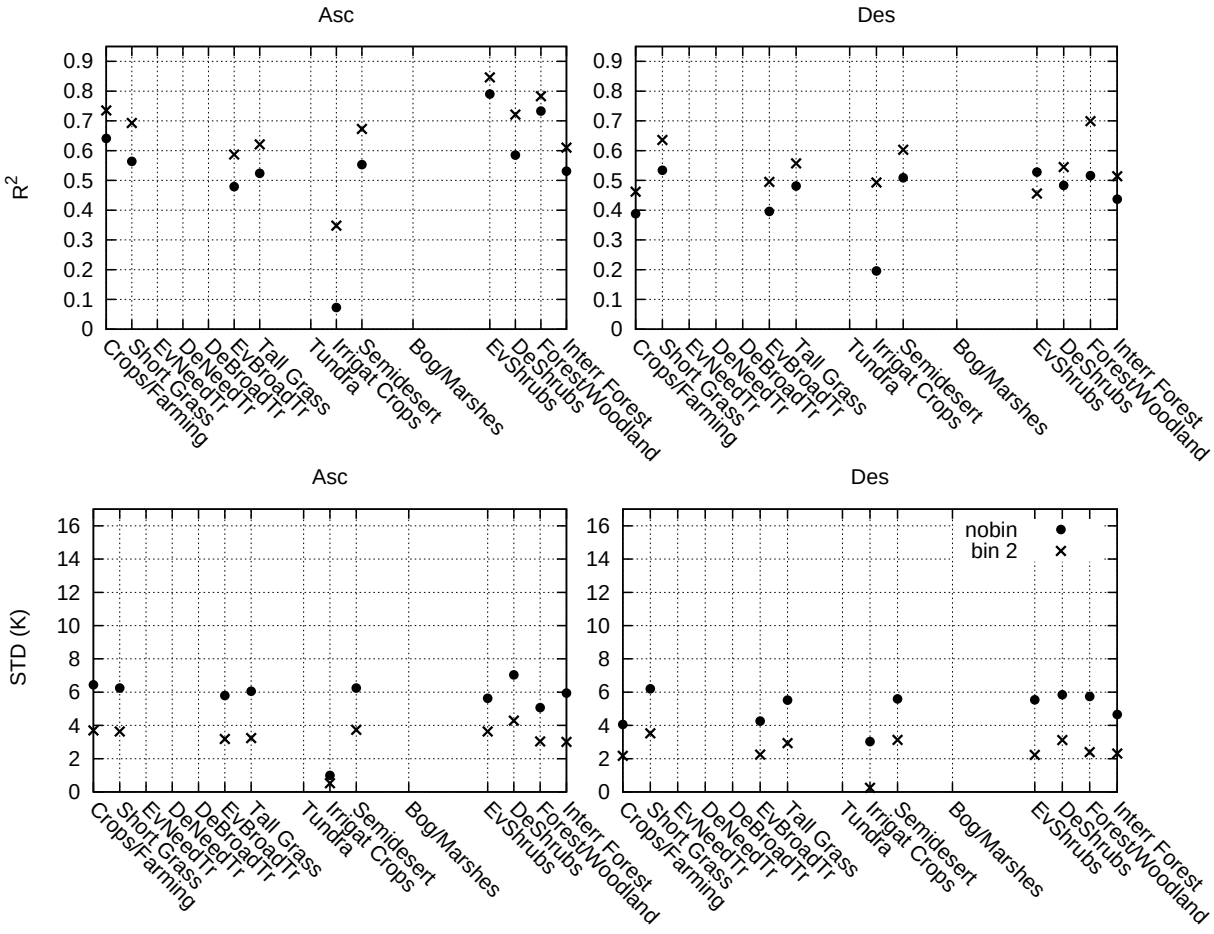

Figure 13: As in Fig. 12 but for the YY polarisation.

5.3 Angular noise

For data assimilation purposes is important to understand whether some incidence angles are more affected by noise than others. In this context, the mean residues of the observed brightness temperatures to a second order polynomial fit were computed per incidence angle, grouping observations in angles of 2 degrees. Fig.14 shows the results separately per continents and per type of orbit. The left panels corresponds to the XX polarisation and the right panel to the YY polarisation. Overlapped to the XX polarisation plots are also the number of observations collected for each angular bin of 2 degrees. It is observed that the bin [42-44] degrees has in all cases the maximum number of observations, as 42.5 degrees is the incidence angle with the maximum number of views. The smallest incidence angles are in all cases those affected with larger noise. A trend is for the observed noise to increase for the largest incidence angles, many of which are within the EA-FOV, an area of lower quality.

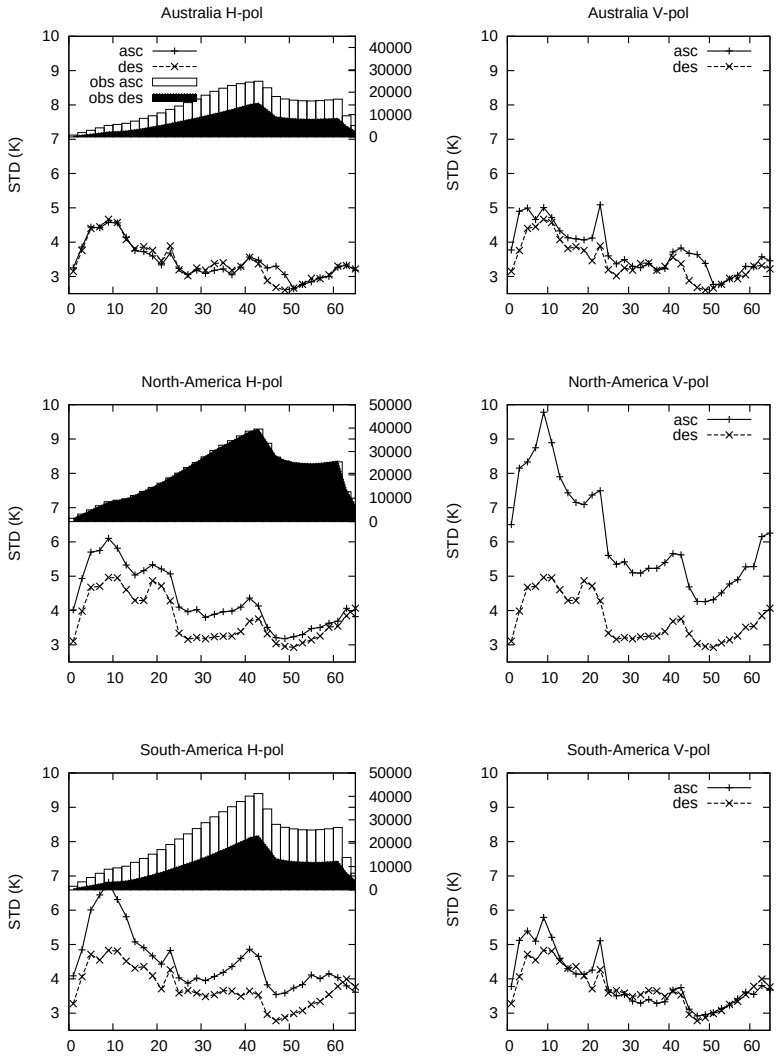

Figure 14: Mean STD of the observations fitted to a 2nd order polynomial as a function of the incidence angle of the observations, the 1 December 2010. Left panel is for the XX polarisation, right panel for the YY polarisation. Top figures are for Australia, middle figures North America and bottom figures for South America. The number of observations as a function of the incidence angle are also overlapped on the left panel figures, for ascending (empty vertical bars) and descending orbits (black vertical bars).

6 Conclusions

This report demonstrates the ability of the angular binning as a simple but effective method for reducing noise from the SMOS observed brightness temperatures. This method can be validated through a two-step process: 1/ a pixel-by-pixel polynomial fit to all observed brightness temperatures collected over a single satellite pass, 2/ repeat step 1 but averaging observations in angular bins of different size. The key statistical variables to be analysed are the coefficient of determination of the polynomial fit and the standard deviation of the residues to the fitted curve. In this study, Europe and Asia were avoided as they are strongly contaminated by RFI and the results could be misleading. Mean values over continental areas were computed in order to obtain statistical representative values and to filter out local outliers, for example caused by local sources of RFI. A second order polynomial regression model was chosen as reference to describe the angular behaviour of the observations. Higher order polynomials can obtain slightly better results than a second order polynomial, but this is due to a better fit of the noise to the regression model. Therefore they should be avoided.

It was found that by averaging observed brightness temperatures over the same node in angular bins of different size, the noise of the observations was reduced. The optimal bin size is 2 degrees, as for this bin the polynomial regression model explains better the angular signature of the observations while decreasing the noise. From a general perspective, this method has the potential to decrease noise from SMOS observed brightness temperatures by 3 K. Another advantage is that this method reduces the volume of data and also eases an operational implementation, which makes this method suitable to be implemented within the Integrated Forecasting System of ECMWF.

In general it was found better results for the XX polarisation, but some differences are found in North America, likely linked to the differences of soil covered by snow between December and June. Although the XX polarisation is more sensitive to soil moisture variations, its higher variability is not caused by a larger noise. It has greater skill to capture soil moisture variations than the YY polarisation. However, the difference in noise between both polarisations is reduced after binning observations and varies from a few tenths of kelvin to more than 1 K when a 2 degrees bin size is applied. Near the coasts spurious signals are obtained, as the contribution from open water surfaces is embedded in the SMOS observations. The results presented in this study also indicate that, in general, ascending orbits are more noisy than descending, whereas this conclusion may change if snow covered areas are included in the computations.

This study showed that there is not any evidence of any type of soil texture or vegetation cover type over which observations are more noisy. Slightly better representation of the low vegetation types by a 2^{nd} order polynomial was shown, as over these pixels a greater sensitivity to the soil water content is present in the angular L-band signal. It was also found that incidence angles below 20 degrees are the noisiest ones, but also the largest incidence angles are more noisy too. This information is important for data assimilation studies, as only the best observations should be assimilated. Also, the flat angular signature obtained for snow covered areas and dense vegetated forests (therefore losing the sensitivity to the soil moisture) produces low correlation and in many cases large noise values. For the soil moisture analysis these areas will also be masked out prior to assimilation, and a snow and dense forest mask will be applied to the observations.

Finally, the introduction of the RFI flags in the BUFR product received at ECMWF (see [7]) and the use of quality flags of the observations, will help to further enhance the efficiency of the binning approach by rejecting contaminated or low quality data before binning the observations.

Acknowledgements

This work is funded under the ESA-ESRIN contract number 4000101703/10/NL/FF/fk. The authors would like to thank Lars Isaksen from ECMWF for his useful advice and comments, and Matthias Drusch and Susanne Mecklenburg, both ESA staff, for their contribution in the project definition.

7 Appendix 1: Maps of the binning effect for ascending orbits.

7.1 December case

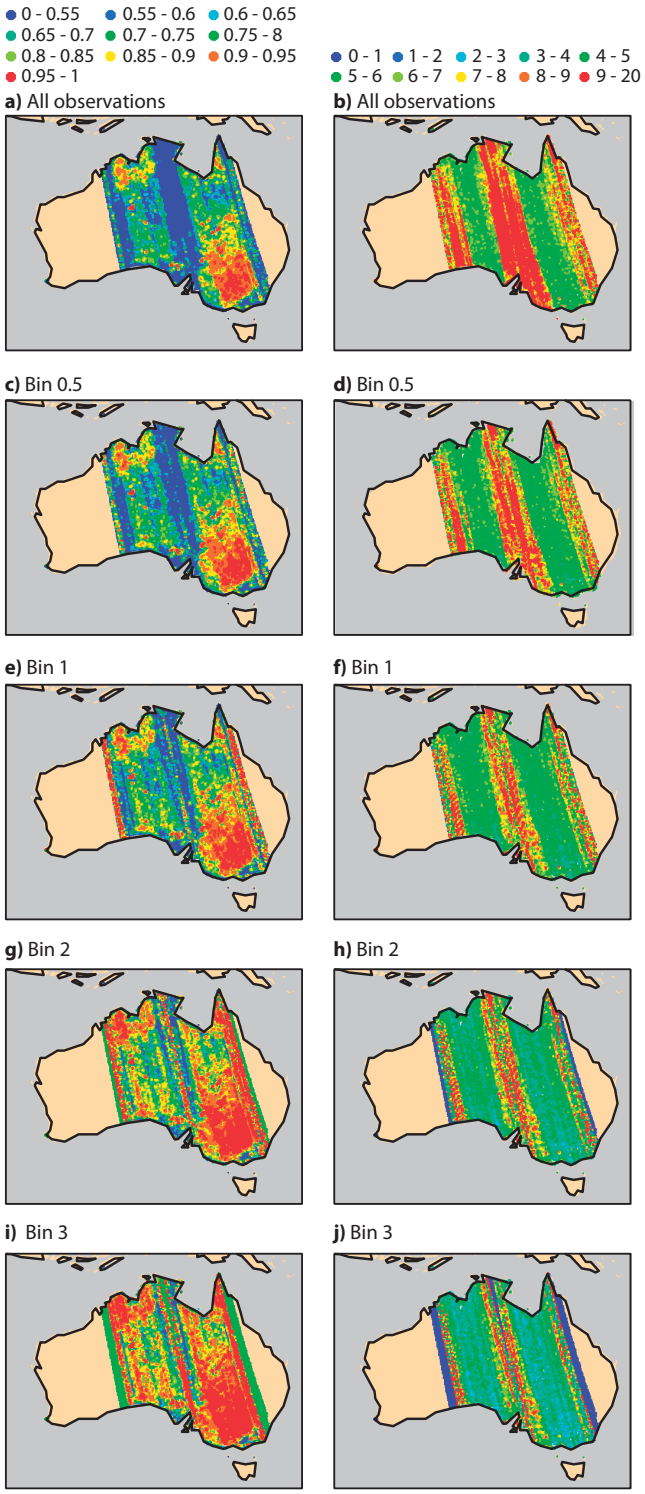

Figure 15: Coefficient of determination (r^2) and STD of the residues (in K) between the SMOS TB angular signature and its 2nd order polynomial regression model for the YY polarisation, the 1 December 2010.

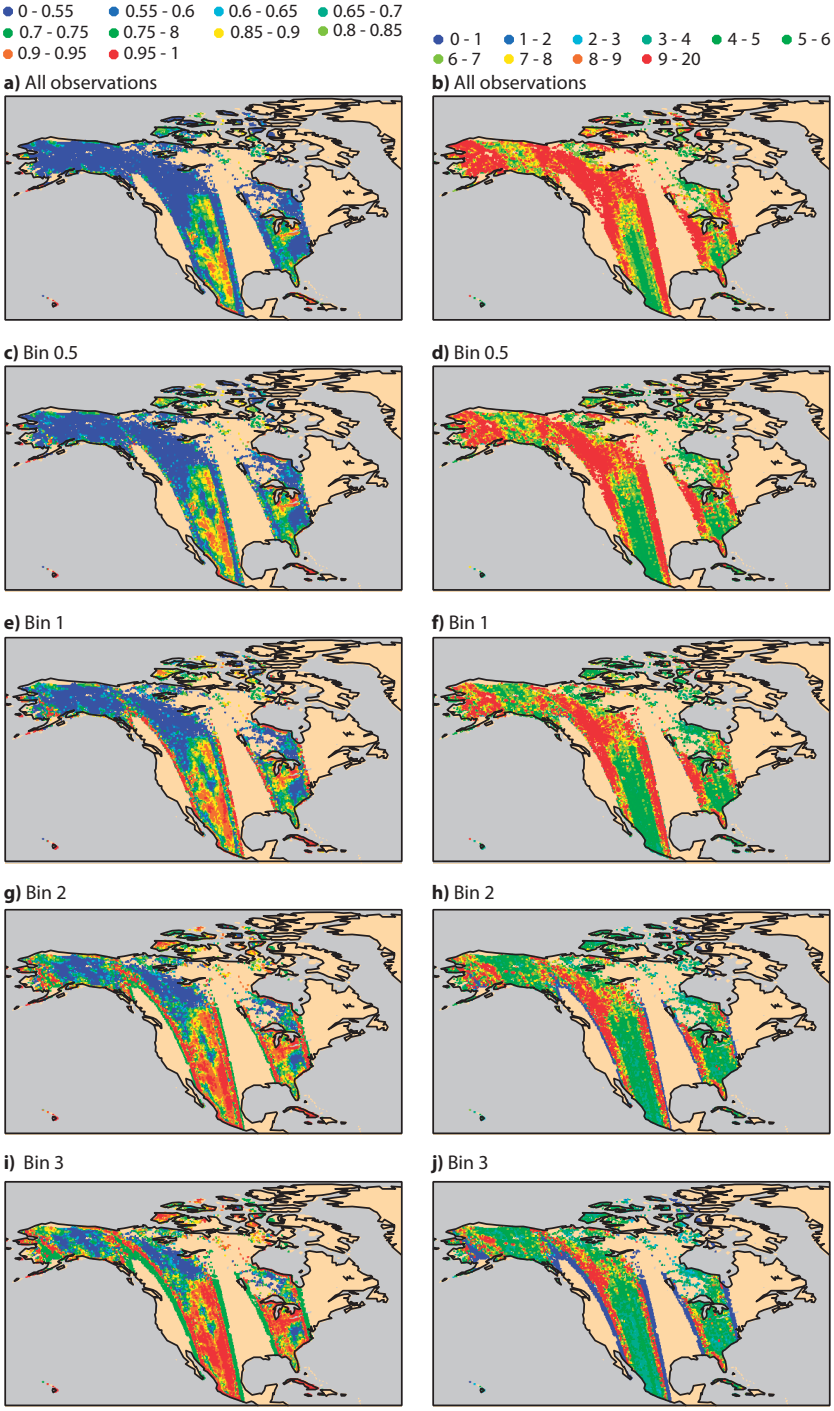

Figure 16: As in Fig. 15, but for North America and the XX pol.

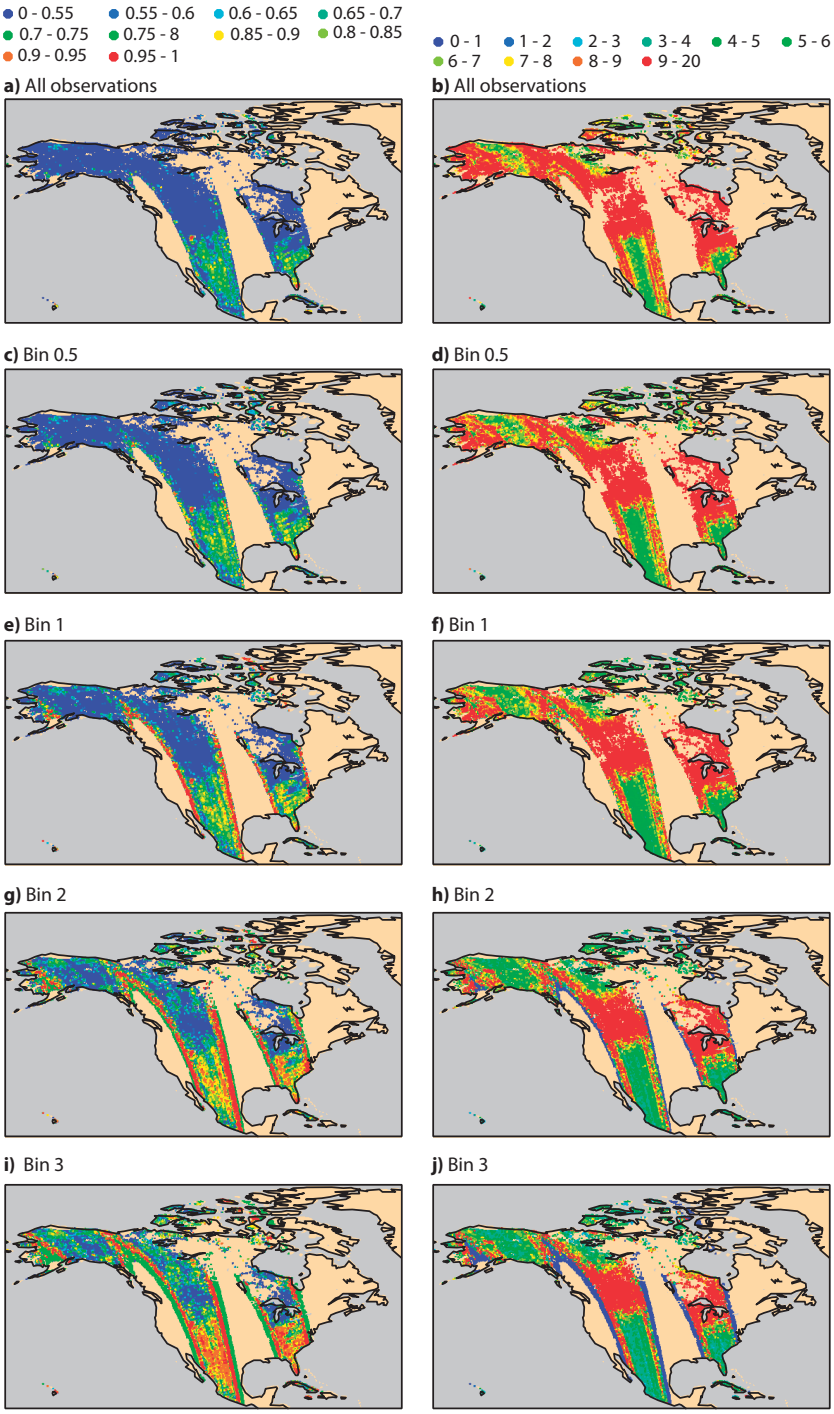

Figure 17: As in Fig. 15, but for North America and the YY pol.

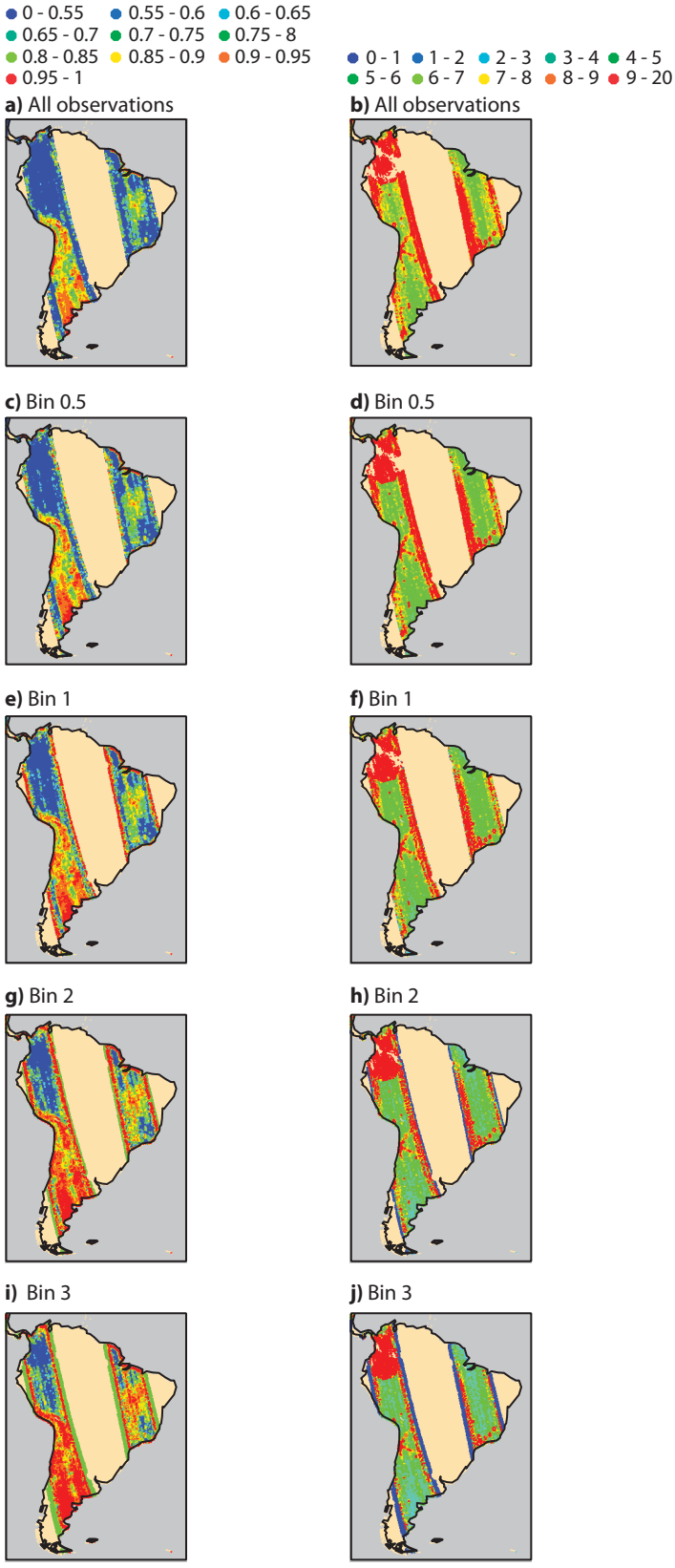

Figure 18: As in Fig. 15, but for South America and the XX pol.

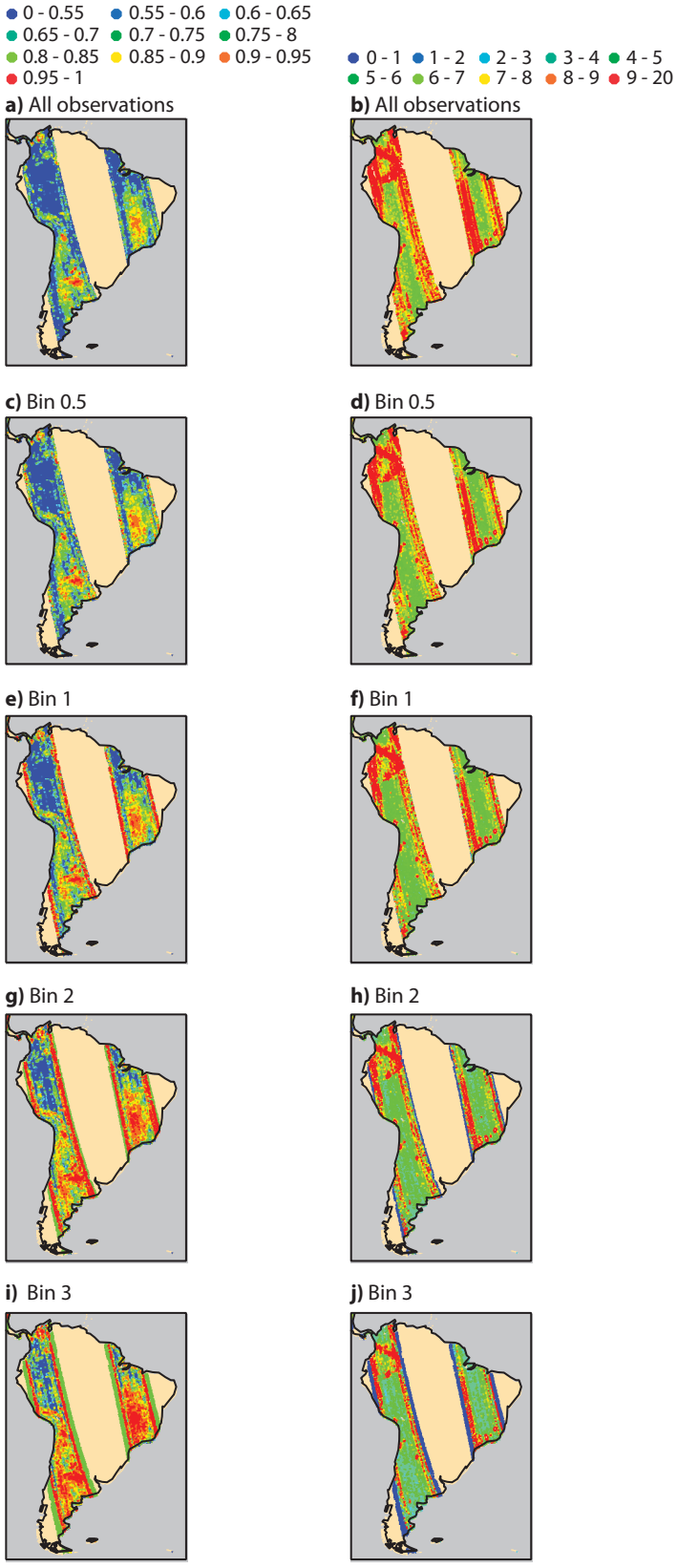

Figure 19: As in Fig. 15, but for South America and the YY pol.

7.2 June case

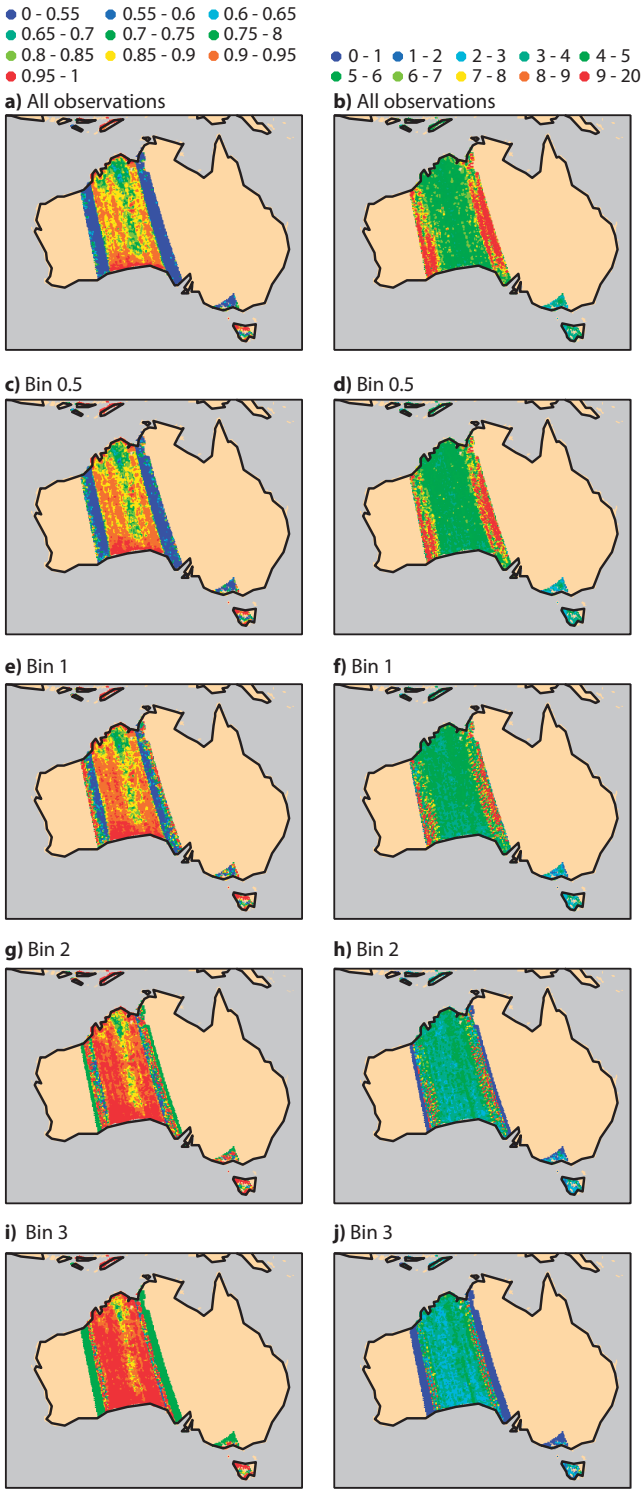

Figure 20: Coefficient of determination (r^2) and STD of the residues (in K) between the SMOS TB angular signature and its 2nd order polynomial regression model for the XX polarisation, the 1 June 2011.

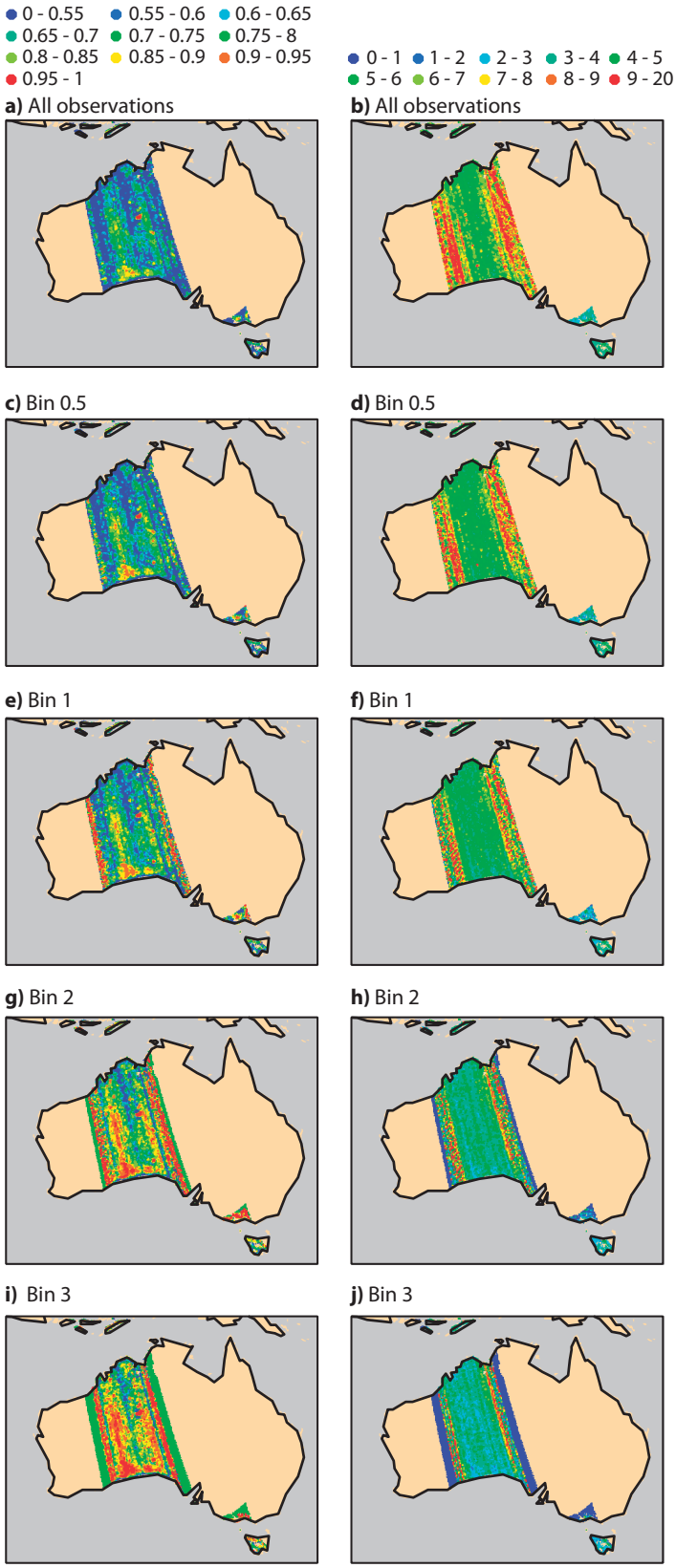

Figure 21: As in Fig. 20, but for YY pol.

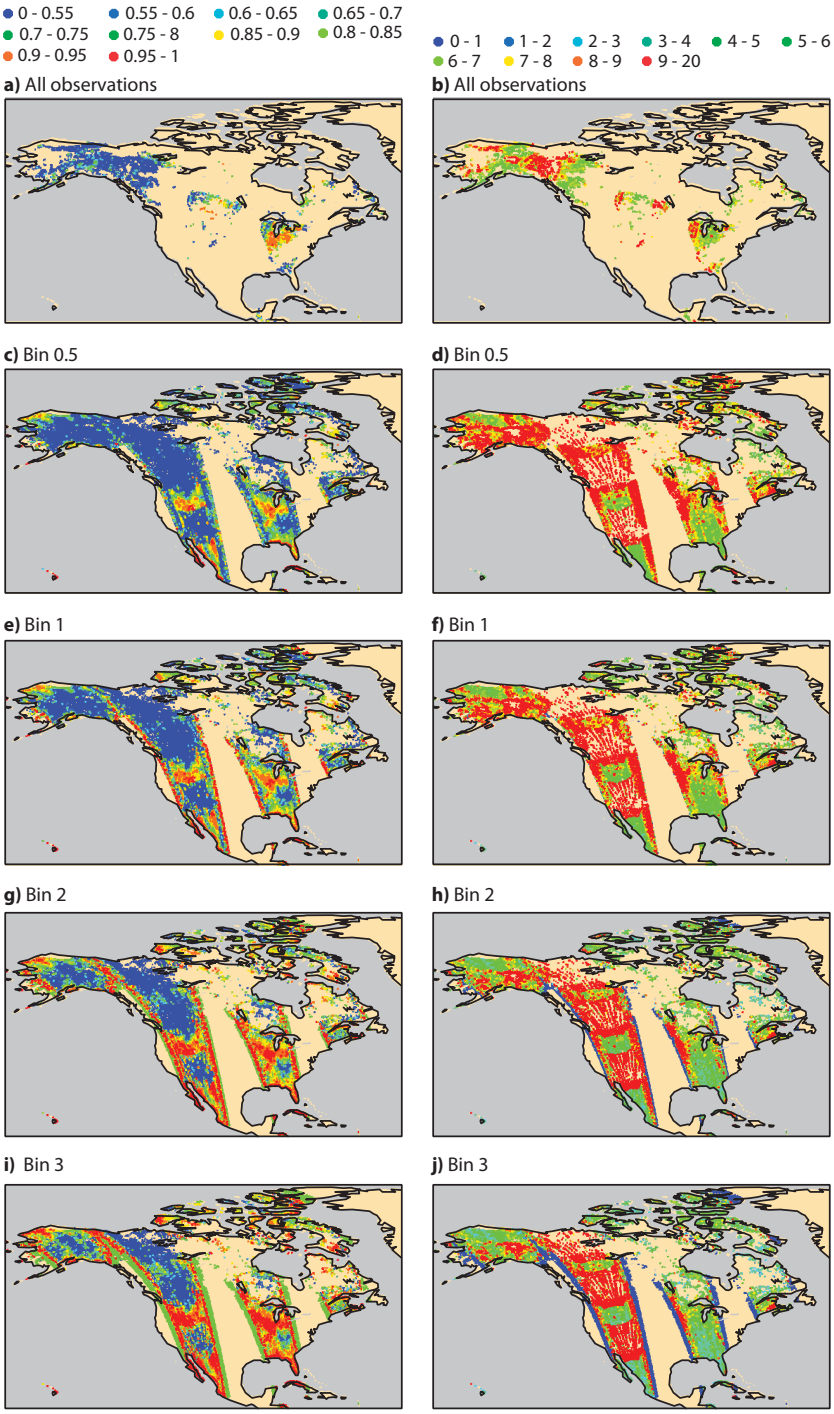

Figure 22: As in Fig. 20, but for North America and the XX pol.

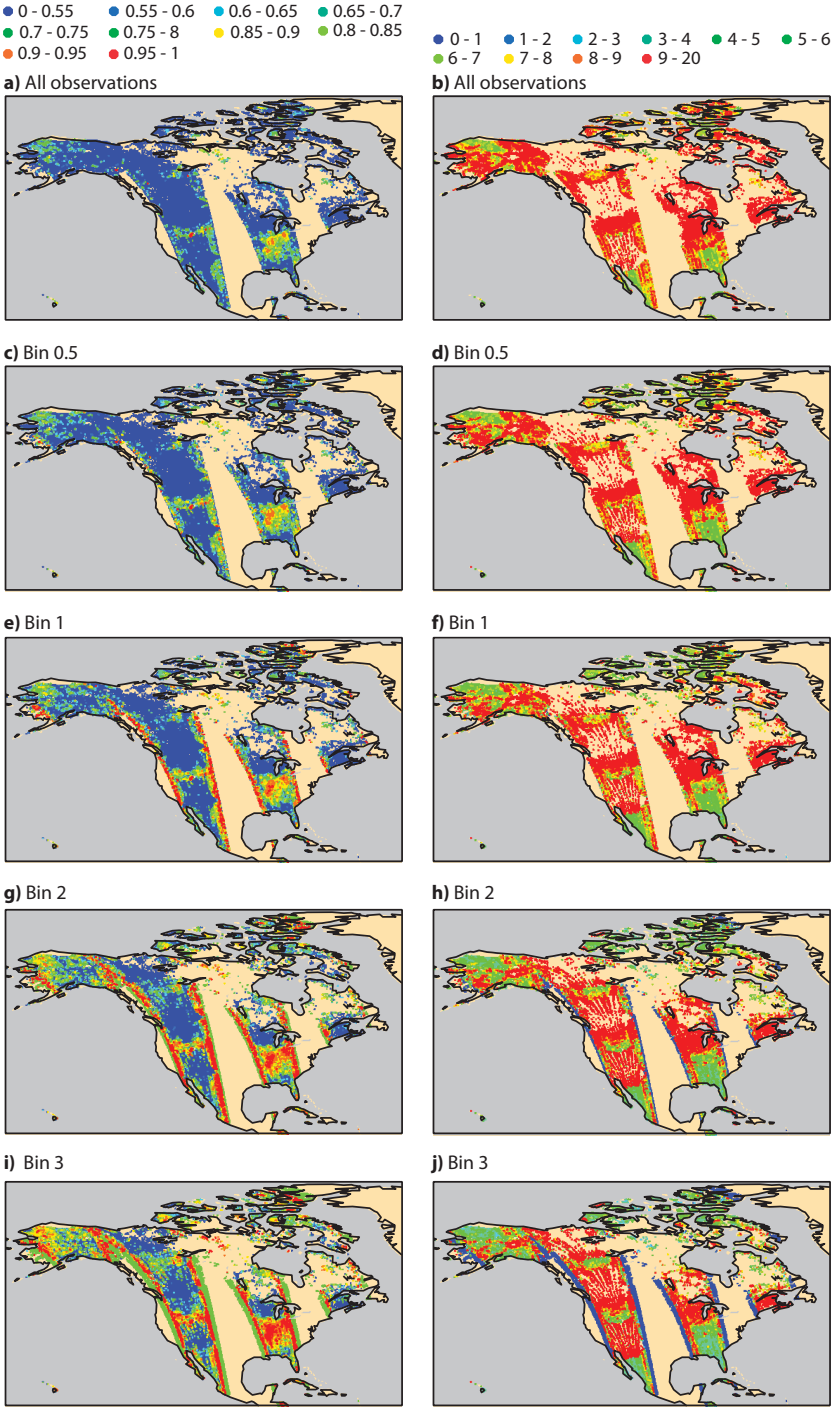

Figure 23: As in Fig. 20, but for North America and the YY pol.

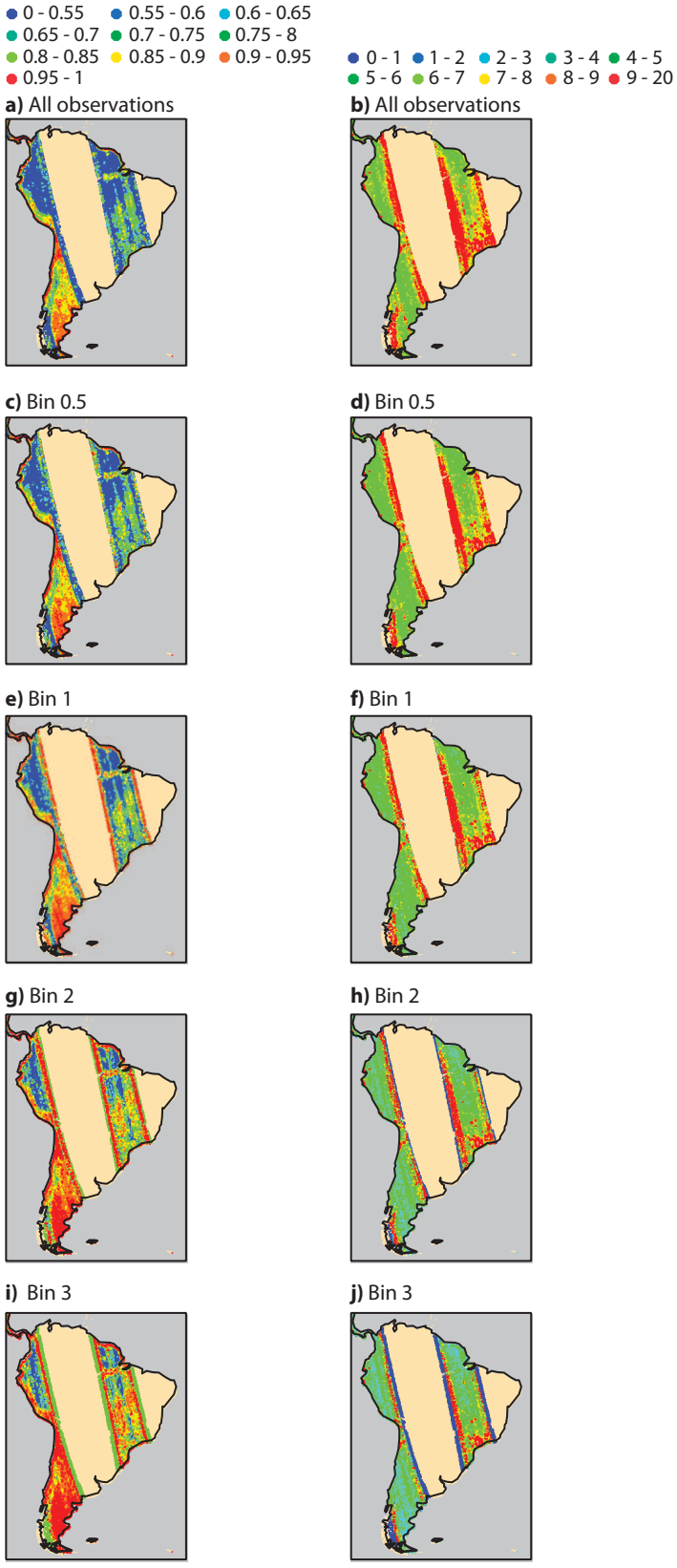

Figure 24: As in Fig. 20, but for South America and the XX pol.

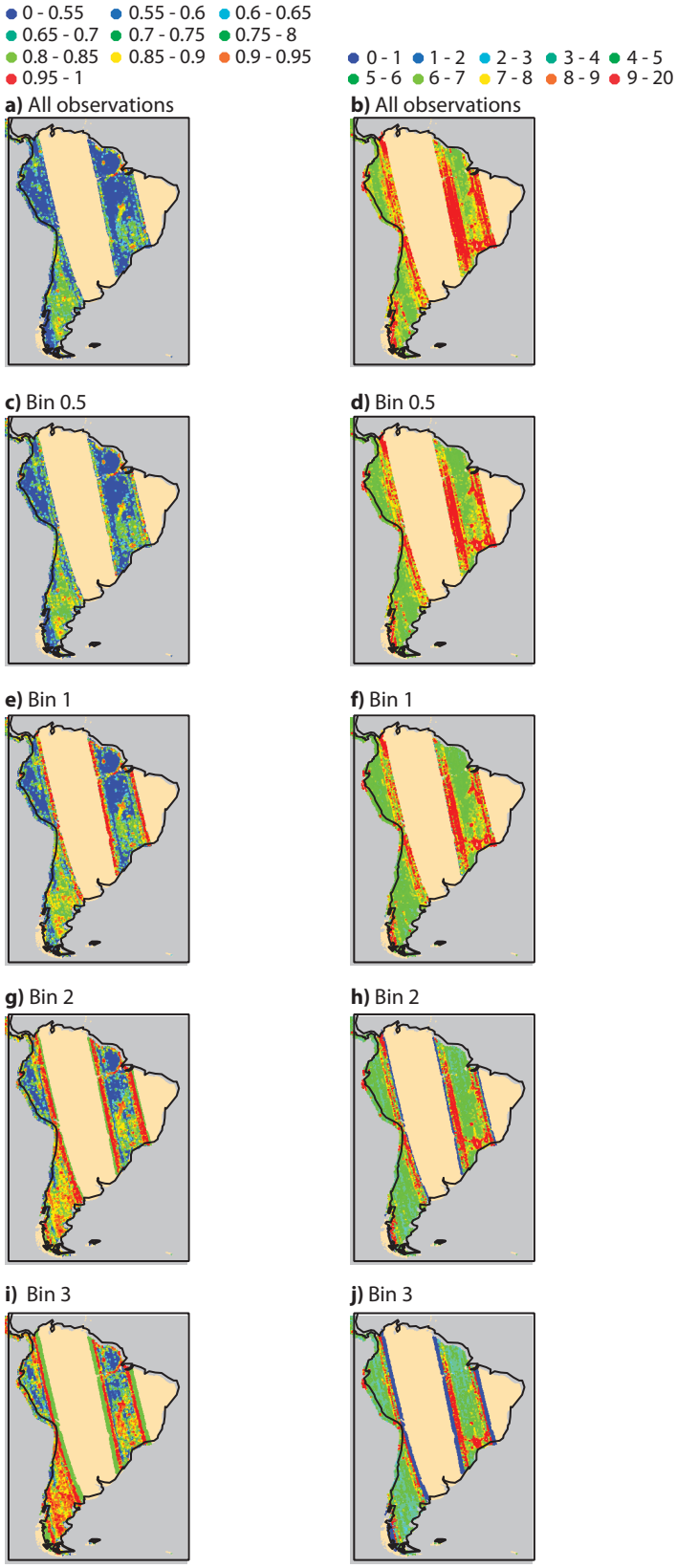

Figure 25: As in Fig. 20, but for South America and the YY pol.

8 Appendix 2: Maps of the binning effect for descending orbits.

8.1 December case

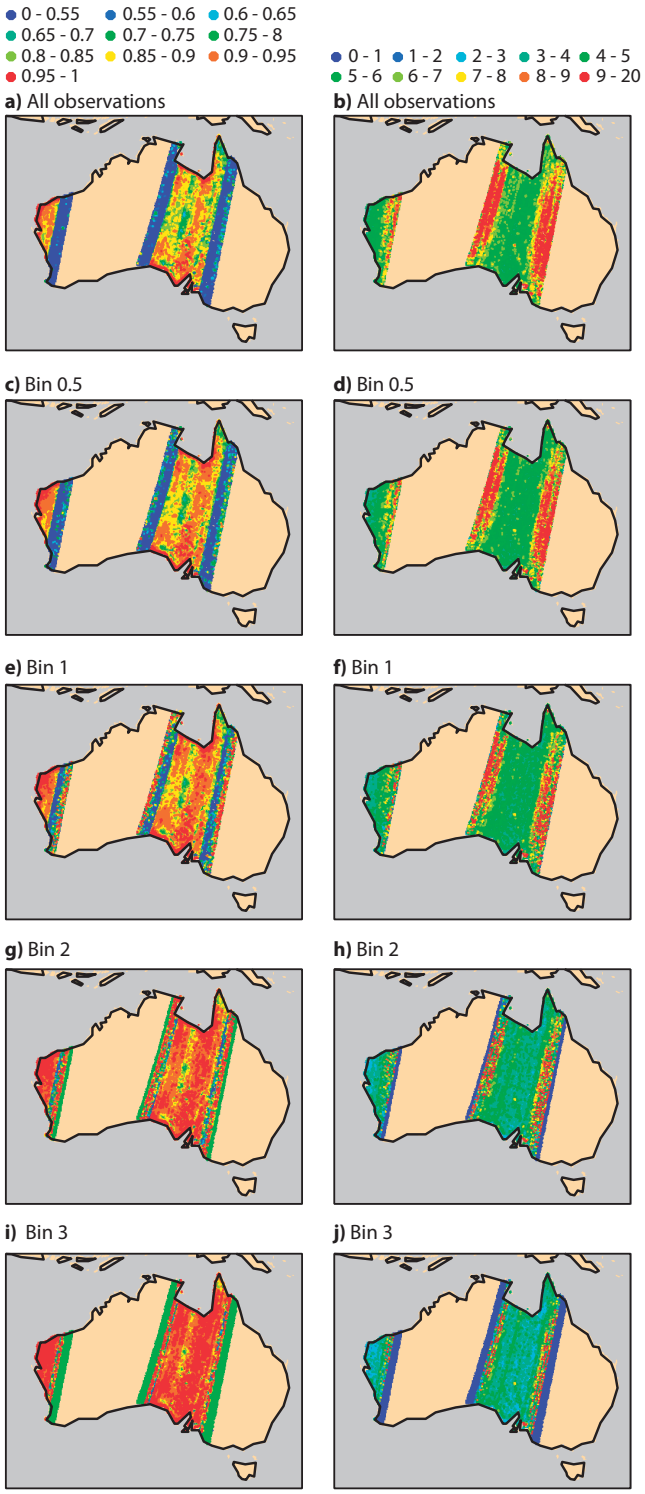

Figure 26: Coefficient of determination (r^2) and STD of the residues (in K) between the SMOS TB angular signature and its 2nd order polynomial regression model for the XX polarisation, the 1 December 2010.

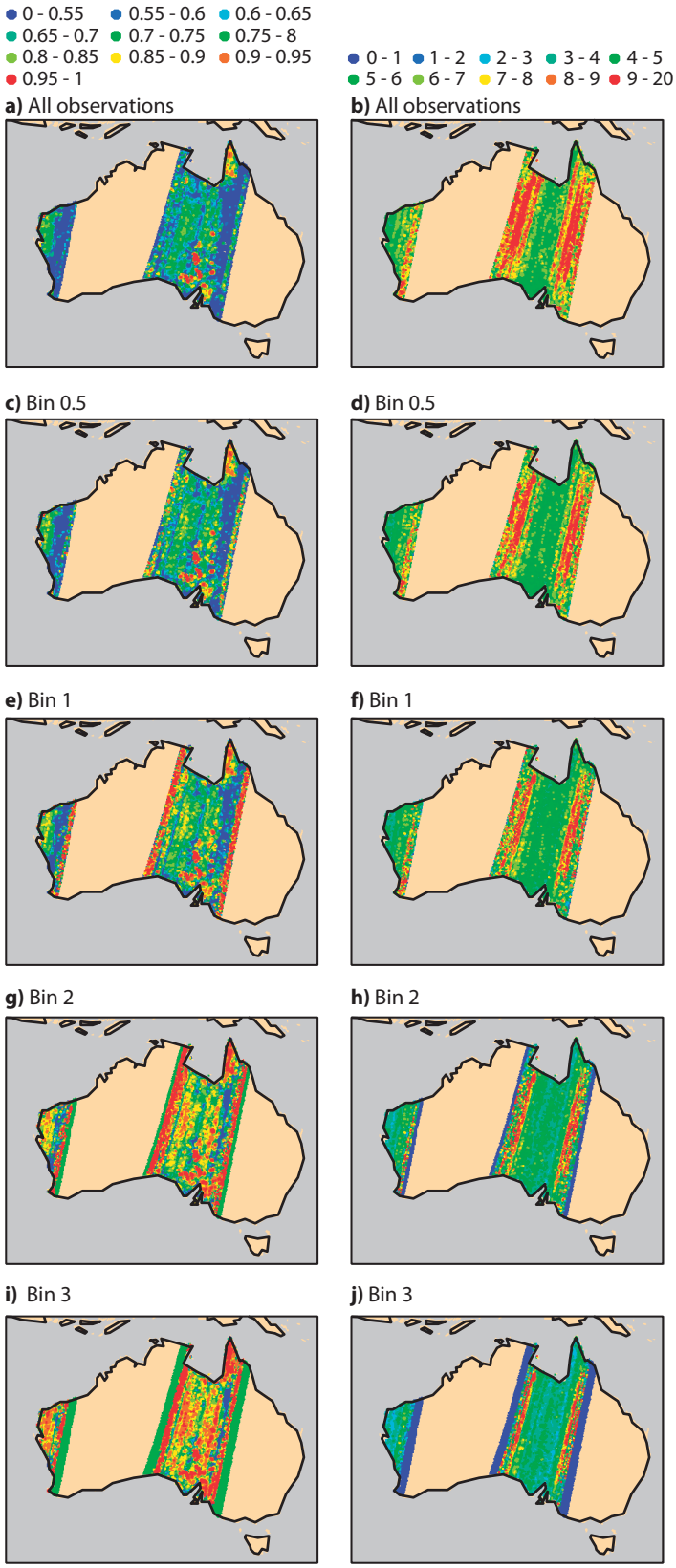

Figure 27: As in Fig. 26, but for the YY pol.

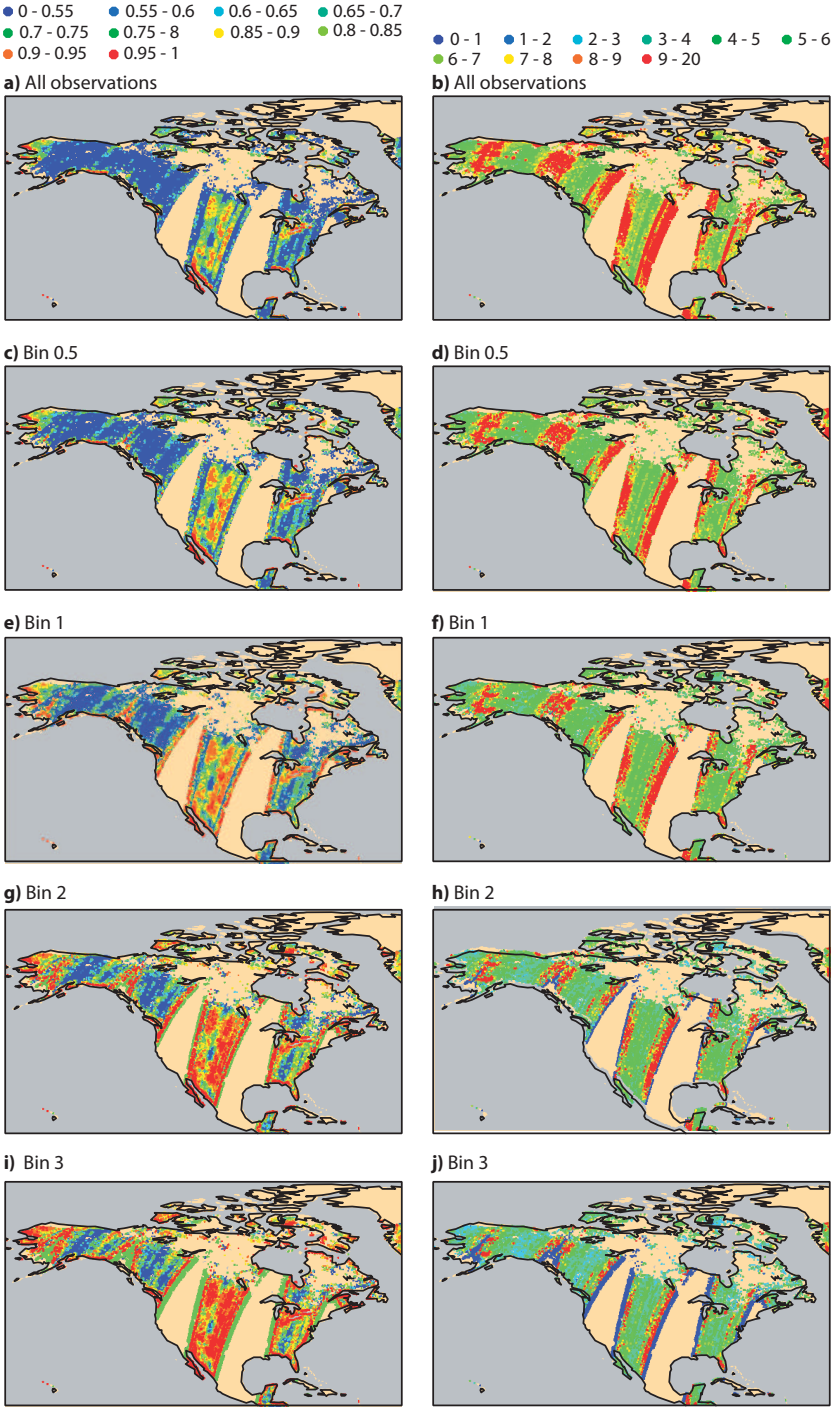

Figure 28: As in Fig. 26, but for North America and the XX pol.

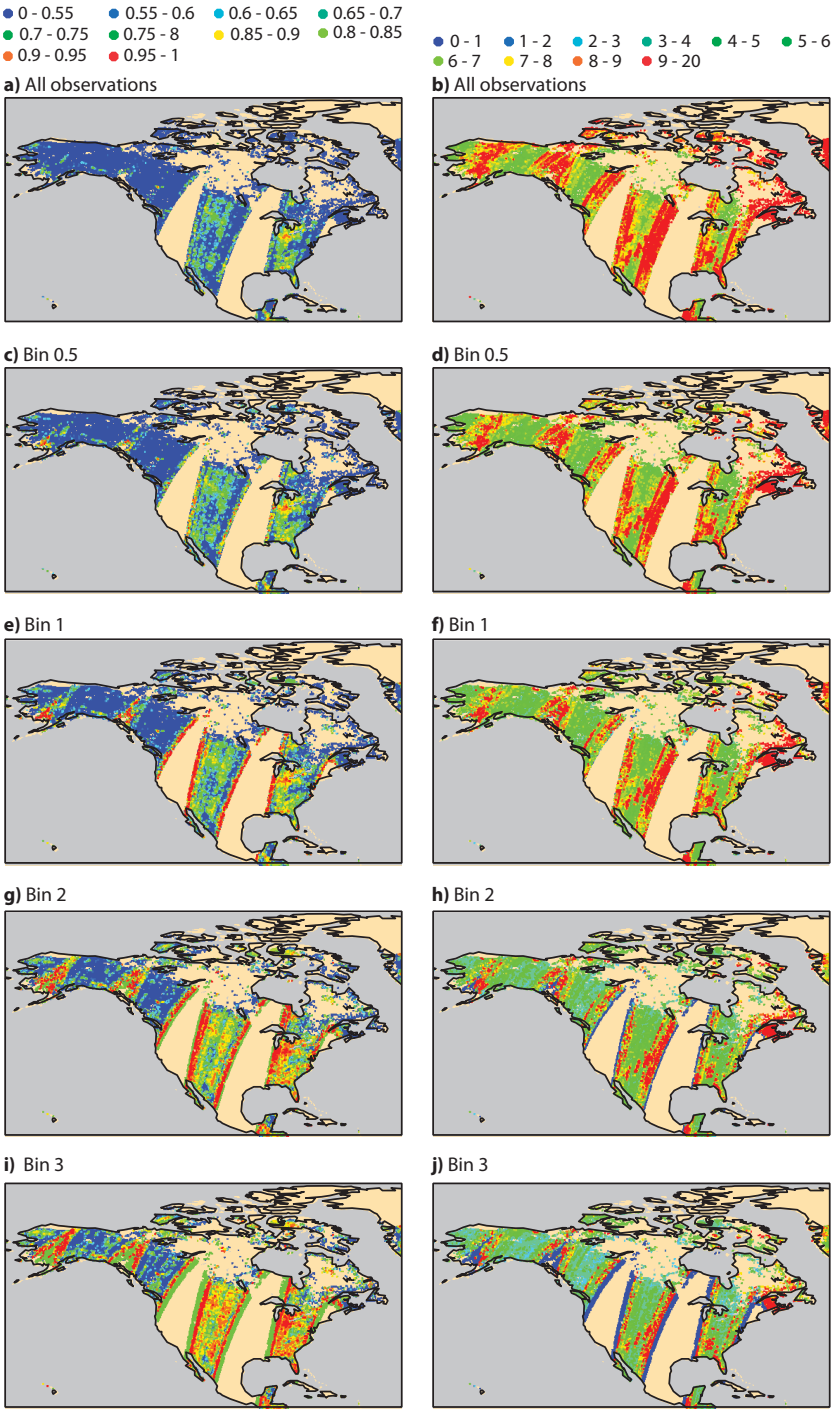

Figure 29: As in Fig. 26, but for North America and the YY pol.

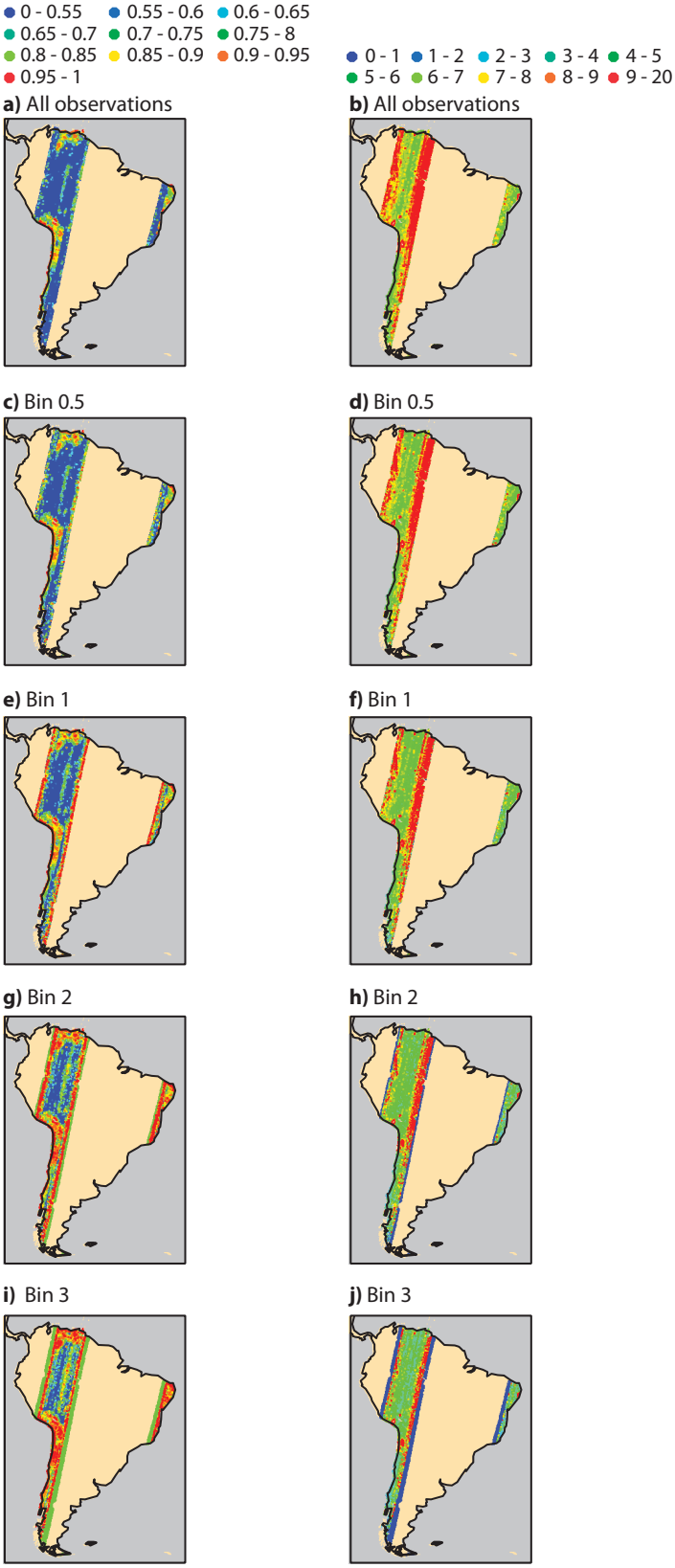

Figure 30: As in Fig. 26, but for South America and the XX pol.

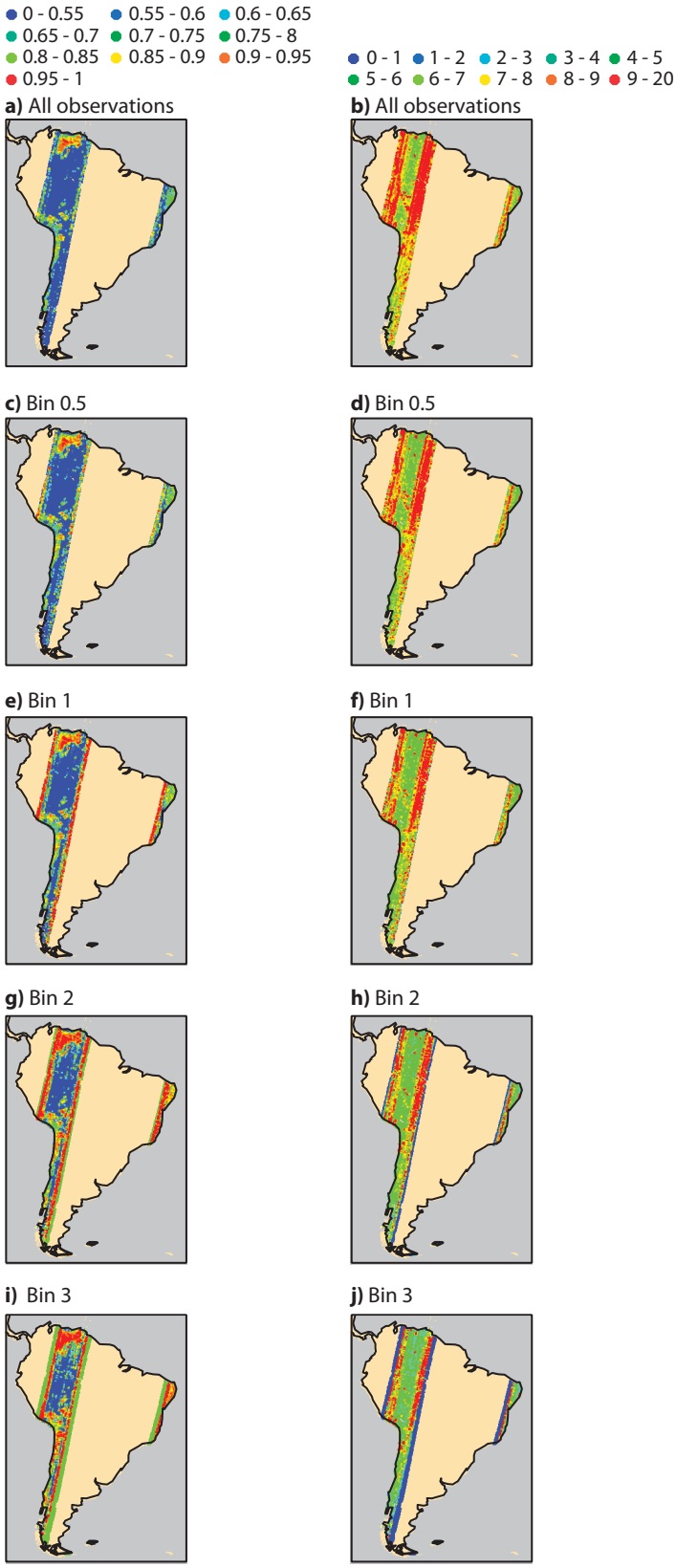

Figure 31: As in Fig. 26, but for South America and the YY pol.

8.2 June case

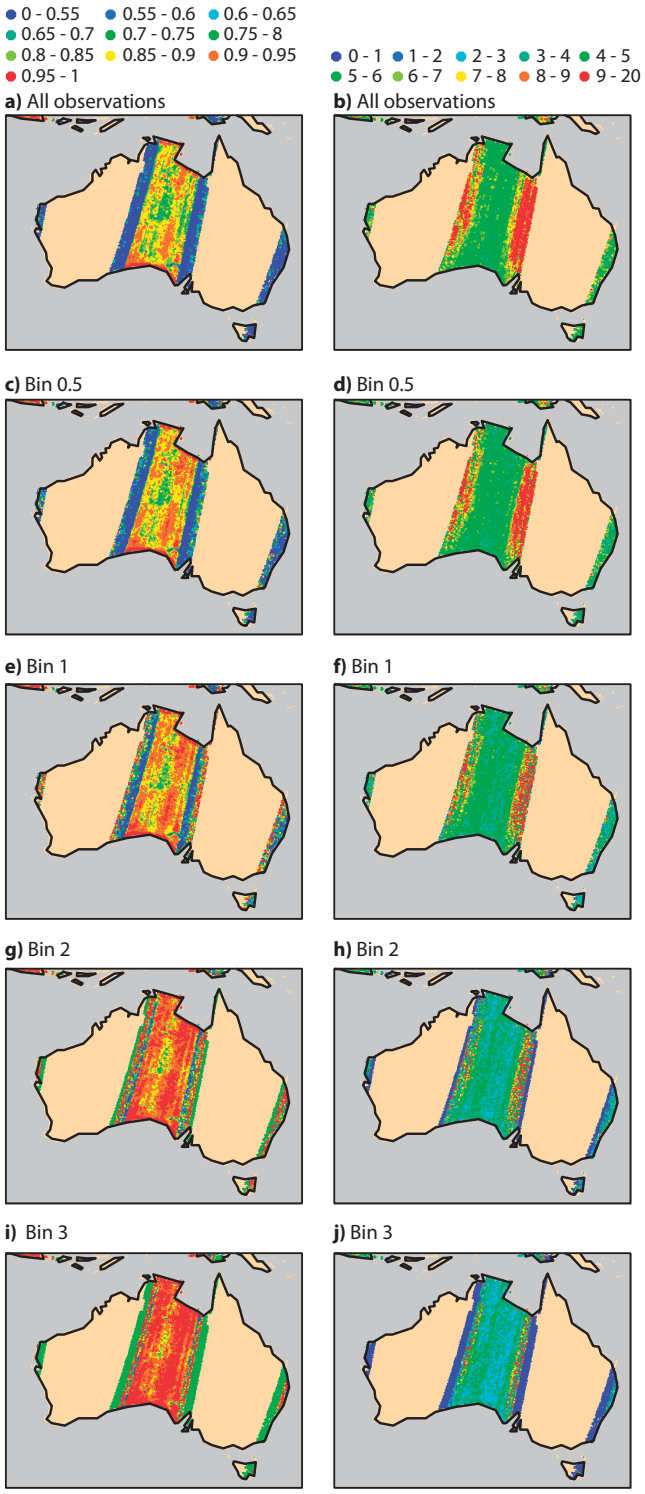

Figure 32: Coefficient of determination (r^2) and STD of the residues (in K) between the SMOS TB angular signature and its 2nd order polynomial regression model for the XX polarisation, the 1 June 2011.

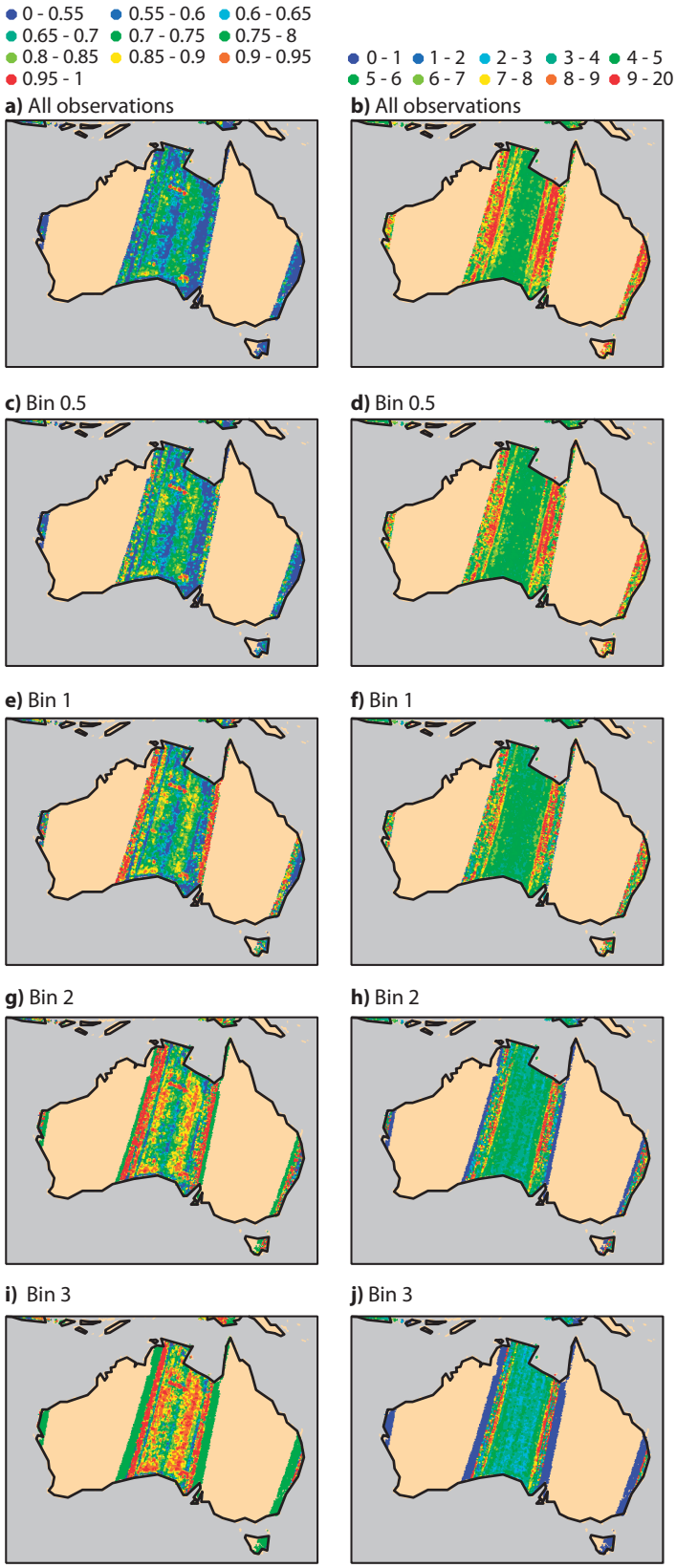

Figure 33: As in Fig. 32, but for YY pol.

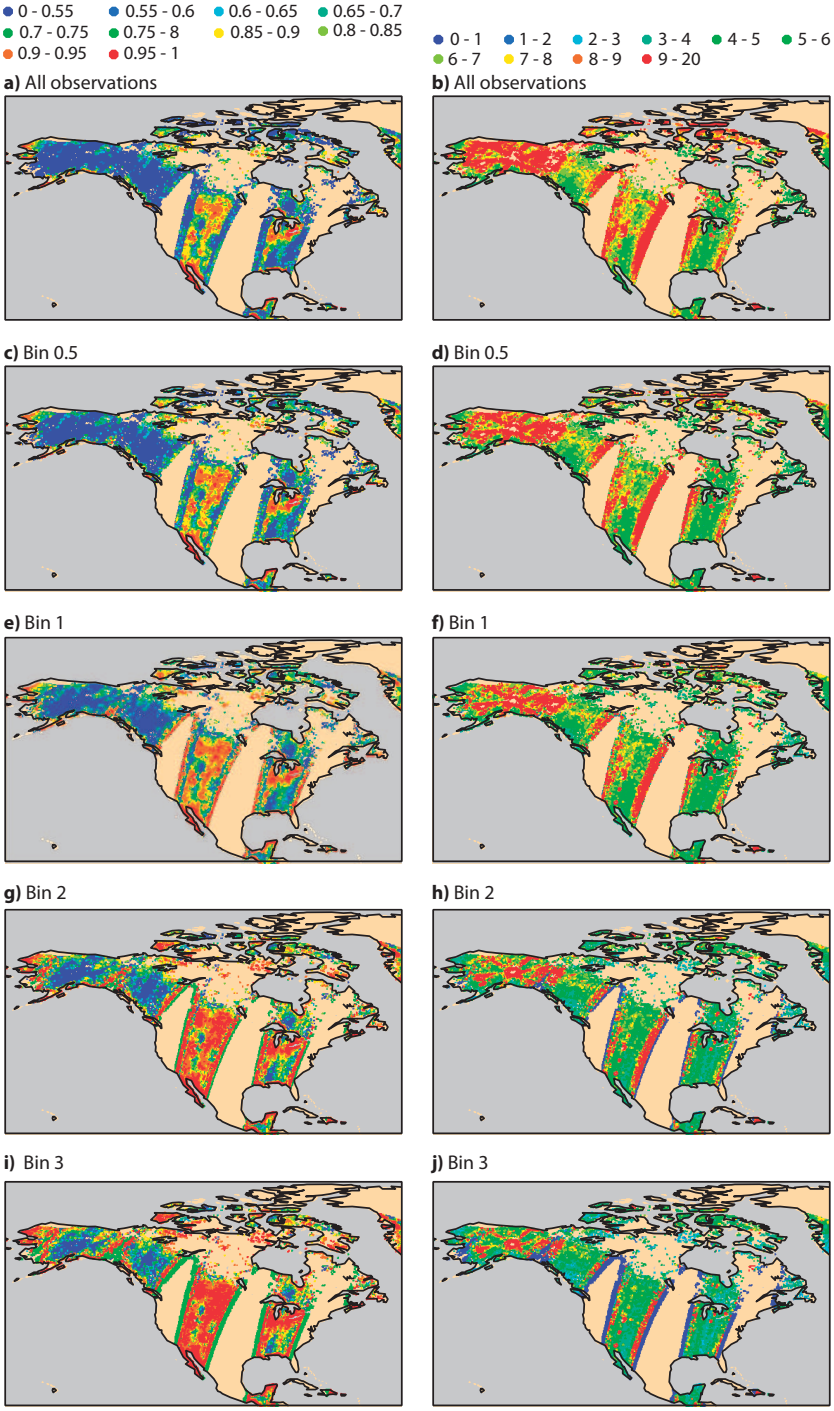

Figure 34: As in Fig. 32, but for North America and the XX pol.

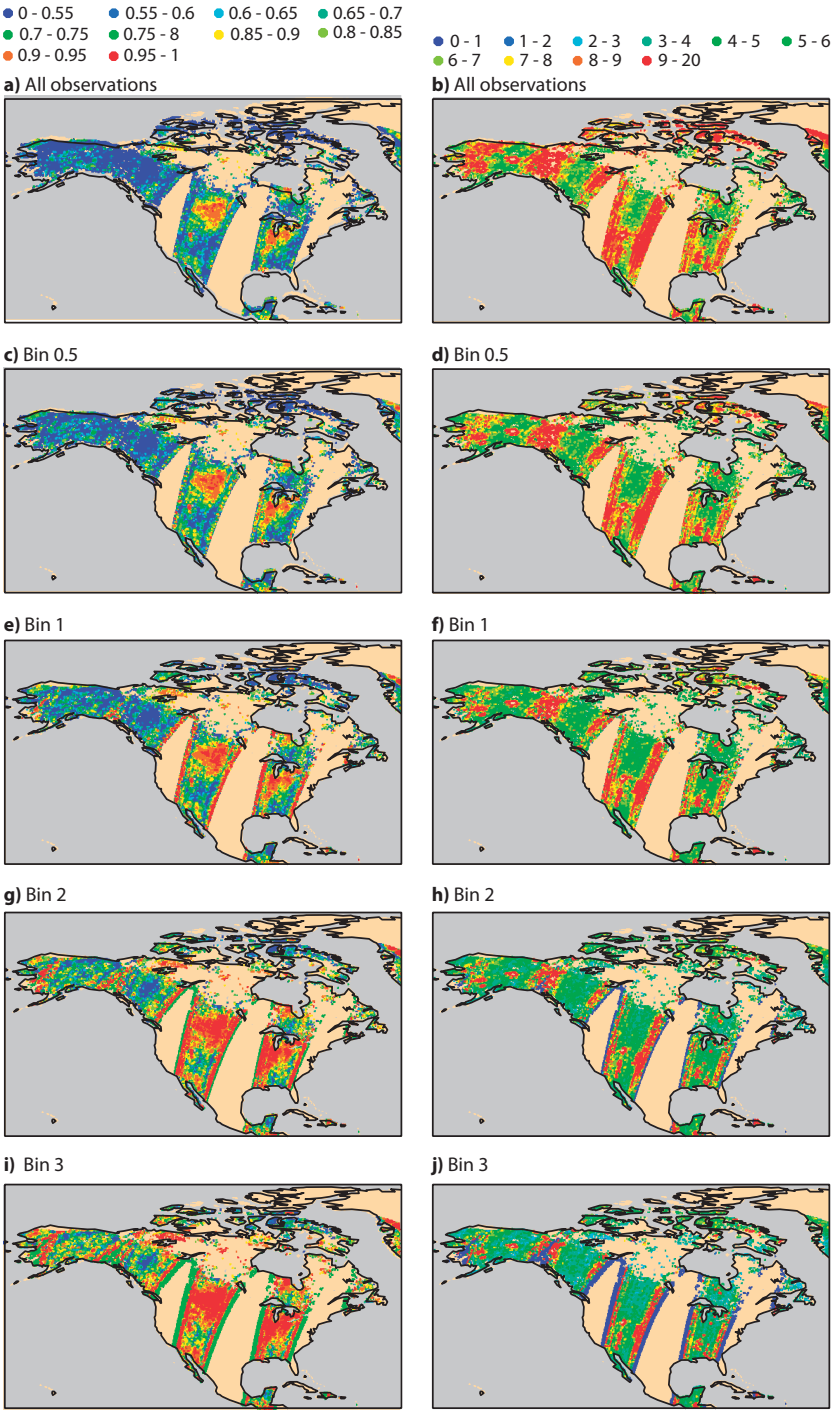

Figure 35: As in Fig. 32, but for North America and the YY pol.

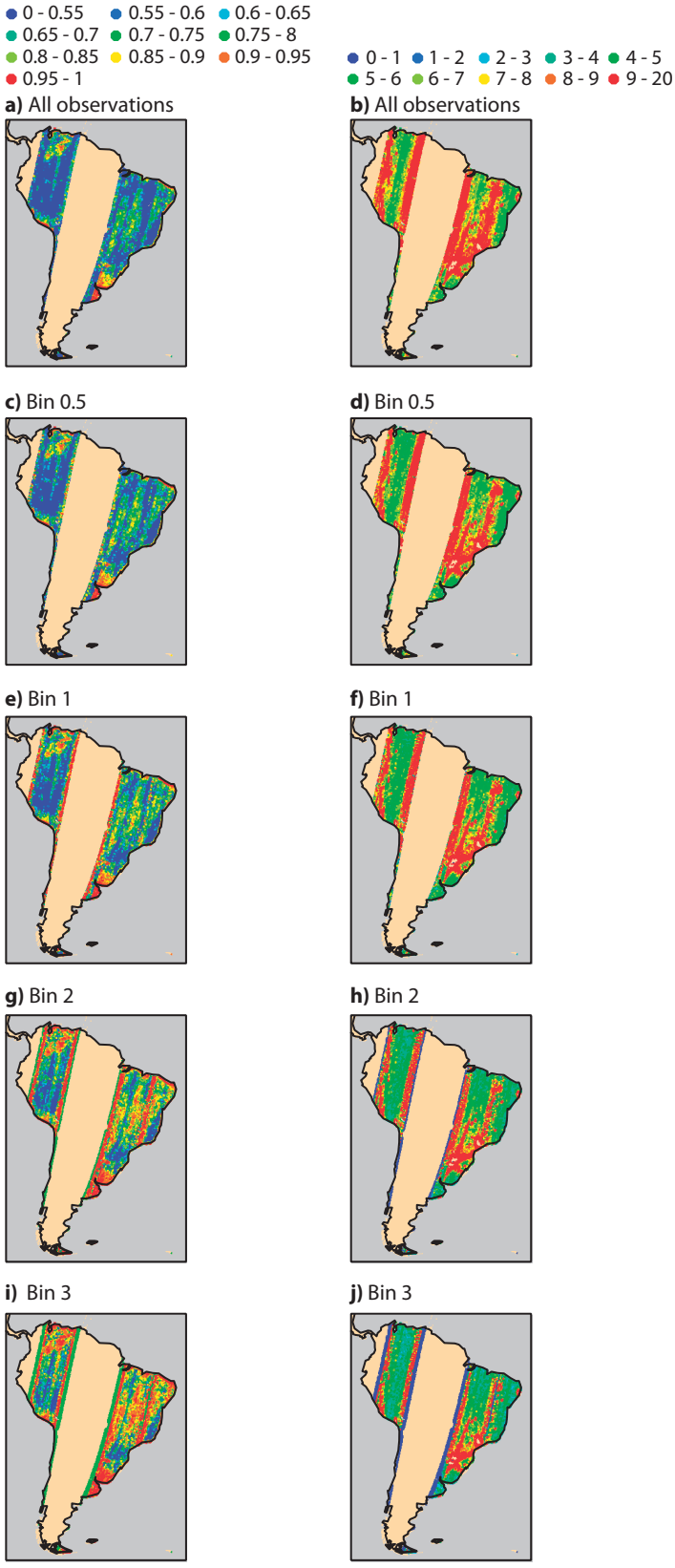

Figure 36: As in Fig. 32, but for South America and the XX pol.

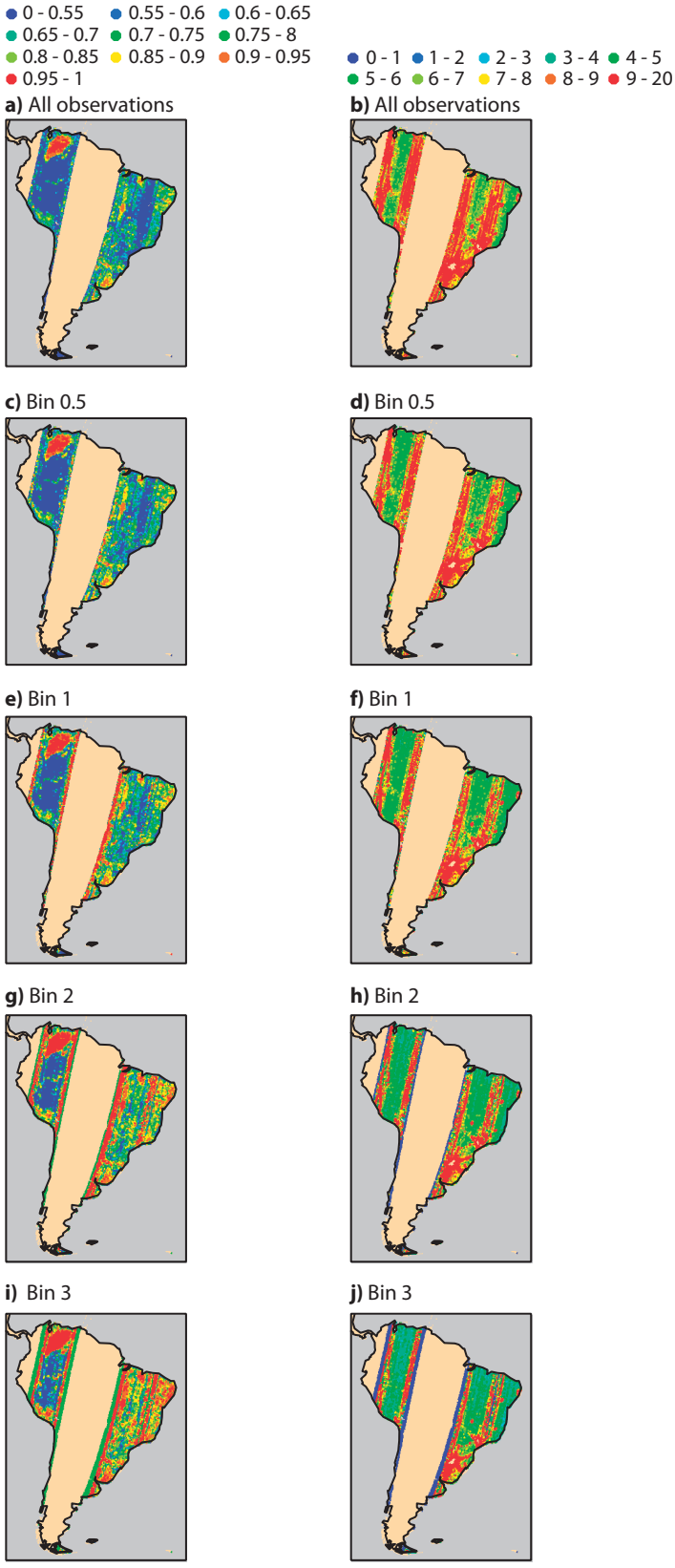

Figure 37: As in Fig. 32, but for South America and the YY pol.

9 Appendix 3: Soil texture types results

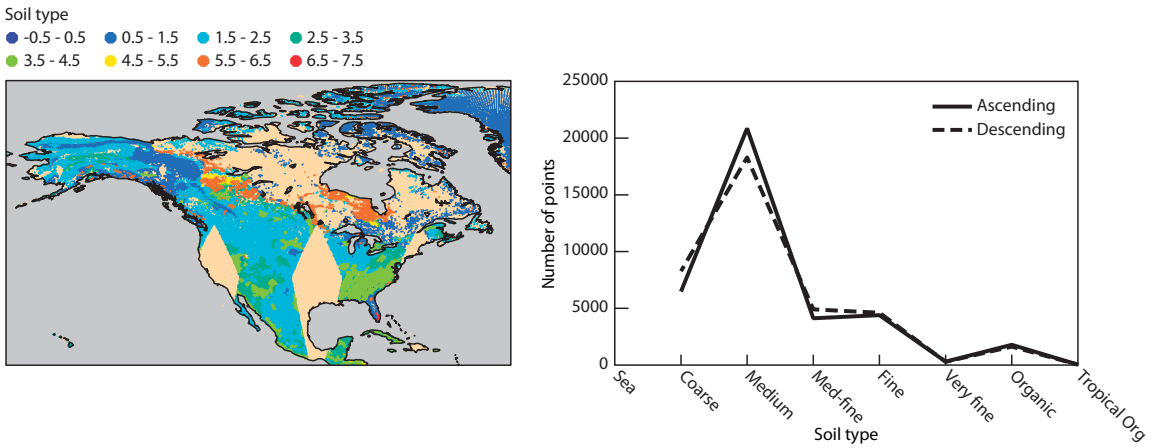

Figure 38: Soil type map and number of observations per soil type the 1 December 2010 in North America. Legend color texture code: 0=Sea, 1=Coarse, 2=Medium, 3=Medium-fine, 4=Fine, 5=Very fine, 6=Organique, 7=Tropical-organique.

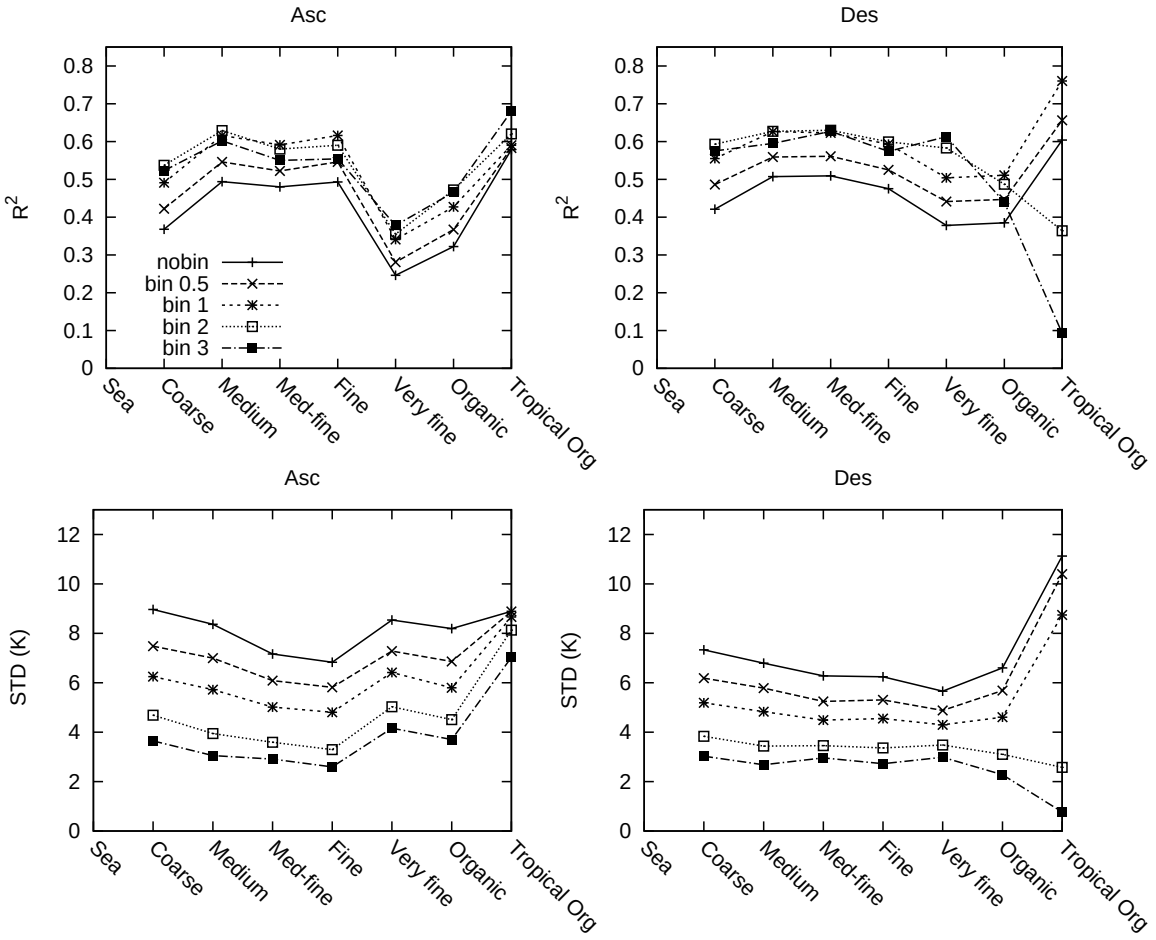

Figure 39: Mean r^2 and STD (in K) per type of soil texture class for North America, between SMOS T_B and its 2^{nd} order polynomial fit, for the XX polarisation, the 1 December 2010.

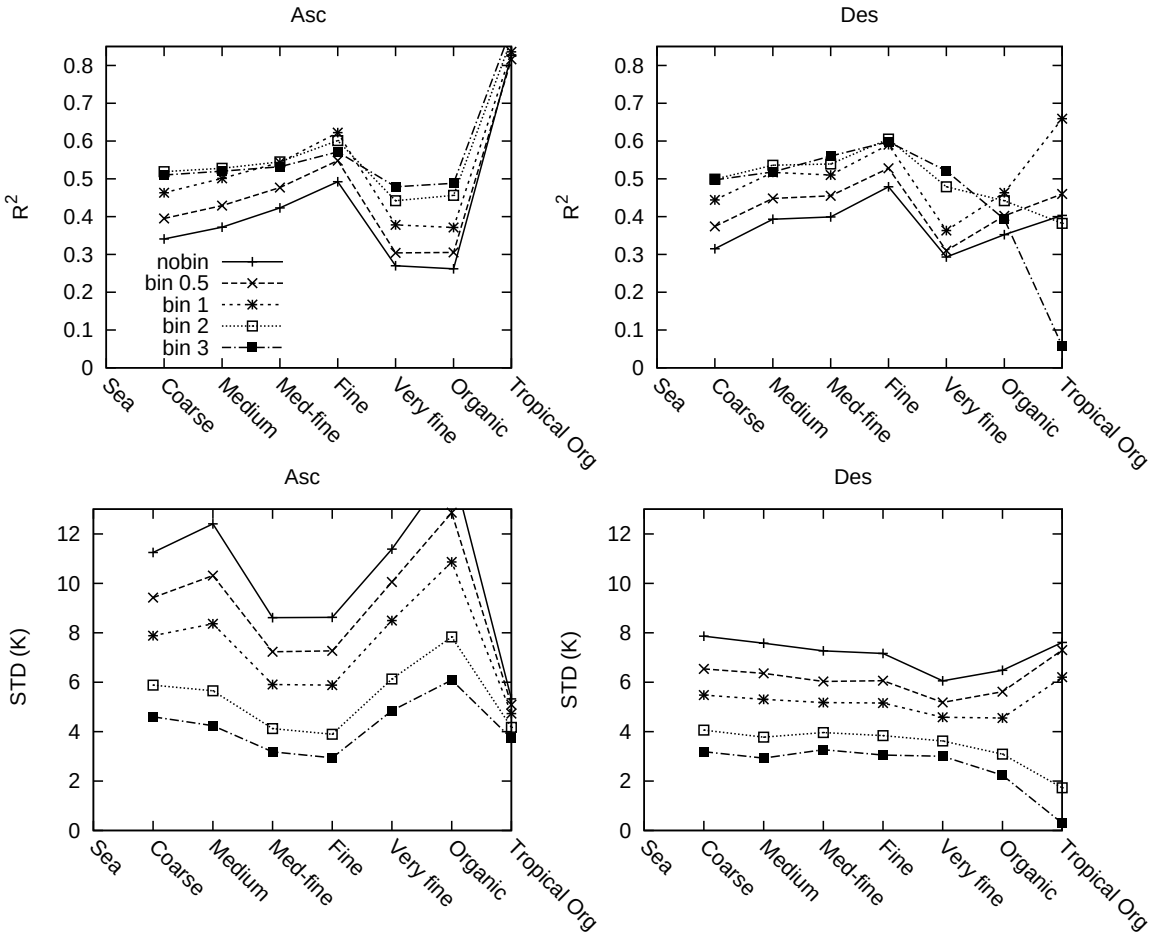

Figure 40: As in Fig. 39 but for the YY polarisation.

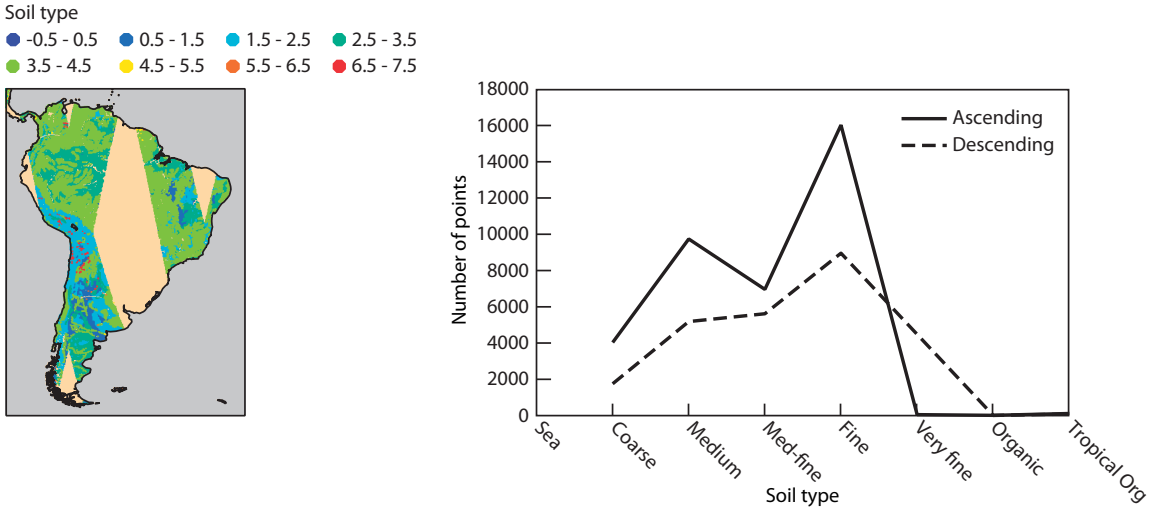

Figure 41: Soil type map and number of observations per soil type the 1 December 2010 in South America. Legend color texture code: 0=Sea, 1=Coarse, 2=Medium, 3=Medium-fine, 4=Fine, 5=Very fine, 6=Organique, 7=Tropical-organique.

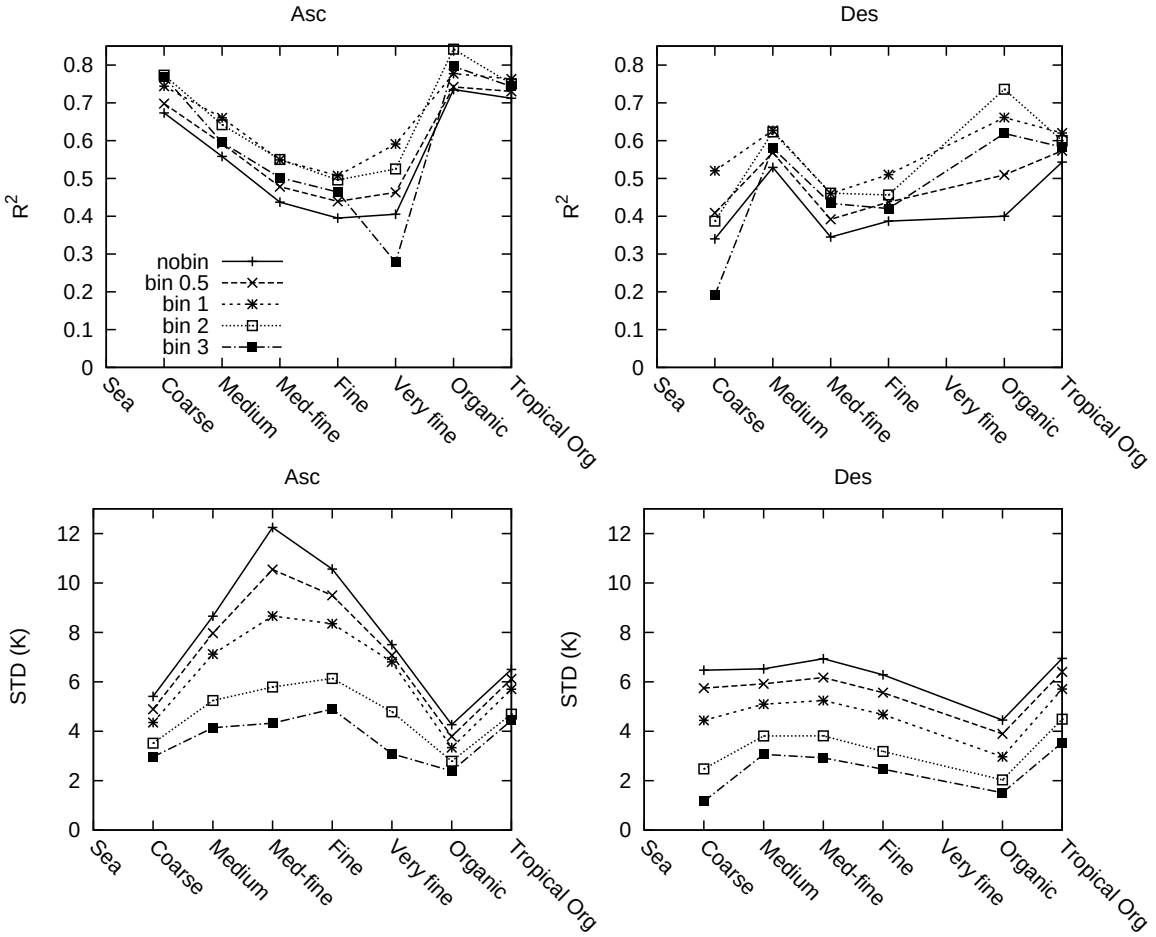

Figure 42: Mean r^2 and STD (in K) per type of soil texture class for South America, between SMOS T_B and its 2^{nd} order polynomial fit, for the XX polarisation, the 1 December 2010.

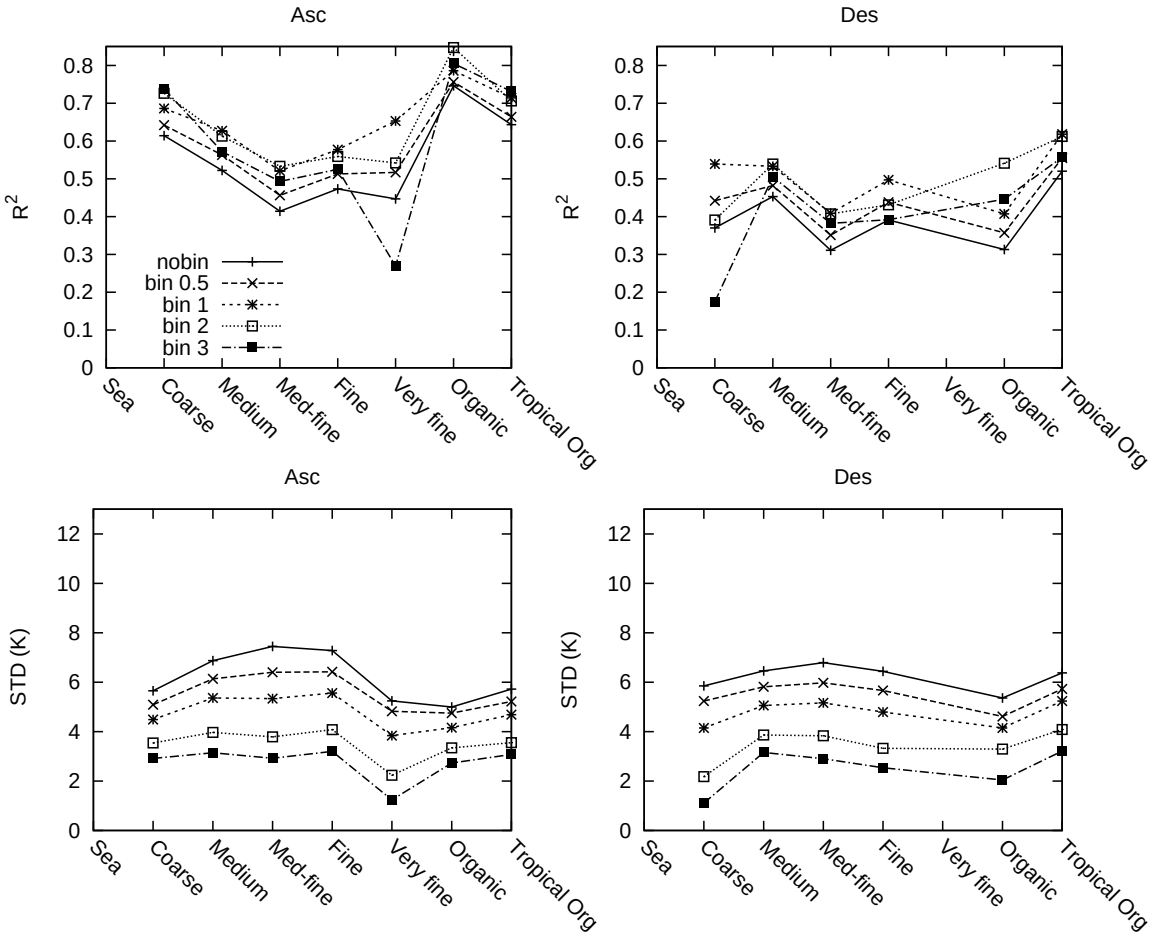

Figure 43: As in Fig. 42 but for the YY polarisation.

10 Appendix 4: Vegetation cover types results

10.1 December 2010

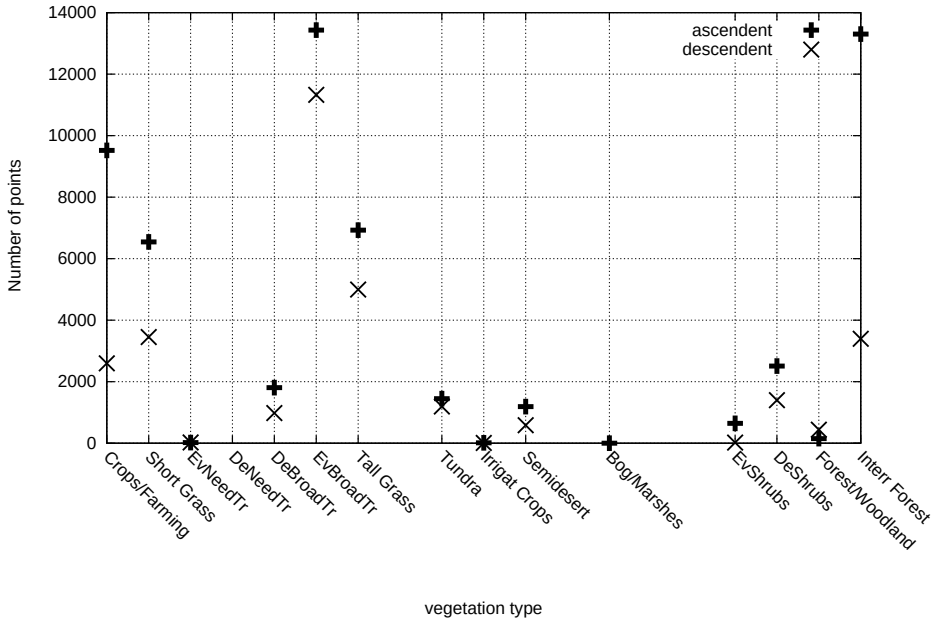

Figure 44: Number of measurements per vegetation biome classe the 1 December 2010 in South America.

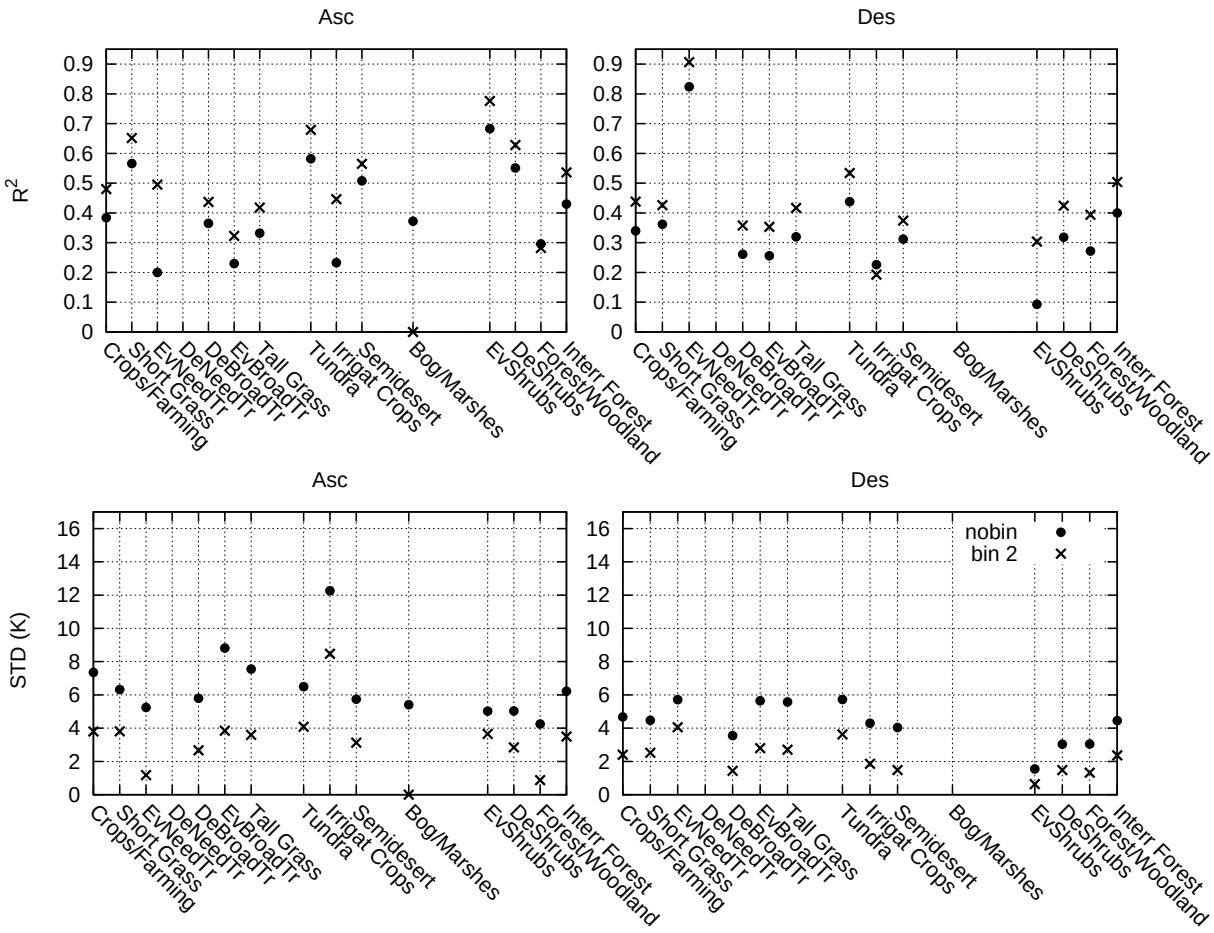

Figure 45: Mean r^2 and STD (in K) per type of vegetation classe in South America, between SMOS T_B and its 2^{nd} order polynomial fit, for the XX polarisation the 1 December 2010.

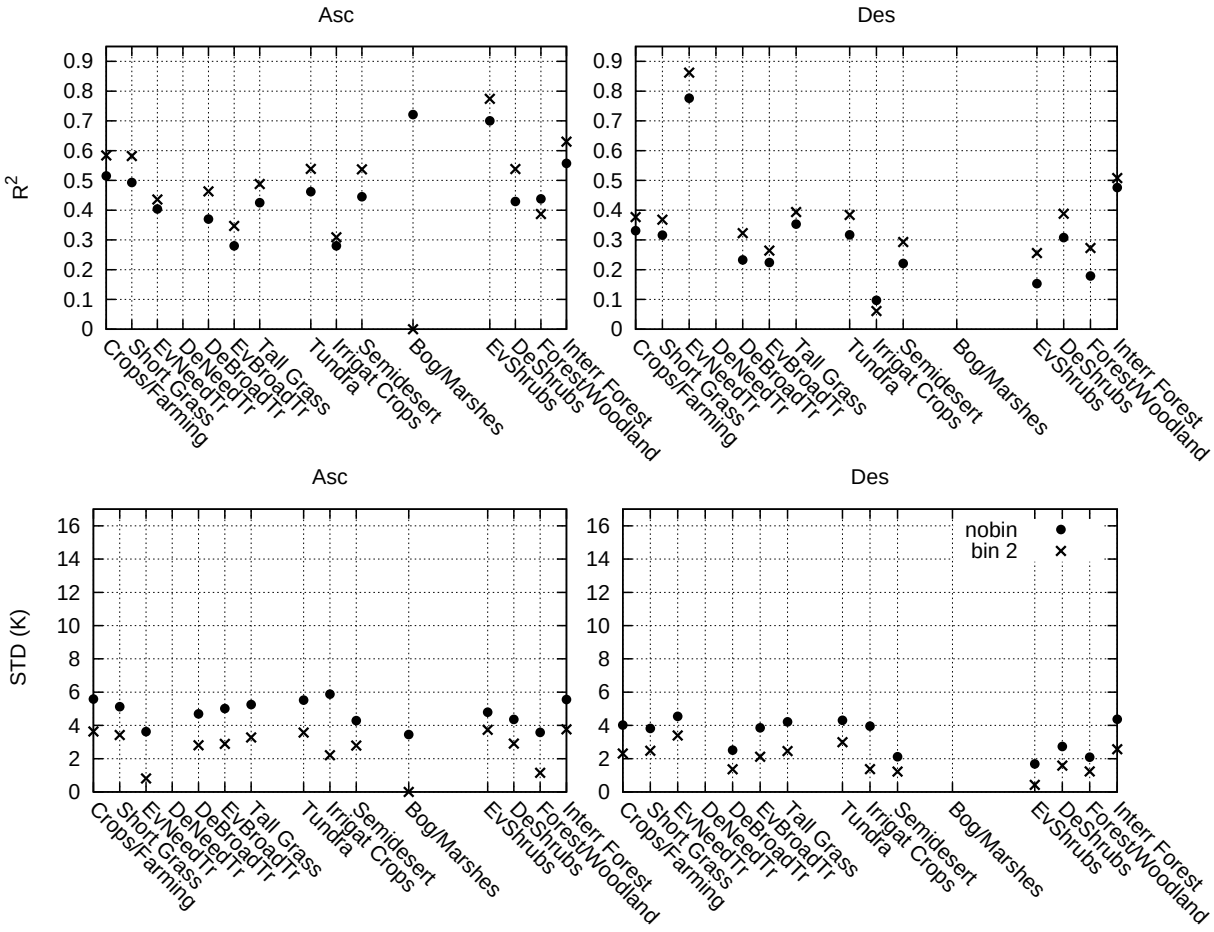

Figure 46: As in Fig. 45 but for the YY polarisation.

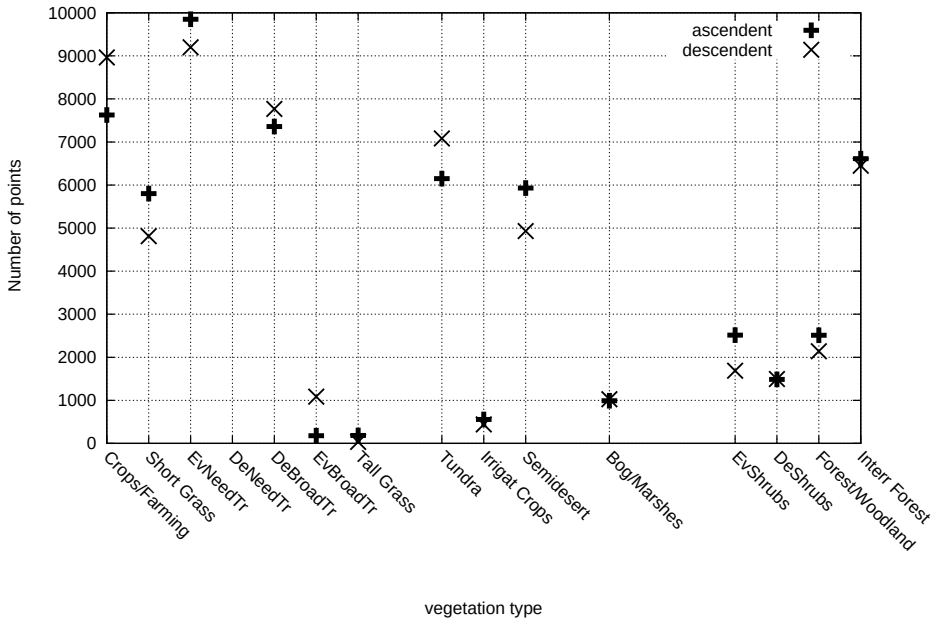

Figure 47: Number of measurements per vegetation biome classe the 1 December 2010 in North America.

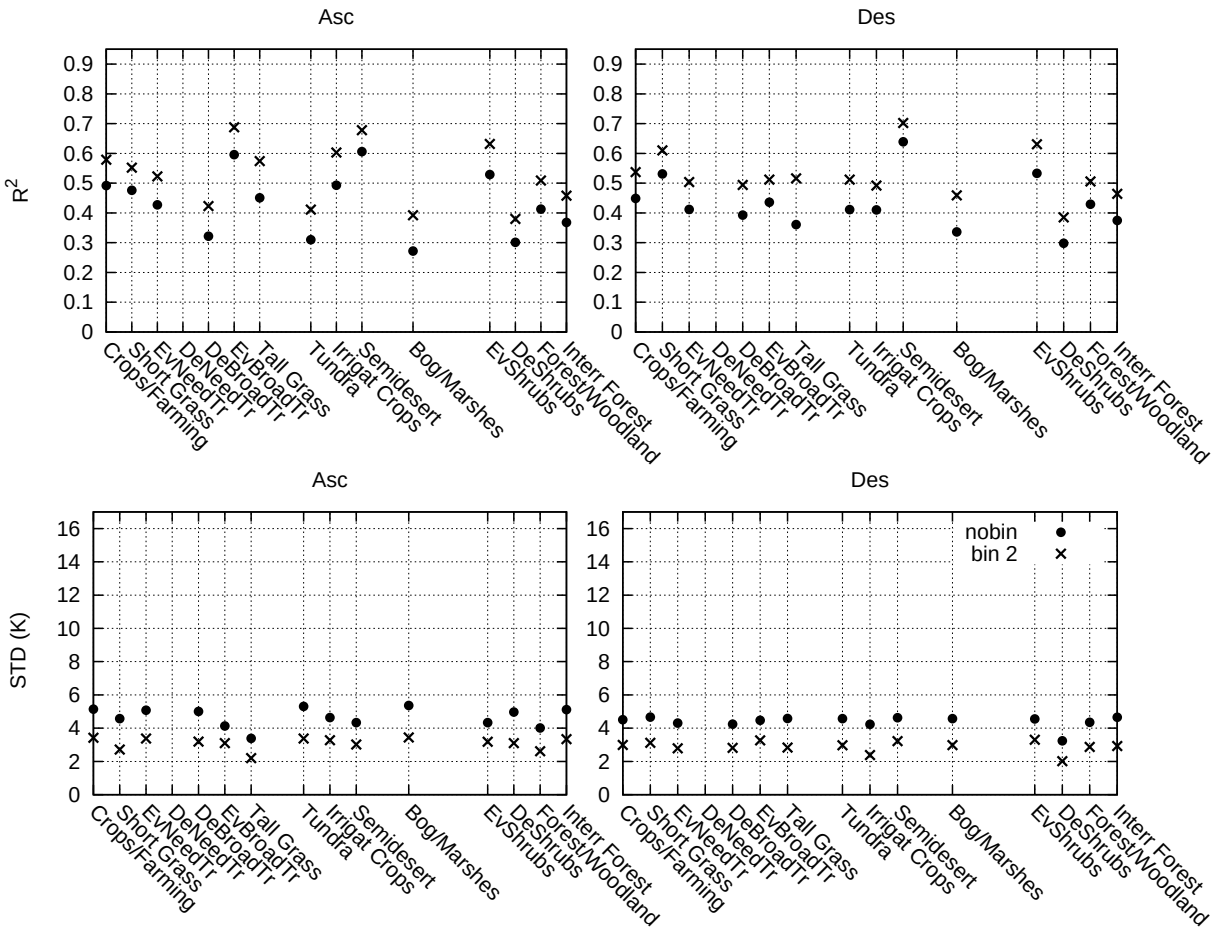

Figure 48: Mean r^2 and STD (in K) per type of vegetation classe in North America, between SMOS T_B and its 2^{nd} order polynomial fit, for the XX polarisation the 1 December 2010.

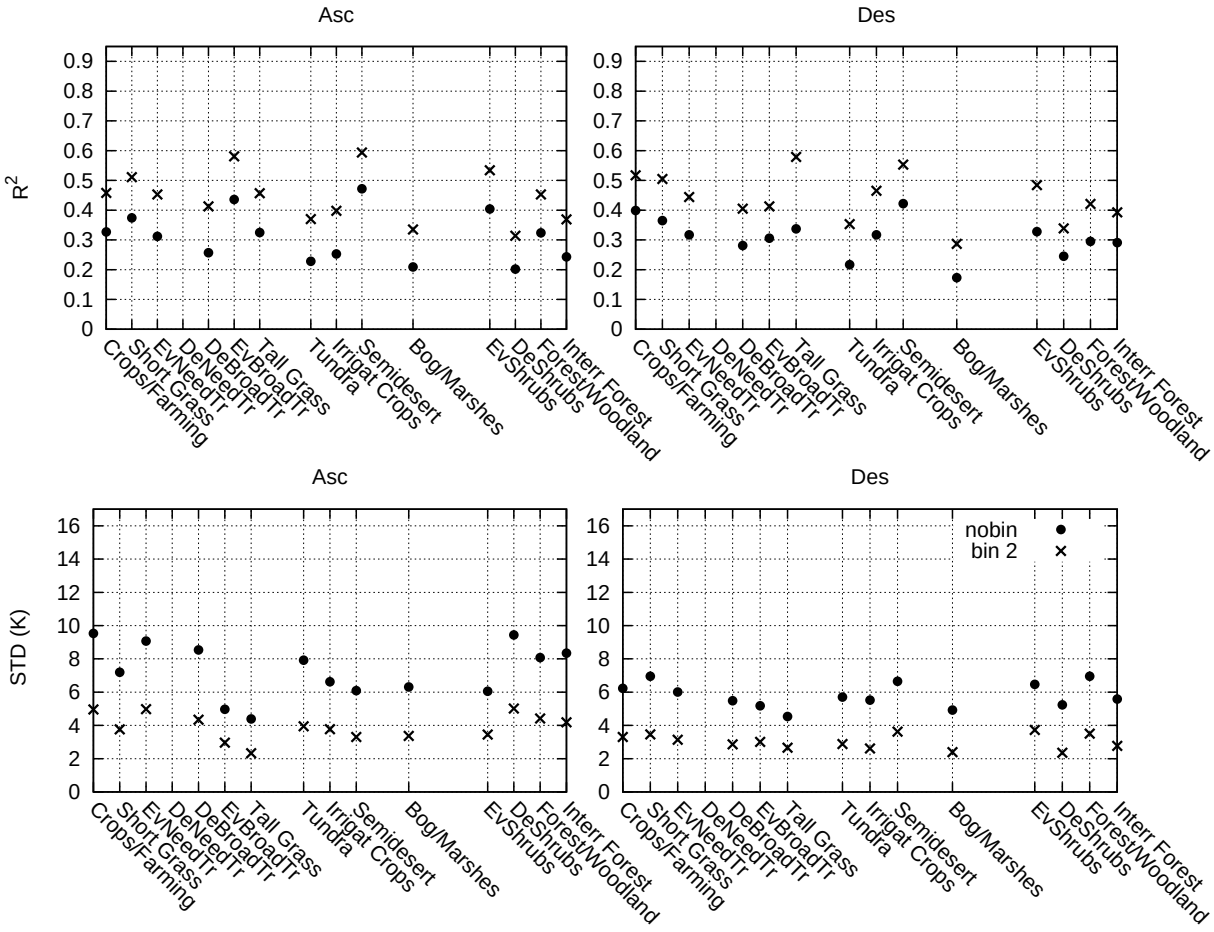

Figure 49: As in Fig. 48 but for the YY polarisation.

10.2 June 2011

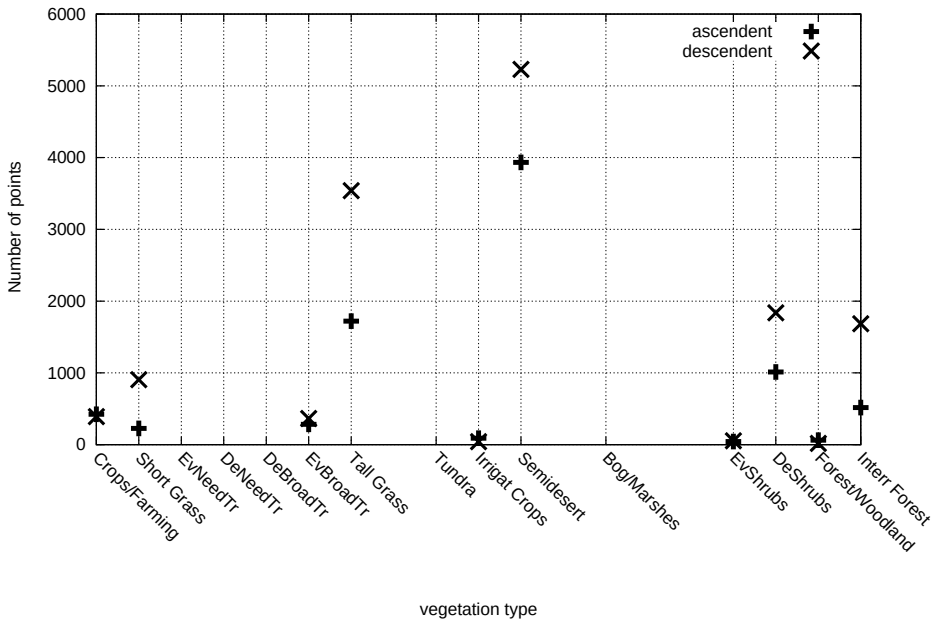

Figure 50: Number of measurements per vegetation biome classe the 1 June 2011 in Australia.

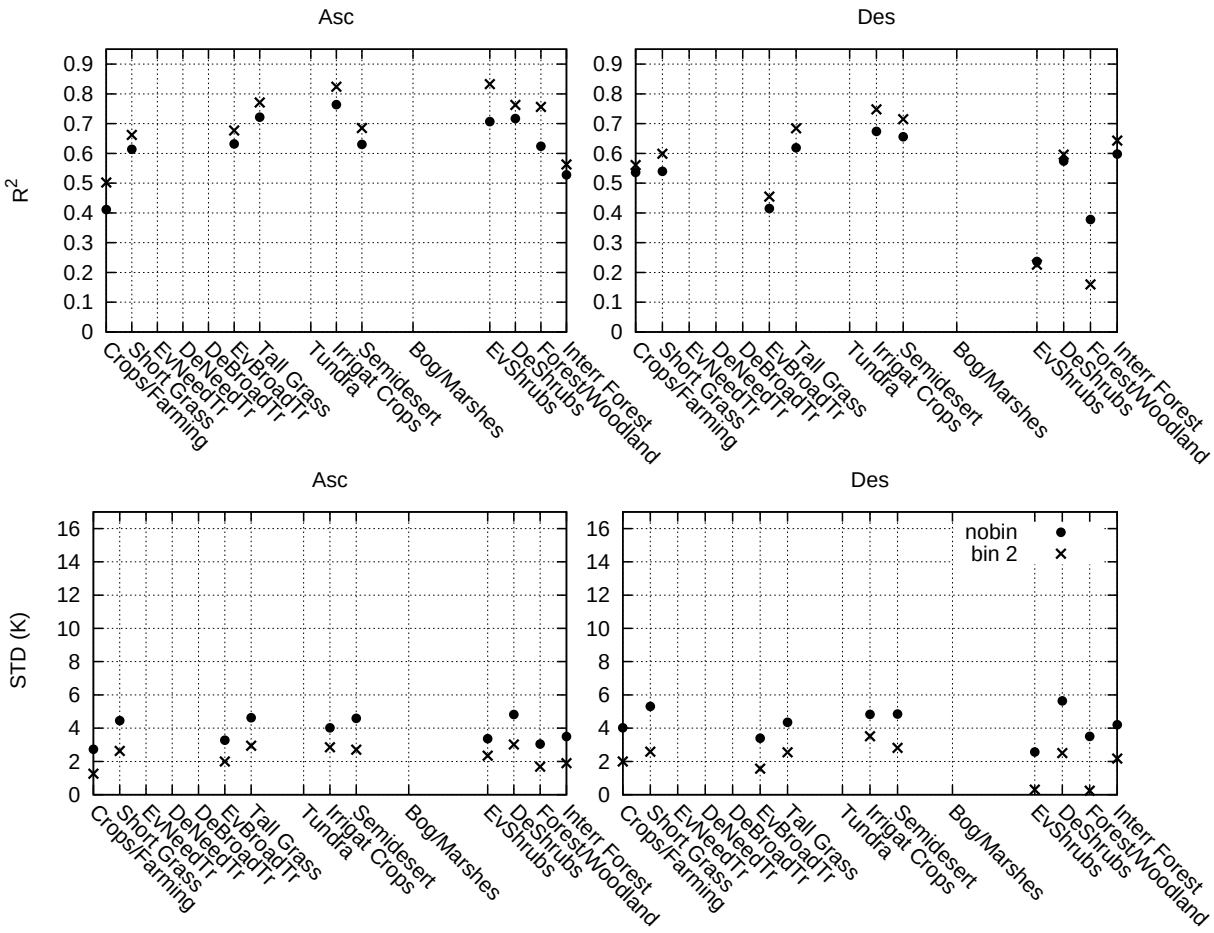

Figure 51: Mean r^2 and STD (in K) per type of vegetation classe in Australia, between SMOS T_B and its 2^{nd} order polynomial fit, for the XX polarisation the 1 June 2011.

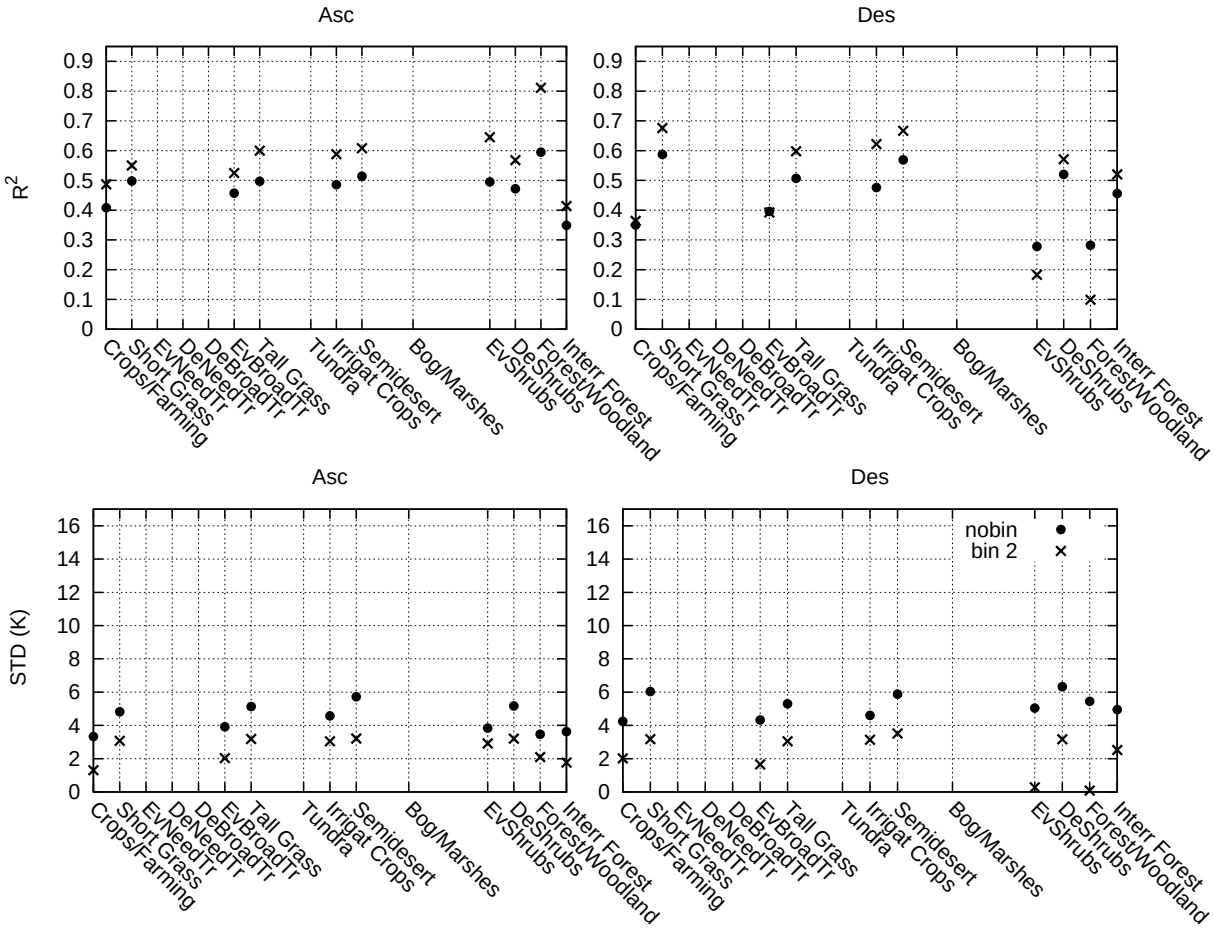

Figure 52: As in Fig. 51 but for the YY polarisation.

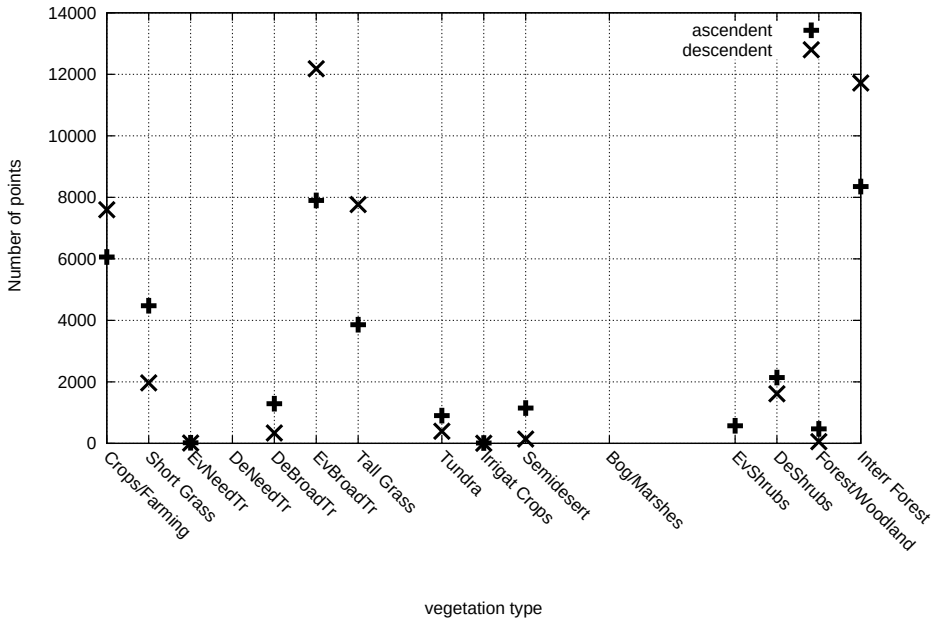

Figure 53: Number of measurements per vegetation biome classe the 1 June 2011 in South America.

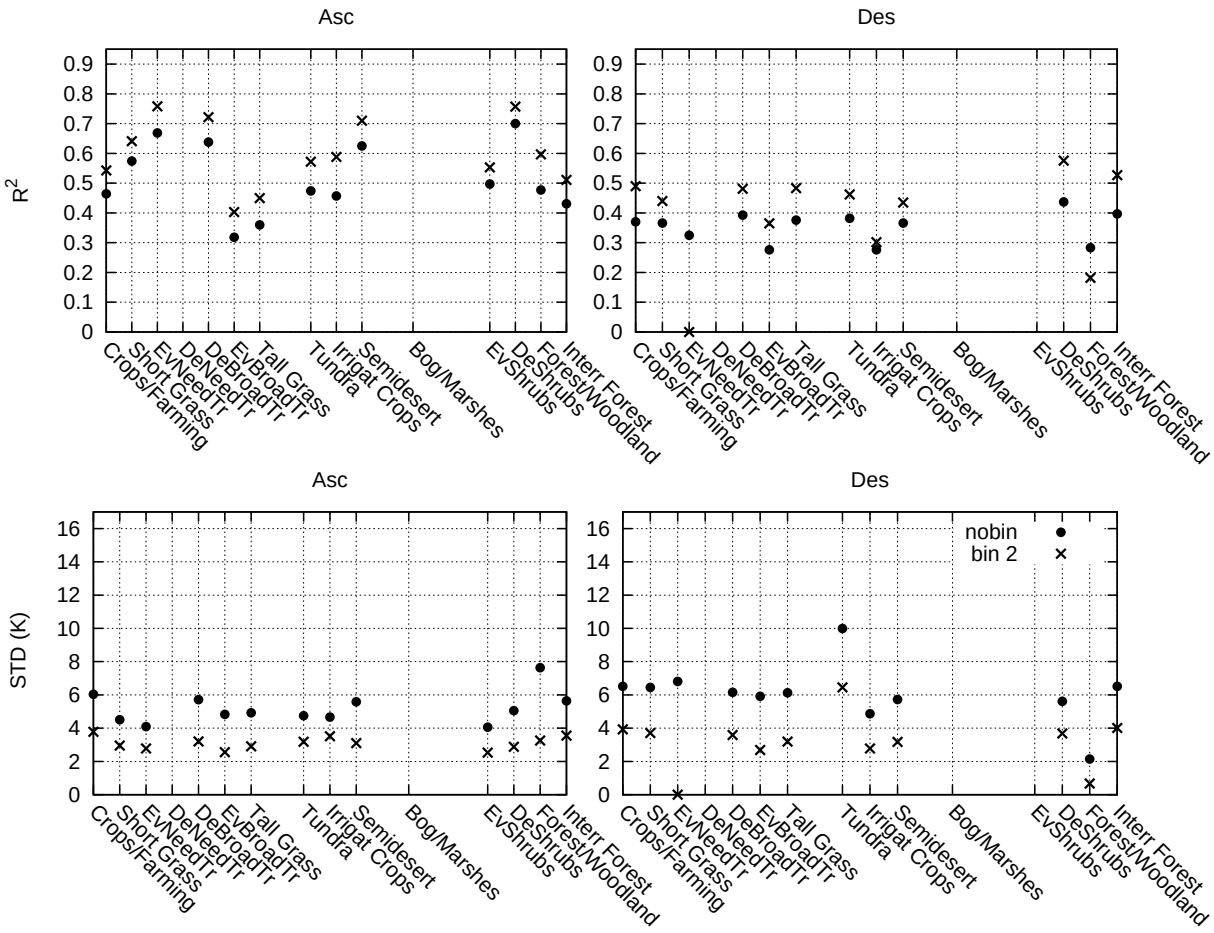

Figure 54: Mean r^2 and STD (in K) per type of vegetation classe in South America, between SMOS T_B and its 2^{nd} order polynomial fit, for the XX polarisation the 1 June 2011.

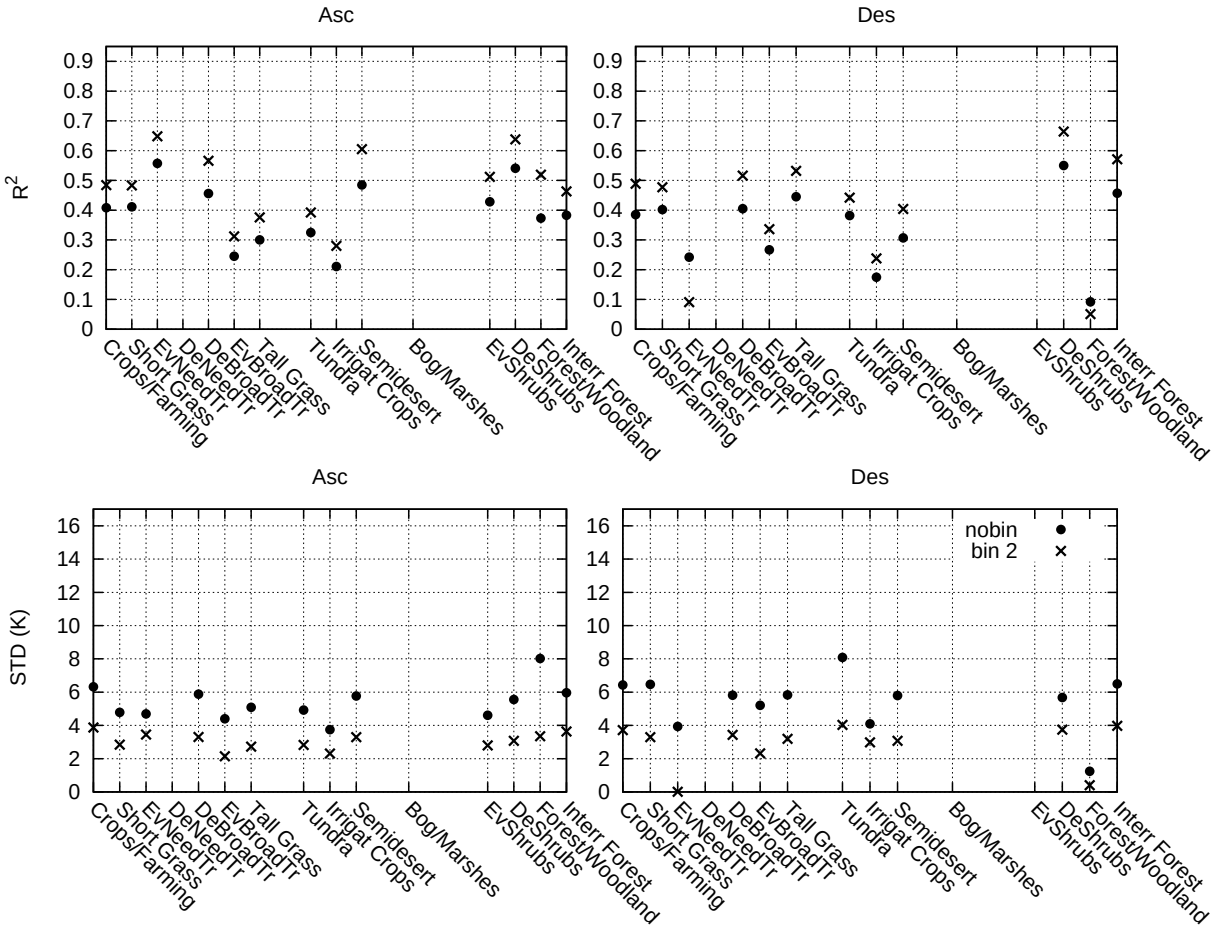

Figure 55: As in Fig. 54 but for the YY polarisation.

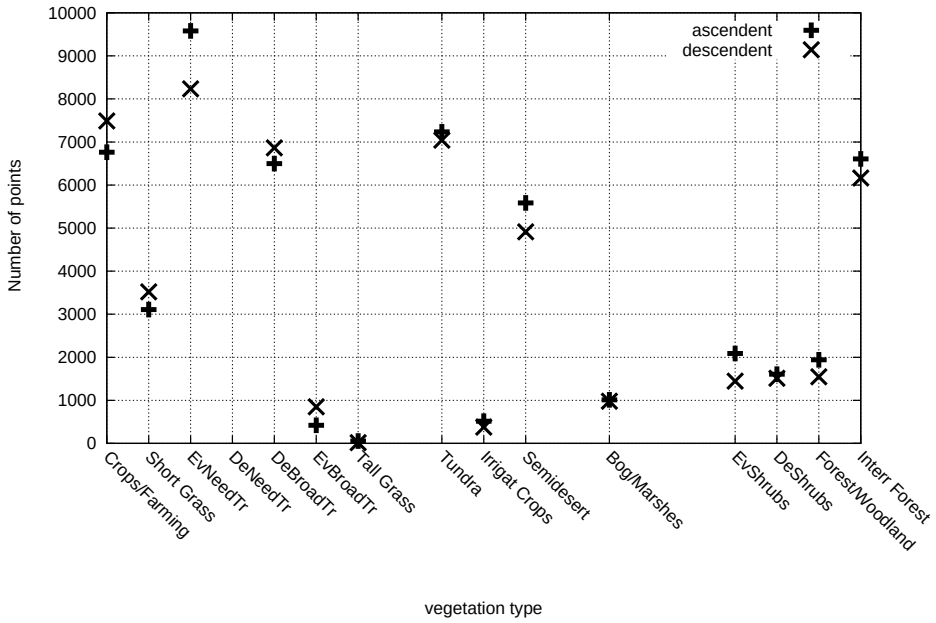

Figure 56: Number of measurements per vegetation biome classe the 1 June 2011 in North America.

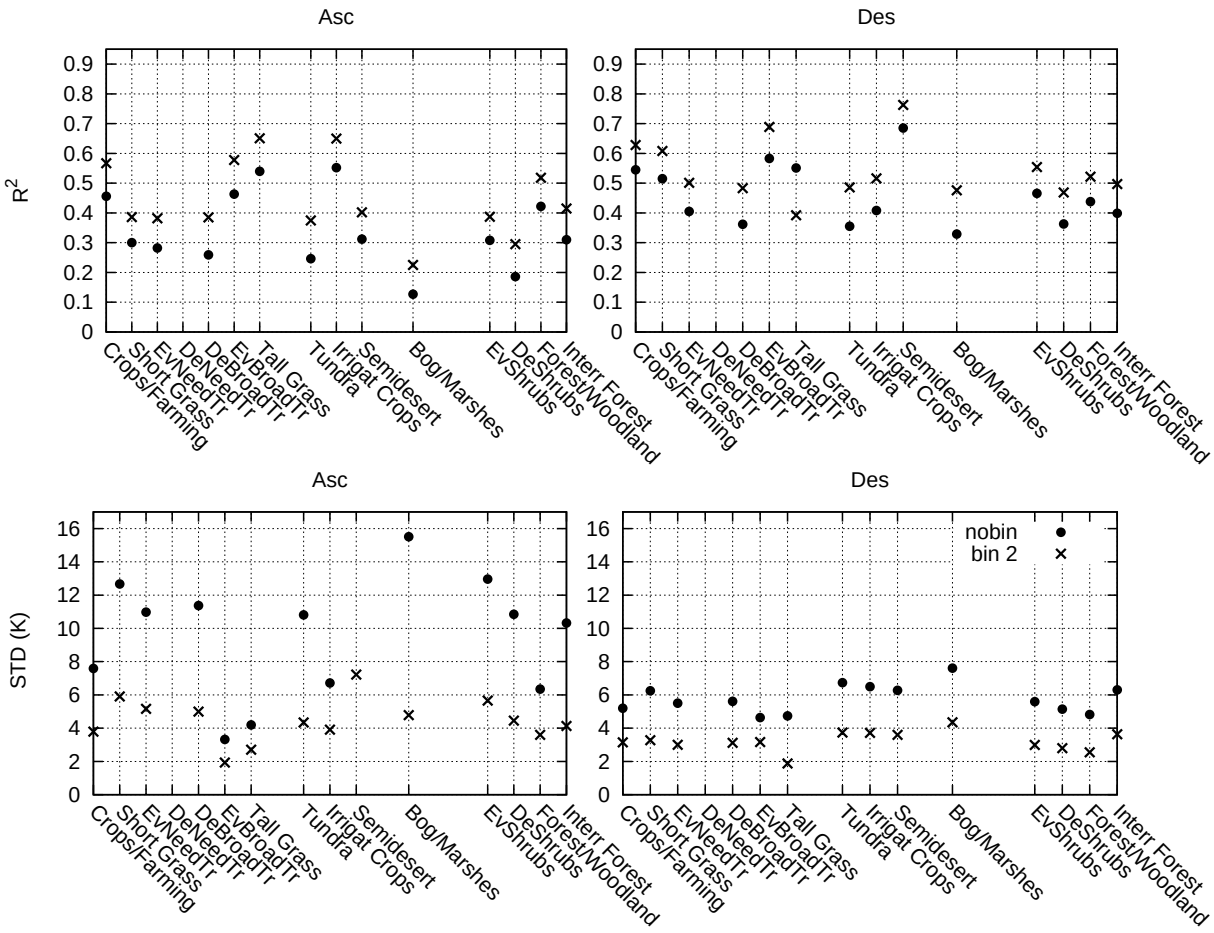

Figure 57: Mean r^2 and STD (in K) per type of vegetation classe in North America, between SMOS T_B and its 2^{nd} order polynomial fit, for the XX polarisation the 1 June 2011.

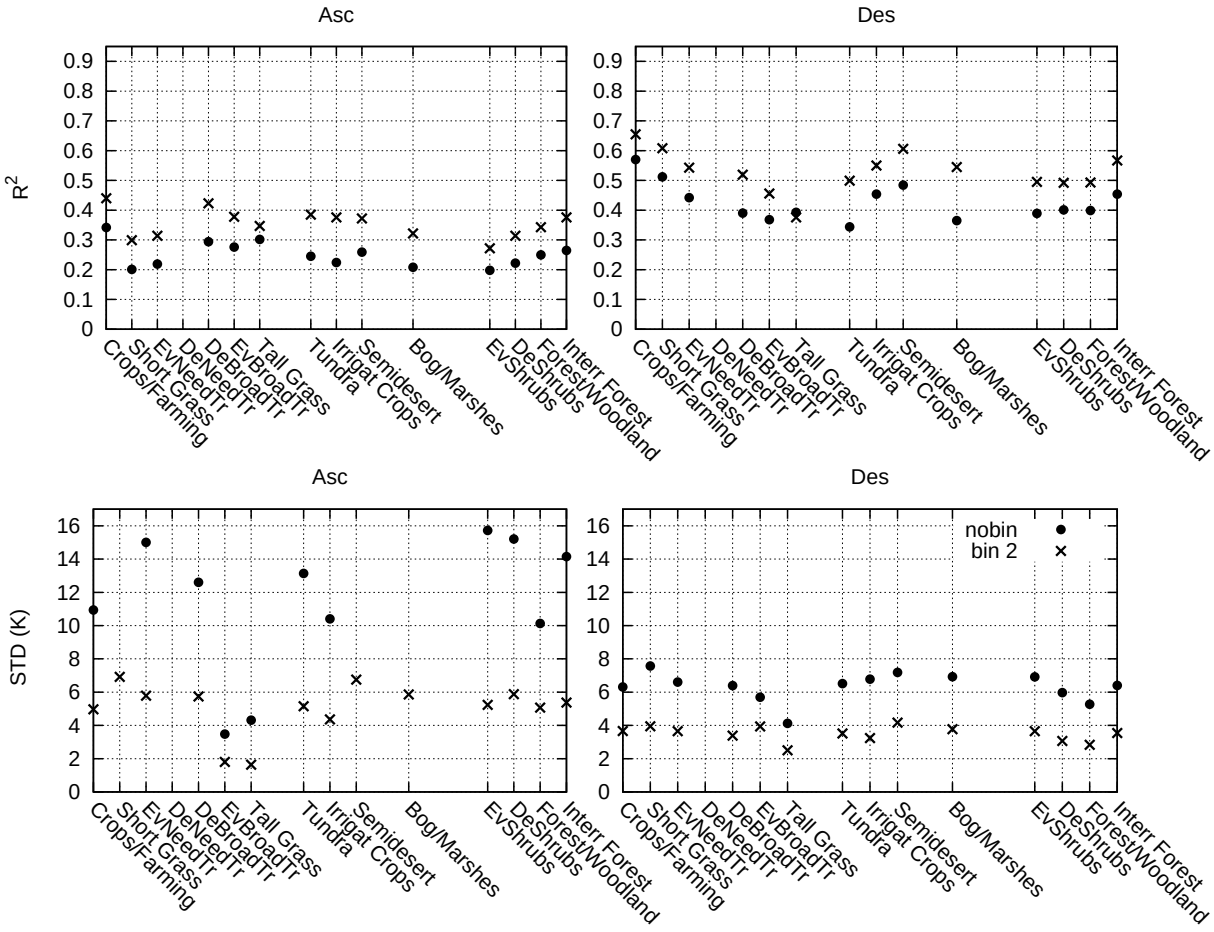

Figure 58: As in Fig. 57 but for the YY polarisation.

11 References

References

[1] G. Balsamo, P. Viterbo, A. Beljaars, B. van den Hurk, M. Hirschi, A.K. Betts, and K. Scipal. A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, 10:623–643, 2009. doi: 10.1175/2008JHM1068.1.

[2] J. Font, A. Camps, and J. Ballabrera-Poy. Microwave aperture synthesis radiometry: Setting the path for (operational) sea salinity measurement from space. *Remote Sensing of European Seas*, pages 223–238, 2008.

[3] Y. Kerr, P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger. Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission. *IEEE Trans. Geosc. Remote Sens.*, 39(8):1729–1735, 2001.

[4] T.R. Loveland, B.C. Reed, J.F. Brown, D.O. Ohlen, Z. Zhu, L. Youing, and J.W. Merchant. Development of a global land cover characteristics database and IGB6 DISCover from the 1 km AVHRR data. *International Journal of Remote Sensing*, 21:1303–1330, 2000.

[5] G. Macelloni, M. Brogioni, P. Pampaloni, A. Cagnati, and M.R. Drinkwater. Domex 2004: An experimental campaign at dome-c antarctica for the calibration of spaceborne low-frequency microwave radiometers. *IEEE Trans. Geosc. Remote Sens.*, 44(10):2642–2653, 2006.

[6] J. Randa, J. Lahtinen, A. Camps, A. Gasiewski, M. Hallikainen, V.D. Leine, M. Martin-Neira, J. Piepmeier, P. Rosenkranz, C. Ruf, J. Shiue, and N. Skou. Recommended terminology for microwave radiometry. Technical report, National Institute of Standards and Technology, U.S. Department of Commerce, 2008. Tech. Note 151.

[7] J.M. Sabater, P. de Rosnay, and A. Fouilloux. Milestone 2 Technical Note, Part I:operational preprocessing chain, Part II:collocation software development, Part III: Offline monitoring suite. Technical report, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, September 2010.

[8] J.M. Sabater, T. Wilhemsson, and P. de Rosnay. Tech note- Part II- WP1200: Report on Data Thinning. Technical report, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, August 2011.

[9] M.J. Schervish. P values: What they are and what they are not. *The American Statistician*, 50:203–206, 1996. doi:10.2307/2684655.

[10] F. Torres, I. Corbella, A. Camps, N. Duffo, and M. Vall-llossera. Error budget map to SRD (System Requirements Document) PRS. Project: Image Validation Support and SEPS Development, Validation and Delivery for SMOS PLM, Phase C/D, ESA Doc. Ref. SO-TN-UPCPLM-0007, version 7.0. Technical report, 2005.