esa ECMWF

ESA Contract Report

ESA contract 4000127267/19/NL/AF

Contract Report to the European Space Agency

GRDS test-bed plan

Authors: Pete Weston, Patricia de Rosnay and Stephen English
Contract officer: Antonio Martellucci

November 2020

Series: ECMWF - ESA Contract Report

A full list of ECMWF Publications can be found on our web site under:

http://www.ecmwf.int/publications/

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

Abbreviations

BUFR .... Binary Universal Form for the Representation of meteorological data
CMEM .... Community Microwave Emissivity Modelling platform
ECFS .... ECMWF's File Storage system
ECMWF .... European Centre for Medium-range Weather Forecasts
ESA .... European Space Agency
GRDS .... Ground RFI Detection System
IFS .... Integrated Forecast System
LDAS .... Land Data Assimilation Scheme
LSM .... Land Surface Model
MIRAS .... Microwave Imaging Radiometer using Aperture Synthesis
NRT...................... Near Real Time
NWP..................... Numerical Weather Prediction
QC .... Quality Control
RDA .... Research and Development Aerospace
RFI........................ Radio Frequency Interference
SAPP .... Scalable Acquisition and Pre-Processing system
SMOS .... Soil Moisture and Ocean Salinity
ZBT .... Zenithal Blue Technologies

1. Introduction

This document proposes a plan for testing the updated radio frequency interference (RFI) screening from the ground RFI detection system (GRDS) developed by Zenithal Blue Technologies (ZBT) and Research and Development in Aerospace (RDA). Data from the European Space Agency (ESA) Soil Moisture and Ocean Salinity (SMOS) mission will be used to assess the RFI screening. At the European Centre for Medium-range Weather Forecasts (ECMWF) SMOS data is routinely monitored by comparing it to high-quality short-range numerical weather prediction (NWP) forecasts. Experiments using this monitoring system will be run to compare the current data quality with the quality of the newly GRDS screened data. The details of the data to be used, experimental setup and analyses to be performed are described in this document.

2. SMOS monitoring at ECMWF: current status

2.1. Operational monitoring of the NRT SMOS brightness temperatures

The GRDS test-bed will be based on the operational ECMWF SMOS monitoring system (Muñoz-Sabater et al, 2013) which is part of the ECMWF integrated forecasting system (IFS). The monitoring system runs twice per day at 00UTC and 12UTC and compares SMOS observations with a high-quality stable reference state provided by the short-range operational forecast fields from the IFS. The resolution of the model fields used is $T_{CO}1279$ which is roughly equivalent to each model grid box being 9km x 9km in size. By comparison the SMOS field of view is approximately 50km in diameter. The observation locations are interpolated to model grid point locations and the Community Microwave Emissivity Modelling platform (CMEM) is used as the observation operator to convert model fields of temperature, humidity etc. into equivalent brightness temperatures. This then allows the modelled brightness temperatures to be subtracted from the collocated observed values to calculate what are known as “first guess departures”. Analysing the statistical distributions of these first guess departures is a key part of assessing the quality of the SMOS observational data. The samples for the statistical analysis can then be split up temporally and geographically as well as by instrument characteristics such as polarisation, incidence angle etc. The various monitoring plots are published online and can be seen at https://www.ecmwf.int/en/forecasts/quality-our-forecasts/monitoring/smos-monitoring. Because the short-range operational forecast, which the SMOS data are compared against, is so stable, any changes in the first guess departure statistics will indicate changes to the quality of the SMOS data which could represent instrument anomalies, changes in calibration, changes to the screening or improvements in the processing algorithms. In the context of this project the aim is to assess the change in first guess departure statistics due to the changes to the radio frequency interference (RFI) flags associated with the data.

2.2. Current use of RFI flags

Currently bits 1 and 4 of the SMOS information flags in the SMOS input BUFR files provide information on whether a SMOS observation is affected by RFI or not. As part of the SMOS pre-processing and pre-screening at ECMWF any SMOS observation which is affected by RFI as indicated by these flags is not processed further and first guess departure statistics are not calculated for these observations.

In a monitoring context this is done to avoid effects of RFI skewing the first guess departure statistics. To identify problems with the SMOS data a stable time series of first guess departures is required. RFI can vary in time and space and if not properly screened could lead to changes to the first guess departures which could be mis-interpreted as a problem with the Microwave Imaging Radiometer using Aperture Synthesis (MIRAS) instrument onboard the SMOS satellite. SMOS brightness temperatures are not directly assimilated into any ECMWF model but if they were and RFI-affected observations were assimilated then this could lead to anomalous analysis increments applied to the initial model state and consequent degraded forecasts. These are two reasons why it is important to screen out RFI affected observations.

The current monitoring plots, e.g. figure 1, show large areas of consistently increased standard deviation of first guess departures over parts of East Africa, the Middle East and East Asia. It is thought that RFI which is not detected by the existing RFI flags is causing these effects on the data which suggests that the current RFI screening is sub-optimal. Hence the need for updated RFI screening algorithms as provided by the GRDS.

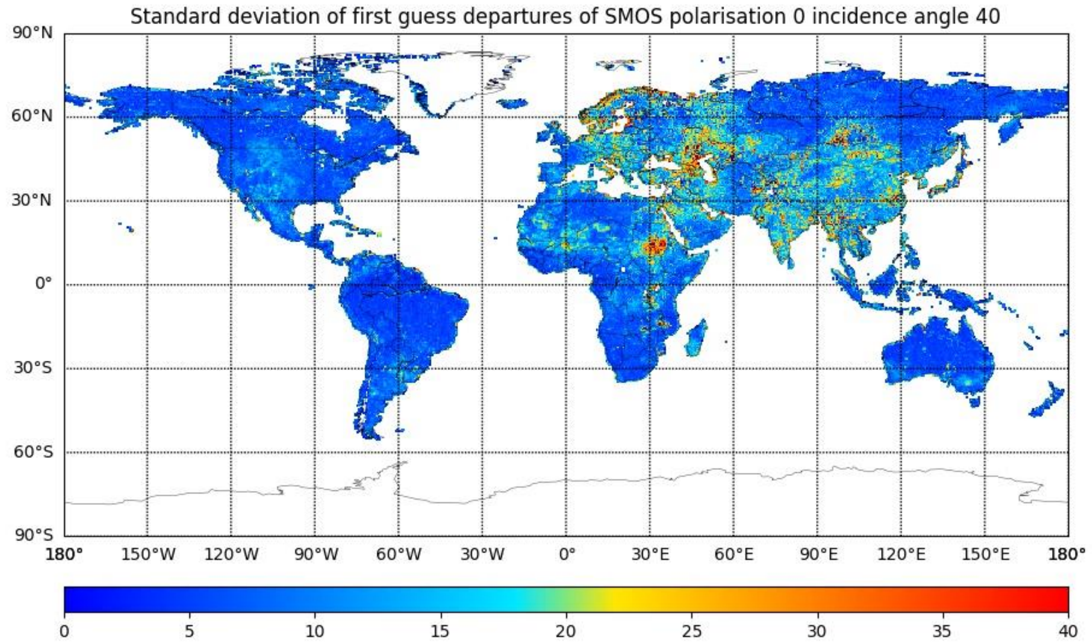

Figure 1: Operational SMOS brightness temperature monitoring: standard deviation of SMOS first guess departures (K) at H polarisation and incidence angles of 40°±0.5° from 22nd August 2019 to 21st September 2019.

3. Plan for testing the new GRDS-based screening

This section describes the test-bed plan in detail including which datasets will be used, the experimental setup and the analyses to be performed. It will also describe various metrics to be used to measure the performance of the GRDS compared to the existing RFI screening.

3.1. Experimental setup

Prior to the experiments being run the ECMWF Forecast Department will run their acquisition and pre-processing system, SAPP, to process the GRDS generated SMOS BUFR files into a format which can be read into the IFS and the pre-processed BUFR files will be stored on ECMWF's File Storage system (ECFS). The control experiment (CTRL) will use the NRT SMOS level 1 brightness temperature product produced using v6.20 of the SMOS processor. This is the product that has been operationally monitored at ECMWF from May 2015 to present. The plan is to relax the existing RFI screening in CTRL whilst retaining the information on which observations have been flagged for RFI. In this way the full sample of data can be filtered and analysed in areas of RFI and no RFI.

The test experiment (TEST) will use the same NRT SMOS level 1 brightness temperature product with the RFI flags adjusted so that they use the outputs from the GRDS. We currently only use bits 1, 4, 5 and 9 in the SMOS BUFR flag table (de Rosnay et al, 2015) so any of the other 9 bits will be available for use by the GRDS RFI flagging. Again, all RFI screening will be switched off whilst retaining all of the new RFI flagging information to allow analysis of the full sample of both RFI affected and non RFI affected observations. In this way the various RFI bits representing different GRDS algorithms can be assessed separately to ascertain which ones are more important than others. This will help to consolidate and fine tune the use of the algorithms in the GRDS and devise an optimal strategy for using the 2 bits available for RFI flagging for operational products.

Both CTRL and TEST experiments will run the ECMWF IFS cycle 46r1, the currently operational configuration, at a horizontal resolution of T_L511 (~40km grid spacing), which matches better the SMOS FOV size than the operational resolution ($T_{CO}1279$, ~9km grid spacing). The experiments will be run over a period of one month between 1st July 2019 and 31st July 2019 in order to gain statistically significant results in the comparisons. See appendix A for a justification on the use of a one month experimental period. Initially a shorter period of data will be requested so that it can be verified that SAPP and the IFS can successfully ingest and process the GRDS generated data. From a technical standpoint only a very small a volume of data such as a single snapshot or orbit will be required. However, it may be useful to have a slightly larger sample, such as a day of data, which can also be used to briefly sanity check the quality of the GRDS data to potentially warn of any problems or sub-optimalities with the GRDS algorithms. This could lead to significant refinements being made to the GRDS algorithms and thresholds on short timescales before a larger sample of data is sent.

3.2. Analyses to be performed

A number of statistical analyses will be performed to compare the CTRL and TEST experiments. These analyses of both experiments will be performed on the following sample of SMOS data:

- Over land - defined by the nearest model grid having a land-sea mask value greater than 0.5

- Avoiding snow and frozen surfaces - defined by the nearest model grid point having a snow depth of less than 1cm and a 2 metre temperature greater than 0°C

Firstly, and most simply, the number of observations where RFI is detected will be analysed to see how this number of observations changes with the new GRDS compared to the existing detection. These numbers can be assessed globally as well as looking at the geographical distribution over the globe to see whether there are specific areas that are more affected than others, see figure 2 for example. The temporal variability of the number of RFI detections can also be assessed. In particular to see whether there are diurnal cycles in any parts of the world and also whether there are any longer-term trends over the one month period. Using the different bits in the flag table, the number of observations with RFI detected by each different algorithm can also be analysed which will give an indication of the consistency between the different algorithms.

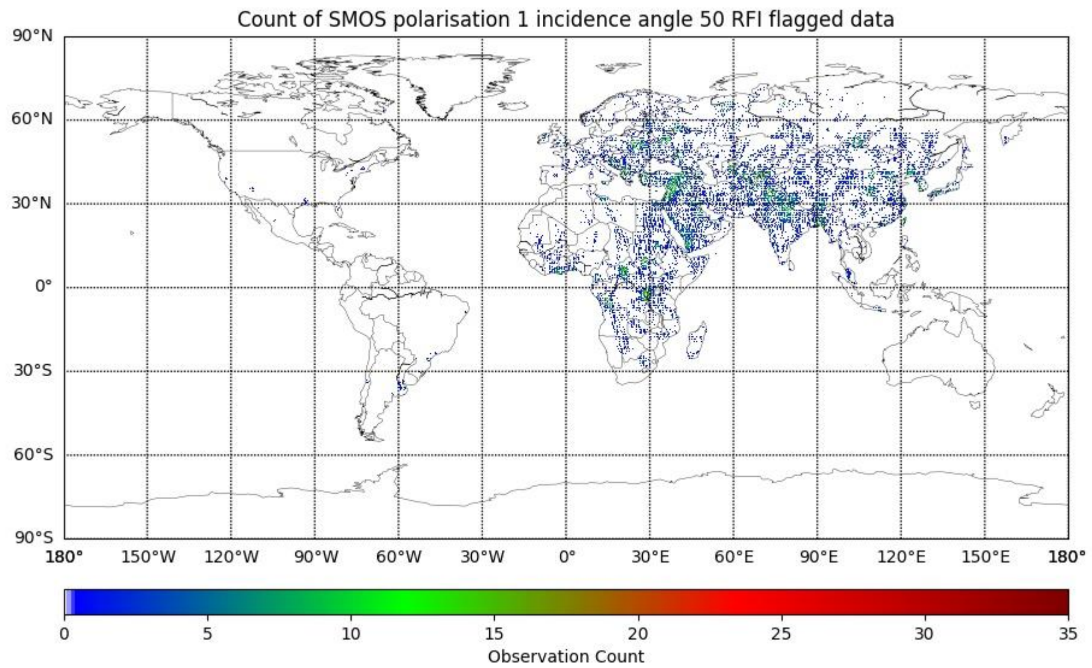

Figure 2: Map showing geographical distribution of current RFI flagged SMOS observations for V polarisation and incidence angles of 50°±0.5° during the period 22nd August to 21st September 2019

Secondly, and most importantly, differences in first guess departure statistics of the screened and unscreened data between the CTRL and TEST experiments will be assessed. This will include calculating the global mean and standard deviations as well as plotting histograms of the first guess departures, see figure 3 for example, to analyse the distribution and statistics of first guess departures where RFI is detected or not.

The first guess departure statistics can be split up geographically by calculating the statistics using samples from different latitude-longitude grid boxes over the whole globe. This data can be displayed on map plots to identify the worst affected regions and the largest changes coming from the new GRDS generated data. These plots can be used to do a more regional analysis of the GRDS as it is likely to perform differently depending on the type and magnitude of RFI detected which will vary by country and/or continent. It will be also important to make sure there are no false alarm detections over areas which appear to be free of any RFI.

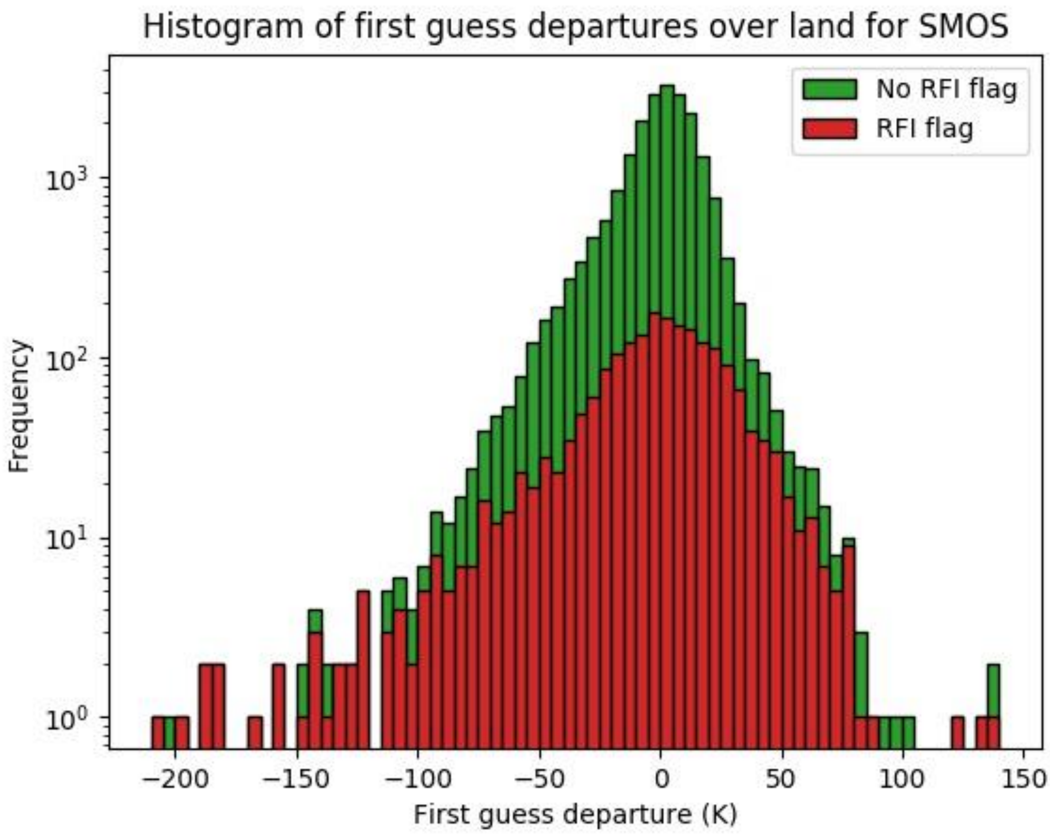

Figure 3: Histogram of SMOS first guess departures over land at V polarisation and incidence angles of 30°±0.5° grouped by those where no RFI is detected (green bars) and those where RFI is detected (red bars) for 22nd August 2019, note the logarithmic y axis

First guess departure statistics can also be calculated for different temporal samples over the one month period to produce timeseries plots, see figure 4 for an example. These plots can be used to identify shorter periods or cases where RFI is more or less prevalent. They can also be used to assess periods when the GRDS screening is performing better or worse prompting more in-depth investigations of individual observations and whether they are being correctly screened for RFI or not.

In addition to the geographical and temporal sampling suggested above, the data can also be sub-sampled using MIRAS instrument characteristics such as the polarisation and incidence angles because some types of RFI may affect some polarisations or incidence angles more than others. It's also possible that the GRDS algorithms may work better for some polarisations or incidence angles than others. Effects such as this should be detectable in the first guess departures when sub-sampled by different polarisations and incidence angles.

Each of the above analyses can be repeated when taking account of different combinations of the various bits in the flag table indicating RFI detection from the different statistical algorithms present in the GRDS. For example, a conservative approach would be to consider an observation is affected by RFI if any of the flags are set. A more aggressive approach would be to only consider an observation is affected by RFI if all the flags are set. Analysing the number of flagged observations, observed brightness temperatures and first guess departure statistics from these different samples will mean the importance of each of the individual algorithms can be assessed. This can then be used to help ZBT and RDA tune the various parameters and thresholds in each of the individual algorithms to help optimise performance, for example by reducing false alarms and misses. This information will also be important in consolidating the outputs from the various GRDS algorithms into the 2 available bits in the flag table whose specific purpose is to flag for RFI affected observations.

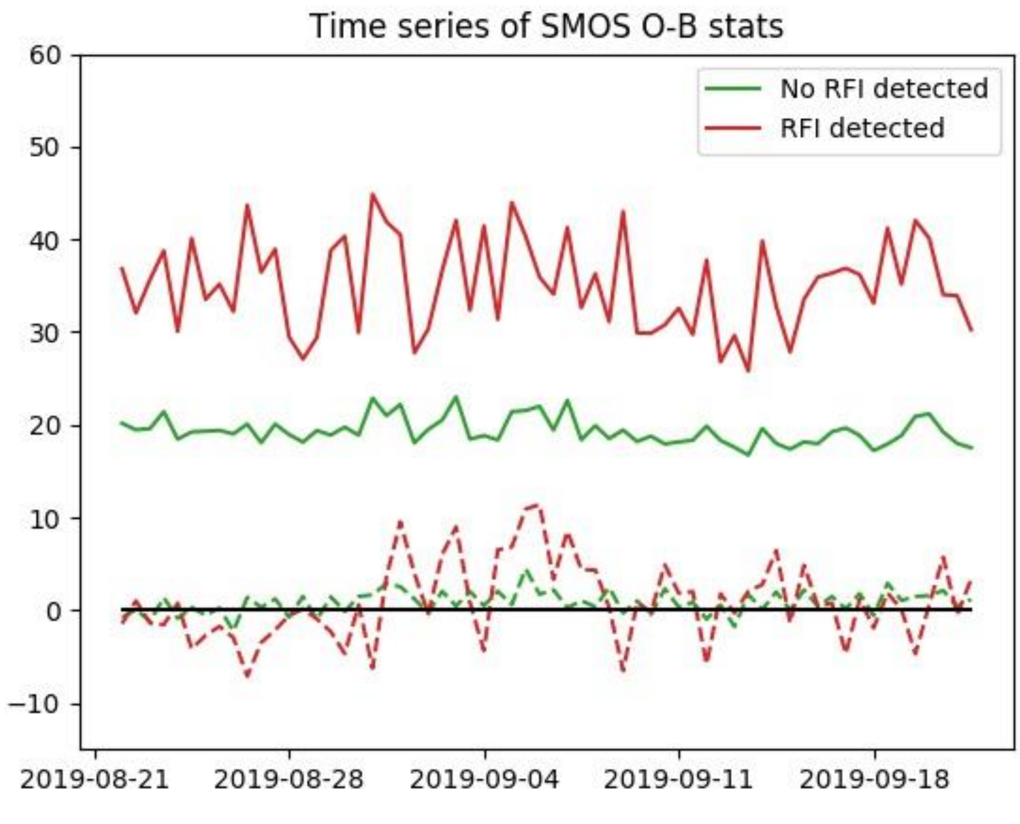

Figure 4: Time series of first guess departure statistics (standard deviations - solid lines; mean - dashed lines) for SMOS data over land at H polarisation and incidence angles of 40°±0.5° between 22nd August and 21st September 2019

4. Example analysis assessing impact of existing RFI flags 1 and 4

This section presents an example analysis of the effect on the first guess departure statistics of screening SMOS observations based on existing RFI flags 1 and 4. The aim is to illustrate the sensitivity of the first guess departure statistics to different RFI detection methods which will be useful to put the analysis of the new GRDS generated data into context. It will also provide a quantitative baseline performance of the existing RFI screening which the new GRDS based detection can be compared to.

4.1. Description of RFI flags 1 and 4

Flags 1 and 4 in the SMOS BUFR files indicate the presence of RFI contamination in the SMOS observations. RFI flag 1 is defined as "Pixel is affected by RFI effects as identified in the AUX_RFILST or it has exceed-ed the BT thresholds". RFI flag 4 as "Measurement is affected by the tails of a point source RFI as identified in the AUX RFI list (tail width is dependent on the RFI expected BT, from each snap-shot measurements, corresponding to 0.16 of the radius of the RFI circle flagged)" (de Rosnay et al., 2015).

4.2. First guess departure statistics analysis

The following analysis compares the effectiveness of the two flags by looking at the first guess departure statistics when different combinations of the RFI flags are used for screening.

Polarisation	Incidence angle	RFI flags used	Count	Mean first guess departure	Standard deviation first guess departure
H	30	No RFI flag	1266266	0.666131	20.60744
H	30	RFI flag 1	1253312	0.630706	20.10639
H	30	RFI flag 4	1163622	0.716496	18.94802
H	30	RFI flags 1 & 4	1157704	0.653375	18.75561
H	40	No RFI flag	1804532	0.862498	21.43371
H	40	RFI flag 1	1781520	0.791168	20.82175
H	40	RFI flag 4	1645178	0.917298	19.57969
H	40	RFI flags 1 & 4	1635041	0.824759	19.33832
H	50	No RFI flag	1773137	-1.85727	23.33054
H	50	RFI flag 1	1742073	-1.98795	22.54074
H	50	RFI flag 4	1602852	-1.74379	21.17273
H	50	RFI flags 1 & 4	1589710	-1.88479	20.85981
V	30	No RFI flag	1268756	0.567879	18.91359
V	30	RFI flag 1	1253313	0.569852	18.3275
V	30	RFI flag 4	1137194	0.739428	16.56255
V	30	RFI flags 1 & 4	1131331	0.682387	16.40153
V	40	No RFI flag	1803846	0.768256	18.33007
V	40	RFI flag 1	1775291	0.769286	17.51192
V	40	RFI flag 4	1596449	0.916137	15.57377
V	40	RFI flags 1 & 4	1585912	0.834985	15.34465
V	50	No RFI flag	1765039	-1.58334	18.01973
V	50	RFI flag 1	1725480	-1.54273	16.79772
V	50	RFI flag 4	1540539	-1.36695	14.61544
V	50	RFI flags 1 & 4	1526918	-1.46431	14.31039

Table 1: Global SMOS first guess departure statistics for different combinations of polarisation, incidence angle and RFI screening. The sample of data used in the statistics is from 22nd August 2019 to 21st September 2019

Table 1 shows that for every combination of polarisation and incidence angle, each level of RFI screening reduces the number of observations passing the screening as well as the standard deviation of first guess departures with the largest reductions coming from the RFI flag 4 check. The effect on the mean first guess departures is small. The average fractions of observations screened out by RFI flag 1 and 4 are 1.6% and 10.3% respectively. From the definitions of the flags this makes sense as the tails of the RFI sources (flag 4) will cover a wider area than the peaks of the RFI sources (flag 1). Also, the reductions in standard deviations of first guess departures are significantly larger for screening based on RFI flag 4 than for screening based on RFI flag 1. For H polarisations the reductions for RFI flags 1 and 4 are 2-4% and 8-9% respectively. For V polarisations the reductions are 3-7% and 12-20% respectively.

Figure 5 shows that the large areas of increased standard deviation of first guess departures over East Africa, the Middle East, Eastern Europe and India are slightly reduced but not removed by screening based on just RFI flag 1. However, when using screening from RFI flag 4 the standard deviation of first guess departures are markedly improved in these areas. When using both RFI flags 1 and 4 there is a marginal improvement over just using RFI flag 4.

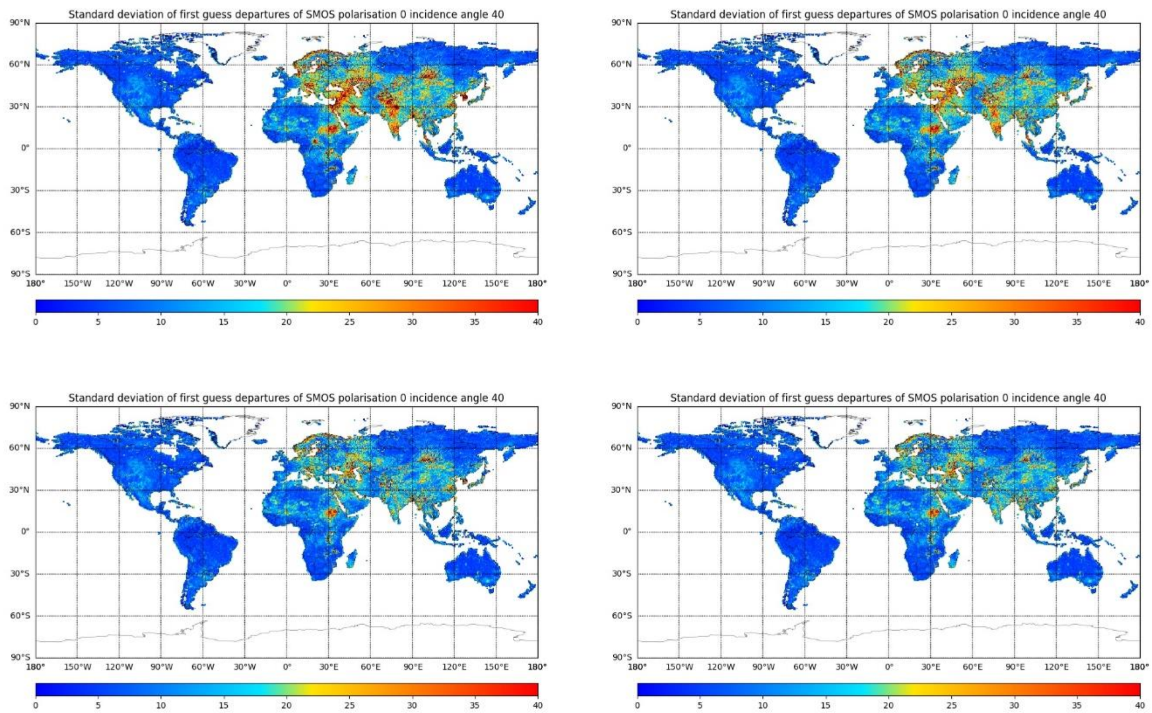

Figure 5: Maps of standard deviation of SMOS first guess departures with different RFI screening applied. Upper left: no RFI screening; upper right: RFI flag 1 screening; lower left: RFI flag 4 screening; lower right: RFI flags 1 & 4 screening. The sample of data for all plots is the same and comes from between 22nd August 2019 and 21st September 2019

Figure 6 shows the effect of screening based on the two RFI flags on the SMOS first guess departure distribution. It shows that screening purely based on RFI flag 1 reduces the tails slightly but has very little effect on the overall distribution. Screening based on RFI flag 4 has a larger effect, reduces the tails significantly and making the histogram more symmetric. Screening based on both RFI flag 1 and

4 results in an almost identical distribution to just screening based on RFI flag 4 but does result in an additional slight reduction in the positive tail.

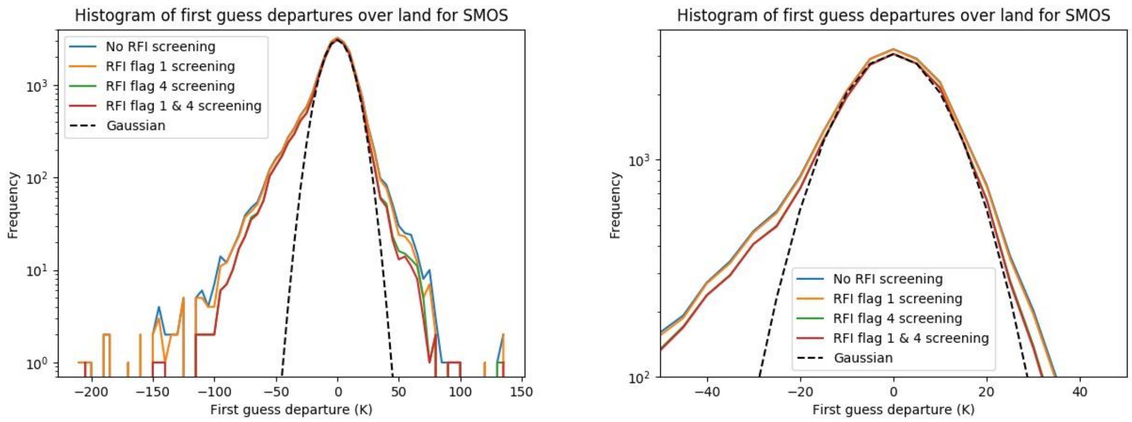

Figure 6: Histograms of SMOS first guess departures over land at V polarisation and incidence angles of 30°±0.5°. The different levels of screening are: no screening (blue); RFI flag 1 screening (orange); RFI flag 4 screening (green); and RFI flag 1 & 4 screening (red). The black dashed line shows the Gaussian pdf which best fits the peak of the SMOS distributions. The right panel shows a zoomed version of the right panel. The data is from 22nd August 2019, note the logarithmic y axis

By comparing to the sample Gaussian pdf it can be seen that, even after screening, the distribution is still far from Gaussian in the tails. However, the distribution of screened SMOS observations is almost perfectly Gaussian around the peak of the distribution, as shown in the zoomed in histogram. The match is almost perfect for first guess departures within ±15K of 0K and on the positive side it remains close for first guess departures up to 30K. However, there are still large tails on both sides of the distribution for first guess departures outside of the range -15K to +30K. This indicates that there are still larger numbers of SMOS observed values that don't agree with the equivalent background values than one would expect from a purely Gaussian distribution. This suggests there are still many SMOS observations affected by undetected RFI.

Notably the Gaussian pdf which provides the best match of the SMOS distribution near the peak in figure 3 has a standard deviation of 11K. This provides an estimate of the true SMOS first guess departure standard deviation (for this polarisation and incidence angle) without the effects of RFI. Table 1 shows that, even with the tightest available screening from RFI flags 1 & 4, the global standard deviation for V polarisation and incidence angle of 30° is 16.4K which is significantly larger than the estimate of 11K for non-RFI contaminated observations. This shows that there is almost certainly still significant undetected RFI affecting even the filtered observations. There will also be contributions from errors in CMEM and representation errors but the asymmetric nature of the histograms in figure 3 point to RFI as a leading cause of the larger errors and remaining large tails.

Appendix A - Justification of the use of one month for the experimental period

Originally, in the RFI4EO project proposal and first version of the test-bed plan, ECMWF requested 3 months of sample data with the new GRDS RFI screening to assess the effectiveness of the new screening compared to the existing screening. During the development of the GRDS it has become clear that producing 3 months of data will be problematic due to the sheer amount of time it will take to run the GRDS and produce the SMOS BUFR files required for input into ECMWF's assimilation system. Also, if any problems are found in the algorithms then it may become unfeasible to re-run the processing due to the running time. Therefore, we have proposed reducing the data sample considered to a single month. Clearly this will reduce the amount of time for the GRDS to process the data and, if problems are found, it will take less time to re-run. In addition, using a period of one month is also more consistent with the operational monitoring of satellite data for NWP performed at ECMWF. However, the use of a smaller sample may reduce the reliability and robustness of the results. To assess whether this is the case for the June-July-August (JJA) 2019 originally chosen, an additional analysis of the existing RFI flags in the operational SMOS data has been carried out.

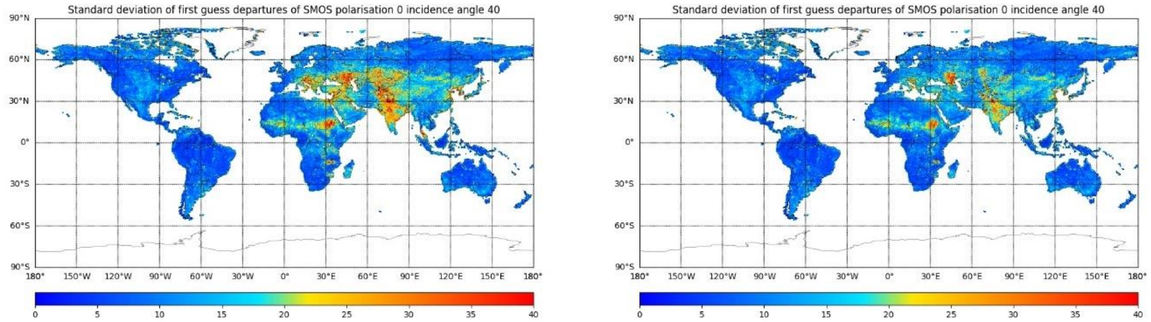

Figure 1: Maps of standard deviation of SMOS first guess departures (H polarisation, 40 degree incidence angle) with different RFI screening applied. Left: RFI flag 1 screening; right: RFI flags 1 & 4 screening. The sample of data for both plots is the same and comes from between 1st June 2019 and 31st August 2019

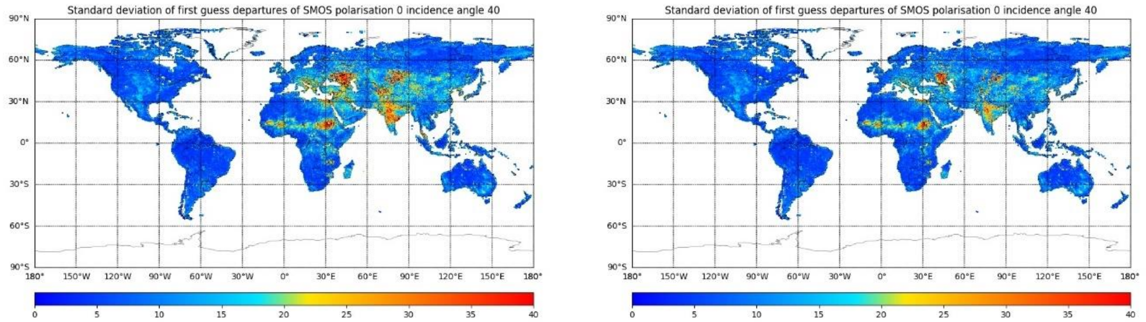

Figure 2: As figure 1 except the sample of data for both plots is the same and comes from between 1st July 2019 and 31st July 2019

The control experiment runs at T_L511 (~40km) and covers the period from 1st June 2019 to 31st August 2019. First-guess departures are calculated for SMOS throughout this period. The use of existing RFI flags in bit numbers 1 and 4 of the SMOS BUFR flag table are compared.

Figures 1 and 2 show that the first guess departure statistics are very similar when considering the whole JJA period and just July. In particular, the same conclusions would be drawn by comparing the level of RFI screening for both samples, that using RFI flags 1 and 4 together reduces the standard deviation of first guess departures compared to using only RFI flag 1. The main regions affected by the additional use of RFI flag 4 are South-East Europe, the Middle East and central Asia which can be seen in both figure 1 and 2.

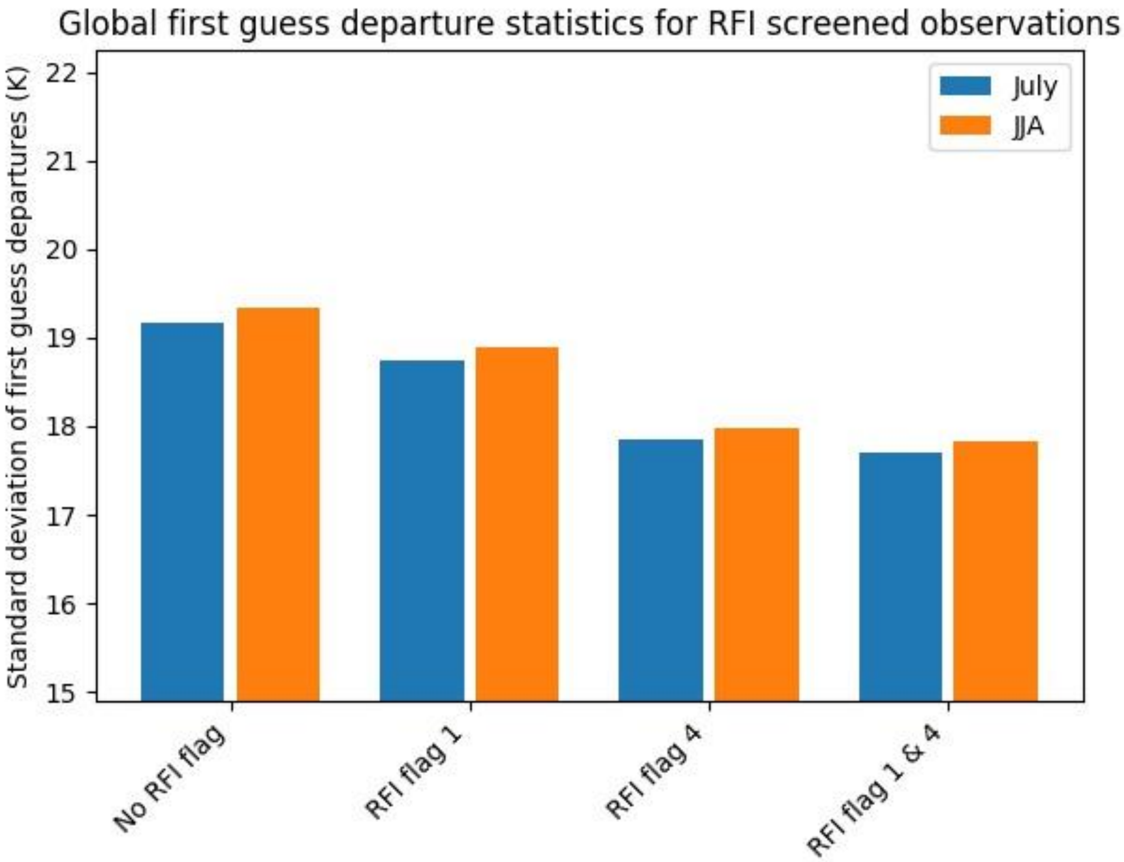

Figure 3: Global standard deviation of SMOS first guess departures (H polarisation, 30 degree incidence angle) with different levels of RFI screening with statistics accumulated between 1st June 2019 and 31st August 2019 (orange bars); and between 1st July 2019 and 31st July 2019 (blue bars).

Figure 3 quantifies the global standard deviation of first guess departures when screening based on different RFI flags and compares these statistics for the entire June-July-August (JJA) 2019 period and just July 2019. It shows that for all levels of RFI screening the global standard deviation of first guess departures is very slightly larger in JJA than it is in just July. These variations are expected over the year due to different levels of RFI and also different geophysical conditions which SMOS is sensitive to. However, the key aspect is that the difference between the statistics for JJA and July are consistent at around 0.15K across all levels of RFI screening. Therefore, if the same analysis of the

RFI flags was done based on either sample, the same conclusions would be drawn, i.e. that the best results are obtained when using both RFI flags 1 & 4, with RFI flag 4 having a more significant effect on the first guess departures than RFI flag 1.

Given these findings for the period of interest we believe that there is no risk to the reliability and robustness of results based on a 1 month sample or a 3 month sample. Therefore, for pragmatic reasons, we recommend the use of 1 month sample to assess the effectiveness of the GRDS RFI screening algorithms in the context of the RFI4EO project.

References

de Rosnay, P., M. Dragosavac, M. Drusch, A. Gutiérrez, M. Rodríguez López, N. Wright, J. Muñoz Sabater, Raffaele Crapolicchio: SMOS NRT BUFR specification, 2015. https://earth.esa.int/documents/10174/1854583/SMOS_NRT_BUFR_Specification

Muñoz Sabater J., M. Dahoui, P. de Rosnay, L. Isaksen: Technical Note, Phase II, WP1100: "*SMOS Monitoring Report number III: Dec-2011 - Dec-2012*". ECMWF ESA/ESRIN Contract 4000101703/10/NL/FF/fk, April 2013