ESA CONTRACT REPORT

Contract Report to the European Space Agency

Technical Note – Phase II – WP1200
SMOS Report on Data Thinning

Joaquín Muñoz Sabater
Tomas Wilhelmsson
Patricia de Rosnay
Lars Isaksen

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

Contract Report to the European Space Agency

Tech Note - Phase-II - WP1200:
SMOS Report on Data Thinning

Authors: Joaquín Muñoz Sabater,
Tomas Wilhelmsson,
Patricia de Rosnay,
Lars Isaksen
Technical Note - Phase-II - WP1200
ESA/ESRIN Contract 4000101703/10/NL/FF/fk

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

August 2011

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

Distribution list:

ESA/ESRIN
Luc Govaert
Susanne Mecklenburg
ESA ESRIN Documentation Desk

SERCO
Raffaele Crapolicchio

ESA/ESTEC
Matthias Drusch

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. SMOS is the second Earth's Explorer mission of the ESA Living Planet Programme. ECMWF is receiving the Near Real Time (NRT) product, which is currently used to monitor SMOS brightness temperatures in NRT. This product collects several dozens of multi-angular and multi-polarized measurements of the same target at a relatively global fine spatial resolution. A consequence is that the data product handled at ECMWF has a very high volume. The current operational capabilities of the ECMWF Integrated Forecasting System (IFS) do not make it possible to use the whole SMOS dataset within the monitoring suite or in an assimilation experiment. On the contrary (as it happens for any other source of satellite data), only a reduced subset of observations can be used within the IFS. Therefore, a strong data thinning filter needs to be applied to the data before it is ingested in the IFS. This report discusses several options to thin SMOS data and make it compatible with the IFS structure, while keeping a subsample of representative and good quality data for monitoring and assimilation purposes.
This is the technical note, phase II, workpackage number 1200 of the ESA Request for Quotation RfQ 3-13053/09/NL/FF/FvK. It is produced as part of the second phase of the monitoring and data assimilation study with SMOS data at ECMWF, and it is complimentary of reports TN-PII-WP1300 and TN-PII-WP1400 as described in the contract document number 4000101703/10/NL/FF/fk (Oct 2010 - Jan 2013).

1 Introduction

In the Soil Moisture and Ocean Salinity (SMOS) mission, the interferometric technique applied makes it possible to observe the same area of the Earth's surface under different views, thus providing multi-angular and multi-polarized observations of the same scene at different time stamps (1). The product used at the European Centre for Medium-Range Weather Forecasts (ECMWF) is the Near Real Time (NRT) product, which constitutes geographically sorted swath-based maps of brightness temperatures. This product is arranged in an equal area grid system called ISEA 4H9 (Icosahedron Snyder Equal Area grid with Aperture 4 at resolution 9) (5). For this grid, the centres of the cell grids, or nodes, are at equal distances of 15 km over land, whereas over oceans the grid has coarser resolution. Depending on the location of the observed target within the field of view of the instrument, up to 150 records of brightness temperatures in L-band between 0 and 65 degrees can be provided per node. Thus, the angular resolution of the observations is very high. This is a crucial component of the SMOS mission design as it opens the possibility to investigate, among other things a) the potential of this dataset to access and improve soil moisture information for Numerical Weather Prediction models and b) the angular noise filtering (addressed in the technical note TN-PII-WP1300).

The SMOS measuring principle results in the production of a very large volume of data which cannot all be ingested in the Integrated Forecasting System (IFS). The daily volume of SMOS data arriving at the ECMWF archives in NRT can be up to more than 8 Gb, which is by far one of the greatest sources of satellite data received at ECMWF. This amount of data cannot all be introduced in the IFS for just one single satellite instrument, taking into account that data from many other satellites is used simultaneously. For SMOS, it is estimated that only 5 to 10% of the initial data volume can realistically be ingested in the IFS. Data thinning is therefore a mandatory step and also prevents redundant and poor quality observations to go through the assimilation system. It is also a critical step in so far as it selects which data from the original files will be monitored, but also which data could be used to potentially correct the soil moisture state through an assimilation experiment. This report addresses different strategies to use and implement a simple, but efficient, thinning filter for SMOS data compatible with the IFS.

2 Thinning approaches

2.1 Pseudorandom thinning scheme. Implementation in cycle 36r1

The implementation of SMOS data in the IFS was a challenging task due to the special characteristics of this new type of satellite data (see for instance (4)). Although there are many different ways to thin SMOS data, for monitoring purposes the thinned subset of observations should be representative and keep the same statistical characteristics as the original dataset. In (2) a simple thinning approach was proposed and implemented in the IFS as a first approach to monitor the observations in NRT. It made it possible to use the data in the IFS almost inmediately after the satellite was launched on 2 November 2009, by creating an internal ODB (Operational Data Base) for SMOS which complies with the technical capabilities of the IFS. At this stage, the objective of the thinning approach was to provide a subset of SMOS data to efficiently and routinely monitor the observed brightness temperatures in NRT. It consisted of keeping only 1 out of 10 observations (based on a pseudorandom selection). It also filtered data affected by hard Radio Frequency Interference (RFI) by rejecting observations out of the 50-350 K range of brightness temperatures. The main advantages of this filter were that it had very good control over the size of the thinned dataset and it kept all incidence angles. This also allowed flexibility for further processing, for example for the averaging in angular bins at later steps. Furthermore, the implementation was very simple and only one parameter was needed to control this filter.

The simple thinning approach described above in this section was implemented in the IFS cycle 36r1. It was implemented using a parallel environment in order to make it compatible with the current operational computational constraints of ECMWF. It is able to process around 4 Gb of data (approximately 150 millions observations) in just one minute if eight processors are used in parallel. Fig. 1 shows a snapshot of the Supervisor Monitor Scheduler (SMS) system, which controls the flow of the different tasks in the IFS. The SMOS data thinning task was implemented within the *obs* (observations) family, and in particular within the '*presmos*' job of the '*prepare_obs*' subfamily. It consists of a series of Fortran routines and shell scripts which process the data in the same format as it is received, i.e., in BUFR format. After this step the filtered SMOS dataset is converted to ODB for further handling and processing in the IFS. All the following versions of the thinning approach are implemented in the same job of the *obs* family.

2.2 Angular thinning scheme. Implementation in cycle 36r4

Although the simple thinning approach described in section 2.1 was a practical way to monitor SMOS data, it has also some drawbacks. The main disadvantage was that there was no control over which observations were filtered out as it was based on a pseudorandom selection of the observations, with the probability of selection greater at larger incidence angles. It also produced an internal ODB with many observations which were not used at later steps. In order to select observations for monitoring on the basis of the viewing angle, and to prevent unused observations increasing the volume of the internal ODB, a new thinning approach was developed and implemented in the IFS. For each model grid point and for each monitored polarization, six different angles were selected. These are multiples of 10, from 10 to 60 degrees, which cover well the whole range of incidence angles. The remaining incidence angles were rejected by the thinning filter. An offset of 0.5 degrees centered on each monitored incidence angle assures that at least one observation is found within each angular bin. In reality, several nodes of the SMOS grid are associated with each model grid point if the model spatial resolution is T799 ($\sim$25 km) or coarser, whereas at T1279 ($\sim$16 km) the model spatial resolution is quite close to that of the SMOS grid. Within each angular bin, only one observation was selected and flagged as active, and this corresponded to the closest observation to the model grid point. The total estimated size

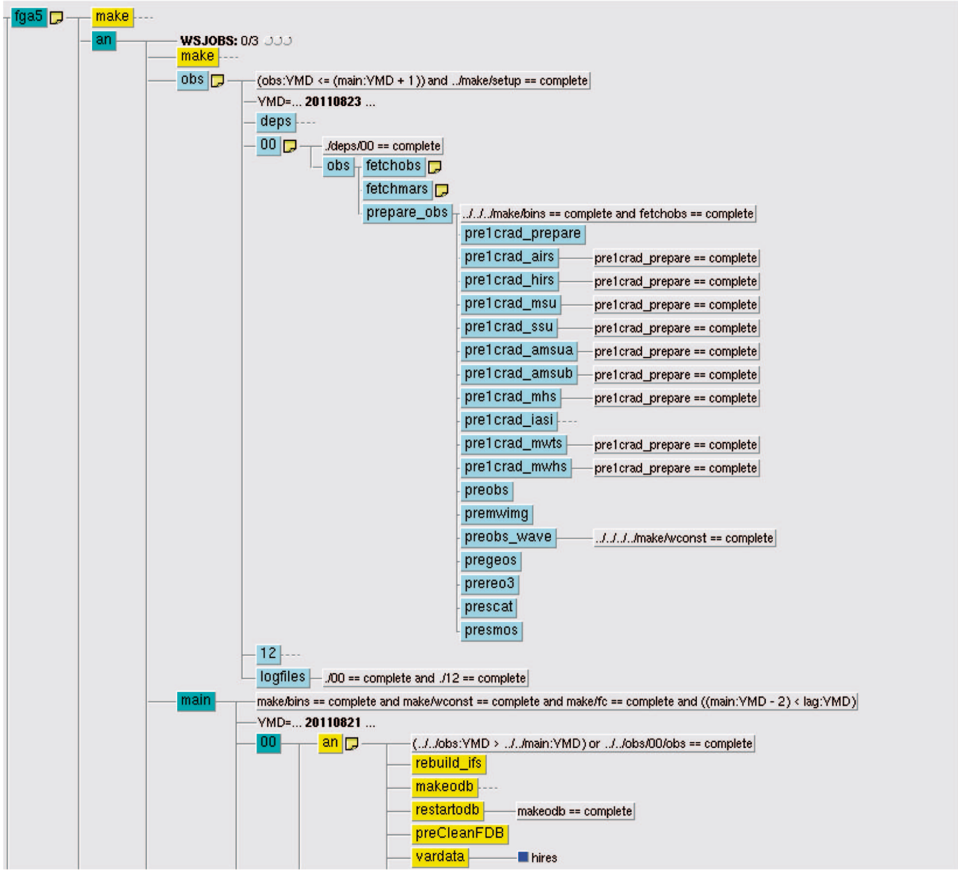


Figure 1: View of the Supervisor Monitor Scheduler (SMS) controlling the preparation of the observations for the analysis. Any thinning filter applied to SMOS data is implemented within the 'presmos' job.

of the thinned subset of observations varies typically between 5 and 8% of the original one. This makes it an affordable fraction of the IFS computational resources, while allowing better control over which observations are monitored. Furthermore, it guarantees that each model grid point will be associated with at least one observation within each angular bin, if a measurement was acquired at that angle. This thinning filter has been running in NRT since November 2010 in the current SMOS monitoring suite (see (3)).

2.3 Flexible thinning scheme. Implementation in cycle 37r1

While the previous thinning approaches were focused rather on the monitoring of SMOS brightness temperatures, the main objective at ECMWF is to evaluate the impact that the multi-angular SMOS dataset has on improving the soil moisture data assimilation system performance and skill. For that, a more flexible thinning filter is needed. Although the current monitoring suite uses six incidence angles and two polarization states, for research purposes different incidence angles to those currently monitored might be needed as input to the IFS. This requires a more flexible approach able to use different data thinning configurations for different assimilation experiments. Note that each research experiment always uses the original BUFR files as input, and the corresponding data thinning configuration and creation of the ODB is specific to that experiment. In other words, running experiments using different thinning configurations does not affect other experiments, nor does it the choice of viewing angles to be used for data assimilation purposes.
The new thinning approach was implemented in cycle 37r1 and it is based on a namelist, from which the characteristics of the thinned observation dataset are specified. The following characteristics can be specified:

- geographical area (*OBNLAT*, *OBWLON*, *OBSLAT* ,*OBELON*),
- number of incidence angles (*NO_ANGLES*),
- monitored incidence angles (*INCIDENCE_ANGLE*),
- offset around each selected angle (*DELTA*),
- selected polarizations (*POLARISATION*),
- percentage of water fraction allowed within the pixel (according to the SMOS water fraction field) (*WATER_FRACTION*),
- extended or alias-free field of view (FOV) area of the snapshot (*SMOS_IF*).

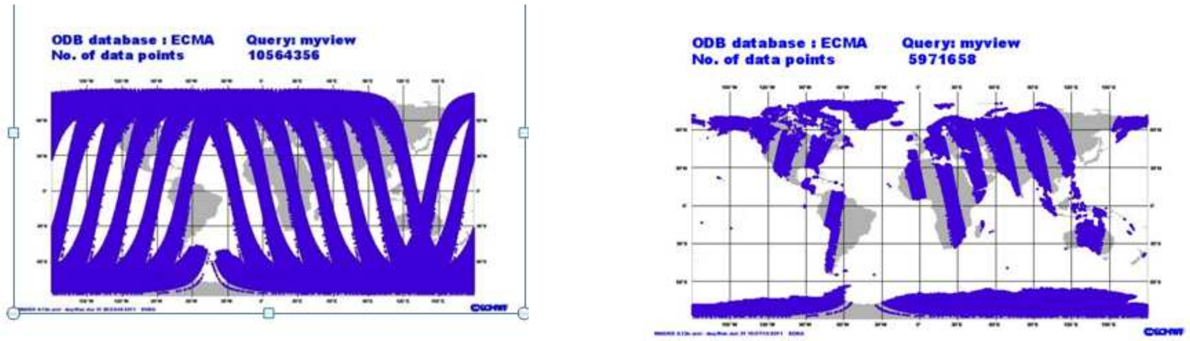
This approach is equivalent to the thinning approach of section 2.2, if the following variables of the namelist are set as follows:

- *NO_ANGLES*=6,
- *INCIDENCE_ANGLE*=10,20,30,40,50,60,
- *DELTA*=0.5,
- *POLARISATION*=0,1 (HH pol =0, VV-pol =1),
- *WATER_FRACTION*=100 (representing land and water pixels),
- *SMOS_IF*=1 (representing the extended-alias FOV).

2.4 Other approaches

This report addresses the necessity of greatly reducing the original volume of the SMOS data files before ingestion in the IFS. With the reduced dataset, an internal ODB specific to the SMOS observations is created for each experiment and used in different routines through the IFS code. As a general rule, the smaller the volume of data used in the IFS, the smaller the size of the ODB, and therefore the lowest computational resources will be used and the faster the experiment will run. The minimum target should be 10% of the initial volume, but for computational reasons it is desirable to use a percentage significantly lower than 10%. In Fig. 3 the results obtained for different thinning configurations based on a 12h assimilation window are shown.

It corresponds to data from 21h00 on 30 November to 9h00 on 1 December 2010 (see Fig. 2.4). For this period most of the observed data arrived at ECMWF in NRT, corresponding to a total of more than 150 million observations. As can be seen in Fig. 3, with the current monitoring suite (6 incidence angles with an 0.5 degree offset) less than 7% of the original data volume is used. Although the data reduction is important, there is still a significantly large number of observations in the SMOS ODB. By using the monitoring suite but rejecting pixels with more than 50% water content (so only keeping land points), only 4% of the total original volume remains in the IFS. If on top of that, only data within the alias-free FOV is used (of better quality) then the percentage is further reduced to 3%, corresponding to about 9% of the total data over land. By increasing the offset parameter for the monitoring suite, the percentage of data filtered increases quickly. Although some of these configurations could be used, either for monitoring brightness temperatures or for assimilation experiments, the main disadvantage is that it is still limited the number of views that can be assimilated simultaneously, as otherwise the ODB will get too large to be used in the IFS. The use of auxiliary data in BUFR space (such



(a) All data

(b) Only land data

Figure 2: SMOS orbits the 30 November 2010

as snow and ice covered areas or highly dense forests, where the data will be rejected for assimilation), or the use of the RFI flag of the BUFR product (once it has been implemented), will help further to thin SMOS data before use in the IFS.

All data 00Z analysis 20101201 (2100-0900)	**154.364.953 (100%)**
Water pixels + HHV / HVV	97.024.125 **(62.85%)**
Land pixels + AFOV	30.073.313 **(19.48%)**
Monitoring suite (6 angles +/- 0.5)	10.564.356 **(6.85%)**
Monitoring suite (6 angles +/- 0.5) + land pixels	5.983.758 (3.87% or **10.41 % over land**)
Monitoring suite (6 angles +/- 0.5) + land pixels + AFOV	4.923.337 (3.19% or **8.58% over land)**
Monitoring suite (6 angles +/- 1) + land pixel	11.979.065 (7.76% or **20.89% over land**)
Monitoring suite (6 angles +/- 2) + land pixels	23.728.031 (**15.37%** or **41.37% over land**)
Monitoring suite (6 angles +/- 3) + land pixels	35.189.567 (**22.79%** or **61.34% over land**)

Figure 3: Several thinning configurations based on SMOS data between 21h00 on 30 November 2010 and 9h00 on 1 December 2010.

2.5 SMOS Discrete Global Grid (DGG) to ECMWF reduced Gaussian grids

The observations found within the NRT product are collocated to the ISEA 4H9 grid nodes (5). Over continental surfaces these nodes are separated by approximately 15 km. This grid has a total of 2,621,450 nodes, over which in most cases several dozens of observations are found. As mentioned above, the volume of the resulting processed files is very large and in most cases this is a product difficult to integrate in current NWP systems. Furthermore, the 15 km spatial resolution NRT product will necessitate to take account of horizontal correlations, as the SMOS footprint is of about 40 km. By collocating the SMOS grid to a lower resolution ECMWF reduced Gaussian grid, a much lighter subproduct in a quasi-uniform grid could be produced which would be easier to handle for other centres as the number of nodes and, in consequence, the size of the product, would be significantly reduced.

In Fig. 4 a table including the conversion of the SMOS DGG to three ECMWF grids of different spectral resolution is shown. For example, by collocating the SMOS NRT product to the ECMWF T159 grid (∼125 km), the resulting number of nodes in the NRT subproduct would be reduced to 1.36% of its original size. With finer grids, using the T511 (∼40 km) the number of nodes would be reduced to 13.3%, and to 32.2% at T799 (∼25 km). However, for assimilation purposes, only nodes over land surfaces would be considered, and even at T799 less than 10% of the original nodes would be used, making it acceptable for the IFS. During the SMOS Quality Working Group meeting held in Rome in March 2011, it was agreed that ECMWF would provide a list of the SMOS DGG point identifiers which would be collocated to the ECMWF T511 reduced Gaussian grid points. The method used is the nearest neighbour. The European Space Agency (ESA) will produce a lighter NRT product based on the ECMWF grid, which will have approximately the same resolution as the SMOS observations, but with the added value of being a much lighter product, approximately only 4% of its original volume.

In Fig. 5 it is shown how the algorithm to reduce the SMOS DGG works. For the sake of clarity, the coarse ECMWF T159 reduced Gaussian grid is overlapped with the SMOS DGG, for latitudes near the equator and longitudes near the meridian zero. Black dots correspond to the SMOS DGG nodes, whereas crosses correspond to the ECMWF model grid points. The closest SMOS node to an ECMWF model grid point is displayed within a circle. Fig. 6 shows the result for the areas around the poles. This approach was validated to choose the nearest SMOS measurements to the model grid, and it will be part of the next IFS cycle 37r4/38r1.

ECMWF reduced Gaussian Grid	Number of nodes	% of DGG grid	% of DGG over land (threshold = 50%)
T159	35.718 **(100%)**	**1.36%**	**10.452 (0.4%)**
T511	348.528 **(100%)**	**13.3%**	**101.197 (3.9%)**
T799	843.490 **(100%)**	**32.2%**	244.428 **(9.32%)**

Figure 4: Reduction of the SMOS DGG to several ECMWF reduced Gaussian grids.

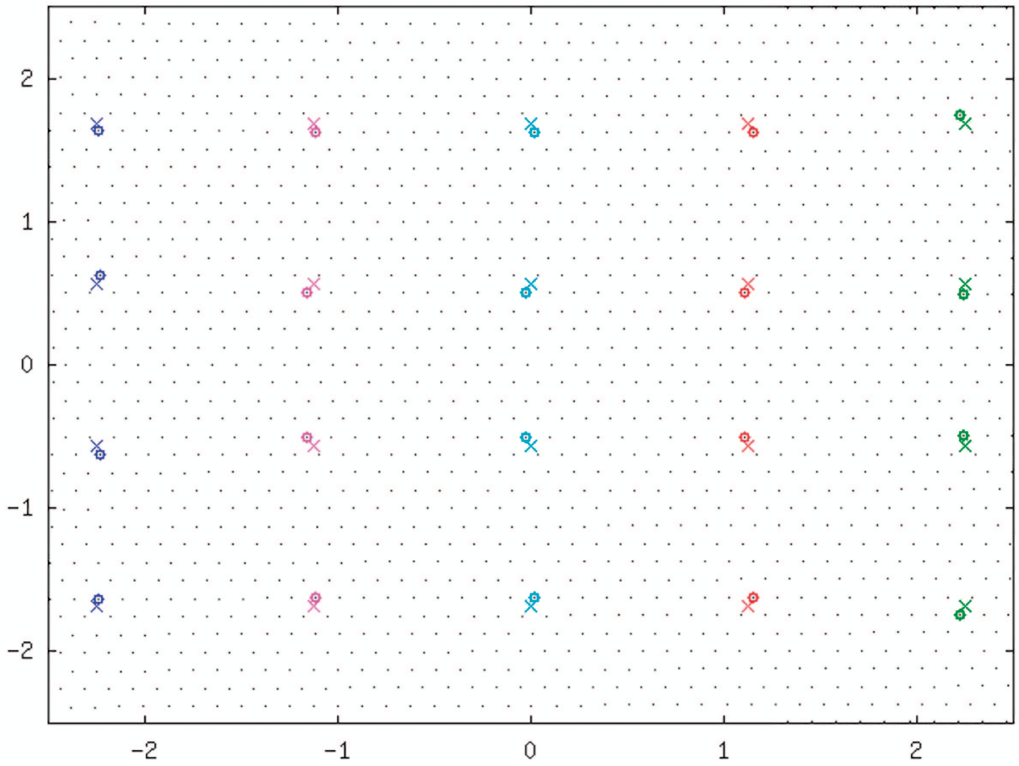


Figure 5: Nearest SMOS DGG nodes (circled black dots) to the ECMWF T159 reduced Gaussian grid points (coloured crosses) for an Equatorial area. The horizontal axis is the geographical longitude and the vertical axis is the geographical latitude, in degrees.

3 Conclusions

ECMWF is receiving the SMOS NRT product, which due to the special characteristics and objectives of this type of satellite data, results in a very large dataset. Due to system constraints and in order to avoid poor quality or redundant observations in the IFS, a thinning filter must be applied to the observations at early stages. However, the optimal thinning filter to be applied to the data is not obvious. For example, possible approaches may include averaging the data within predefined angular bins, discarding snapshots where a grid point has a brightness temperature value exceeding a given threshold, or removing noisy data at low incidence angles. Several approaches have been analysed with different objectives and implemented in different IFS cycles since the SMOS launch date. All these approaches are compatible with the current IFS computational limitations.
In general it was estimated that less than 10% of the initial data volume can be used in the IFS, but for optimality reasons it is desirable to reduce it even further. Several approaches have been implemented in the IFS since the ECMWF cycle 36r1 (since November 2009). The current one makes it possible to use a very flexible configuration for the input dataset in the IFS for both monitoring of brightness temperatures and assimilation experiments. This approach permits the investigation of assimilating different multi-angular configurations of the observations for the accurate access to soil moisture. For research experiments, a lighter NRT product collocated to the ECMWF T511 reduced Gaussian grid will be used, making it much easier to handle and more efficient from the computational point of view, but also avoiding undesirable horizontal correlations between adjacent observations.

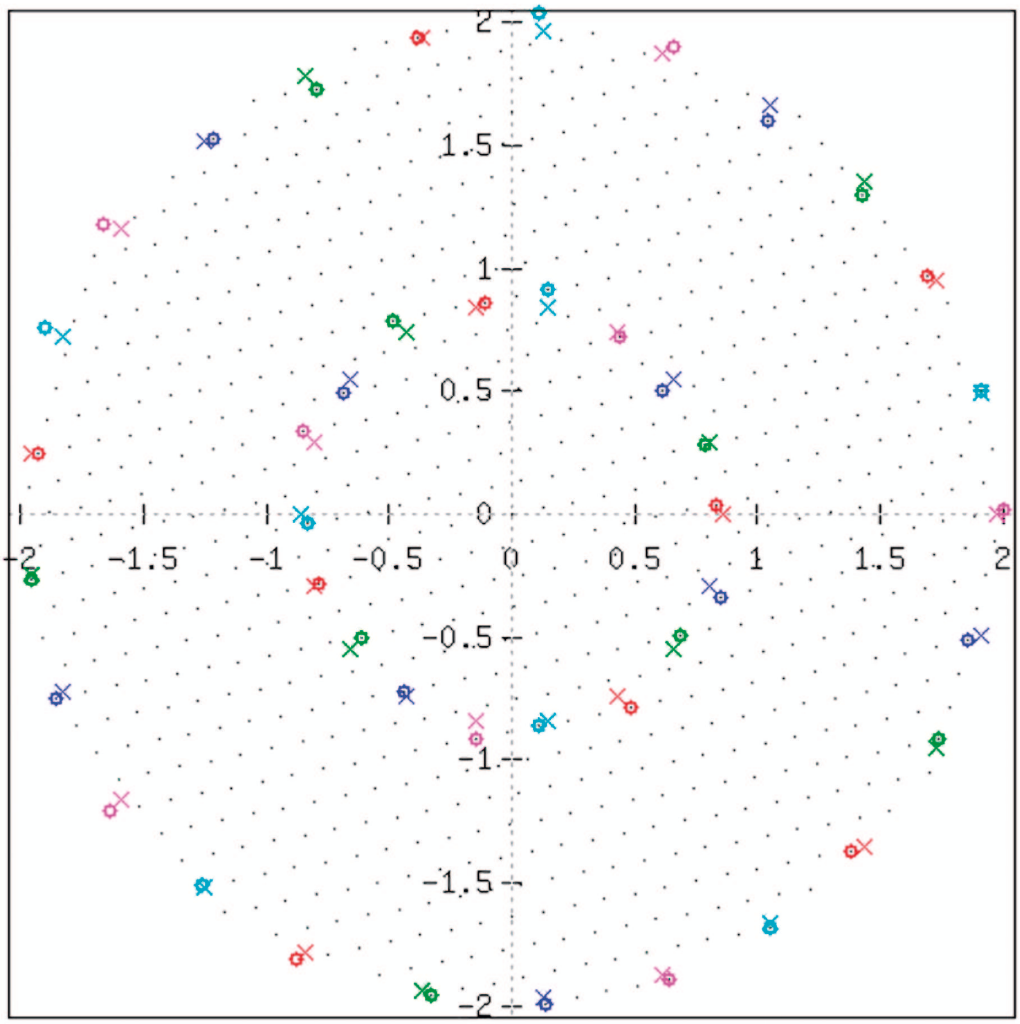


Figure 6: As in Fig. 5 but for the polar areas.

Acknowledgements

This work is funded under the ESA-ESRIN contract number 4000101703/10/NL/FF/fk. The authors would like to thank Milan Dragosavac and Anne Fouilloux at ECMWF for their support in the implementation of the thinning approaches, and Matthias Drusch and Susanne Mecklenburg, both ESA staff, for their contribution in the project definition.

4 References

References

[1] Y. Kerr, P. Waldteufel, J.-P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M.J. Escorihuela, J. Font, N. Reul, C. Gruhier, S. Juglea, M. Drinkwater, A. Hahne, M. Martin-Neira, and S. Mecklenburg. The SMOS mission: New tool for monitoring key elements of the Global Water Cycle. *Proc. IEEE*, 98(5):666–687, 2010.

[2] J.M. Sabater, A. Fouilloux, and P. de Rosnay. Technical implementation of SMOS data in the ECMWF Integrated Forecasting System. *Geosc. Remote Sens. Letters*, 2011. in press.

[3] J.M. Sabater, P. de Rosnay, and M. Dahoui. SMOS Continuous Monitoring; Part I. Technical report, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, February 2011.

[4] J.M. Sabater, P. de Rosnay, A. Fouilloux, M. Dragosavac, and A. Hofstadler. Milestone 1 Technical Note 1 Part II: IFS interface. Technical report, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, November 2009.

[5] K. Sahr, D. White, and A. J. Kimerling. Geodesic discrete global grid systems cartography. *Cartography and Geographic Information Science*, 30(2):121–134, 2003.