Technical Implementation of SMOS Data in the ECMWF Integrated Forecasting System

Joaquín Muñoz Sabater, Member, IEEE, Anne Fouilloux, and Patricia de Rosnay

Abstract—The launch of the Soil Moisture and Ocean Salinity (SMOS) satellite of the European Space Agency opens the way to using a new type of satellite data that are very sensitive to soil moisture for numerical weather prediction. The European Centre for Medium-Range Weather Forecasts (ECMWF) has developed an operational chain which makes it possible to process SMOS data in near real time (NRT) and compare it with a model equivalent. This process has been very challenging. The main reasons are the particular characteristics of the SMOS observation system and the large volume of data. Despite these obstacles, SMOS data are being processed successfully in NRT within the ECMWF Integrated Forecasting System (IFS). The ultimate objective is to assimilate these data in the IFS. It is expected to have an impact on the weather forecast at short and medium ranges. Prior to assimilation experiments, the quality of the data has to be assessed. This can be done through monitoring activities. Monitoring is a routine task performed with all satellite data, and among other things, it makes it possible to localize temporal (or spatial) bias or drifts in the data, thus providing NRT reports to the calibration and validation teams, which can act accordingly. In this letter, the implementation of SMOS data in the ECMWF IFS for monitoring purposes is discussed. The system was developed using a simulated file for the NRT processor, and it was tested using real data from the first year since the launch date.

Index Terms—European Centre for Medium-Range Weather Forecasts (ECMWF), implementation, monitoring, Soil Moisture and Ocean Salinity (SMOS).

I. INTRODUCTION

The successful launch of the Soil Moisture and Ocean Salinity (SMOS) satellite of the European Space Agency [1] is already providing an unprecedented new source of remotely sensed data that are sensitive to soil moisture over land and salinity over the oceans. Soil moisture has been extensively identified as a critical land variable due to its strong influence in the exchanges of water, energy, and carbon fluxes at the interface between the soil, vegetation, and the lowest level of the atmosphere [2]. A good estimation of soil moisture has a direct impact on precipitation and air temperature predictability at short and medium ranges [3], [4]. Passive L-band measurements are the most suitable ones for soil moisture retrievals. In this microwave region, attenuation from clouds and vegetation is smaller than that at higher frequencies [5]. Nonetheless, the heavy cost and technological challenge of arranging a large antenna in L-band have prevented an earlier spatial L-band mission. For SMOS, an antenna of approximately 8 m in diameter would be necessary to meet the spatial resolution requirements of the mission [1], making the cost prohibitive with the current technology. In SMOS, this problem is overcome by applying the interferometric technique. Instead of one large antenna, 69 little receivers installed in three deployable Y-shaped arms of 3.5-m length collect the 54 radiation emitted by the Earth’s surface between 1400 and 1542 MHz. The phase difference measured between the individual receivers makes it possible to reconstruct an image which meets the science requirements, i.e., volumetric soil moisture with an accuracy of 0.04 m³/m³ and a spatial resolution of 40–50 km [6]. As a numerical weather prediction (NWP) center, the European Centre for Medium-Range Weather Forecasts (ECMWF) is receiving a near-real-time (NRT) product, which is automatically recovered from the SMOS Data Processing Ground Segment. To fully take advantage of this product, ECMWF has implemented this new data type in the Integrated Forecasting System (IFS), which opens the possibility to monitor and assimilate the data within the IFS. This is a challenging task. In SMOS, the interferometric technique applied makes it possible to observe the same area under different views, thus providing multiangular and multipolarized observations of the same scene at different time stamps. Up to 150 records of brightness temperatures (\(T_B\)) between 0° and 65° are provided per observed area. The angular resolution of the observations is very high. This measuring principle has the following two consequences: 1) the distribution of a unique data set with new features very different to any other source of satellite data used for NWP and 2) the production of a very large volume of data which cannot all be ingested in the IFS. Therefore, the data volume must be reduced significantly before a model equivalent is computed and compared with the observations. The previous characteristics of the SMOS observation system have raised great concern about the feasibility of integrating SMOS data in a complex NWP system. In this letter, the feasibility of its operational use is demonstrated. The main steps involved in the design of the SMOS data process chain are presented. The objective is to set up the structure necessary to operationally produce a comparison between a subgroup of SMOS observations and their model equivalent, as this is the main input for an assimilation scheme. The chain developed was tested with the first few of available data and run successfully for one year.

II. DATA PRODUCT USED AT ECMWF

The product used at ECMWF is the NRT which constitutes geographically sorted swath-based maps of \(T_B\). The geolocated product received at ECMWF is arranged in an equal area grid.
system called ISEA 4H9 (Icosahedron Snyder Equal Area grid
with aperture 4 at resolution 9) [7]. For this grid, the centers of
the cell grids are at equal distances of 15 km over land, with
a standard deviation of 0.9 km. Over oceans, the grid has a
99% coarser resolution, which is half of the resolution over land, as
oceans are more homogeneous than continental surfaces. The
data are organized in messages. Each message corresponds to
a snapshot where the integration time is 1.2 s, as this is the
time in which all correlations of a single scene are measured.
On average, each message contains around 4800 observations
over land if the instrument runs in dual-polarization mode. In
this running mode, data set records are generated alternately
each 1.2 s at horizontal (HH) and vertical (VV) polarizations.
In full-polarization mode, all four Stokes parameters of the
radiation are collected by the antennas during four consecutive
integrations. Thus, the polarization state of the radiation is fully
111 described in this mode. Since the end of the commissioning
phase, the instrument on-board SMOS has been operating only
113 in full-polarization mode.

III. IMPLEMENTATION

In this section, the main steps and challenges involved in the
implementation of SMOS data in the IFS are addressed. Before
SMOS data can be assimilated, the data go through a series of
118 tasks which have the objective of validating and comparing the
119 observations with a simulated value, thus producing an input
120 for the soil moisture assimilation scheme. These tasks can be
121 classified into the following two large groups:

1) data prescreening;
2) computations in the model grid.

A. Data Prescreening

First, NRT raw data (see Section II) processed at the
European Space Astronomy Centre in Madrid (Spain) are re-
trieved and slightly modified to feed the prescreening tasks. 128
These tasks perform quality control checks.

1) Generic checks: Files which do not contain crucial header
information are rejected. It is checked that date and time
are complete, geographic coordinates are not missing, and
instrument data correspond to SMOS data.

2) The validity of data is checked: Individual observations
are checked to be in a correct geographical position, and
the incidence angle is checked to be in the range of physically reasonable
values (not lower than 50 K and not greater than 350 K),
which is also a practical hard radio-frequency interfer-
ence (RFI) filter.

3) Data are thinned to reduce the volume of SMOS data
processed within the IFS.

Data thinning is a critical step insofar as it selects not
only which data from the original files will be monitored but
also which data could be used to correct the soil moisture
state through an assimilation experiment. The daily volume of
SMOS data arriving in ECMWF archives in NRT is of about
8 GB, which is by far one of the greatest sources of satellite data
received at ECMWF. This amount of data cannot be introduced
in the IFS for just one single satellite instrument, taking into
account that data from many other satellites are used simulta-
neously. For SMOS, only 5%–10% of the initial data volume

Fig. 1. Histograms of brightness temperatures for an SMOS-simulated file for
December 17, 2010. (a) Histogram of $T_B$ for testing file. (b) Histogram of $T_B$
for EXP1. (c) Histogram of $T_B$ for EXP2.

...
Fig. 2. (Blue circles) Nearest SMOS observations to the (black dots) ECMWF T159 model grid points. The values are the distances between grid points and the nearest SMOS observation, in meters.

by using EXP2, data can be averaged in angular bins at later steps. Furthermore, the implementation is very simple, and only one parameter is needed to control this filter. For EXP1, the offset parameter has to be optimized each time to control the number of observations filtered, and not all angular geometries can be monitored. Thus, the initial implementation of a filter type as in EXP2 was preferred for monitoring purposes. Due to operational constraints, this process has to be completed very quickly. On average, 12 h of data are processed in almost 1 min using eight processors in parallel. This makes SMOS prescreening tasks fully compatible with the current ECMWF operational system.

B. Computations in the Model Grid

The implementation of SMOS data monitoring in the IFS was carried out in the model grid. This has several advantages.

1) All the background fields necessary to simulate \( T_B \) at the top of the atmosphere are computed and available in the model grid point space. Thus, it avoids interpolating physical quantities to an observation location.

2) Other satellite data that are sensitive to soil moisture, such as AMSR-E data in C-band, are also available in the model space, making a comparison with other satellite data possible. The subset of observations which are selected after the prescreening filters undergoes a two-step process.

a) Observations are brought into the model grid point space at the required model resolution by using the nearest neighbor technique. At the same time, a mask of the flags containing information of the grid point is created. Fig. 2 shows the SMOS observations for the simulated file on December 17, 2010, collocated to the nearest grid point. For the sake of clarity, a magnified area over the south Indian coast is shown, and a rather coarse grid T159 is used (~125 km). The distance limit beyond which observations are rejected was set to 10,000 m. The number of observations monitored depends on the model grid resolution and the distance limit parameter. At T799 (~25 km) or higher spectral resolution, SMOS observations within the distance limit were found for all grid points (not shown).

b) For each observation flagged as being the closest to a model grid point, a background value is simulated with an L-band forward model operator. Thus, the innovation vector is obtained as the main input for a soil moisture analysis. The forward model interfaced to the IFS is the Community Microwave Emission Model [8], [9], as explained in [10].

IV. RESULTS WITH REAL DATA

The operational chain described in Section III was tested with real \( T_B \) data from the commissioning phase. No distinction between ascendant and descendent orbits was made. These data sets were the following:

Data set 1) November 28, 2009;
Data set 2) December 15, 2009;
Data set 3) December 20, 2009;

A. Prescreening Results and Model Equivalent

The quality checks listed in Section III-A were tested with data sets 1), 3), and 4). Fig. 3 shows the number of individual \( T_B \) values rejected as a function of the first 18,000 snapshots. This corresponds to the first 6 h of collected data for these days. During the commissioning phase, the SMOS operational mode was alternated between dual and full polarizations so as to test and select the optimal mode for the instrument. Data sets tested in this section contain data in both modes. Thus, for comparison purposes, only snapshots with fewer than 5000 subsets were used because they correspond to pure HH- or VV-polarization integrations. This figure clearly shows that the number of rejected radiances is the greatest for November 28, when the data were not yet calibrated, whereas the number is significantly reduced for December 20 and January 16. Table I shows a quantitative comparison between the three data sets. It
TABLE I
NUMBER OF OBSERVATIONS REJECTED FOR 6 h OF DATA DURING THE EARLY QUALITY CHECK PHASE

<table>
<thead>
<tr>
<th>Date</th>
<th>snapshots</th>
<th>subsets</th>
<th>rejections</th>
<th>% rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>28-11-2009</td>
<td>17940</td>
<td>28203176</td>
<td>147185</td>
<td>0.52</td>
</tr>
<tr>
<td>20-12-2009</td>
<td>17592</td>
<td>28739029</td>
<td>58967</td>
<td>0.21</td>
</tr>
<tr>
<td>16-01-2010</td>
<td>15347</td>
<td>24322415</td>
<td>34386</td>
<td>0.14</td>
</tr>
</tbody>
</table>

B. Continuous Monitoring of SMOS Observations

The feasibility of using the previous chain operationally was tested with several months of real data. SMOS data were monitored by systematically producing global maps of NRT $T_B$ for different incidence angles (0, 10, 20, 30, 40, 50, and 60) and for both HH- and VV-polarization modes. Plots for 2010 are available at http://www.ecmwf.int/research/ESA_projects/SMOS/monitoring/2010/2010.html. A simple inspection of these figures makes it possible to observe not only the angular $40^\circ$ evolution of the data for each polarization state but also a $90^\circ$ significant improvement in the quality of the observations, especially after the instrument was calibrated. Fig. 5 shows $T_B$ at $40^\circ$ incidence angle and VV-polarization state for days 1298 (top), 3 (middle), and 4 (bottom). It shows a clear evolution in the quality of the data, from day 1 (top) to day 4 (bottom). The day in November (top) is shown to be very noisy. These 301 data were received within the first two weeks of the instrument “Switch-On Phase,” which obviously had not benefited yet from a good calibration. In December, a major calibration event took place, and the difference in the product is quite significant when comparing the top and middle figures. Improvements are present almost everywhere. The data are even better for 307 January 16, although this needs a closer examination and quantification to confirm it (see Table I). Results for the HH polarization were equivalent. As SMOS is a research mission, it was also important to check the correct functioning of the novel polarimetric product. Hence, it was checked that the $T_B$ values got colder 312 with increasing the incidence angle for HH polarization, and the opposite behavior was shown for the VV polarization (see, for example, the online plots), with both displaying values within an acceptable physical range. It is also an objective of data monitoring activities to report on possible spatial or temporal effects on the data: The most outer sides of the satellite track look colder than the inner part, which is due to the extended 316 alias-free field of view, of lower quality than data closer to the center of the track. There is still residual RFI over Europe, the Middle East, and Asia, which is particularly straightforward to see when the data look very “red” and noisy. The data are of 320 good quality overall, except in these areas. The SMOS data were relatively better quality over the whole of America, Australia, and southern Africa.

V. SUMMARY

Integrating a new type of satellite data with innovative features in a complex NWP system and making it fully compatible with an NRT structure is a very challenging task. This is the case for the SMOS NRT product arriving at ECMWF. The 328 large volume of SMOS data and the large number of angular views per pixel have raised concerns about the feasibility of using this product in an operational NWP context. Despite these challenging features, the operational use and the feasibility of 334
a grid point has a $T_B$ value exceeding a given threshold, or removing noisy data at low incidence angles. Certainly, the development of an enhanced thinning filter deserves further investigation, and it is a key activity as assimilation experiments may depend strongly on the selected subsample of data.

Forward modeling on the model grid (using the geometry of the nearest SMOS observation) proved to be efficient and fast if several processors were used simultaneously. The analysis of the first batch of real data using this structure proved to be useful. It suggested a clear enhancement in the data quality during the first months of the mission, both qualitatively and quantitatively. There are still strong sources of RFI remaining in Europe and Asia, whereas a visible data degradation is still observed at the edges of the satellite track over oceans at high incidence angles. The systematic production of these plots in NRT is an excellent way to monitor the data just a few hours after sensing time and to quickly inform calibration and validation teams about trends or drifts in the data. In this context, SMOS data monitoring will be supported through the provision of the statistics of the modeled values in the grid point space, where first-guess departures are computed. These statistics are currently being obtained for several weeks of data, and their analysis will be presented in a follow-up paper totally devoted to this aspect.

**ACKNOWLEDGMENT**

The authors would like to thank M. Dragosavac, I. Mallas, and A. Hofstadler for their support with the production of the preprocessed data within the near-real-time chain and the anonymous reviewers for the useful comments.

**REFERENCES**


AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1 = This sentence was reworded for clarity. Please check if the intended meaning was retained.
AQ2 = “ESA” and “ESRIN” were defined as “European Space Agency” and “European Space Research Institute,” respectively. Please check if appropriate.
AQ3 = The financial support statement was reworded to conform to IEEE style. Please check if appropriate.
AQ4 = Occurrences of the unit “Gb” were changed to “GB” (for gigabyte). Please check if appropriate.
AQ5 = List items “1/” and “2/” were captured as a displayed list for clarity in presentation. Please check if appropriate.
AQ6 = Please provide the expanded form of the acronym “AMSR-E.”
AQ7 = This sentence was reworded for clarity. Please check if the intended meaning was retained.

END OF ALL QUERIES