Validation of the acquisition algorithms for FSSCat's Soil Moisture product

Author: Inés Terraza Palanca
Faculty of Physics, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Yolanda Sola, Adrian Perez-Portero, Adriano Jose Camps

This essay presents a validation of two months (October-November 2020) of soil moisture retrieved by the FSSCat mission by comparing it with the ESA Climate Change Initiative mission (ESA CII) dataset. A data processing pipeline, along with various statistical tests, was designed to detect disparities between the two datasets. The results with RMSE, Bias, and ubRMSE revealed notable discrepancies in some regions, such as Russia, with values of 0.1 m^3/m^3 for the ubRMSE. In general, FSSCat's data has underestimated measurements compared to ESA CII's dataset. These discrepancies can be attributed to instrumental errors, the presence of ice in certain regions, and uncertainties in the re-gridding method.

I. INTRODUCTION

Soil moisture (SM) refers to the water content present in the soil, typically expressed as [m^3/m^3] of the total soil volume. It plays a critical role in Earth's climate system, influencing various physical and biological processes including the water cycle, plant growth, and energy balance [1].

The SM value is one of the essential climate variables (ECVs), as it directly impacts the exchange of water and energy between the land surface and the atmosphere. High SM levels result in increased evaporation, leading to higher humidity close to the ground and, in some cases, cloud cover [2]. These changes influence the absorption and reflection of solar radiation by the Earth's surface, ultimately affecting temperature and climate patterns.

SM also plays a key role in the carbon cycle, as it affects the growth and productivity of plants, which absorb carbon dioxide from the atmosphere through photosynthesis. In addition, changes in SM can also have effects on weather patterns, such as droughts and floods [3].

On the other hand, in the field of climate studies, the measurement of SM is of utmost importance, as it provides valuable insights into the state of Earth's water cycle and the potential consequences of climate change [1]. Remote sensing techniques, as employed in missions like ESA CII and FSSCat, which will be later explained, have revolutionized the ability to measure soil moisture over large areas, enabling researchers to monitor temporal and spatial variations in SM levels.

This study aims to validate soil moisture data obtained from FSSCat by comparing it with ESA CII in Europe. The data processing was conducted at the UPC Nanosat Lab of the Universitat Politècnica de Catalunya (UPC), which designed the mission, and processed the acquired data.

II. FSSCAT MISSION OVERVIEW

The FSSCat mission was the winner of the 2017 Copernicus Masters Competition. It is an innovative mission consisting of two 6U nano-satellites carrying two scientifc payloads. The payloads onboard are the Flexible Microwave Payload 2 (FMPL-2) for ^{3}Cat-5/A and a GNSS reflectometer (GNSS-R) and the HyperScout-2, hyperspectral camera, in ^{3}Cat-5/B [4].

The mission was launched the 3rd of September 2020, on Vega Flight VV16, and injected into a 535-km synchronous orbit. The CubeSats were fully operational, and the scientific requirements were met. Throughout the autumn months, maps of soil moisture and sea salinity were generated.

This essay aims to verify and compare specifically the data retrieved for SM after the commissioning phase, from the 1st of October to the 1st of December. Despite the measurements have a global coverage, the study only focuses on the region of Europe.

A. Instrumentation: FMPL-2 and HyperScrout

^{3}Cat-5/A carried FMPL-2, designed and implemented by the UPC, which combined an L-band microwave radiometer and a GNSS-R in a single instrument. Since L-band (1-2 GHz) has a strong sensitivity to the change of surface SM and can more easily penetrate the atmosphere and vegetation canopy, it has been widely used as the main SM remote sensing frequency band in the satellite-based radiometer and radar missions [5].

On the other hand, the GNSS-R component of FMPL-2 relies on the conventional GNSS-R (cGNSS-R) technique. The latter involves capturing signals reflected by the Earth's surface, which generate signal wavefronts determined by the reflecting surface. Each signal wavefront exhibits a unique delay (τ) and Doppler (ν) based on the geometry [6]. In flat surfaces, the majority of reflected power originates from the region around the specular point. The Specular Reflection Point refers to the location with the shortest path between the satellite, the surface, and the antenna. However, as the surface roughness increases, the antenna receives signal wavefronts scattered around the specular point [7], which implies it has undergone multiple reflections in the soil.

SM data is retrieved by the combination of instruments in FSSCat/A and FSSCat/B by means of pixel downscaling [7]. Retrieving SM using GNSS-R is a complex task, as the surface roughness makes SM recovery difficult. The latter was accomplished by using Artificial Neural Networks (ANN) with a resolution of 1km [8], and the final result was used for this study.

III. DATA VALIDATION WITH ESA CII

The validation of the SM of FSSCat was performed with the ESA Climate Change Initiative mission (CCI). The latter is a long-term project led by the *European Space Agency* (ESA) aimed at providing ECVs related to the Earth's climate system.
The ESA CCI mission focuses on improving the understanding of the Earth's climate system by providing accurate, high-quality data on various aspects of the climate, such as sea level, sea ice content, greenhouse gases, land cover, and ocean color. The data is collected through a range of satellite sensors, and processed using advanced algorithms to ensure its accuracy and consistency over time.
This project lasted from 1971 to 2020, resulting in a substantial amount of data for processing. For the comparison, data from the *combined* dataset [9] was selected, covering the time period of day 18536 (1st of October 2020) to day 18597 (1st of December). The *combined* dataset incorporates various passive missions such as SMOS or SMAP. The project and specific missions are described in [9]. This dataset has been extensively used in previous studies [10], and offers numerous attributes for analysis. These attributes include day/night information, frequency bands, flags for identifying inconsistencies in soil, and a spatial resolution of 25 km. As the FSSCat dataset had fewer variables, the following section details the data processing approach, including the selected attributes and grid used for the comparison between the datasets.

IV. DATA PROCESSING

In order to process the data efficiently, a processing pipeline has been programmed using the Rust programming language, together with SQL databases making use of Geospatial Information Systems (GIS). The software is structured into two distinct executables, each accompanied by a library and configuration files specific to the FSSCat mission.
Each executable is responsible for specific tasks, as data ingestion, processing of raw data into different levels, and generating the necessary output. This modular approach ensures that the software remains independent, allowing uninterrupted processing even if one of the executables unexpectedly stops. The following paragraphs provide detailed explanations of the two executables, outlining their respective roles and functionalities.
The first executable, in charge of data ingestion, plays a crucial role in the data processing pipeline. Its primary function is to receive and ingest the raw data obtained from the satellites. The software takes the raw data and stores it in a database in SQL. Additionally, the metadata associated with each file is also stored for later processing. This metadata includes information such as the instrument used, the timestamp of the data, geospatial coordinates, and other mission-specific details. Consequently, the data obtained from [8] for FSSCat and from [9], with *netcf4* format, was inserted into the database as part of the first level of data processing. By organizing the ingestion process in this manner, the software effectively handled the initial stages of data storage and metadata integration, preparing the data for subsequent processing.
The second executable is in control of the processing chains for all levels. Table I shows the overall flow of the data processing pipeline. Firstly, the data was formatted into latitude, longitude, and timestamp values, and stored in new tables within the database. As mentioned earlier, only data from October and November was used. Moreover, a specific region was selected, in order to focus the analysis on a smaller, more detailed area. For this study, the chosen region of interest was Europe. Moving on to a higher level of processing, L1, the data underwent a re-gridding of 1°× 1° (100km × 100km). The latter was necessary to reduce the large amount of data obtained from both missions and to optimize the processing time for subsequent levels.

ESA	FSSCat	Processing level
L0	L0	Data from [9] (ESA) and [8] (FSSCat)
L1	L1	Regrid of 1°× 1°
L2		RMSE, BIAS for each lat, lon, day
L2A,L2B		RMSE, BIAS for each lat, lon, month
L3		Mean over time of soil moisture for each lat, lon
L4		Density of data for each pixel

TABLE I: Overall flow of the data processing pipeline explained.

From the latter table, SM products from ESA and FSSCat were spatially and temporally matched. The Bias and Root Mean Square Error (RMSE) values were calculated for each product. RMSE measures the overall SM uncertainty, and bias the overestimation or underestimation of one of the datasets compared to the other. These metrics have been widely used in validation studies and are defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{N}} \tag{1}$$

$$\text{Bias} = \sum_{i=1}^{N} (y_i - x_i) \tag{2}$$

where x is SM FSSCat, y is SM ESA, and N denotes the number of data pairs. Another valuable test necessary for the last analysis was Unbiased Root Mean Square Error (ubRMSE), which is commonly used in statistical analysis and model evaluation to measure the accuracy of predictions or estimates, accounting for bias.

$$\text{ubRMSE} = \sqrt{\text{RMSE}^2 - \text{Bias}^2} \tag{3}$$

The latter step to process the output data was done in *Python* to represent the figures and maps.

V. ANALYSIS AND RESULTS

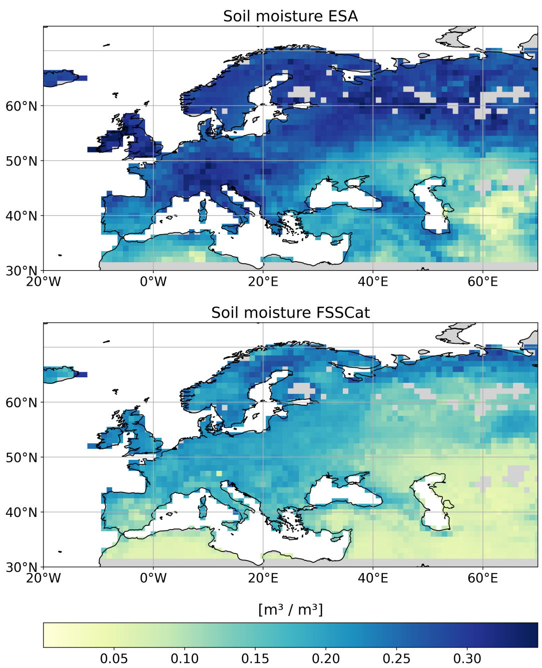

FIG. 1: Averaged soil moisture of the data of ESA and FSSCat with resolution of 1°× 1°. The average has been calculated for all the measurement period. To differ easily between maps, the pixels which FSSCat did not provide sufficient data were not represented in either of the maps.

Figure 1 shows the averaged soil moisture for the months of October and November obtained from ESA and FSSCat respectively. While the values varied among ESA and FSSCat datasets, the consistent pattern persisted across European regions.

In the northern countries, higher soil moisture values were observed, which aligns with [12] who reported increased precipitation in the region. The elevated rainfall can be partially attributed to temperature differences between the colder continental air and the relatively warmer oceanic air in Northern Europe, and also the North Atlantic Oscillation (NAO) in its positive phase, which is more common in winter [13]. Furthermore, the lower temperatures during this season result in reduced evapotranspiration, which contributes to SM accumulation.

Conversely, lower soil moisture values were observed in the southwestern regions. This finding is consistent with the results reported in [12]. Although, these countries are still affected by higher precipitations due to Mediterranean cyclones in autumn, the temperatures are still higher than in the north, contributing to water loss from the soil through evaporation and plant transpiration.

On the other hand, it should be noted that this analysis of FSSCat only focuses on the cold season. Consequently, different outcomes would be expected during the warmest semester, when soil moisture content is generally at its lowest [1]. European summer months are often characterized by higher temperatures, and lower precipitation levels in certain regions, droughts in the south, among other factors. These conditions result in reduced SM content in that period [2].

Furthermore, after analyzing datasets in Figure 1, it was noted that the data from FSSCat is considerably underestimated compared to the SM in ESA CII. To visualize easily the complementarity between the data, scatter plots were represented for each month.

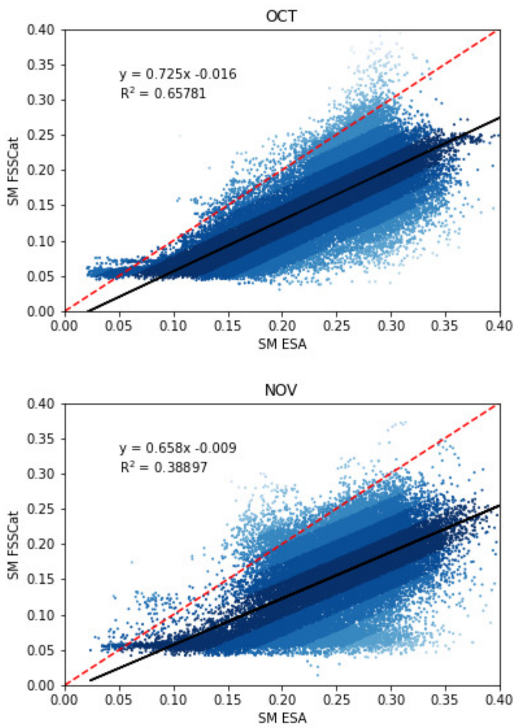

FIG. 2: Comparison of SM of FSSCat in the y-axis and x-axis for ESA CII for the months October and November. The linear regression was also represented.

Figure 2 shows two graphs which exhibit significant differences, indicating that the data for the month of November may be less reliable. This factor must be taken into consideration when conducting the RMSE and Bias tests to identify any inconsistencies. Furthermore, for higher SM values of ESA, the ones for FSSCat are more underestimated.
To facilitate a more comprehensive data analysis RMSE was calculated in Figure 3, where the values have a range of 0.0023 to 0.243 m^3/m^3, with the Alps and Russia being the most affected and discrepant areas compared to the reference values of ESA.

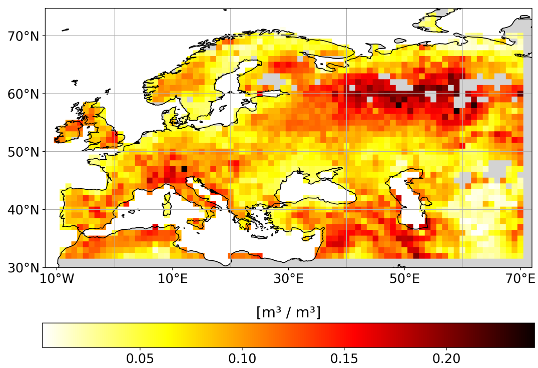

FIG. 3: RMSE, the reference was taken from ESA CII and the compared value from FSSCat. A mask (grey) is also used for the lack of data.

Several hypotheses were considered to find the causality of these disparities. At first, it was concluded that the areas with more ice would be less reliable. One reason attributed to these phenomena would be the combination of MWR and GNSS-R. As GNSS-R relies on the scattering of signals from different types of surfaces to gather information, ice surfaces have a lower land roughness. Land surfaces with varying roughness provide more diverse scattering patterns, which can lead to a richer set of measurements and more detailed information retrieval. This argument was discarded due to the lack of discrepancy in the Scandinavian Peninsula. Although, there is a considerable lack of data for higher latitudes, it is not enough to conclude that surfaces with ice give less trustworthy results. On the other hand, the region below Anatolian Peninsula has also discrepant results in humidity.
Further research was done in order to provide more information about the disparities. The Bias test was represented to find out the overestimated or underestimated areas. Figure 4 confirms our expectations, showing underestimation in most areas, except for regions below the Caspian Sea. As observed in Figure 1, FSSCat generally exhibits lower SM values. However, the region in Russia stands out in Figure 4, displaying a notable bias of -0.2 m^3/m^3 compared to the reference. These results could be attributed to potential instrumental errors or inaccuracies in the measurement devices, as well as the decrease

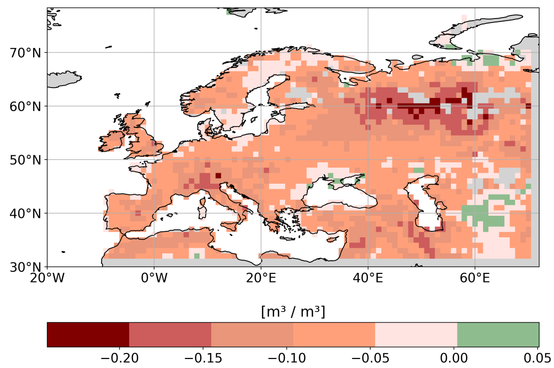

FIG. 4: Bias, between the reference value of ESA and the value obtained from FSSCat

in the amount of data obtained for higher latitudes.
To conclude the analysis, an additional test was performed using ubRMSE [8], a scale-independent metric that enables direct comparison of errors between different datasets. Although, ubRMSE does not consider the variability and range of the measured variable as RMSE, it is more interpretable and consistent for comparing accuracy. Based on Figure 5, the values of ubRMSE range from 0 to 0.1 m^3/m^3. This finding is significant as it indicates a high level of compatibility between the two datasets. It is particularly valuable because regions that showed significant disparities in Figure 3, such as Russia, are not prominent in the ubRMSE analysis. This suggests that the region may exhibit higher variability in data in that region, which is effectively accounted in ubRMSE metric. Lastly, the decrease in the amount of data for the northern region is also directly affected, and leads to more areas not being represented, which is the reason to have more regions of Russia in grey.

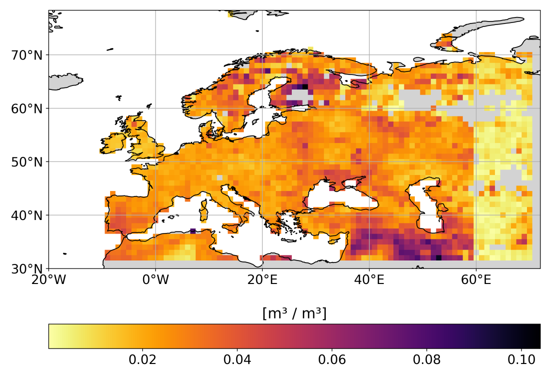

FIG. 5: UbRMSE, the reference value of ESA and the value obtained from FSSCat

VI. CONCLUSIONS

This study has aimed to investigate and validate the soil moisture data obtained from the FSSCat mission by

comparing it with the ESA CII mission. A data processing pipeline was implemented, with steps such as regridding the datasets to a common resolution of 1°× 1°. The latter could lead to some uncertainties within each pixel of the grid, that should be taken into account. The analysis revealed notable underestimation of soil moisture values in the FSSCat dataset, particularly in regions such as the Alps, Anatolian Peninsula, and Russia. These findings can be attributed to various factors, including instrumental errors and the challenges of retrieving soil moisture in certain geographical areas. To study discrepancies within the datasets, the several metrics were used, which emphasized some regions, as the Russia or Anatolian Peninsula, with higher variability. Additionally, a brief examination of soil moisture patterns during the cold season was conducted, revealing higher values in northeastern Europe compared to the southern regions. These findings are consistent with expected SM dynamics in different geographical areas.
Finally, it is important to note that studying SM presents challenges due to its rapid variability within short time intervals. To enhance data validation, this study proposes, for future investigation, expanding the observation periods and focus on days instead of months. Additionally, since there is no dataset with the true real values of SM, comparing multiple datasets with the triple-collocation technique [14] would offer a more unbiased perspective for evaluating SM data.

Acknowledgments

My journey in the UPC Nanosat Lab has been enriching in many ways. My deepest gratitude for Adrián Perez-Portero and Adriano José Camps for the support and patience during my stay in the lab. Lastly, thank you to Yolanda Sola for leading me through the meteorological part of the thesis and the four years of the degree.

[1] Almendra-Martín, L., et al. (2022). Analysis of soil moisture trends in Europe using rank-based and empirical decomposition approaches. Global and Planetary Change, 215, 103868. https://doi.org/10.1016/j.gloplacha.2022.103868
[2] Zampieri, M., et al. (2009). Hot European Summers and the Role of Soil Moisture in the Propagation of Mediterranean Drought. Journal of Climate, 22(18), 4747-4758. https://doi.org/10.1175/2009jcli2568.1
[3] Chew, C., et al. (2020). Description of the UCAR/CU Soil Moisture Product. Remote Sensing, 12(10), 1558. https://doi.org/10.3390/rs12101558
[4] Camps, A., et al. (2021). FSSCat Mission Description and First Scientific Results of the FMPL-2 Onboard 3CAT-5/A. https://doi.org/10.1109/igarss47720.2021.9554589
[5] Piepmeier, J. R., et al. (2017). SMAP L-Band Microwave Radiometer: Instrument Design and First Year on Orbit. IEEE Transactions on Geoscience and Remote Sensing, 55(4), 1954-1966. https://doi.org/10.1109/tgrs.2016.2631978
[6] Zavorotny, V. U., et al. (2000). Scattering of GPS signals from the ocean with wind remote sensing application. IEEE Transactions on Geoscience and Remote Sensing, 38(2), 951-964. https://doi.org/10.1109/36.841977
[7] Camps, A., et al. (2021b). FSSCat Mission Description and First Scientific Results of the FMPL-2 Onboard 3CAT-5/A. https://doi.org/10.1109/igarss47720.2021.9554589
[8] Munoz-Martin, J., et al. (2021). Soil Moisture Estimation Synergy Using GNSS-R and L-Band Microwave Radiometry Data from FSSCat/FMPL-2. Remote Sensing, 13(5), 994. https://doi.org/10.3390/rs13050994
[9] Dataset Record: ESA Soil Moisture Climate Change Initiative (SoilMoisturecci): COMBINED product, Version 07.1. (2023).
[10] Gruber, A., et al. (2019). Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. Earth System Science Data, 11(2), 717-739. https://doi.org/10.5194/essd-11-717-2019
[11] Li, W., et al. (2022). Exploration of Multi-Mission Spaceborne GNSS-R Raw IF Data Sets: Processing, Data Products and Potential Applications. Remote Sensing, 14(6), 1344. https://doi.org/10.3390/rs14061344
[12] Cammalleri, C., et al. (2016). Recent temporal trend in modelled soil water deficit over Europe driven by meteorological observations. International Journal of Climatology, 36(15), 4903-4912. https://doi.org/10.1002/joc.4677
[13] Jaagus, J., et al. (2021). Long-term changes in drought indices in eastern and central Europe. International Journal of Climatology, 42(1), 225-249. https://doi.org/10.1002/joc.7241
[14] Pablos, M., et al. (2021). Correlated Triple Collocation to Estimate SMOS, SMAP and ERA5-Land Soil Moisture Errors. https://doi.org/10.1109/igarss47720.2021.9553378