Comments from the Reviewers and the corresponding revision

SM2RAIN-ASCAT (2007–2018): GLOBAL DAILY SATELLITE RAINFALL FROM ASCAT SOIL MOISTURE

by Luca Brocca, Paolo Filippucci, Sebastian Hahn, Luca Ciabatta, Christian Massari, Stefania Camici, Lothar Schüller, Bojan Bojkov, Wolfgang Wagner

We thank the reviewers for their appreciation of our study and for the valuable suggestions that helped us to clarify and improve the manuscript. A detailed answer to each comment is reported in the sequel clarifying the procedure used for developing the global SM2RAIN-ASCAT data record and the obtained results. In Italic, we have reported the reviewers comments, in blue, the detailed replies, and in red, the text changed and/or added in the revised manuscript to address reviewers' comments.

As general reply to all the reviewers, we would like to underline that the paper goal is to present and describe the SM2RAIN-ASCAT global rainfall data record and to perform a comparison with state-of-the-art global rainfall products. We do not want to show a comprehensive assessment of the product. Indeed, we believe that researchers other than the product developers should perform the assessment and the validation of the dataset. This clarification has been added in the revised manuscript at the end of the Introduction section (see lines 154-160):

"We underline that the paper goal is to present and describe the SM2RAIN-ASCAT quasi-global rainfall data record and to perform a comparison with state-of-the-art global rainfall products. We do not want to show a comprehensive assessment of the product. Indeed, we believe that researchers other than the product developers should perform the validation of the dataset. Even better, we stress the importance of performing the validation by using the datasets in hydrological or agricultural applications (e.g., flood prediction and agricultural water management)."

Indeed, the dataset is made freely available and a first paper has been already published by Paredes-Trejo et al., 2019 (doi:10.3390/rs11091113) who have assessed the accuracy of the SM2RAIN-ASCAT data record in Brazil. Even better, we stress the importance of performing the validation by using the data record in hydrological or agricultural applications. The comparison with raingauges or any reference dataset could be misleading, mainly when the rainfall products include the ground observed information used for their derivation. Another independent paper has been just submitted by Mazzoleni et al. (2019) who have performed the hydrological validation of SM2RAIN-ASCAT in 8 large basins worldwide showing that the product outperforms all the other satellite-only rainfall dataset. The paper preprint is available on EarthArXiv at https://eartharxiv.org/v2r7c/.

Anonymous Referee #1

This study provides detailed descriptions of the SM2Rain product and several evaluation results. Overall, the study would be useful for current and future SM2Rain users, and fits the scope of ESSD. However, I do find that the manuscript misses several key information in SM2Rain production and evaluation.

We thank the reviewer for her/his appreciation of our study and for the valuable suggestions that helped us to clarify and improve the manuscript. A detailed answer to each comment is reported in the sequel clarifying better the procedure used for developing the global SM2RAIN-ASCAT product.

1. Line 24 – 29: The statement here is too strong. I agree that SM2Rain is a useful product in some aspects. However, I have not seen strong evidences that SM2Rain substantially outperforms other merged products, e.g., MSWEP v2.0. Additionally, soil moisture retrievals prior 2002 have very low data quality. I personally doubt if good precipitation can be derived from these soil moisture data sets. Hence, I also suspect whether SM2Rain "is suited to build long-term consistent rainfall".

The reviewer is right; the SM2RAIN-ASCAT rainfall product will hardly outperform merged products, mainly if the comparison with raingauges is carried out (see the general answer above). Apart the possibility to include SM2RAIN-ASCAT in merged products, we believe that a strong added-value of SM2RAIN-ASCAT is its expected availability in the next 25 years, with already 12 years of data available, and its independence with respect to the others state-of-the-art satellite rainfall products (e.g., GPM IMERG, PERSIANN, CMORPH). The sentence was misleading as we intended to say that a long-term SM2RAIN-ASCAT dataset, starting from 2007, and ensured until mid-2040s, can be built with the proposed approach. The sentence has been modified in the revised manuscript, accordingly (see lines 29-33):

“We exploit here the Advanced SCATterometer (ASCAT) on board three Metop satellites, launched in 2006, 2012 and 2018, as part of the EUMETSAT Polar System programme. The continuity of the scatterometer sensor is ensured until mid-2040s through the Metop Second Generation Programme. By applying SM2RAIN algorithm to ASCAT soil moisture observations, a long-term rainfall data record will be obtained, starting in 2007 until mid-2040s.”

2. Line 147: A global map of ASCAT temporal sampling frequency would be helpful.

A global map of ASCAT temporal sampling frequency has been added in the Appendix (see figure A1 and lines 218-220):

“(see ***Figure A1*** for the mean daily revisit time of ASCAT in the period 2007-2012 with only Metop-A and the period 2013-2018 with Metop-A+B)”

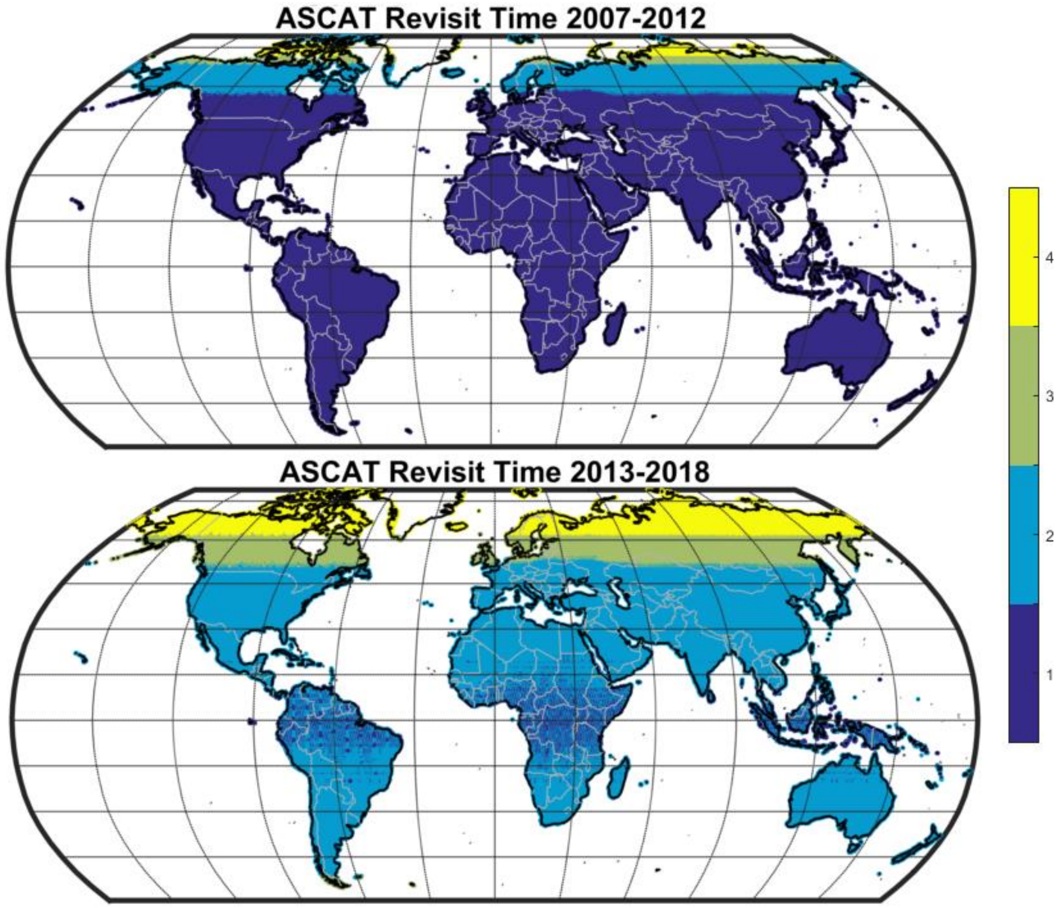


Figure A1. Mean daily revisit time [days] of ASCAT soil moisture observations for the period 2007-2012 (only Metop-A, top panel) and for the period 2013-2018 (Metop-A+B, bottom panel).

3. Line 201 – 203: I'm wondering if there are any risks of increasing false rainfall events by linear interpolation?

The reviewer is right; linear interpolation may increase the risk of false rainfall events, and future research will be addressed to mitigate this problem. The comment has been added in the revised manuscript at lines 222-223:

"The interpolation may increase the risk of false rainfall events, but it is a required step to obtain accumulated rainfall over a fixed duration."

4. Line 242: The authors state that runoff at 20km grid is negligible. Can you provide some rainfall-runoff simulation works to support this hypothesis?

With the runoff assumption, we are saying that surface runoff is expected to be negligible at larger spatial scales due to the possibility that locally generated surface runoff (e.g., over impervious surfaces) can re-infiltrate into more permeable areas in the same pixel. Of course, this hypothesis can be not valid in some areas, but we have indirectly validated this hypothesis as we have hardly seen the ASCAT soil moisture signal to be saturated for more than one day. Therefore, surface runoff due to saturated soil is expected to occur very rarely at 20 km scale. This aspect has been clarified in the revised manuscript showing the number of days the ASCAT soil moisture signal is saturated for more than one day (see lines 269-277):

"We have indirectly tested this hypothesis by counting the number of days the ASCAT soil moisture product is higher than 99.5 percentile for two (or more) consecutive days in the period 2007-2018. We have found that the number of consecutive days in which the soil is saturated is equal to 4 days (median value on a global scale) over 12 years, with 90% of land pixels with values lower than 12 days (i.e., 1 day per year). The occurrence of higher values is limited to very few areas in the tropical forests and over Himalaya (see Figure A2)."

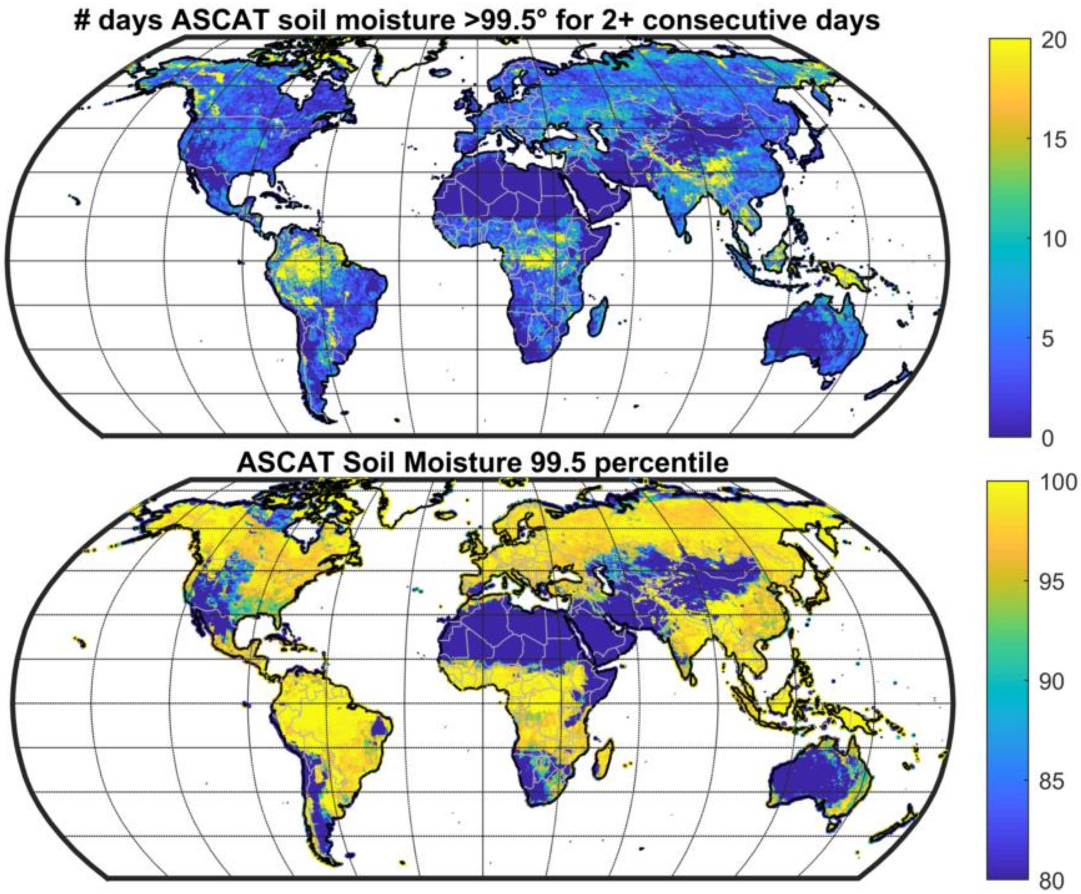


Figure A2. Number of days in which ASCAT soil moisture observations are close to saturation (>99.5 percentile, top panel) for 2 (or more) consecutive days in the period 2007-2018. Please note that the upper value is set to 20 days as in most of land areas the occurrence is very low (90% of land pixel with values lower than 12 days over 12 years). In the bottom panel the soil moisture values at 99.5 percentile (in the period 2007-2018) are shown.

5. Line 249: I'm a little bit confused by equation 2. First, the authored stated that ET is negligible. Then, why it is still considered in equation 2? Second, it seems g(t) and e(t) should be plus in sign, according to equation (1)?

The reviewer is right; g(t) and e(t) must be plus in sign, thanks for spotting the error that we have corrected in the revised manuscript (see line 281). In previous applications of SM2RAIN, we have assumed ET negligible, during rainfall, but in this study we wanted to test the possibility to include the

ET term and to assess its impact for rainfall estimation through SM2RAIN. For that, we left the e(t) component in equation (2) but we have used this formulation only at the analysis over 1009 points.

6. Line 261: e(t) is calculated using ERA5 ET. The ERA5 ET is expected to depend on ERA5 precipitation. For instance, a dry period seen by ERA5 (precipitation deficiency) will lead to low ET. Therefore, the authors should discuss the dependency of ERA5 and SM2Rain rainfall product, particularly when TC is considered in the later part of the paper.

We agree with the reviewer, some dependencies between ERA5 ET and precipitation may occur. However, we underline that in the selected configuration (see lines 434-436 in the revised manuscript) the ERA5 ET is not used and, hence, this dependency is excluded. Moreover, in Triple Collocation Analysis application we didn't consider ERA5, to avoid any dependency between the products. This point has been underlined better in the revised manuscript (see lines 509-510):

"In TC analysis we have not considered ERA5 purposely to avoid any dependency between the products."

7. Line 291: What is the reference rainfall?

The reference rainfall is the one used for the calibration of SM2RAIN parameter values and the climatological correction factor. In the section "4.1 Selection of the best SM2RAIN processing configuration at 1009 points", we have used ground-based rainfall observations as reference and it has been clarified in the revised manuscript (see lines 375-377):

"The ground-based high quality rainfall observations available for the four regions are used as reference data (ground truth) in this analysis."

As stated at lines 440-443, for the global SM2RAIN-ASCAT dataset production we have used ERA5 rainfall dataset as reference.

"As reference dataset for calibrating the filtering, SM2RAIN, and post-processing parameter values, the ERA5 rainfall has been used mainly because of its higher spatial resolution compared to GPCC (36 km versus 100 km)."

And in the Introduction at lines 146-153:

"As reference datasets we have used high-quality local raingauge networks from 2013 to 2017 in the United States, Italy, India and Australia for the assessment at 1009 points and for the regional assessment. Three additional global datasets have been considered: the latest reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5, the gauge-based Global Precipitation Climatology Centre (GPCC), and the GPM IMERG product (Early Run version). ERA5 has been used for the generation of the quasi-global SM2RAIN-ASCAT data record; GPCC and GPM IMERG have been considered for the TC analysis."

8. Section 3.4: Please also specify the error model used in this TC analysis.

As in Massari et al. (2017), we have used an additive error model in TC analysis, it has been clarified in the revised manuscript at lines 337-340:

"In this study, we have implemented the same procedure as described in Massari et al. (2017), i.e., by implementing an additive error model at daily time scale, and we refer the reader to this study for the analytical details."

9. Section 3.5: Equations of these scores will be helpful here.

The equations have been added in the Table A1 of the Appendix of the revised manuscript.

"For a complete description of the performance scores, see Table A1 in the Appendix."

Table 1. Equations used for the performance scores. For the continuous scores, P_{ref} is the reference dataset (e.g., ground observations, ERA5) and P_{est} is the estimated dataset (e.g., SM2RAIN-ASCAT, GPM-ER), *cov* is the covariance operator, σ is the standard deviation operator, $\sum$ is the summation operator, and *N* is the sample size. For the categorical scores, *H* is the number of successfully predicted events by a given rainfall product, *F* the number of non-events erroneously predicted to occur, and *M* the number of actual events that are missed.

Performance Score	**Score symbol**	**Equation**
	Continuous scores	
Pearson's correlation	R	$R = \frac{cov(P_{est}, P_{ref})}{\sigma(P_{est})\sigma(P_{ref})}$
Root Mean Square Error	RMSE	$RMSE = \sqrt{\frac{\sum(P_{est} - P_{ref})^2}{N}}$
Temporal Variability Ratio	STDRATIO	$STDRATIO = \frac{\sigma(P_{est})}{\sigma(P_{ref})}$
Bias	BIAS	$BIAS = \frac{\sum(P_{est} - P_{ref})}{N}$
	Categorical scores	
False Alarm Ratio	FAR	$FAR = \frac{F}{H + F}$
Probability of Detection	POD	$POD = \frac{H}{H + M}$

Threat Score	TS	$TS = \frac{H}{H+F+M}$

10. Line 367: I'm still not clear which product is used as a reference to correct SM2Rain.

In the section "4.1 Selection of the best SM2RAIN processing configuration at 1009 points", the ground-based rainfall observations are used as reference. Differently, in the sections "4.2 Generation of SM2RAIN-ASCAT dataset" and "4.3 Regional and global assessment of SM2RAIN-ASCAT dataset", the ERA5 rainfall is used as reference. It has been clarified better in the revised manuscript (see reply to comment (7)).

11. Section 4.1 and Line 385: It's un clear how SM2Rain parameters were calibrated (determined) and extended to the global scale.

SM2RAIN parameter values are calibrated point-by-point by using the reference rainfall as target (see reply to comment 10 for the definition of reference rainfall). As objective function we have used the minimization of the RMSE between SM2RAIN-ASCAT and reference rainfall datasets. There is no linkage between the local scale and global scale calibration, as different reference rainfall and data periods are used in the two calibrations. It has been clarified in the revised manuscript (see lines 312-314):

"SM2RAIN parameter values are calibrated point-by-point by using the reference rainfall as target. As objective function, we have used the minimization of the RMSE between SM2RAIN-ASCAT and reference rainfall."

12. Figure 5: SM2Rain is calibrated against ERA5. Therefore, the consistency of ERA5 and SM2Rain only suggests how well ASCAT was fitted to ERA5. The authors should be clear that this is not suggesting the accuracy or the performance of SM2Rain (Line 414 – 415).

The reviewer is right; we have removed the terms "performance" and "accuracy" from this section to avoid misunderstanding (see lines 461-478). Of course, we expect better performance in the areas in which the consistency is higher, but the preliminary assessment of SM2RAIN-ASCAT is performed in section 4.3. We have underlined in the revised manuscript that Figure 5 shows the consistency of ERA5 and SM2RAIN-ASCAT (see lines 462-466):

"Therefore, ***Figure 5*** illustrates the consistency between SM2RAIN-ASCAT and ERA5, and it is not intended to assess the performance of the data record (even though we expect better accuracy in areas where the agreement is higher)."

13. Line 431 – 433: SM2Rain show better performances relative to which product? It seems that SM2Rain's R is much lower than the other three in Figure 6 a and b.

Here, we wanted to highlight the regions where SM2RAIN-ASCAT is performing better, not with respect to other products, but only across to different regions (note that we have added different performance metrics in the Figure 6 of the revised manuscript). It has been clarified in the revised manuscript (see lines 484-488):

"By focusing on the SM2RAIN-ASCAT data record performance over the different regions, it shows better performance in Italy (median R=0.67) and United States (median R=0.62), almost comparable with the other datasets; while in Australia and India R-values are lower (median R=0.61 and 0.59)."

14. Following the comment above, SM2Rain was derived by calibrations against ERA5. However, its performances are consistently lower than ERA5. Then, what's the contribution/ value of SM2Rain?

There are several important differences between SM2RAIN-ASCAT and ERA5. The most important difference is the possibility to provide SM2RAIN-ASCAT rainfall in near real time (e.g., with latency lower than 6 hours), while ERA5 is provided with a latency of weeks. Therefore, SM2RAIN-ASCAT can be used in many applications that require rainfall data with short latency, whereas ERA5 (or GPCC) cannot be used. It has been underlined at lines 492-494:

"We highlight also that differently from SM2RAIN-ASCAT and GPM-ER, GPCC and ERA5 have a latency of weeks to months and, hence, these products cannot be used for near real time applications."

Moreover, we should underline that ERA5 is using ground observations and in the regions analysed in Figure 6 a dense coverage of ground stations is available. Differently, in poorly gauged areas (e.g., Africa and South America) a lower performance of ERA5 might be expected.

15. Line 446: What products are used for TC analysis? Massari used ERA. However, I don't think this is appropriate for this study. SM2Rain here is calibrated against ERA, and they may have cross-correlated errors.

The reviewer is right; we didn't use ERA5 but GPCC, GPM Early Run and SM2RAIN-ASCAT as stated at lines 507-510:

"On a global scale, the TC approach has been implemented by using the triplet SM2RAIN-ASCAT, GPM-ER and GPCC, by considering the common period 2015-2018. In TC analysis we have not considered ERA5 purposely to avoid any dependency between the products."

Anonymous Referee #2

An update of a satellite soil moisture-based rainfall dataset (SM2RAIN-ASCAT) is presented. The paper is fairly well written but paints an overly rosy picture of the dataset. Both the dataset and the validation exhibit a number of serious issues which must be addressed before the paper can be published.

We thank the reviewer for the valuable suggestions that helped us to clarify and improve the manuscript. A detailed answer to each comment is reported in the sequel.

(1) The peak underestimation issue has not been resolved in the current release of the dataset, as revealed by both the low STDRATIO values (Figure 3) and the time series comparison (Figure 4). This major issue has been highlighted in two large precipitation dataset evaluations that have been ignored in the present study (https://www.hydrol-earth-syst-sci.net/21/6201/2017/ and https://www.hydrolearth-syst-sci.net/23/207/2019/). It is important that previously identified issues are addressed or at least discussed.

The reviewer is right; SM2RAIN-ASCAT has underestimation issue that has not been resolved completely. It has been clearly underlined in the revised manuscript (see below). However, we want to stress that the climatological correction partly addresses this issue. A more specific CDF correction can be used for addressing the target (e.g., daily CFD matching), but we have preferred not to implement to avoid a strong dependency between SM2RAIN-ASCAT rainfall data and the reference dataset (indeed, also the reference might be wrong, particularly in poorly gauged regions).

Lines 48-51:

"Limitations of SM2RAIN-ASCAT data record consist in the underestimation of peak rainfall events, in the occurrence of spurious rainfall events due to high frequency soil moisture fluctuations that might be corrected with more advanced bias correction techniques."

Lines 387-390:

"Very good statistics have been obtained in terms of RMSE and BIAS but a tendency to underestimate the observed rainfall variability (median STDRATIO=0.60) and medium-high probability of false alarm (median FAR=0.53). The other scores are similar, or slightly lower than those obtained through GPM-ER."

Lines 501-506:

"As shown also in ***Figure 3***, the SM2RAIN-ASCAT data record has limitations in reproducing the variability of rainfall (low STDRATIO) mainly due underestimation issues. Moreover, FAR values of SM2RAIN-ASCAT are quite high highlighting the difficulties in removing the problem of high frequency soil moisture fluctuations wrongly interpreted by SM2RAIN as rainfall events."

Lines 552-555:

"Limitations of SM2RAIN-ASCAT data record consist in: 1) the underestimation of peak rainfall events, 2) the presence of spurious rainfall events due to high frequency soil moisture fluctuations, 3) the estimation of liquid rainfall only (snowfall cannot be estimated), and 4) the possibility to estimate rainfall over land only."

(2) The CDF correction is based on the REF data and is thus not independent, giving the dataset an unfair advantage compared to GPM-ER in Figure 3.

In figure 3, we have used GMP-ER as a state-of-the-art reference, not to perform a comparison between the datasets. As mentioned above, the paper is not intended to perform a comprehensive assessment of SM2RAIN-ASCAT dataset, or its comparison in terms of accuracy with respect to other products. We only want to show that SM2RAIN-ASCAT is performing similarly to state-of-the-art products and, hence, can be a valuable alternative for applications using rainfall observations as input. It has been specified better in the revised manuscript at lines 154-160:

"We underline that the paper goal is to present and describe the SM2RAIN-ASCAT quasi-global rainfall data record and to perform a comparison with state-of-the-art global rainfall products. We do not want to show a comprehensive assessment of the product. Indeed, we believe that researchers other than the product developers should perform the validation of the dataset. Even better, we stress the importance of performing the validation by using the datasets in hydrological or agricultural applications (e.g., flood prediction and agricultural water management)."

(3) The RMSE metric should not be used for the evaluation of precipitation datasets at daily time scales as it yields misleading results (makes it seem datasets with underestimated peaks such as SM2RAIN are better). This is due to the high skewness of the precipitation distribution and the prevalence of temporal mismatches between estimated and observed precipitation peaks. The problem is illustrated in the paper by Figure 3, which shows a higher RMSE value (i.e., "worse" performance) for the bias- and CDF-corrected SM2RAIN product (BC-CDF) than for any of the uncorrected SM2RAIN products.

We agree with the reviewer that the RMSE statistic has some limitations in evaluating precipitation datasets. Indeed, we have used different statistics and in the revised manuscript we have performed the evaluation with multiple statistics also for Figure 6. Anyhow, RMSE is used in many papers evaluating precipitation datasets (and we don't believe they are all wrong), and it suffers from the same limitation of any single score; an assessment by using multiple scores is needed. As mentioned above, we believe that the real validation should be performed using the rainfall products in the hydrological or agricultural applications. These aspects have been underlined in the revised manuscript as shown in the reply of comment (3).

(4) Only correlation and RMSE statistics are presented for the performance evaluation in Figure 6. Please remove the RMSE for the previously mentioned reason and add other metrics, such as variability ratio, bias, hit/miss ratio, frequency of wet days, peak magnitude, etc. for a more thorough performance evaluation.

In the revised manuscript, we have added multiple statistics in Figure 6, similarly to Figure 3.

(5) The TC evaluation only takes into account the monthly correlation – just one aspect of dataset performance (monthly temporal dynamics). Hence the TC evaluation alone cannot be used to conclude

whether a particular dataset is better or worse (as is done in the last paragraph of the abstract: "SM2RAIN-ASCAT dataset provides better performance better than GPM and GPCC in the data scarce regions of the world"). Other aspects should also be considered.

TC analysis is performed at daily time scale, not monthly time scale. Therefore, we believe TC analysis provides information on the accuracy of the different rainfall products at daily time scale, it has been clarified in the revised manuscript at lines 507-509:

"On a global scale, the TC approach has been implemented by using the triplet SM2RAIN-ASCAT, GPM-ER and GPCC, by considering the common period 2015-2018 and at daily time scale."

(6) "The recent "bottom up" approach that uses satellite soil moisture observations for estimating rainfall through the SM2RAIN algorithm is suited to build long-term and consistent rainfall data record as a single polar orbiting satellite sensor is used." If this is true, why does the dataset span such a short period (2007-2018)? All datasets listed in Table 1 (excluding IMERG) span a longer period. This statement should be revised.

The statement has been revised as we intended to say that a long-term SM2RAIN-ASCAT dataset, starting from 2007, and ensured until mid-2040s, can be built based the proposed approach. Sorry for the misunderstanding that has been corrected in the revised manuscript (see lines 29-33):

"We exploit here the Advanced SCATterometer (ASCAT) on board three Metop satellites, launched in 2006, 2012 and 2018, as part of the EUMETSAT Polar System programme. The continuity of the scatterometer sensor is ensured until mid-2040s through the Metop Second Generation Programme. By applying SM2RAIN algorithm to ASCAT soil moisture observations, a long-term rainfall data record will be obtained, starting in 2007 until mid-2040s."

(7) On a related note, the evaluation of https://www.hydrol-earth-systsci.net/21/6201/2017/ (co-authored by the first author of the present study) shows that SM2RAIN-ASCAT performs worst among all precipitation datasets in terms of trend, due to the combination of data from different ASCAT sensors. So are the different ASCAT sensors consistent with each other or not? Has this trend issue been resolved in this SM2RAIN-ASCAT release? If so, this should be shown. If not, this should be communicated to the reader.

The trend issue has been solved as in the previous delivery of the SM2RAIN-ASCAT dataset (preliminary distribution) we did not consider appropriately the availability of two ASCAT sensors (Metop-A and -B) after 2013. The dual calibration performed in this study (see lines 448-450) has been carried out exactly to address this issue. It has been clarified in the revised manuscript at lines 450-452:

"The dual calibration has solved the issue in terms of long-term trend that has been found in previous application of SM2RAIN to ASCAT soil moisture data (Beck et al., 2017)."

(8) In the interest of transparency the abstract should mention that the presented SM2RAIN dataset i) is limited to liquid precipitation (snowfall is not present in the dataset), ii) exhibits spurious drizzle, iii) underestimates extremes (as demonstrated by Figures 3 and 4 of the paper), and iv) potentially suffers

from intercalibration issues (see comment (7)). If any of these problems have been fixed in the current release of SM2RAIN-ASCAT, this should be shown in the paper.

As suggested by the reviewer, we have clearly communicated the limitations of SM2RAIN-ASCAT dataset in the abstract of the revised manuscript. Limitations and strengths of the SM2RAIN-ASCAT dataset have been clearly illustrated (see replies to comment (1)). In the interest of transparency, we have made the SM2RAIN-ASCAT product freely available, and also the dataset at 1009 points that we have used for selecting the best configuration to develop the product. All the input and test datasets used in the paper are freely available and the analysis can be easily performed by the reader (note that also SM2RAIN code is made available on GitHub).

(9) "The limitations of the bottom up approach are the possibility to estimate only terrestrial rainfall and its dependence on land characteristics (e.g., low accuracy for dense vegetation coverage and complex topography, Broccaet al., 2014)." The other limitations (spurious drizzle, underestimation of extremes, and intercalibration issues) should also be mentioned here.

Limitations and strengths of the SM2RAIN-ASCAT dataset have been clearly illustrated in the revised manuscript (see replies to comment (1)).

(10) To my understanding the regional evaluation is performed using daily accumulations, while the triple collocation (TC) analysis is performed using monthly accumulations – correct? To avoid confusion, please state the time scale of each specific evaluation/analysis in both the abstract and the captions of all figures.

All the analyses have been performed at daily time scale and it has been clarified in the revised manuscript (lines 507-509):

"On a global scale, the TC approach has been implemented by using the triplet SM2RAIN-ASCAT, GPM-ER and GPCC, by considering the common period 2015-2018 and at daily time scale."

(11) Version numbers should be assigned to the different SM2RAIN-ASCAT releases, to avoid confusion. I know there have been at least two releases. Which one is this?

The first official version of the SM2RAIN-ASCAT dataset should be considered the one presented in this paper. Indeed, the dataset has been published on Zenodo and a DOI (digital object identifier) has been assigned to the dataset to avoid confusion.

(12) Please add ERA5 to Figure 3 and make it easier to see the differences among the boxes, either by reducing the range of the y-axes or by exanding the size of the y-axes.

In the revised manuscript, we have added ERA5 and we have also improved figure readability.

(13) The intro/methods part of the abstract is a bit too long, while the results/discussion part is a bit too short (just three sentences).

The abstract of the revised manuscript has been revised accordingly.

(14) "the surface runoff rate, i.e., the water that does not infiltrate into the soil and flows at the surface to the watercourses, is much lower than the rainfall rate, mainly if equation (1) is applied at coarse spatial resolution (20 km), i.e., with satellite soil moisture data." This statement does not make sense to me. Runoff can be equal to rainfall if the soil is saturated, at all scales – from hillslope to catchment.

With the runoff assumption, we are saying that surface runoff is expected to be negligible at larger spatial scales due to the possibility that locally generated surface runoff (e.g., over impervious surfaces) can re-infiltrate into more permeable areas in the same pixel. Of course, this hypothesis can be not valid in some areas, but we have indirectly validated this hypothesis as we have hardly seen the ASCAT soil moisture signal to be saturated for more than one day. Therefore, surface runoff due to saturated soil is expected to occur very rarely at 20 km scale. This aspect has been clarified in the revised manuscript showing the number of days the ASCAT soil moisture signal is saturated for more than one day (see lines 269-277):

"We have indirectly tested this hypothesis by counting the number of days the ASCAT soil moisture product is higher than 99.5 percentile for two (or more) consecutive days in the period 2007-2018. We have found that the number of consecutive days in which the soil is saturated is equal to 4 days (median value on a global scale) over 12 years, with 90% of land pixels with values lower than 12 days (i.e., 1 day per year). The occurrence of higher values is limited to very few areas in the tropical forests and over Himalaya (see Figure A2)."

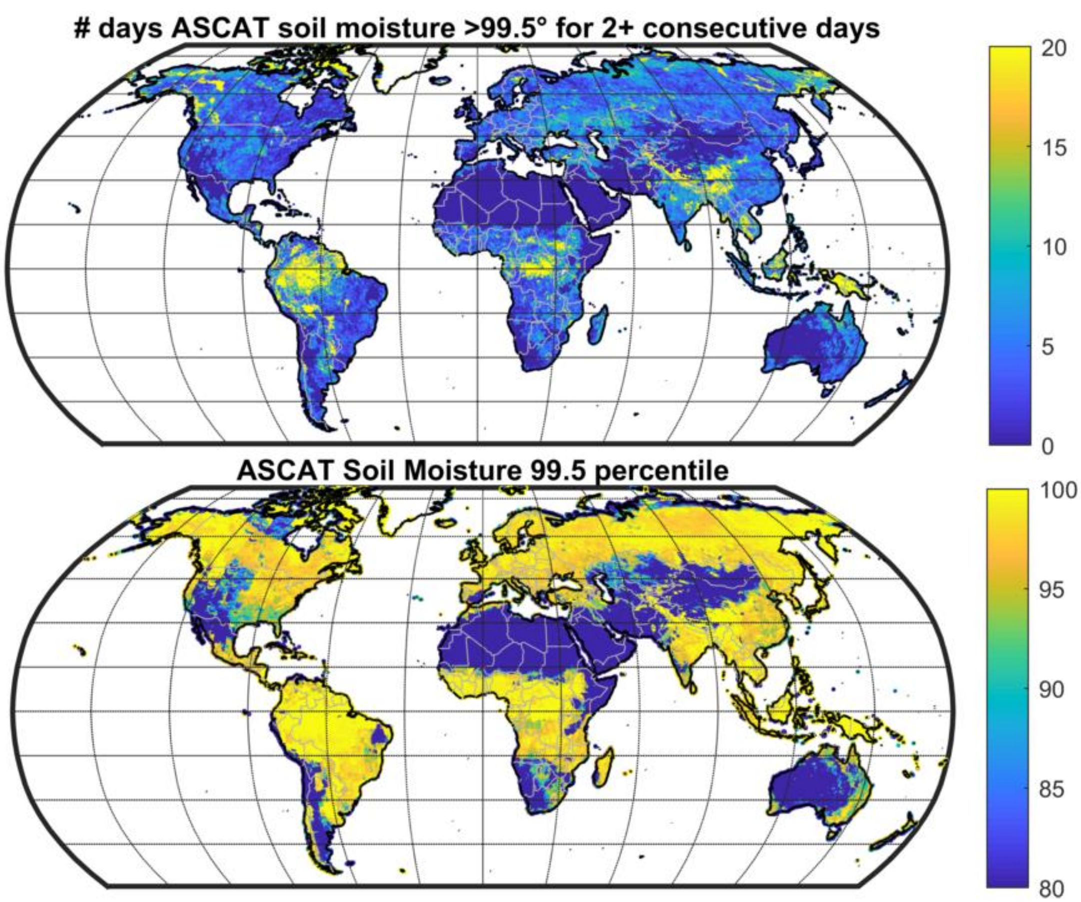


Figure A2. Number of days in which ASCAT soil moisture observations are close to saturation (>99.5 percentile, top panel) for 2 (or more) consecutive days in the period 2007-2018. Please note that the upper value is set to 20 days as in most of land areas the occurrence is very low (90% of land pixel with values lower than 12 days over 12 years). In the bottom panel the soil moisture values at 99.5 percentile (in the period 2007-2018) are shown.

Anonymous Referee #3

This study provides the descriptions and validation results of the 11-year (2007-2018) SM2RAIN-ASCAT global rainfall dataset. Overall, the study fits the scope of ESSD, the paper is well written and the presentation quality is very good. I think that the dataset has great potentials for different applications, as also stated by the authors, especially in specific regions of the world where it seems to outperform other products based on different approaches (e.g., GPM IMERG Early Run). However, there are several aspects that need to be addressed before the paper can be published.

We thank the reviewer for her/his appreciation of our study and for the valuable suggestions that helped us to clarify and improve the manuscript. A detailed answer to each comment is reported in the sequel.

Line 34: What do the author mean by"operationally available in NRT"? This is not a crucial aspect for the dataset presented in this paper.

The reviewer is right; the sentence has been removed by the abstract of the revised manuscript. It is due to a parallel activity we are performing for producing a NRT SM2RAIN-ASCAT product.

Line 36: It is important to note that it is not global as it does not provide rainfall over water bodies, and it is limited to the avalability and quality of soil mosture data. This should be clearly stated also in the conclusions.

The reviewer is right; SM2RAIN-ASCAT product is not global and it has been clarified in the abstract and in the conclusions of the revised manuscript (see lines 34-36):

"The paper describes the recent improvements in data pre-processing, SM2RAIN algorithm formulation, and data post-processing for obtaining the SM2RAIN-ASCAT quasi-global (only over land) daily rainfall data record at 12.5 km sampling from 2007 to 2018."

And lines 552-555:

"Limitations of SM2RAIN-ASCAT data record consist in: 1) the underestimation of peak rainfall events, 2) the presence of spurious rainfall events due to high frequency soil moisture fluctuations, 3) the estimation of liquid rainfall only (snowfall cannot be estimated), and 4) the possibility to estimate rainfall over land only."

Line 42: Please, specify "the IMERG Global Precipitation Measurement (GPM) mission product

The text has been modified, accordingly (see lines 38-42):

"Moreover, an assessment on a global scale is provided by using the Triple Collocation technique allowing us also the comparison with the latest ECMWF reanalysis (ERA5), the Early Run version of the Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG), and the gauge-based Global Precipitation Climatology Centre (GPCC) products."

Line 75: Rainfall is not "measured" from space. Precipitation retrieval based on "topdown" approaches is very complex due to interaction of the radiation emitted by the Earth's surface with gases, and liquid *and solid hydrometeors within the clouds. For example, passive microwave retrieval techniques need to account for variability of all these elements (e.g., surf. emissivity and temperature, water vapour content, cloud water content, sizes, shapes, density, 3-D distribution of liquid, solid and mixed-phase hydrometeors).*

The term "measured" will be removed from the revised manuscript, even though it's matter of terminology. Every measurement is affected by errors. The text has been changed (see lines 76-78):

"The standard methods for estimating rainfall from space are based on instantaneous measurements obtained from microwave radiometers, radars, and infrared sensors (Kidd and Levizzani, 2011)."

Line 76-78: Please, rephrase this sentence: "these methods are based on inversion techniques where the upwelling radiation (or backscattered signal for radars) is related to the surface precipitation rate".

Accordingly, the sentence has been revised (see lines 78-81):

"These methods are based on inversion techniques where the upwelling radiation (or backscattered signal for radars) is related to the surface precipitation rate, i.e., a "top down" approach (Brocca et al., 2014)."

Line 133 (and Line 176, and Line 327): Please, clarify what you mean by "1009 points". Are these 12.5km x1.2km grid boxes? What do you mean by "uniformly distributed? How have they been selected? How many "points" are selected in each region? How are the raingauge measurements treated to be associated to each "point"?

The 1009 points are uniformly distributed over a regular grid with spacing of 1.5°. Each point is considered representative of a 0.25° x 0.25° box; the selection is carried out for reducing the computational time for running the different SM2RAIN configurations. The numbers of points for each region is based on the size of the region (328 points in Australia, 163 in India, 55 in Italy, and 463 in the United States). Ground observations and GPM-ER data are regridded by spatial averaging measurements contained over each 0.25° x 0.25° box. All these details have been reported in the revised manuscript (see lines 365-371):

"We have selected 1009 points uniformly distributed over a regular grid with spacing of 1.5°. Each point is considered representative of a 0.25° x 0.25° box. The selection is carried out for reducing the computational time in running the different SM2RAIN configurations. The numbers of points for each region is depending on the size of the region: 328 points in Australia, 163 in India, 55 in Italy, and 463 in the United States. Ground observations, GPM-ER and ERA5 data are regridded by spatial averaging measurements contained over each 0.25° x 0.25° box."

Line 203-205: it is not clear how the 12 hour sampling of the ASCAT soil moisture product is used to obtain the daily (24 hour) SM2RAIN rainfall prouct.

The 24-hour accumulated rainfall is obtained by summing the two 12-hour accumulated rainfall data obtained for each day, it has been specified in the revised manuscript at lines 227-229:

"The 24-hour accumulated rainfall is obtained by summing the two 12-hour accumulated rainfall data obtained for each day."

Line 282-284: correction of the overall bias can be very effective for mitigating errors in all products. It should be pointed out by the authors if SM2RAIN-ASCAT dataset presented in this paper is the same product that would be obtained operationally in NRT (see also Line 34). If this is not the case, in my opinion, for a fair comparison, the IMERG GPM Final Run should be used instead of the Early Run in this study. Otherwise, the authors should explain clearly why the GPM Early Run is used in this study. Although I understand that IMERG Final Run can not be used for TC, I recommend to show the results of SM2RAIN-ASCAT compared to IMERG Final Run.

The SM2RAIN-ASCAT dataset presented in the paper is the same product that would be obtained operationally in NRT. The climatological correction is performed with constant parameter values and, hence, it can be implemented in NRT. We note that a climatological correction is performed in several satellite rainfall datasets delivered in NRT (e.g., 3B42RT, IMERG ER, PERSIANN CCS, CMORPH CRT). It has been clarified in the revised manuscript at lines 319-322:

"Specifically, we refer here to a static correction procedure that once calibrated for a time period can be applied in the future periods, also for operational real time productions. We note that a climatological correction is performed in several satellite rainfall datasets delivered in near real-time (e.g., GPM Early Run)."

Line 314: Optimal value for FAR is 0, not 1. Please, correct.

The reviewer is right; we have corrected the error in the revised manuscript, thanks for spotting the mistake.

Line 316: Please, motivate the choice of 0.5 mm/day (and not a lower value > 0 mm/day) as rainfall/no rainfall threshold.

As mentioned in the manuscript, the threshold is selected in order to exclude spurious events that might be due to rainfall interpolation\regridding in the reference datasets.

Line 379-380: How many points are used to compute these averages in each region? Are "problematic" areas for soil moisture retrieval (complex orography, highly vegetated, ecc.) included among the 1009 points used here?

All points in each region are used, i.e., 328 points in Australia, 163 in India, 55 in Italy, and 463 in the United States. The "problematic" areas are included as 1009 points are randomly selected; no masking has been carried out in this analysis.

Line 389: Why R and RMSE are considered "more important"? Please, justify this choice.

We believe that R and RMSE are the two most important statistics for evaluating precipitation datasets after performing several assessment studies of different datasets. However, we acknowledge that the selection of the statistics could be arbitrary and in the revised manuscript we have added multiple statistics at Figure 6 (similarly to Figure 3) to provide a more comprehensive assessment of the products.

Line 400-401: It is not clear what periods is used for the calibration in the two separate time frames. I assume that the calibration is not carried out for the whole periods.

In the development of the global SM2RAIN-ASCAT dataset the calibration is performed for the whole periods. Indeed, we do not want to perform calibration and validation against ERA5. As mentioned above, the validation should be performed with independent datasets, and even better by using the product for applications.

Line 404-408: it is not clear what the authors mean by distinguishing "in space" and "in time".

In space, we mean a fixed spatial mask over which we are aware of the lower performance of the ASCAT soil moisture product, and consequently of SM2RAIN-ASCAT rainfall product. In time, we have considered a temporally variable mask that flags observations with soil temperature, obtained from ERA5, lower than 3°C. It has been specified better in the revised manuscript (see lines 453-459):

"In space (i.e., a fixed spatial mask), we have used the committed area mask developed for ASCAT soil moisture product (PVR 2017), a frozen probability mask and a topographic complexity mask. In time (i.e., a temporally variable mask), we have considered the soil temperature data from ERA5 and flagged the observations with soil temperature values between 0°C and 3°C as temporary frozen soil and below 3°C as frozen soil. As many applications require continuous data, we have preferred to flag the data instead of removing them with an expected loss of accuracy."

Line 411-413: ERA-5 is used for calibration. It is not fair to use this dataset to create this map, and show R and RMSE.

The reviewer is right; Figure 5 shows the consistency of ERA5 and SM2RAIN-ASCAT and not the "accuracy" or the "performance" of the product, these terms has been removed from this section of the revised manuscript. Of course, we expect better performance in the areas in which the consistency is higher, but the preliminary assessment of SM2RAIN-ASCAT is performed in section 4.3. See lines 462-466:

"Figure 5 shows R and RMSE values between SM2RAIN-ASCAT and ERA5 in a single map. Therefore, Figure 5 illustrates the consistency between SM2RAIN-ASCAT and ERA5, and it is not intended to assess the performance of the data record (even though we expect better accuracy in areas where the agreement is higher)."

Line 468: Please, specify what is the committed area for ASCAT products (not ASCAT).

The reviewer is right; the committed area refers to the ASCAT soil moisture product and not to ASCAT.

The committed area has been built from ASCAT soil moisture product developers to indicate the areas in which the quality of soil moisture retrieval is expected to be good; it has been specified in the revised manuscript (see lines 453-455):

"we have used the committed area mask developed for the ASCAT soil moisture product (i.e., the area in which the ASCAT soil moisture retrievals are expected to be good, PVR 2017)"

Minor corrections:

Line 46: correct: "provides better performance better"

The text has been modified, accordingly.

Line 100: correct "has the advantage of requiring"

The text has been modified, accordingly.

Line 138-139: please specify which datasets have been used for the TC, what for the regional assessment, and what for global assessement.

The datasets used for the three analyses has been specified in the revised manuscript (see lines 146-153):

"As reference datasets we have used high-quality local raingauge networks from 2013 to 2017 in the United States, Italy, India and Australia for the assessment at 1009 points and for the regional assessment. Three additional global datasets have been considered: the latest reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5, the gauge-based Global Precipitation Climatology Centre (GPCC), and the GPM IMERG product (Early Run version). ERA5 has been used for the generation of the quasi-global SM2RAIN-ASCAT data record; GPCC and GPM IMERG have been considered for the TC analysis."

Line 190: Please, correct: "spatially averaging"

The text has been modified, accordingly.

Line 392-393: please correct this sentence. Something is missing, or maybe remove "," after "filtering".

The sentence has been revised, see lines 440-443:

"As reference dataset for the calibration of the parameter values of the pre-processing (filtering), of SM2RAIN, and of the post-processing, the ERA5 rainfall has been used mainly because of its higher spatial resolution compared to GPCC (36 km versus 100 km)."

Line 402-403: Please, correct this sentence.

The sentence has been corrected, see lines 452-453:

"Finally, we have flagged rainfall observations in space and time when the data are supposed to be less reliable."

SM2RAIN-ASCAT (2007-2018): global daily satellite rainfall from ASCAT soil moisture

Luca Brocca[1*], Paolo Filippucci[1], Sebastian Hahn[2], Luca Ciabatta[1], Christian Massari[1], Stefania Camici[1], Lothar Schüller[3], Bojan Bojkov[3], Wolfgang Wagner[2]

[1]{Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, Italy}

[2]{Department of Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria }

[3]{European Organisation for the Exploitation of Meteorological Satellites, Darmstadt, Germany}

August 2019

Eliminato: March

Re-Submitted to: *Earth System Science Data*

* Correspondence to: Ph.D. Luca Brocca, Research Institute for Geo-Hydrological Protection, National Research Council, Via della Madonna Alta 126, 06128 Perugia, Italy. Tel: +39 0755014418 Fax: +39 0755014420 E-mail: luca.brocca@irpi.cnr.it.

Abstract

Long-term gridded precipitation products are crucial for several applications in hydrology, agriculture and climate sciences. Currently available precipitation products suffer from space and time inconsistency due to non-uniform density of ground networks and the difficulties in merging multiple satellite sensors. The recent "bottom up" approach that exploits satellite soil moisture observations for estimating rainfall through the SM2RAIN algorithm is suited to build consistent rainfall data record as a single polar orbiting satellite sensor is used.

We exploit here the Advanced SCATterometer (ASCAT) on board three Metop satellites, launched in 2006, 2012 and 2018, as part of the EUMETSAT Polar System programme. The continuity of the scatterometer sensor is ensured until mid-2040s through the Metop Second Generation Programme. Therefore, by applying SM2RAIN algorithm to ASCAT soil moisture observations, a long-term rainfall data record will be obtained, starting in 2007 until mid-2040s. The paper describes the recent improvements in data pre-processing, SM2RAIN algorithm formulation, and data post-processing for obtaining the SM2RAIN-ASCAT quasi-global (only over land) daily rainfall data record at 12.5 km sampling from 2007 to 2018. The quality of SM2RAIN-ASCAT data record is assessed on a regional scale through the comparison with high-quality ground networks in Europe, United States, India and Australia. Moreover, an assessment on a global scale is provided by using the Triple Collocation technique allowing us also the comparison with the latest ECMWF reanalysis (ERA5), the Early Run version of the Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG), and the gauge-based Global Precipitation Climatology Centre (GPCC) products.

Results show that the SM2RAIN-ASCAT rainfall data record performs relatively well both at regional and global scale, mainly in terms of root mean square error when compared to other products. Specifically, SM2RAIN-ASCAT data record provides performance better than IMERG and GPCC in the data scarce regions of the world, such as Africa and South America. In these areas, we expect the larger benefits in using SM2RAIN-ASCAT for hydrological and agricultural applications. Limitations of SM2RAIN-ASCAT data record consist in the underestimation of peak rainfall events and in the presence of spurious rainfall events due to high frequency soil moisture fluctuations that might be corrected in the future with more advanced bias correction techniques.

The SM2RAIN-ASCAT data record is freely available at https://doi.org/10.5281/zenodo.2591215.

Eliminato: obtained from rain gauges, remote sensing and meteorological modelling
Eliminato: uses
Eliminato: long-term and
Eliminato: ¶
Eliminato: W
Eliminato: on European operational weather satellites
Eliminato: B
Eliminato: can
Eliminato: also operationally available in near real time
Eliminato:
Eliminato: dataset
Eliminato: (
Eliminato: -
Eliminato:)
Eliminato: dataset
Eliminato: other global products such as
Eliminato: European Centre for Medium-Range Weather Forecasts
Eliminato: Global Precipitation Measurement (GPM) mission
Eliminato: dataset
Eliminato: datasets
Eliminato: dataset
Eliminato: better
Eliminato: GPM
Eliminato: ¶
Eliminato: dataset

Keywords: Rainfall, Soil moisture, ASCAT, SM2RAIN, Remote Sensing.

1 Introduction

Rainfall is ranked the first among the Essential Climate Variable by the Global Climate Observing System (GCOS) as it represents the most important variable in many applications in geosciences (Maggioni and Massari, 2018). Long-term rainfall records are essential for drought monitoring (e.g., Forootan et al., 2019), water resources management (e.g., Abera et al., 2017) and climate studies (e.g., Herold et al., 2016; Pendergrass and Knutti, 2018) while near real-time rainfall data are needed for the mitigation of the impacts of natural disasters such as floods and landslides (e.g., Wang et al., 2107; Camici et al., 2018; Brunetti et al., 2018; Kirschbaum and Stanley, 2018). Additional applications in which near real-time rainfall plays a crucial role are weather forecasting, agricultural planning, vector-borne and waterborne diseases (e.g., Rinaldo et al., 2012; Thaler et al., 2018).

Three different techniques can be used for estimating rainfall: ground measurements, meteorological modelling and remote sensing. Ground measurements are based on rain gauges and meteorological radars (Lanza et al., 2009), but also new approaches such as microwave links are being developed (e.g., Overeem et al., 2011). These measurements guarantee high accuracy but suffer in many regions from limited spatial coverage (Kidd et al., 2017). Alternatively, meteorological models are used to estimate rainfall mainly in areas without ground reliable observations (Ebert et al., 2007), e.g., reanalysis. The uncertainties associated with these estimates can be large, mainly in areas where ground observations are scarce (Massari et al., 2017a). Therefore, to fill the gaps in the spatial coverage of ground measurements, and to improve the estimates obtained by models, different remote sensing techniques have been developed in the last 30 years (Hou et al., 2014). The standard methods for estimating rainfall from space are based on instantaneous measurements obtained from microwave radiometers, radars, and infrared sensors (Kidd and Levizzani, 2011). These methods are based on inversion techniques where the upwelling radiation (or backscattered signal for radars) is related to the surface precipitation rate, i.e., a "top down" approach (Brocca et al., 2014).

The most recent and successful example of satellite precipitation estimates is represented by the Integrated Multi-Satellite Retrievals for Global Precipitation Measurement, GPM (IMERG) of the GPM mission (Hou et al., 2014) which provide high spatial (0.1°) and temporal

Eliminato: N

Eliminato: while long-term rainfall record are essential for drought monitoring (e.g., Forootan et al., 2019), water resources management (e.g., Abera et al., 2017) and climate studies (e.g., Herold et al., 2016; Pendergrass and Knutti, 2018)

Eliminato: (e.g., reanalysis)

Eliminato: (Ebert et al., 2007)

Eliminato: measuring

Eliminato: methods are based on the inversion of the atmospheric signals reflected or radiated by atmospheric hydrometeors

Eliminato: GPM

Eliminato: Global Precipitation Measurement (

Eliminato:)

(30-minute) resolution and quasi-global coverage (+/-60°). To obtain such resolution and coverage, the IMERG products use a constellation of polar and geostationary satellite sensors operating in the microwave and infrared bands. However, the use of multiple sensors has some problems, including: the inconsistency between rainfall estimates from different sensors (intercalibration problem), the difficulties in collecting observations from multiple space agencies (i.e., problem of delivering the products in near real-time), and the high costs for the operation and the maintenance of the overall constellation. Moreover, as the top down approach requires the merging of instantaneous rainfall measurements from multiple sensors, the failure of one of them may imply a significant degradation in the accuracy of accumulated rainfall estimate due to the high temporal variability of rainfall (Trenberth and Asrar, 2014).

In recent years, a new "bottom up" approach has emerged that uses satellite soil moisture observations to infer, or to correct, rainfall over land (Brocca et al., 2013a; Crow et al., 2009; Pellarin et al., 2013; Wanders et al., 2015). The major difference between the bottom up and top down approaches is in the type of measurement; i.e., accumulated rainfall with the bottom up method and instantaneous rainfall rates with the top down method. This difference makes the two approaches highly complementary and their integration has been already successfully tested and demonstrated in several recent studies (e.g., Brocca et al., 2016; Ciabatta et al., 2017; Chiaravallotti et al., 2018; Massari et al. 2019). When accumulated rainfall estimates are needed (e.g., daily rainfall), the bottom up approach has the advantage of requiring a much lower number of measurements and, hence, of satellite sensors. The limitations of the bottom up approach are the possibility to estimate only terrestrial rainfall and its dependence on land characteristics (e.g., low accuracy for dense vegetation coverage and complex topography, Brocca et al., 2014).

Eliminato: to

Eliminato: e

The bottom up approach has been applied over a range of scales: global (Crow et al., 2011; Brocca et al., 2014; Ciabatta et al., 2018), continental (Wanders et al., 2015; Brocca et al., 2016), and local (Massari et al., 2014; Brocca et al., 2015; Román-Cascón et al., 2017) scale. Moreover, different satellite soil moisture products have been considered including SMOS (Soil Moisture Ocean Salinity mission, Brocca et al., 2016), ASCAT (Advanced SCATterometer, Brocca et al., 2017), AMSR-E (Advanced Microwave Scanning Radiometer, Crow et al., 2009), and SMAP (Soil Moisture Active and Passive, Koster et al., 2016; Tarpanelli et al., 2017; Zhang et al., 2019). First studies employing satellite rainfall estimates obtained through the bottom up approach for hydrological and water resources applications have been recently published (e.g., Ciabatta et al., 2016; Abera et al., 2017; Brunetti et al., 2018; Camici et al., 2018). These studies have highlighted the large potential of this technique as a complimentary and useful approach for estimating rainfall from space, and have also shown its main limitations. Specifically, the temporal resolution and the accuracy of satellite soil moisture products play a fundamental role in determining the accuracy of the bottom up rainfall estimates.

In this study, we describe the newly developed SM2RAIN-ASCAT rainfall data record covering the period 2007-2018 and characterized by a spatial/temporal sampling of 12.5 km/1-day. The new SM2RAIN-ASCAT data record is obtained from the application of SM2RAIN algorithm (Brocca et al., 2014) to the ASCAT soil moisture data records H113 and H114 provided by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF). It is the first SM2RAIN-ASCAT data record available at the same spatial resolution as the ASCAT soil moisture product (previous data records have been under-sampled at 0.5- and 1-degree resolution). Moreover, we have included the latest improvements in pre- and post-processing of soil moisture and rainfall data as well as in the SM2RAIN algorithm. The main differences with the SM2RAIN-CCI rainfall data record (Ciabatta et al., 2018) are the input soil moisture product (the input of SM2RAIN-CCI is the European Space Agency Climate Change Initiative Soil Moisture, ESA CCI soil moisture, product, Dorigo et al., 2017), and the time coverage (SM2RAIN-CCI spans the period 1998-2015). Technically, the use of the same satellite sensor in SM2RAIN-ASCAT data record is preferable to ensure consistency between soil moisture estimates over time to which the SM2RAIN algorithm is highly sensitive.

The purpose of this study is twofold. As a first objective, we have applied SM2RAIN algorithm at 1009 points uniformly distributed (with spacing of 1.5°) in the United States, Italy, India and Australia for testing different configurations of data pre-/post-processing and SM2RAIN model equation. This analysis has allowed us to select the best configuration that is implemented on a global scale for obtaining the SM2RAIN-ASCAT data record. The second objective is the assessment of the global scale SM2RAIN-ASCAT data record through the comparison with reference datasets and by exploiting the Triple Collocation (TC) approach (Massari et al., 2017a). As reference datasets we have used high-quality local raingauge networks from 2013 to 2017 in the United States, Italy, India and Australia for the assessment at 1009 points and for the regional assessment. Three additional global datasets have been

Eliminato: dataset
Eliminato: dataset
Eliminato: , and i
Eliminato: dataset
Eliminato: dataset
Eliminato: dataset
Eliminato: dataset
Eliminato: dataset
Eliminato: dataset
Eliminato: dataset
Eliminato: dataset
Eliminato: and t

considered: the latest reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5, the gauge-based Global Precipitation Climatology Centre (GPCC), and the GPM IMERG product (Early Run version). ERA5 has been used for the generation of the quasi-global SM2RAIN-ASCAT data record; GPCC and GPM IMERG have been considered for the TC analysis.

We underline that the paper goal is to present and describe the SM2RAIN-ASCAT quasi-global rainfall data record and to perform a comparison with state-of-the-art global rainfall products. We do not want to show a comprehensive assessment of the product. Indeed, we believe that researchers other than the product developers should perform the validation of the dataset. Even better, we stress the importance of performing the validation by using the datasets in hydrological or agricultural applications (e.g., flood prediction and agricultural water management).

Eliminato: and to perform a preliminary assessment of the product with respect to other state-of-the-art global rainfall products

2 Datasets

Nine different datasets have been collected for this study which are based on remote sensing, ground observations and reanalysis. Refer to ***Table 1*** for a summary of the datasets.

The main input dataset for producing SM2RAIN-ASCAT data record is the ASCAT soil moisture data record provided by the "EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF)" (http://hsaf.meteoam.it/). ASCAT, currently on board Metop-A (launched on October 2006), Metop-B (September 2012) and Metop-C (November 2018) satellites, is a scatterometer operating at C-band (5.255 GHz) and, by using the TU Wien algorithm (Wagner et al., 2013) the H SAF provides a soil moisture product characterized by 12.5 km spatial sampling. The temporal sampling is varying as a function of latitude and the number of satellites: by using Metop-A only a daily sampling is obtained, by using Metop-A and Metop-B two observations per day are available at mid-latitudes. Here we have used the H SAF ASCAT soil moisture data record (using Metop-A and Metop-B) available through the product H113 (PUM, 2018) covering the period 2007-2017 and its extension product H114 for the year 2018.

Eliminato: dataset

Eliminato: from

Eliminato: to

Three datasets obtained from the latest reanalysis of ECMWF, i.e., ERA5, have been used. ERA5 reanalysis is characterized by a spatial resolution of ~36 km and hourly temporal resolution. ERA5 is available from the Copernicus Climate Change service and the datasets cover the period 1979 to present. We have extracted hourly observations for the period 2007-

2018 for three variables: evaporation, soil temperature for the first layer (0-7 cm) and total rainfall (computed as the difference between total precipitation and snowfall). Evaporation data are used as additional input to the SM2RAIN algorithm and soil temperature data for masking periods with frozen soils. Total rainfall has been considered as a benchmark for the calibration of global SM2RAIN parameter values (see next section).

Ground-based rainfall datasets from regional networks have been also collected including the Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation in the United States, the gridded rainfall data provided by ~3000 stations of the National Department of Civil Protection in Italy (Ciabatta et al., 2017), the India Meteorological Department (IMD, http://www.imd.gov.in/pages/services_hydromet.php) rainfall observations in India, and the Australia Water Availability Project (AWAP, http://www.bom.gov.au/jsp/awap/rain/index.jsp) gridded rainfall data in Australia. These datasets have been used firstly for the selection of the optimal configuration of SM2RAIN implementation. For that, 1009 points uniformly distributed over the four regions have been selected. Secondly, the regional networks have been used for the assessment of the global SM2RAIN-ASCAT rainfall product at regional scale.

The ERA5 and local rainfall datasets have been regridded over the ASCAT grid (12.5 km) through the nearest neighbouring method and resampled at daily time scale as accumulated rainfall from 00:00 to 23:59 UTC. The ERA5 evaporation and soil temperature data are also regridded to the same grid and aggregated at daily scale as accumulated and average value from 00:00 to 23:59 UTC, respectively.

For the global assessment of SM2RAIN-ASCAT, two additional rainfall datasets have been considered: Global Precipitation Climatology Centre (GPCC) Full Data Daily Product (Schamm et al., 2015) and GPM IMERG Early Run product (Hou et al., 2014), hereinafter referred to as GPM-ER. Due to the availability of GPM-ER from April 2014, the global analysis has been carried out in the 4-year period from January 2014 to December 2018. Moreover, for the global inter-comparison all the datasets (SM2RAIN-ASCAT, ERA5, GPCC, and IMERG-ER) have been regridded at 0.25-degree resolution by spatially averaging the pixels contained in each 0.25-degree cell for SM2RAIN-ASCAT and GPM-ER, and by selecting the nearest pixel for ERA5 and GPCC.

3 Methods

In the following, the methodology used for obtaining the SM2RAIN-ASCAT data record is described. Specifically, three steps are carried out (see ***Figure 1***): 1) surface soil moisture data pre-processing, 2) SM2RAIN algorithm, and 3) rainfall data post-processing. Different configurations for the data pre-/post-processing and for the SM2RAIN model equation are considered; the details are given in ***Table 2***.

Eliminato: dataset

3.1 Soil moisture data pre-processing

The ASCAT surface soil moisture product is provided as relative soil moisture (between 0 and 1) at the overpass time of the satellite sensor (see ***Figure A1*** for the mean daily revisit time of ASCAT in the period 2007-2012 with only Metop-A and the period 2013-2018 with Metop-A+B). For the application of SM2RAIN algorithm, data should be equally spaced in time and hence, we have linearly interpolated in time soil moisture observations every 24 hours, 12 hours and 8 hours. The interpolation may increase the risk of false rainfall events, but it is a required step to obtain accumulated rainfall over a fixed duration. In a preliminary test (not shown for brevity), we have tested the three sampling frequencies with the baseline formulation for SM2RAIN (***equation 6***, see below). The best performances have been obtained with 12 hours sampling, particularly from 2013 to 2018 in which both Metop-A and -B are available. Therefore, 12 hours sampling has been used in the following analyses. The 24-hour accumulated rainfall is obtained by summing the two 12-hour accumulated rainfall data obtained for each day.

One of the major problems in using satellite soil moisture observations for rainfall estimation is related to the high frequency fluctuations caused by measurement and retrieval errors. If positive, such fluctuations are interpreted erroneously as rainfall by SM2RAIN algorithm. Therefore, satellite surface soil moisture data need to be filtered before being used as input into SM2RAIN. In previous studies, the exponential filtering has been considered (Wagner et al., 1999). The exponential filter, also known as Soil Water Index (SWI), has been used for filtering surface soil moisture time series as a function of a single parameter, T, i.e., the characteristic time length. In this study, we have tested two additional filtering methods. The first one is an extension of the exponential filter in which the T parameter is assumed to be varying with soil moisture as proposed in Brocca et al. (2013b). Specifically, T decreases with increasing soil moisture through a 2-parameter power law. Therefore, the data are filtered more during dry conditions. The third approach that we have implemented is a discrete wavelet filter (similar to Massari et al., 2017b). The discrete wavelet filter cuts the higher frequencies of the signal, typically characterized by noises, over a threshold selected through the principle of Stein's Unbiased Risk at multiple levels. We have found the Daubechies wavelets to be the most appropriate functions because their shape and the shape of the soil moisture signal is similar. Therefore, we have implemented a Daubechies-based wavelet filter in which the filtering level is optimized.

For all the filtering approaches, the parameter values of the filters have been optimized point-by-point in order to reproduce the reference rainfall observations.

3.2 SM2RAIN algorithm and calibration

The SM2RAIN algorithm is based on the inversion of the soil water balance equation and allows to estimate the amount of water entering the soil by using as input soil moisture observations from in situ or satellite sensors (e.g., Brocca et al., 2013a; 2014; 2015; Koster et al., 2016; Ciabatta et al., 2017; Massari et al., 2014). Specifically, the soil water balance equation can be described by the following equation (over non-irrigated areas):

$$nZ\frac{dS(t)}{dt} = p(t) - g(t) - sr(t) - e(t) \qquad (1)$$

where n [-] is the soil porosity, Z [mm] is the soil layer depth, $S(t)$ [-] is the relative saturation of the soil or relative soil moisture, t [days] is the time, $p(t)$ [mm/day] is the rainfall rate, $g(t)$ [mm/day] is the drainage (deep percolation plus subsurface runoff) rate, $sr(t)$ [mm/day] is the surface runoff rate and $e(t)$ [mm/day] is the actual evapotranspiration rate.

For estimating the rainfall rate, ***equation (1)*** is applied only during rainfall periods and, hence, some of the components of the equation can be considered as negligible. For instance, the actual evapotranspiration rate during rainfall is quite low due to the presence of clouds and, hence, the absence of solar radiation. Similarly, the surface runoff rate, i.e., the water that does not infiltrate into the soil and flows at the surface to the watercourses, is much lower than the rainfall rate, mainly if ***equation (1)*** is applied at coarse spatial resolution (20 km), i.e., with satellite soil moisture data. Indeed, most of water becomes runoff flowing in the subsurface, and also the part that does not infiltrate, due to for instance impervious land cover or soil, may re-infiltrate downstream within a pixel at 20 km scale. We have indirectly tested this hypothesis by counting the number of days the ASCAT soil moisture product is higher than 99.5 percentile

Eliminato: almost

for two (or more) consecutive days in the period 2007-2018. We have indirectly tested this hypothesis by counting the number of days the ASCAT soil moisture product is higher than 99.5 percentile for two (or more) consecutive days in the period 2007-2018. We have found that the number of consecutive days in which the soil is saturated is equal to 4 days (median value on a global scale) over 12 years, with 90% of land pixels with values lower than 12 days (i.e., 1 day per year). The occurrence of higher values is limited to very few areas in the tropical forests and over Himalaya (see *Figure A2*).

Eliminato: We have indirectly tested this hypothesis by counting the number of times the ASCAT soil moisture signal is higher than 99 percentile for two (or more) consecutive days. We have found that only 0.1% of times the soil is saturated for 1+ days, and the occurrence is limited to few areas worldwide (see

Following the indications obtained in Brocca et al. (2015), we have assumed the surface runoff rate, *sr(t)*, as negligible (i.e., Dunnian runoff) and we have rearranged ***equation (1)*** for estimating the rainfall rate:

$$p(t) = nZ\frac{dS(t)}{dt} + g(t) + e(t) \tag{2}$$

Eliminato: –

Eliminato: –

In this study, we have considered different formulations for equation (2) by varying the drainage rate as:

$$g(t) = K_s S(t)^m \tag{3.1}$$

$$g(t) = K_s S(t)^{\lambda+1}\left[1-\left(1-S(t)^{\frac{\lambda+1}{\lambda}}\right)^{\frac{\lambda}{\lambda+1}}\right]^2 \tag{3.2}$$

$$g(t) = K_s S(t)^{\tau}\left[1-\left(1-S(t)^{\frac{1}{m}}\right)^{m}\right]^2 \tag{3.3}$$

where K_s [mm/day] is the saturated hydraulic conductivity, m [-] and λ [-] are exponents related to the pore size distribution index, and τ is the tortuosity index. Specifically, the three equations represent the hydraulic conductivity - soil moisture formulation by Brooks-Corey (3.1), van Genuchten (3.2), and Mualem-van Genuchten (3.3).

The actual evapotranspiration rate has been considered as an additional input, together with soil moisture, here obtained from ECMWF reanalysis ERA5:

$$e(t) = K_c\, ET_{ERA5}(t) \tag{4}$$

where $ET_{ERA5}(t)$ [mm/day] is the actual evapotranspiration rate obtained from ERA5 reanalysis and K_c [-] is a correction factor for taking into account potential bias in ERA5 estimates.

Moreover, we have considered an additional formulation in which Z is a function of soil moisture taking into account the different penetration depth of satellite sensors as a function of wetness conditions:

$$Z = Z[0.1 + (1 - S(t)^{c})] \tag{5}$$

where c exponent determines the rate of decrease of penetration depth with increasing soil moisture.

Accordingly, we have used different formulations for equation (2) that are compared with the baseline equation used in previous studies (e.g., Brocca et al., 2014):

$$p(t) = Zn\frac{dS(t)}{dt} + K_s S(t)^{m} \tag{6}$$

In synthesis, we have investigated 3 different configurations (total of 5 options) for: 1) selecting the best equation for the drainage rate (***equations 3***), 2) testing the possibility to include the evapotranspiration component (***equation 4***), and 3) testing the use of a variable penetration depth with soil moisture conditions (***equation 5***). Each new configuration has been compared with the baseline (***equation 6***) in order to select the best configuration for SM2RAIN algorithm (see ***Figure 1***). For all configurations, negative rainfall values, that might occur during some dry-down cycles, have been set equal to zero.

SM2RAIN parameter values are calibrated point-by-point by using the reference rainfall as target. As objective function, we have used the minimization of the RMSE between SM2RAIN-ASCAT and reference rainfall.

3.3 Rainfall data post-processing

The use of satellite soil moisture observations for obtaining rainfall estimates is affected by errors in the input data and in the retrieval algorithm SM2RAIN. The correction of the overall bias in the climatology is a simple and effective approach for mitigating part of such errors. Specifically, we refer here to a static correction procedure that once calibrated for a time period can be applied in the future periods, also for operational real time productions. We note that a climatological correction is performed in several satellite rainfall datasets delivered in near real-time (e.g., GPM-Early Run). We have implemented two different approaches for climatological correction: 1) a cumulative density function (CDF) matching approach at daily time scale, and 2) a monthly correction approach. Specifically, the implemented CDF matching approach is a 5-order polynomial correction as described in Brocca et al. (2011) for matching

Eliminato: -

Eliminato: products

the CDF of estimated rainfall with respect to reference rainfall, in which the CDF are computed over the whole calibration period at daily time scale. The monthly correction approach computes the monthly ratios between the climatology of estimated and reference rainfall, i.e., 12 correction factors per pixel. Then, the SM2RAIN-estimated rainfall is multiplied for the monthly correction factors to obtain the climatologically corrected SM2RAIN-estimated rainfall.

3.4 Triple collocation analysis

For the global assessment of satellite, reanalysis and gauge-based rainfall products we have used the Triple Collocation (TC) technique. TC can theoretically provide error and correlations of three products (a triplet) given that each of the three products is afflicted by mutually independent errors. Therefore, in principle, TC can be used for assessing the quality of satellite products without using ground observations (Massari et al., 2017a). In this study, we have implemented the same procedure as described in Massari et al. (2017), i.e., by implementing an additive error model at daily time scale, and we refer the reader to this study for the analytical details. In synthesis, by using the extended TC method firstly proposed by McColl et al. (2014), it is possible to estimate the temporal correlation, R_{TC}, of each rainfall product in the triplets with the truth.

Eliminato:

3.5 Performance scores

Several metrics have been used to assess the product performance during the validation period. As continuous scores we have computed the Pearson's correlation coefficient (R), the root mean square error (RMSE), the mean error between estimated and reference rainfall (BIAS), and the ratio of temporal variability of estimated and reference rainfall (STDRATIO). Continuous scores have been computed on a pixel-by-pixel basis by considering 1 day of accumulated rainfall. Moreover, three categorical scores, i.e. Probability of Detection (POD), False Alarm Ration (FAR) and Threat Score (TS), have been computed. POD is the fraction of correctly identified rainfall events (optimal value POD=1), FAR is the fraction of predicted events that are non-events (optimal value FAR=0), while TS provides a combination of the other two scores (optimal value TS=1). The categorical assessment is carried out by considering a rainfall threshold of 0.5 mm/day (instead of 0 mm/day) in order to exclude spurious events that might be due to rainfall interpolation\regridding in the reference datasets. For a complete description of the performance scores, see ***Table A1*** in the Appendix.

Eliminato: 1

Eliminato: categorical

Eliminato: the reader is referred to Brocca et al. (2014)

4 Results

The results are split in three parts: 1) selection of the optimal configuration of SM2RAIN through the assessment at 1009 points, 2) generation of global SM2RAIN-ASCAT rainfall data record, and 3) regional assessment of SM2RAIN-ASCAT with gauge-based rainfall datasets and global assessment through TC.

4.1 Selection of the best SM2RAIN processing configuration at 1009 points

As a first step we have co-located satellite soil moisture data from ASCAT soil moisture H113+H114, ground-based rainfall observations and actual evapotranspiration data from ERA5 in space and time at 1009 points. We have selected 1009 points uniformly distributed over a regular grid with spacing of 1.5°. Each point is considered representative of a 0.25° x 0.25° box. The selection is carried out for reducing the computational time in running the different SM2RAIN configurations. The numbers of points for each region is depending on the size of the region: 328 points in Australia, 163 in India, 55 in Italy, and 463 in the United States. Ground observations, GPM-ER and ERA5 data are regridded by spatial averaging measurements contained over each 0.25° x 0.25° box. These datasets are made freely available here (https://doi.org/10.5281/zenodo.2580285, Brocca, 2019) for those interested to test alternative approaches for rainfall estimation from ASCAT soil moisture. Specifically, we have considered the period 2013-2016, 2013-2014 for the calibration and 2015-2016 for the validation; in the sequel only the results in the validation period are shown. The ground-based high quality rainfall observations available for the four regions are used as reference data (ground truth) in this analysis. The reference configuration, REF, as used in previous SM2RAIN applications (e.g., Brocca et al., 2014), uses the SWI for data filtering, the SM2RAIN formulation as in ***equation (6)***, and no climatological correction. Results in the validation period are shown in ***Figure 2A*** in terms of temporal R against reference data. As shown, the median R for all points is equal to 0.60, with better results in Italy (median R=0.67, see boxplots) and similar results in the other 3 regions (median R=0.60 and 0.59). These results are in line with previous studies (e.g., Ciabatta et al., 2017; Tarpanelli et al., 2017) carried out in Italy and India and highlight the potential of ASCAT soil moisture observations to provide daily rainfall estimates. ***Figure 3*** (first column) shows the results for the different performance metrics; in the last two columns the results obtained with GPM-ER and ERA5 are shown for comparison. Very good statistics have been obtained in terms of RMSE and BIAS but a tendency to underestimate the observed rainfall variability (median STDRATIO=0.60) and medium-high

Eliminato: dataset

Eliminato: , instead of the whole region, in order to test multiple configuration at reasonable computational cost probability of false alarm (median FAR=0.53). The other scores are similar, or slightly lower than those obtained through GPM-ER and ERA5.

Eliminato: 53

The first test has been dedicated to the filtering of soil moisture data by using three approaches: 1) SWI, i.e., the REF configuration, 2) SWI with T varying with soil moisture, SWI-Tvar, and 3) the discrete wavelet filtering, WAV. ***Figure 3*** shows in the first three columns the summary of the performance scores highlighting that the SWI-Tvar configuration is performing the best, even though the differences with REF configuration are small. ***Figure 2b*** shows the R map for SWI-Tvar configuration while in ***Figure 2c*** the differences in R-values with REF are displayed. Improved performance in terms of R is visible over most of the pixels except in central Australia.

The second test has been performed on the SM2RAIN equation by using different drainage functions (VGEN and MUA configurations), by adding the evapotranspiration component (EVAP), and by considering the variability of sensing depth, *Z*, with soil moisture (ZVAR). VGEN, MUA and ZVAR configurations are characterized by lower performances than REF (see ***Figure 3***, columns 4, 5 and 7), even though MUA and ZVAR incorporate an additional parameter to be calibrated (and, hence, better performance was expected). The addition of evapotranspiration brings a slight improvement with respect to REF (see ***Figure 3***, column 6), with results similar to SWI-Tvar. Larger improvements are obtained over areas in which evapotranspiration is more important, e.g., in the south-western United States and central western Australia. In India and Italy, the results are very similar to REF. However, EVAP configuration requires actual evapotranspiration data from ERA5 as additional input and such data might be not available for the implementation of the processing algorithm in an operational context.

Eliminato: ; the details of the different configurations are given in ***Table 2***

Eliminato: s

The final test has been done by applying the daily CDF matching, BC-CDF, and monthly correction factors, BC-MON, for correcting the climatological bias in SM2RAIN-derived rainfall estimates; results are shown in columns 8 and 9 of ***Figure 3***. For these two configurations, the improvements with respect to REF are evident but with different magnitude for the different scores. BC-CDF improves significantly STDRATIO, TS and FAR with a slight deterioration in R and RMSE. BC-MON shows the highest R-values among all configurations with the larger improvements in India, Italy and United States. However, the improvement in terms of STDRATIO, TS and FAR is less important than BC-CDF. Therefore, depending on which score is assumed more important, one of the two configurations can be selected. If compared with GPM-ER, BC-CDF and BC-MON configurations show similar results with larger positive differences, in terms of RMSE, BIAS, STDRATIO and POD; R values are slightly better for GPM-ER that is also much better in terms of TS and FAR. Similar findings can be summarized in the comparison with ERA5, even though ERA5 is performing the best in terms of R, STDRATIO, FAR, and TS among all configurations.

Figure 4 shows time series of rainfall averaged over the four regions as obtained from ground observations and from BC-MON configuration. The agreement of spatially averaged rainfall with observations is high with R-values greater than 0.83, and very low BIAS. Moreover, regional scale rainfall events are correctly reproduced both in terms of timing and magnitude.

Eliminato: and KGE>0.67

4.2 Generation of SM2RAIN-ASCAT data record

Eliminato: dataset

Based on the tests performed in the previous paragraph, we have selected the best configuration using SWI-Tvar for filtering, Brooks-Corey function for losses, and the monthly correction approach for climatological correction. The addition of evapotranspiration component, even though showing some improvements, has been not used in view of an operational implementation of the method. The monthly correction approach has been selected as R and RMSE scores have been considered more important based on previous investigations on the assessment of satellite rainfall products (e.g., Massari et al., 2017).

The selected configuration has been applied on a global scale to 839826 points over which ASCAT soil moisture observations are available. As reference dataset for the calibration of the parameter values of the pre-processing (filtering), of SM2RAIN, and of the post-processing, the ERA5 rainfall has been used mainly because of its higher spatial resolution compared to GPCC (36 km versus 100 km). However, we have also tested the use of the two datasets for calibration at randomly chosen 20000 points which showed that the estimated rainfall in the two calibration tests is very similar. For instance, the median R between the two SM2RAIN-ASCAT data records is higher than 0.90 (not shown for brevity). This result clearly demonstrate that the selection of reference dataset has a small influence on SM2RAIN-derived rainfall that is mostly driven from soil moisture temporal fluctuations. Additionally, considering the improved ASCAT coverage after 2013, the calibration has been split from 2007 to 2012 (Metop-A) and from 2013 to 2018 (Metop-A and -B). The dual calibration has solved the issue in terms of long-term trend that has been found in previous application of SM2RAIN to ASCAT

Eliminato: calibrating

Eliminato: parameter values

Eliminato: dataset

soil moisture data (Beck et al., 2017). Finally, we have flagged rainfall observations in space and time when the data are supposed to be less reliable. In space (i.e., a fixed spatial mask), we have used the committed area mask developed for the ASCAT soil moisture product (i.e., the area in which the ASCAT soil moisture retrievals are expected to be good, PVR 2017), a frozen probability mask and a topographic complexity mask. In time (i.e., a temporally variable mask), we have considered the soil temperature data from ERA5 and flagged the observations with soil temperature values between 0°C and 3°C as temporary frozen soil and below 3°C as frozen soil. As many applications require continuous data, we have preferred to flag the data instead of removing them with an expected loss of accuracy.

Eliminato: being aware of the regions in which ASCAT soil moisture product is performing not good,

Eliminato: in which the quality of ASCAT soil moisture is found to be good

The SM2RAIN-ASCAT data record so obtained has a spatial sampling of 12.5 km, a daily temporal resolution and covers the 12-year period 2007-2018. ***Figure 5*** shows R and RMSE values between SM2RAIN-ASCAT and ERA5 in a single map. Therefore, ***Figure 5*** illustrates the consistency between SM2RAIN-ASCAT and ERA5, and it is not intended to assess the performance of the data record (even though we expect better accuracy in areas where the agreement is higher). Green light colours indicate very good agreement with high R and low RMSE, orange to red colours indicate low R and high RMSE, while black colour indicates low RMSE and R highlighting areas in which rainfall occurrence and variability is very low (e.g., Sahara Desert, high latitudes). The data record has been found in very good agreement with ERA5 (high R and low RMSE) in western United States, Brazil, southern and western South America, southern Africa, Sahel, southern-central Eurasia, and Australia. The areas in which SM2RAIN-ASCAT is characterized by lower consistency with ERA5 are those with dense vegetation (Amazon, Congo, and Indonesia), with complex topography (e.g., Alps, Himalaya, Andes), at high latitudes and Saharan and Arabian deserts. In these areas it is well-known that the ASCAT soil moisture product has limitations (e.g., Wagner et al., 2013), and generally the retrieval of soil moisture from remote sensing is more challenging. The median R and RMSE values are equal to 0.56 and 3.06 mm/day, with slightly better scores in the period 2013-2018 (R=0.57, RMSE=2.95), thanks to the availability of ASCAT on both Metop-A and Metop-B.

Eliminato: dataset

Eliminato: in which

Eliminato: performance

Eliminato: dataset

Eliminato: to perform accurately

Eliminato: accuracy

4.3 Regional and global assessment of SM2RAIN-ASCAT data record

Eliminato: dataset

By using all the pixels included in the four regions (Italy, United States, India and Australia), for a total of 29843 points, the new SM2RAIN-ASCAT rainfall data record has been compared with reference rainfall observations in ***Figure 6***, by considering the whole period 2007-2018. Specifically, the box plots of different performance metrics (the same of ***Figure 3***)

Eliminato: dataset

Eliminato: R and RMSE scores

are shown and compared with the results obtained through GPCC, ERA5, and GPM-ER. By focusing on the SM2RAIN-ASCAT data record performance over the different regions, it shows better performance in Italy (median R=0.67) and United States (median R=0.62), almost comparable with the other datasets; while in Australia and India R-values are lower (median R=0.61 and 0.59). In the selected regions, the best product is GPCC (mainly in Australia) followed by ERA5 while GPM-ER and SM2RAIN-ASCAT are performing similarly in terms of R. The better performance of GPCC are expected (gauge-based dataset) and also the very good performance of ERA5 in Italy and Australia thanks to the availability of ground observations for the reanalysis. We highlight also that differently from SM2RAIN-ASCAT and GPM-ER, GPCC and ERA5 have a latency of weeks to months and, hence, these products cannot be used for near real time applications. When considering the RMSE score, the results are quite different with respect to R. SM2RAIN-ASCAT is overall very good being the best (second best) product in United States (India). The ranking of the product is GPCC, SM2RAIN-ASCAT, ERA5 and GPM-ER, with the latter showing high RMSE values in United States and Australia. As obtained in previous studies (Brocca et al., 2016; Ciabatta et al., 2017), the SM2RAIN approach is very good in obtaining low RMSE values thanks to its accuracy in the retrieval of accumulated rainfall. Moreover, the product accuracy is stable over time as it is not as strongly affected by the availability of satellite overpasses as in the top down approach. As shown also in ***Figure 3***, the SM2RAIN-ASCAT data record has limitations in reproducing the variability of rainfall (low STDRATIO) mainly due underestimation issues. Moreover, FAR values of SM2RAIN-ASCAT are quite high highlighting the difficulties in removing the problem of high frequency soil moisture fluctuations wrongly interpreted by SM2RAIN as rainfall events.

On a global scale, the TC approach has been implemented by using the triplet SM2RAIN-ASCAT, GPM-ER and GPCC, by considering the common period 2015-2018 and at daily time scale. In TC analysis we have not considered ERA5 purposely to avoid any dependency between the products. Theoretically, the extended TC approach provides the correlation, R_{TC}, against the underlying "truth". ***Figures 7A and 7B*** show the R_{TC} maps for SM2RAIN-ASCAT and GPM-ER highlighting similar mean values (0.66 and 0.64 for SM2RAIN-ASCAT and GPM-ER, respectively). It should be underlined that the R_{TC} values are higher than those obtained in the comparison with ground observations as theoretically the metric does not contain the error in the reference (Massari et al., 2017a). The spatial pattern of the performance is quite different as demonstrated in ***Figure 7c*** in which the differences between the two R_{TC}

Eliminato: also

Eliminato: In terms of R

Eliminato: SM2RAIN-ASCAT show

maps is shown. Again, these results underline the strong complementarity between bottom up and top down approaches (e.g., Ciabatta et al., 2017; Chiaravallotti et al., 2018). As expected, SM2RAIN-ASCAT performs worse over desert areas, tropical forests and complex mountainous regions. Differently, over plains and low vegetated areas SM2RAIN-ASCAT is performing better than GPM-ER, particularly in the southern hemisphere. Indeed, in Africa and South America SM2RAIN-ASCAT provides good performance (see also ***Figure 7A***) thanks to the capability of the bottom up approach to estimate accumulated rainfall accurately with a limited number of satellite overpasses occurring in these areas, differently from the top down approach used in GPM-ER.

The box plots of R_{TC} shown in ***Figure 7D*** indicates that, overall, GPCC is performing similar to the two satellite products with major differences in the spatial patterns of the performance. SM2RAIN-ASCAT is performing the best among the three products in Africa, South America, central-western United States and central Asia while GPCC is performing the best in the remaining parts except the tropical region in which GPM-ER is performing very good (see ***Figure 8***). If we consider only the committed area of ASCAT (PVR 2017), in which the soil moisture product is supposed to be reliable, the mean value of R_{TC} is equal to 0.72 whereas in the masked area it is equal to 0.59.

5 Data availability

The SM2RAIN-ASCAT data record is freely available at https://doi.org/10.5281/zenodo.2591215 (Brocca et al., 2019).

Eliminato: dataset

6 Conclusions

In this study, we have described the development of a new SM2RAIN-ASCAT rainfall data record highlighting the steps carried out for improving the retrieval algorithm and the pre-/post-processing of the data. The major novelties of the SM2RAIN-ASCAT rainfall data record developed here with respect to previous versions are: 1) application of SM2RAIN at full spatial resolution thus providing a gridded data record with sampling of 12.5 km, 2) improved sampling and filtering of ASCAT soil moisture data, 3) application of monthly climatological correction, and 4) improved calibration strategy.

Eliminato: dataset

Eliminato: dataset

Eliminato: dataset

The SM2RAIN-ASCAT data record has been preliminary assessed at regional (***Figures 4 and 6***) and global (***Figure 5, 7 and 8***) scale in terms of different performance metrics with

Eliminato: dataset

larger emphasis on the temporal correlation, R, and the root mean square error, RMSE. The overall performances are good, mainly in terms of RMSE thanks to the capacity of SM2RAIN to accurately reproduce accumulated rainfall consistently over time. Importantly, SM2RAIN-ASCAT is found to perform even better than ground-based GPCC product over the southern hemisphere in Africa and South America, and in central-western United States and central Asia. Limitations of SM2RAIN-ASCAT data record consist in: 1) the underestimation of peak rainfall events, 2) the presence of spurious rainfall events due to high frequency soil moisture fluctuations, 3) the estimation of liquid rainfall only (snowfall cannot be estimated), and 4) the possibility to estimate rainfall over land only.

Eliminato: also

Eliminato: the estimation of

In the near future, we are going to develop the near real-time version of the SM2RAIN-ASCAT rainfall product that can be used as input for applications such as flood prediction (similarly to Camici et al., 2018 and Massari et al., 2018), landslide prediction (Brunetti et al., 2018) and novel applications for the agriculture and for the water resources management.

Eliminato: The

Eliminato: dataset

Eliminato: can now

Acknowledgments: The authors gratefully acknowledge support from the EUMETSAT through the Global SM2RAIN project (contract n° EUM/CO/17/4600001981/BBo) and the "Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF)" CDOP 3 (EUM/C/85/16/DOC/15).

References

Abera, W., Formetta, G., Brocca, L., Rigon, R.: Modeling the water budget of the Upper Blue Nile basin using the JGrass-NewAge model system and satellite data. Hydrology and Earth System Sciences, 21, 3145-3165, 2017.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrology and Earth System Sciences, 21, 6201-6217.

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., Bittelli, M.: Soil moisture estimation through ASCAT and AMSR-E sensors: an intercomparison and validation study across Europe. Remote Sensing of Environment, 115, 3390-3408, 2011.

Brocca, L., Melone, F., Moramarco, T., Wagner, W.: A new method for rainfall estimation through soil moisture observations. Geophysical Research Letters, 40(5), 853-858, 2013a.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Albergel, C.: Scaling and filtering approaches for the use of satellite soil moisture observations. In: George P. Petropoulos (ed.), Remote Sensing of Energy Fluxes and Soil Moisture Content, CRC Press 2013, Chapter 17, 411-426, ISBN: 978-1-4665-0578-0, 2013b.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., Levizzani, V.: Soil as a natural rain gauge: estimating global rainfall from satellite soil moisture data. Journal of Geophysical Research, 119(9), 5128-5141, 2014, 2014.

Brocca, L., Massari, C., Ciabatta, L., Moramarco, T., Penna, D., Zuecco, G., Pianezzola, L., Borga, M., Matgen, P., Martínez-Fernández, J.: Rainfall estimation from in situ soil moisture observations at several sites in Europe: an evaluation of SM2RAIN algorithm. Journal of Hydrology and Hydromechanics, 63(3), 201-209, 2015.

Brocca, L., Pellarin, T., Crow, W.T., Ciabatta, L., Massari, C., Ryu, D., Su, C.-H., Rudiger, C., Kerr, Y.: Rainfall estimation by inverting SMOS soil moisture estimates: a comparison of different methods over Australia. Journal of Geophysical Research, 121(20), 12062-12079, 2016.

Brocca, L., Crow, W.T., Ciabatta, L., Massari, C., de Rosnay, P., Enenkel, M., Hahn, S., Amarnath, G., Camici, S., Tarpanelli, A., Wagner, W.: A review of the applications of ASCAT soil moisture products. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(5), 2285-2306, 2017.

Brocca, L.: SM2RAIN test dataset with ASCAT satellite soil moisture (Version 1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.2580285, 2019.

Brocca, L., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B., Wagner, W.: SM2RAIN-ASCAT (2007-2018): global daily satellite rainfall from ASCAT soil moisture (Version 1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.2591215, 2019.

Brunetti, M.T., Melillo, M., Peruccacci, S., Ciabatta, L., Brocca, L.: How far are we from the use of satellite data in landslide forecasting? Remote Sensing of Environment, 210, 65-75, doi:10.1016/j.rse.2018.03.016, 2018.

Eliminato: ¶

Camici, S., Ciabatta, L., Massari, C., Brocca, L.: How reliable are satellite precipitation estimates for driving hydrological models: a verification study over the Mediterranean area. Journal of Hydrology, 563, 950-961, 2018.

Chiaravalloti, F., Brocca, L., Procopio, A., Massari, C., Gabriele, S.: Assessment of GPM and SM2RAIN-ASCAT rainfall products over complex terrain in southern Italy. Atmospheric Research, 206, 64-74, 2018.

Ciabatta, L., Brocca, L., Massari, C., Moramarco, T., Gabellani, S., Puca, S., Wagner, W.: Rainfall-runoff modelling by using SM2RAIN-derived and state-of-the-art satellite rainfall products over Italy. International Journal of Applied Earth Observation and Geoinformation, 48, 163-173, 2016.

Ciabatta, L., Marra, A.C., Panegrossi, G., Casella, D., Sanò, P., Dietrich, S., Massari, C., Brocca, L.: Daily precipitation estimation through different microwave sensors: verification study over Italy. Journal of Hydrology, 545, 436-450, 2017.

Ciabatta, L., Massari, C., Brocca, L., Gruber, A., Reimer, C., Hahn, S., Paulik, C., Dorigo, W., Kidd, R., Wagner, W.: SM2RAIN-CCI: a new global long-term rainfall data set derived from ESA CCI soil moisture. Earth System Science Data, 10, 267-280, 2018.

Crow, W.T., Huffman, G.F., Bindlish, R., Jackson, T.J.: Improving satellite rainfall accumulation estimates using spaceborne soil moisture retrievals. Journal of Hydrometeorology, 10, 199-212, 2009.

Crow, W.T., van den Berg, M.J., Huffman, G.J., Pellarin, T.: Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART). Water Resources Research, 47, W08521, 2011.

Dorigo, W.,Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, D., Hamer, P. Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y.Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S.I., Smolander, T., Lecomte, P.: ESA CCI soil moisture for improved earth system understanding: state-of-the art and future directions. Remote Sensing of Environment, 203, 185-215, 2017.

Ebert, E.E., Janowiak, J.E., Kidd, C.: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. Bulletin of the American Meteorological Society, 88(1), 47-64, 2007.

Forootan, E., Khaki, M., Schumacher, M., Wulfmeyer, V., Mehrnegar, N., van Dijk, A.I.J.M., Brocca, L., Farzaneh, S., Akinluyi, F., Ramillien, G., Shum, C.K., Awange, J., Mostafaie, A.: Understanding the global hydrological droughts of 2003-2016 and their relationships with teleconnections. Science of the Total Environment, 650, 2587-2604, 2019.

Herold, N., Alexander, L.V., Donat, M.G., Contractor, S., Becker, A.: How much does it rain over land? Geophysical Research Letters, 43(1), 341-348, 2016.

Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K., Iguchi, T.: The Global Precipitation Measurement (GPM) mission. Bulletin of the American Meteorological Society, 95(5), 701-722, 2014.

Kidd, C., Levizzani, V.: Status of satellite precipitation retrievals. Hydrology and Earth System Sciences, 15, 1109-1116, 2011.

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., Kirschbaum, D. B.: So, how much of the Earth's surface is covered by rain gauges? Bulletin of the American Meteorological Society, 98(1), 69-78, 2017.

Kirschbaum, D., Stanley, T.: Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness. Earth's Future, 6(3), 505-523, 2018.

Koster, R.D., Brocca, L., Crow, W.T., Burgin, M.S., De Lannoy, G.J.M.: Precipitation Estimation Using L-Band and C-Band Soil Moisture Retrievals. Water Resources Research, 52(9), 7213-7225, 2016.

Lanza, L.G., Vuerich, E.: The WMO Field Intercomparison of Rain Intensity Gauges. Atmospheric Research, 94, 534-543, 2009.

Maggioni, V., Massari, C.: On the performance of satellite precipitation products in riverine flood modeling: A review. Journal of Hydrology, 558, 214-224, 2018.

Massari, C., Brocca, L., Moramarco, T., Tramblay, Y., Didon Lescot, J.-F.: Potential of soil moisture observations in flood modelling: estimating initial conditions and correcting rainfall. Advances in Water Resources, 74, 44-53, 2014.

Massari, C., Crow, W., Brocca, L.: An assessment of the accuracy of global rainfall estimates without ground-based observations. Hydrology and Earth System Sciences, 21, 4347-4361, 2017a.

Massari, C., Su, C.-H., Brocca, L., Sang, Y.F., Ciabatta, L., Ryu, D., Wagner, W.: Near real time de-noising of satellite-based soil moisture retrievals: An intercomparison among three different techniques. Remote Sensing of Environment, 198, 17-29, 2017b.

Massari, C., Maggioni, V., Barbetta, S., Brocca, L., Ciabatta, L., Camici, S., Moramarco, T., Coccia, G., Todini, E.: Complementing near-real time satellite rainfall products with satellite soil moisture-derived rainfall through a Bayesian inversion approach. Journal of Hydrology, in press, 2019.

McColl, K.A., Vogelzang, J., Konings, A.G., Entekhabi, D., Piles, M., Stoffelen, A.: Extended triple collocation: estimating errors and correlation coefficients with respect to an unknown target. Geophys. Res. Lett., 41, 6229-6236, 2014.

Overeem, A., Leijnse, H., Uijlenhoet, R.: Measuring urban rainfall using microwave links from commercial cellular communication networks. Water Resources Research, 47(12), doi:10.1029/2010WR010350, 2011.

Pellarin, T., Louvet, S., Gruhier, C., Quantin, G., Legout, C.: A simple and effective method for correcting soil moisture and precipitation estimates using AMSR-E measurements. Remote Sensing of Environment, 136, 28-36, 2013.

Pendergrass, A.G., Knutti, R.: The uneven nature of daily precipitation and its change. Geophysical Research Letters, 45(21), 11980-11988, 2018.

"Product User Manual (PUM) Soil Moisture Data Records, Metop ASCAT Soil Moisture Time Series": Tech. Rep. Doc. No: SAF/HSAF/CDOP3/PUM, version 0.7, 2018.

"Product Validation Report (PVR) Metop ASCAT Soil Moisture CDR products": Tech. Rep. Doc. No: SAF/HSAF/CDOP3/PVR, version 0.6, 2017.

Rinaldo, A., Bertuzzo, E., Mari, L., Righetto, L., Blokesch, M., Gatto, M., Casagrandi, R., Murray, M., Vesenbeckh, S.M., Rodriguez-Iturbe, I.: Reassessment of the 2010–2011 Haiti cholera outbreak and rainfall-driven multiseason projections. Proceedings of the National Academy of Sciences, 109(17), 6602-6607, 2012.

Román-Cascón, C., Pellarin, T., Gibon, F., Brocca, L., Cosme, E., Crow, W., Fernández, D., Kerr, Y., Massari, C.: Correcting satellite-based precipitation products through SMOS soil moisture data assimilation in two land-surface models of different complexity: API and SURFEX. Remote Sensing of Environment, 200, 295-310, 2017.

Schamm, K., Ziese, M., Raykova, K., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U.: GPCC Full Data Daily Version 1.0 at 1.0°: Daily Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. doi:10.5676/DWD_GPCC/FD_D_V1_100, 2015.

Tarpanelli, A., Massari, C., Ciabatta, L., Filippucci, P., Amarnath, G., Brocca, L.: Exploiting a constellation of satellite soil moisture sensors for accurate rainfall estimation. Advances in Water Resources, 108, 249-255, 2017.

Thaler, S., Brocca, L., Ciabatta, L., Eitzinger, J., Hahn, S., Wagner, W.: Effects of different spatial precipitation input data on crop model outputs under a Central European climate. Atmosphere, 9(8), 290, 2018.

Trenberth, K.E., Asrar, G.R.: Challenges and opportunities in water cycle research: WCRP contributions. Surv. Geophys. 35(3), 515-532, 2014.

Wagner, W., Lemoine, G. Rott, H.: A method for estimating soil moisture from ERS scatterometer and soil data. Remote Sens. Environ., 70, 191-207, 1999.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa, J., de Ros- nay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Zuger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Bloeschl, G., Eitzinger, J., Steinnocher, K., Zeil, P., Rubel, F.: The ASCAT soil mois- ture product: a review of its specifications, validation results, and emerging applications. Meteorologische Zeitschrift 22 (1), 5-33, 2013.

Wanders, N., Pan, M., Wood, E.F.: Correction of real-time satellite precipitation with multi-sensor satellite observations of land surface variables. Remote Sensing of Environment, 160, 206-221, 2015.

Wang, Z., Zhong, R., Lai, C., Chen, J.: Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. Atmospheric research, 196, 151-163, 2017.

Zhang, Z., Wang, D., Wang, G., Qiu, J., Liao, W.: Use of SMAP soil moisture and fitting methods in improving GPM estimation in near real time. Remote Sensing, 11, 368, 2019.

Tables

Table 1. List of satellite, ground-based and reanalysis products used in this study (the spatial/temporal sampling used in this study is reported).

Short name	Full name and details	Data source	Spatial/ temporal sampling	Time period	References
			SOIL MOISTURE		
ASCAT	Advanced Scatterometer	satellite	12.5 km/ daily	2007 - present	Wagner et al. (2013)
			RAINFALL		
ERA5	ECMWF	reanalysis	0.25°/ daily	1979 - present	https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview
GPCC	Global Precipitation Climatology Centre Full Data Reanalysis	gauge	1°/ daily	1988 - present	Schamm et al. (2015)
IMERG Early Run	Global Precipitation Measurement	satellite	0.1°/ daily	2014 - present	Hou et al. (2014)
CPC	Climate Prediction Center – United States	gauge	0.5°/ daily	1948 - present	https://www.esrl.noaa.gov/psd/data/gridded/data.unified.daily.conus.html
ITA-DPC	Gauge-based rainfall dataset –Italy	gauge	0.1°/ daily	2008 - present	Ciabatta et al. (2017)
AWAP	Australian Water Availability Project – Australia	gauge	0.05°/ daily	1900 - present	http://www.bom.gov.au/jsp/awap/rain/index.jsp
IMD	India Meteorological Department - India	gauge	0.25°/ daily	1901 - present	http://www.imd.gov.in/pages/services_hydromet.php
			SOIL TEMPERATURE and EVAPOTRANSPIRATION		
ERA5	ECMWF	reanalysis	0.25°/ daily	1979 - present	https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview

Table 2. Configurations used in the paper (SWI: Soil Water Index, BCO: Brooks-Corey, VGE: van Genuchten, MUA: Mualem-van Genuchten, SWI-Tvar: SWI with T varying with soil moisture, WAV: wavelet filtering, CDF: climatological correction with daily cumulative density function matching, MON: climatological correction with monthly correction factors).

Short Name	Filtering	Losses	Evapotranspiration	Depth varying	Climatological Correction
REF	SWI	BCO	no	no	no
SWI-Tvar	SWI-Tvar	BCO	no	no	no
WAV	WAV	BCO	no	no	no
VGEN	SWI	VGE	no	no	no
MUA	MUA	VGE	no	no	no
EVAP	SWI	BCO	yes	no	no
ZVAR	SWI	BCO	no	yes	no
BC-CDF	SWI-Tvar	BCO	no	no	CDF
BC-MON	SWI-Tvar	BCO	no	no	MON

Figures

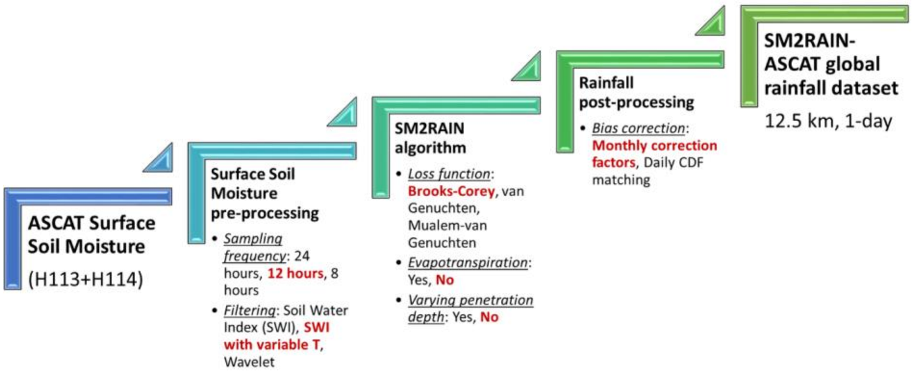


Figure 1. Processing steps for obtaining the SM2RAIN-ASCAT global rainfall data record (2007-2018) from ASCAT surface soil moisture data: pre-processing, SM2RAIN algorithm, and post-processing. Each bullet represents a possible configuration that has been tested, the selected configuration is in red, bold font.

Eliminato: dataset

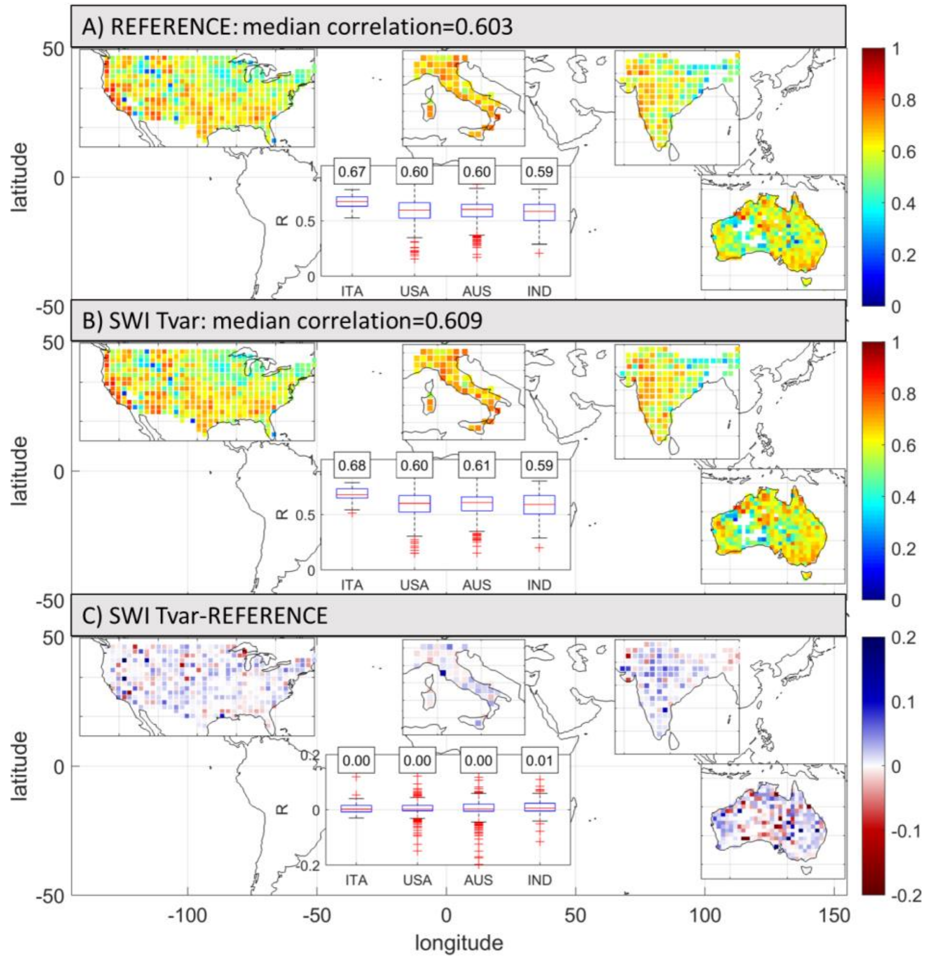
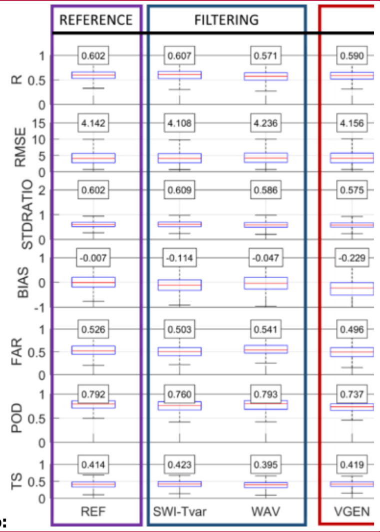


Figure 2. Performance of two different configurations at 1009 points in terms of Pearson's correlation, R [-]. A) R map with reference configuration, B) R map with Soil Water Index (SWI) filtering with variable T as a function of soil moisture, and C) R map difference (B)-(A). The inner box plots show the R values (and R differences) split for different regions.

Figure 3. Performances at 1009 points in terms of Pearson's correlation, R [-], root mean square error, RMSE [mm/day], variability ratio, STDRATIO [-], BIAS [mm/day], false alarm ratio, FAR [-], Probability of Detection, POD [-], and Threat Score, TS [-]. For details of the different configurations see Table 2 (GPM-ER: GPM IMERG Early Run product).

Eliminato:



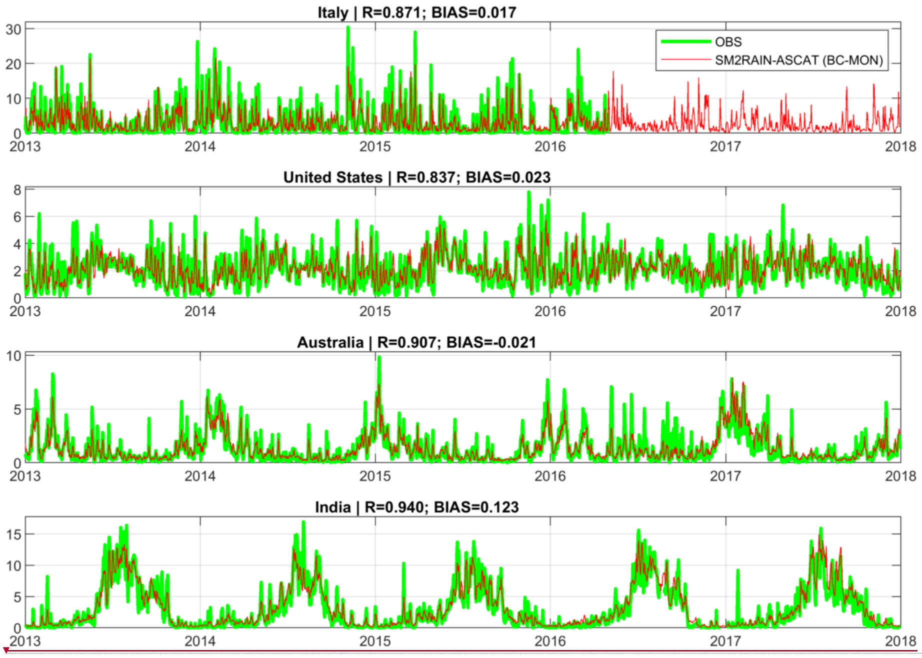
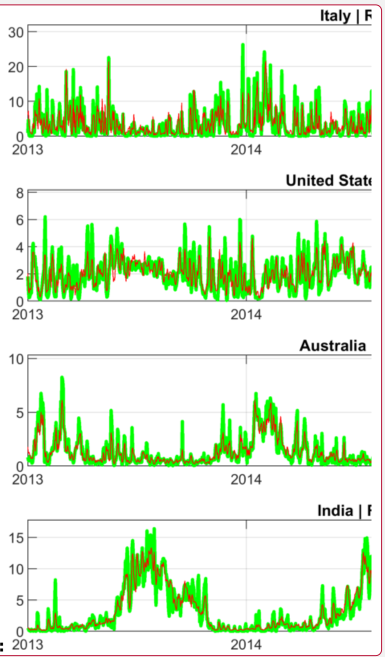


Figure 4. Time series of mean areal rainfall for the four regions for observed data, OBS, and SM2RAIN-ASCAT data record, BC-MON configuration (R [-]: Pearson's correlation, BIAS [mm/day]: mean error).

Eliminato: 

Eliminato: dataset

Eliminato: , KGE [-]: Kling-Gupta efficiency

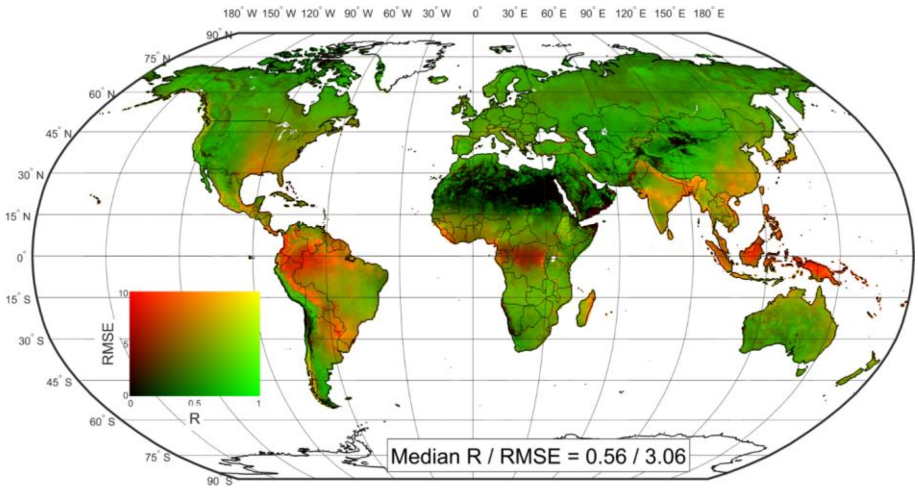


Figure 5. Pearson's correlation, R, and root mean square error, RMSE, map of SM2RAIN-ASCAT data record compared with ERA5 reanalysis dataset used as benchmark (period 2007-2018). The analysis is carried out at 1-day and 12.5 km temporal and spatial resolution. The map shows that SM2RAIN-ASCAT data record is performing well in the western United States, Brazil, southern and western South America, southern Africa, Sahel, southern-central Eurasia, and Australia (green colours).

Eliminato: dataset

Eliminato: dataset

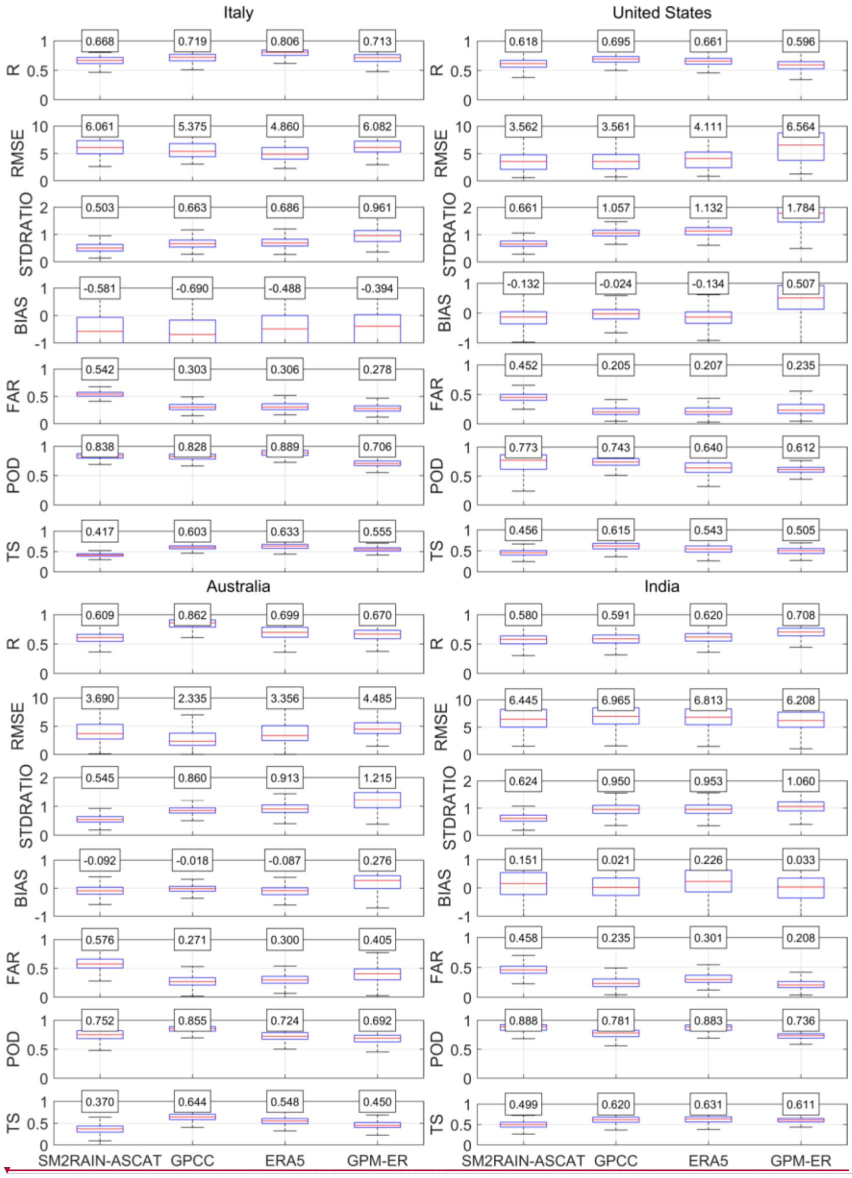
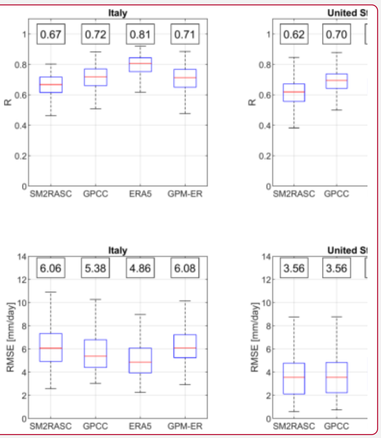


Figure 6. Regional assessment of SM2RAIN-ASCAT rainfall data record and comparison with GPCC, ERA5 and GPM-ER rainfall products. As reference the high-quality ground-based datasets in Italy, United States, India and Australia are used. Results in terms of Pearson's correlation, R [-], root mean square error, RMSE [mm/day], variability ratio, STDRATIO [-], BIAS [mm/day], false alarm ratio, FAR [-], Probability of Detection, POD [-], and Threat Score, TS [-].

Eliminato:



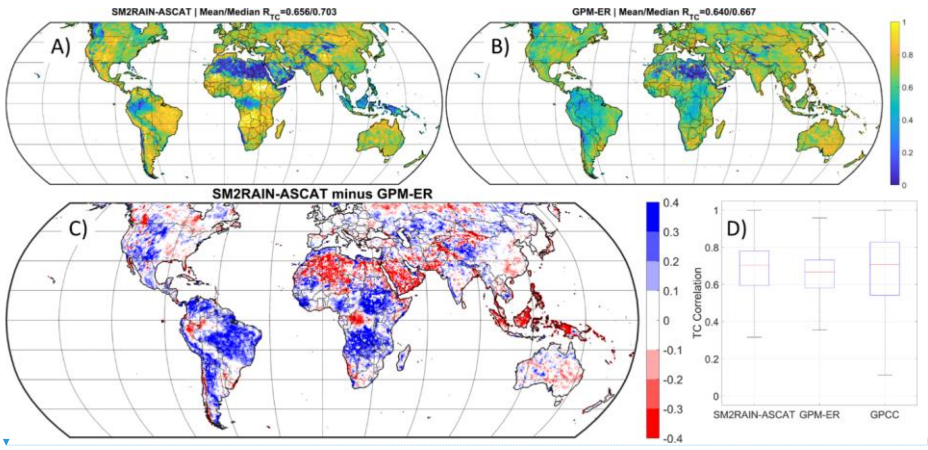


Figure 7. Global triple collocation, TC, results. A) R_{TC} map for SM2RAIN-ASCAT, B) R_{TC} map for GPM-ER, (C) differences between (A) and (B), i.e., blue (red) colours for pixels in which SM2RAIN-ASCAT (GPM-ER) is performing better, and D) box plot of R_{TC} for SM2RAIN-ASCAT, GPM-ER, and GPCC. SM2RAIN-ASCAT is performing significantly better than GPM-ER in South America and Africa (excluding desert and tropical forest areas), elsewhere the two satellite products perform similarly.

Eliminato: **Figure 6.** Regional assessment of SM2RAIN-ASCAT rainfall datasetdata record and comparison with GPCC, ERA5 and GPM-ER rainfall products. As reference the high-quality ground-based datasets in Italy, United States, India and Australia are used. Results in terms of Pearson's correlation, R [-], root mean square error, RMSE [mm/day], variability ratio, STDRATIO [-], BIAS [mm/day], false alarm ratio, FAR [-], Probability of Detection, POD [-], and Threat Score, TS [-]Top panels show the Pearson's correlation, R, and the bottom panels the root mean square error, RMSE. ¶

interruzione pagina

¶

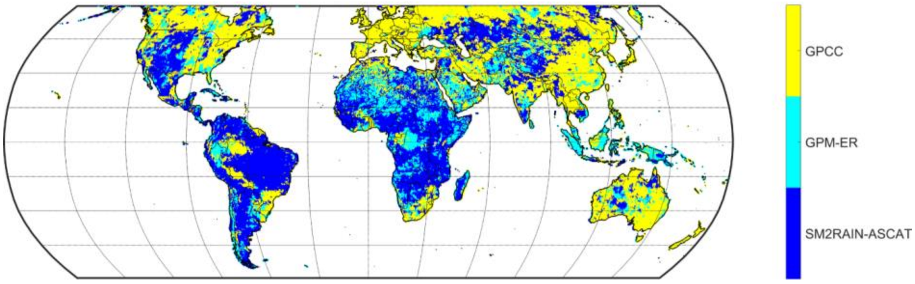


Figure 8. Best performing product based on the results of triple collocation shown in Figure 7. SM2RAIN-ASCAT is performing the best among the three products in Africa, South America, central-western United States and central Asia while GPCC is performing the best in the remaining parts of the northern hemisphere and in Australia. GPM-ER is the best product in the tropical and equatorial region.

Appendix

Table 1. Equations used for the performance scores. For the continuous scores, P_{ref} is the reference dataset (e.g., ground observations, ERA5) and P_{est} is the estimated dataset (e.g., SM2RAIN-ASCAT, GPM-ER), *cov* is the covariance operator, σ is the standard deviation operator, Σ is the summation operator, and *N* is the sample size. For the categorical scores, *H* is the number of successfully predicted events by a given rainfall product, *F* the number of non-events erroneously predicted to occur, and *M* the number of actual events that are missed.

Performance Score	**Score symbol**	**Equation**
		Continuous scores
Pearson's correlation	R	$R = \frac{cov(P_{est}, P_{ref})}{\sigma(P_{est})\sigma(P_{ref})}$
Root Mean Square Error	RMSE	$RMSE = \sqrt{\frac{\sum(P_{est} - P_{ref})^2}{N}}$
Temporal Variability Ratio	STDRATIO	$STDRATIO = \frac{\sigma(P_{est})}{\sigma(P_{ref})}$
Bias	BIAS	$BIAS = \frac{\sum(P_{est} - P_{ref})}{N}$
		Categorical scores
False Alarm Ratio	FAR	$FAR = \frac{F}{H + F}$
Probability of Detection	POD	$POD = \frac{H}{H + M}$
Threat Score	TS	$TS = \frac{H}{H + F + M}$

Eliminato: ¶

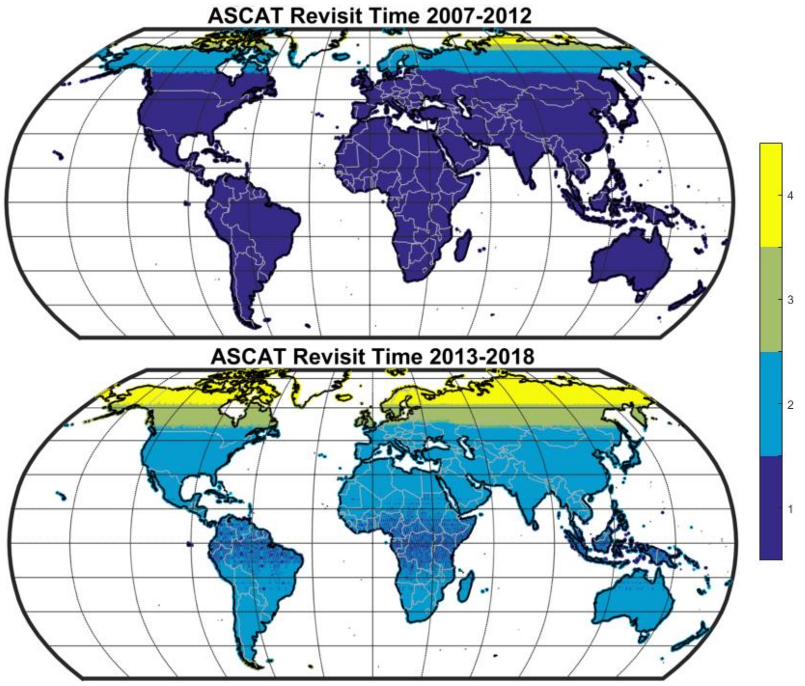
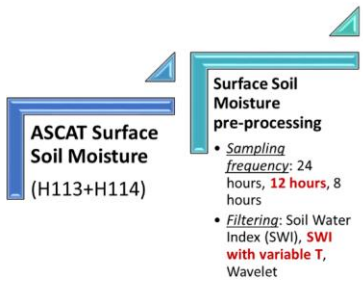


Figure A1. Mean daily revisit time [days] of ASCAT soil moisture observations for the period 2007-2012 (only Metop-A, top panel) and for the period 2013-2018 (Metop-A+B, bottom panel).



Eliminato:

Eliminato: R

Eliminato: left

Eliminato: right

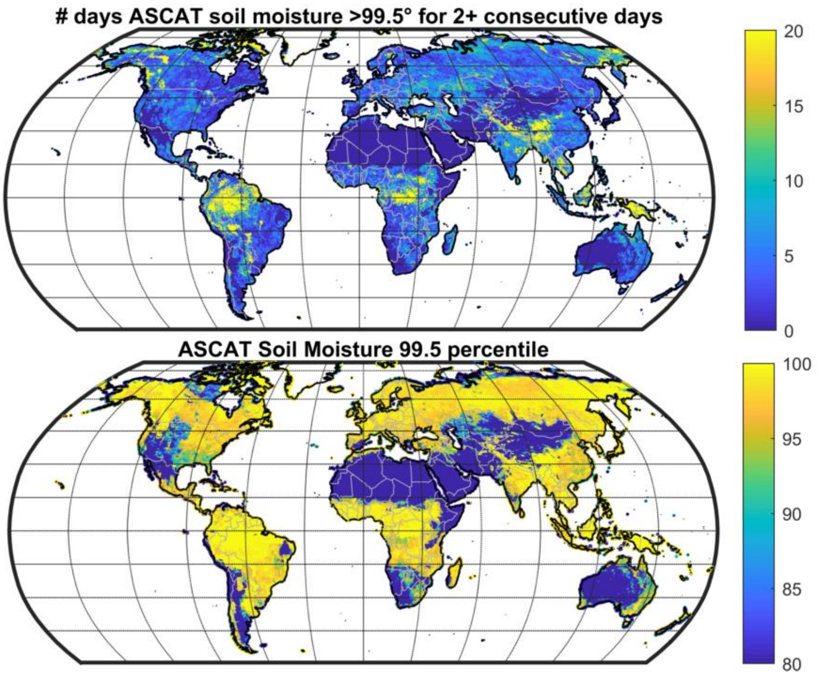


Figure A2. Number of days in which ASCAT soil moisture observations are close to saturation (>99.5 percentile, top panel) for 2 (or more) consecutive days in the period 2007-2018. Please note that the upper value is set to 20 days as in most of land areas the occurrence is very low (90% of land pixel with values lower than 12 days over 12 years). In the bottom panel the soil moisture values at 99.5 percentile (in the period 2007-2018) are shown.