This is the accepted manuscript version of this paper before copy-editing, formatting, technical enhancements and pagination. The finally published version (version of record) is available via https://doi.org/10.1016/j.jhydrol.2021.126837
© 2021. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/

# Toward a self-calibrated and independent SM2RAIN rainfall product

Paolo Filippucci (1, 2, 3), Luca Brocca (1), Christian Massari (1), Carla Saltalippi (2), Wolfgang Wagner (3), Angelica Tarpanelli (1)

1. National Research Council, Research Institute for Geo-Hydrological Protection, Perugia, Italy (paolo.filippucci@irpi.cnr.it)
2. Università degli Studi di Perugia, Department of Civil and Environmental Engineering, Perugia, Italy
3. TUWien (Technische Universität Wien), Department of Geodesy and Geoinformation, Wien, Austria

## Abstract

Rainfall monitoring is fundamental in many hydrological applications such as flood and landslide forecasting and water resources management. In-situ measurements are the traditional data source of rainfall, but the worldwide declining number of stations, their low spatial representativeness and the data access problem limit their use. Satellite products are being widely used as an alternative data source. Among them, SM2RAIN-based products, which exploit the inversion of the water balance equation to derive rainfall from soil moisture observations, have shown relatively good skills for hydrological applications. However, the need of calibrating the SM2RAIN parameter values against a reference represents one important limitation, particularly over data scarce regions.

In this study, we explore the possibility to self-calibrate SM2RAIN and thus to obtain rainfall estimates from the Advanced SCATterometer (ASCAT) soil moisture independently from any reference rainfall dataset. Four parametric relationships relating SM2RAIN parameter values to static descriptors (average rainfall, topography, soil moisture noise) are developed. To develop such relationships, a sample of 1009 points uniformly distributed over the areas covered by rain gauges in Australia, India, Italy and the United States is selected. A global validation of the methodology is conducted by comparing the performances of the parameterized product with the classical product in which the parameter values are estimated by calibration against a reference rainfall dataset. The Final Run of the Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG) precipitation dataset is used for performance assessment, together with the triple collocation techniques by using the gauge-based Global Precipitation Climatology Center (GPCC) product and the Late Run of IMERG.

The aim of the analysis is to obtain an uncalibrated SM2RAIN methodology to retrieve rainfall whose performance are similar to those obtained with calibration. The results at 1009 points show that the performances of the parameterized SM2RAIN product are in line with those of the calibrated one, with an increased capability in the detection of intense rainfall events and an acceptable reduction of the performance according to both Pearson Correlation and Root Mean Square Error indexes. The application of triple collocation confirms these findings on a global scale, showing that the SM2RAIN product outperforms both GPCC and IMERG - Late run estimations in areas characterized by low density of rain gauges and good quality of ASCAT soil moisture retrievals (i.e., Africa and South America).

**Keywords**: Rainfall; Soil Moisture; Remote sensing; SM2RAIN

# 1. Introduction

Floods, drought and landslides are the water related natural hazards that cause the most serious damage to the environment, people and properties. The occurrence of those events is related to climate: wet soil moisture (SM) conditions and intense rainfall are often the drivers of flood and landslide events (Ciabatta et al., 2016). According to International Panel on Climate Change (IPCC) 5th report, climate change is expected to aggravate the occurrence of those phenomena since extreme weather and climate events will step up (IPCC, 2013).

The knowledge of the triggering conditions of hydroclimatic hazards can be used in prediction models in order to help the authorities to prevent or mitigate them (Hannah et al., 2011; Ponziani et al., 2012). The presence of an adequate monitoring network capable of providing accurate precipitation estimation is therefore fundamental not only for water resources management or agricultural planning, but also to reduce the loss of lives and economic damages. However, the rain gauge coverage is declining worldwide and unequally distributed, being concentrated in developed countries (Kidd et al., 2017; Vörösmarty et al., 2001). Moreover, despite being highly accurate, rain gauge stations are not free from errors (Peterson et al., 1998; Villarini et al., 2008).

Remote sensing techniques are currently the only valuable alternative to ground-based networks, as they have demonstrated their potential in the estimation of rainfall at relevant spatial and temporal scales globally (Kidd and Levizzani, 2011). The classical remote sensing-based technique to estimate rainfall is the "top-down" approach (Brocca et al., 2014a), where the upwelling radiation or backscatter from clouds measured by satellite sensors are used to estimate the surface instantaneous precipitation rate. One distinguished example of this type of product is the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG) product of the Global Precipitation Measurement (GPM) mission (Hou et al., 2014), characterized by relatively high spatial and temporal resolutions compared to its predecessors (0.1 degree and 30 minutes, respectively) and global coverage. This was achieved using a new Dual-frequency Precipitation Radar (DPR) and an accurate radiometer, both fundamental to calibrate infrared and microwave data from multiple polar and geostationary satellites. Despite the good level of accuracy achieved, the difficulties in obtaining and inter-calibrating near-real time observations from multiple agencies and the overall high cost for operation and maintenance of the whole satellite constellation are obstacles still to be overcome in order to guarantee the data continuity. Moreover, additional rainfall datasets are still needed to understand residual uncertainties and errors (Massari et al, 2017; Chen et al., 2020) like e.g. seasonal and local bias (Maggioni and Massari, 2018). The integration of IMERG with alternative rainfall products can be also carried out to reduce uncertainties, as in Massari et al. (2020).

The recently introduced "bottom-up" approach points toward addressing these problems by inferring or correcting rainfall estimation over land using SM observations from satellite or gauges. This method provides accumulated rainfall estimates (Crow et al., 2009; Brocca et al., 2013; Pellarin et al., 2013) instead than the instantaneous rate, as for the "top-down" products. Many methods based on this approach share the same limitations, linked to the limits of measuring SM from space: rainfall estimated only over land, low accuracy in presence of dense vegetation or complex topography and difficulties in estimating rainfall in case of soil saturation. Among the "bottom-up" approaches, SM2RAIN (Brocca et al., 2014a) was applied to different satellite SM products over different regions worldwide with satisfying results. Through the inversion of the soil water balance equation, it is capable to obtain the accumulated rainfall occurred between two SM measurements. The method has already been applied to different SM products for local (Brocca et al. 2015; Tarpanelli et al., 2017) and global (Brocca et al., 2019; Massari et al., 2020) analysis. Three global rainfall products based on SM2RAIN were developed: two of them were derived from the use of SM2RAIN alone (SM2RAIN-CCI, Ciabatta et al., 2018; SM2RAIN-ASCAT, Brocca et al., 2019) while the third one was derived from the integration with a "top-down" product, i.e., IMERG Late Run (GPM-SM2RAIN, Massari et al., 2020). Different studies have shown the usefulness of these products for hydrological application such as flood and landslide prediction (Brunetti et al., 2018; Camici et al., 2018; Brocca et al., 2020). In order to obtain accurate rainfall estimates, SM2RAIN parameter values need to be calibrated against a reference rainfall dataset (e.g., gauge-based) with spatial and temporal resolution comparable with those of the SM dataset.

In this paper, we propose a methodology to estimate the SM2RAIN parameter values independently from a reference, i.e., a self-calibrated SM2RAIN product. Four parametric relationships are obtained starting from climatic and land descriptors (e.g. observed mean annual rainfall, topography, soil moisture error) to obtain the four SM2RAIN parameter values. Understanding the relationships of the parameters with these descriptors is a step forward for a better physical understanding of SM2RAIN and the possibility: 1) to obtain an independent rainfall product, i.e. without the need of calibration against a reference dataset, and 2) to apply the method at high resolution (e.g., 1 km as obtained from Sentinel-1 mission, Bauer-Marschallinger et al.,2018; 2019). The methodology is tested, firstly, at 1009 points uniformly distributed (regular grid with a space resolution of 0.25 degrees) over the areas covered by rain gauges in Australia, India, Italy and United States (US). Several datasets globally available including soil texture, evapotranspiration, soil temperature, satellite SM and observed rainfall climatology are collected to be used as predictors. A qualitative and quantitative analysis of the data collected is carried out to identify the descriptors better related to each parameter of SM2RAIN algorithm and to obtain parametric relationships linking SM2RAIN parameter values to the selected predictors. Secondly, the parametric relationships are applied on a global scale for the period 2013-2019 and the performances of the parameterized SM2RAIN product are compared with those resulting from the calibration of SM2RAIN in the same period. Different global rainfall products (GPM IMERG Final Run, GPM IMERG Late Run and GPCC) are considered for performance evaluation, through classical performance metrics computation and Triple Collocation analysis (MASSARI ET AL., 2017). We finally aim to assess whether the self-calibrated SM2RAIN product performances are in line with those of the calibrated product.

# 2. Data

Multiple descriptors are considered for the estimation of SM2RAIN algorithm parameter values through a regression-based regionalization approach (JAKEMAN ET AL., 1992; POST ET AL., 1998; SEFTON AND HOWARTH, 1998; SEIBERT, 1999; WAGENER ET AL., 2004): several datasets are selected describing climatic (rainfall and evapotranspiration) and land (soil texture and soil type, SM, soil temperature, topography and vegetation cover) characteristics. The datasets have been selected for different reasons, including their relation with soil state and their availability worldwide. In the following, the datasets description is provided (see ***Table 1***).

## 2.1. Climatic data

*Regional rainfall datasets*

Regional gauge-derived rainfall datasets were collected for the 1009 points uniformly distributed (0.25-degree resolution) over the areas covered by rain gauges in Australia, Italy, US and India. The regional rainfall datasets are used as reference for SM2RAIN calibration at the points for which the parametric relationships are developed. For each region of the study area, the data are collected for the period 2013-2017. In particular:

- The Australian Water Availability Project (AWAP) rainfall product was downloaded for the Australia region. This gridded dataset is obtained from the interpolation of daily measurements of the

Australian Bureau of Meteorology raingauge network, performed by using an optimized Barnes successive correction technique. Its spatial resolution is about 5 km (0.05-degree) with a daily temporal resolution.

- The rainfall dataset of the Italian Civil Protection Department (ITA - DPC) is an interpolation of more than 3000 rain gauges distributed over the Italian territory. The interpolation is carried out using the Random Generator of Space Interpolations from Uncertain Observations (GRISO, Pignone et al., 2010) algorithm to spatially interpolate the measurements on a grid with about 10 km (0.1-degree) spatial resolution and aggregating the hourly data to the daily time step.
- For the US region, the National Oceanic and Atmospheric Administration Climate Prediction Center (NOAA CPC) Daily US UNIFIED Precipitation was downloaded. This rainfall product is characterized by an improved quality obtained by combining all information sources available at CPC and by taking advantage of the optimal interpolation (OI) objective analysis technique (Xie et al., 2007). Its spatial resolution is about 25 km (0.25-degree) with a daily temporal resolution.
- India region daily rainfall was obtained by downloading the India Meteorological Department (IMD) gridded dataset. This product combines daily rainfall data from 6955 gauges, using the Shepard method (Pai et al., 2014) to interpolate them, and it is characterized by a spatial resolution of about 25 km (0.25-degree).

The mentioned datasets were all temporally interpolated from their local time to 00:00 UTC, accepting the resulting uncertainty to obtain regular time spacing, in order to simplify the intercomparison with satellite-derived products (available at 00:00 UTC).

*Global Rainfall datasets*

Different global rainfall datasets were downloaded to obtain and validate the new SM2RAIN-ASCAT parameterized rainfall product, for the period 2013-2019:

- Global Precipitation Climatology Centre (GPCC) rainfall product (First Guess) is obtained from ~7000 quality controlled stations all over the world (Schamm et al., 2014). Its spatial resolution is 1 degree, with a daily temporal resolution. Since it is based on ground observation, the accuracy of the dataset is greater over region with high gauge density, i.e., Europe and US.

- The IMERG algorithm estimates precipitation over the majority of Earth's surface by inter-calibrating the available Passive Microwave (PMW) satellite precipitation estimates to the Combined Radar-Radiometer precipitation estimates from the GPM mission Core Observatory (GPM-CO) and then by merging and interpolating together these estimates with other precipitation estimates from infrared geostationary sensors (Huffman et al., 2020). Morphing and Kalman filtering interpolation are used to provide the precipitation estimate if no valid microwave data are available. The resulting product spatial resolution is 0.1-degrees, and the temporal resolution is 30 minutes. Three Runs of IMERG are available to the users, based on increasing latency and accuracy: Early Run (IMERG-ER; latency of 4–6 h after observation), Late Run (IMERG-LR; latency 12–18 h) and Final Run (IMERG-FR; latency of about 3 months). Final Run V06 product, with a monthly adjustment based on GPCC, and Late Run V06 product, are used here. In this study, the 30 minutes rainfall data were accumulated to obtain daily precipitation estimates.

- European Centre for Medium-Range Weather Forecasts, ECMWF, Reanalysis 5th Generation (ERA5) provide hourly data of various global atmosphere, land surface and sea-state variables, combining models with observations. It was developed within the Copernicus Climate Change Service (C3S) and it replaces the previous ERA-interim reanalysis product. Its spatial resolution is around 36 km, resampled on a regular 0.25-degree grid, and the temporal resolution is 1-hour (Hersbach et al., 2020). The hourly rainfall was calculated by subtracting the snowfall fraction to the total precipitation and accumulated to daily scale in this study.

*Evapotranspiration*

Hourly evapotranspiration data from ERA5 were obtained for each point of the study area. The hourly data were accumulated on windows of 12 hours centred at 00:00 UTC and 12:00 UTC, to obtain a temporal resolution aligned with the SM datasets (12 hours spacing).

## 2.2. Land data

*Satellite Soil Moisture and Soil Moisture Noise*

Advanced SCATterometer (ASCAT) is an active microwave sensor onboard of MetOp-A (launched 19/10/2006), MetOp-B (launched 17/09/2012) and MetOp-C (launched 07/11/2018) satellites. It uses two sets of three vertically polarized antennae, one on each side of the satellite ground track, and it senses backscatter radiation at 5.255 GHz (C-band). The sensor was originally developed to sense wind speed over oceans, but it turned out to be also sensitive to the amount of water in the soil, leading to the development of one of the longest satellite SM product available nowadays (from 2007 onward). ASCAT retrievals have a spatial resolution of 25 km, sampled at 12.5 km (~0.125°). Relative SM estimates and their related noise were downloaded from EUropean organisation for the exploitation of METeorological SATellites (EUMETSAT) Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF) H115 and H116 products for the period 2013-2019. In these years, the contemporary availability of the satellites MetOp-A and B permitted a sub-daily temporal resolution over most of the Earth (WAGNER ET AL., 2013). When the surface state was indicated as frozen, SM estimates were discarded. ASCAT measurements were linearly interpolated every 12 hours, to obtain regular time spacing. If no data were found within 5 days, each datum in the interval was set to Not a Number (NaN).

*Modelled Soil Moisture*

Hourly SM in the first soil layer (0 - 7 cm) of the ECMWF Integrated Forecasting System data from ERA5-Land were downloaded for the analysis period. ERA5-Land was produced by regridding the land component of the ECMWF ERA5 climate reanalysis with a finer spatial resolution (0.1-degree). SM was subsampled every 12 hours to obtain the same temporal resolution of ASCAT data.

*Topographic data*

Elevation data from Earth topography 5 arc minute (ETOPO5) were downloaded. Although the product is available on a regular grid of 5 -minutes (~0.08°), the resolution of the source data base varies from 5-minute for the ocean floors, USA, Europe, Japan and Australia to 1 degree in data-deficient parts of Asia, South America, northern Canada and Africa.

*Soil Temperature*

Soil Temperature data in the first soil layer (0 - 7 cm) of the ECMWF Integrated Forecasting System data from ERA5-Land were downloaded for the analysis period. The hourly data were subsampled every 12 hours to match the temporal resolution of ASCAT data.

*Soil Composition Data*

The Harmonized World Soil Database v1.2 (Wieder et al., 2014) contains worldwide soil composition information derived from regional and national data. Several soil parameters were downloaded for this analysis for the nominal year of 2000, including soil depth, sand-silt-clay fraction, reference soil depth, carbon content and bulk density, at a spatial resolution of 5 minutes (~0.08°).

*Vegetation Continuous field*

Global fractional vegetation cover data VCF5KYRv001 was downloaded from NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs). The dataset relative to the nominal year of 2015 was downloaded, containing information of tree cover vegetation, bare ground and non-tree cover vegetation area percentage, with a spatial resolution of 0.05°.

# 3. Methods

## 3.1. SM2RAIN

SM2RAIN is an algorithm developed by Brocca et al. (2013; 2014a) to estimate the accumulated rainfall between two SM measurements. This result can be achieved by inverting the soil water balance equation. It was successfully applied to different satellite and in situ SM dataset (Ciabatta et al., 2018; Brocca et al., 2019; Filippucci et al., 2020) offering good results, especially in poorly gauged regions (Massari et al, 2020).

Considering a layer characterized by a soil depth $Z$ [mm] and a soil porosity $n$ [m$^3$/m$^3$], the soil water balance equation can be written as:

$$Zn \, {}^{dSM(t)}/_{dt} = p(t) - r(t) - e(t) - g(t) \tag{1}$$

where $SM(t)$ is the relative SM [-], i.e. the soil moisture saturation fraction, $p(t)$ is the rainfall rate [mm/d], $r(t)$ is the surface runoff rate [mm/d], $e(t)$ the evaporation rate [mm/d] and $g(t)$ the drainage rate [mm/d]. During rainfall events and unsaturated conditions, evaporation and surface runoff rates can be considered negligible (Brocca et al., 2015). Equation (1) can therefore be rewritten, by using Famiglietti and Wood (1994) relationship to express the drainage rate, as:

$$p(t) = Z^* \, {}^{dSM(t)}/_{dt} \; + \; a \, SM(t)^b \tag{2}$$

with $Z^* = Zn$, a [mm/d] is the saturated hydraulic conductivity and b [-] is the exponent of Famiglietti and Wood equation. Remotely sensed SM tends to be noisy and it is sensitive to a thin topsoil layer (few centimetres). Therefore, the exponential filter approach (Wagner et al., 1999; Albergel et al., 2008) is applied to satellite SM observations before their use in equation (2). The estimation of rainfall is therefore obtained by the knowledge of two consecutive SM measurements together with 4 parameters: $Z^*$, $a$, $b$ and $T$, the time constant of the exponential filter. In its standard application, the parameter values are estimated by calibrating SM2RAIN against reference rainfall data with similar spatial and temporal resolution, with the objective of minimizing the Root Mean Square Error (RMSE).

## 3.2. Procedure for the parametric relationship

The methodology used to obtain the four parametric relationships is described here. As a first step, 1009 points uniformly distributed over the areas covered by rain gauges in Australia, India, Italy and US were selected (see ***Figure 1*** and **Brocca et al., 2019**). Each point is representative of an area of 25 x 25 $km^2$, and the spacing between the points is around 1 degree.

Secondly, the climatic and land descriptors were spatially interpolated to the 1009 points. The chosen interpolation methods consists in the nearest neighbour technique for evapotranspiration data, weighted average of the overlapping areas for datasets with a spatial resolution finer than 25 km and weighted average of the four nearest pixels for the remaining datasets. For the time-varying quantities (e.g., rainfall and soil moisture), different statistics were computed for each point to obtain the descriptors. Specifically, the daily annual average and the average number of rainy days per year was calculated for rainfall, therefore obtaining information about the climatology of each point. To obtain them, first the percentage of rainy days (precipitation > 0) and the average precipitation were calculated for each day of the year (DOY) using the available years, then the average values were calculated to obtain the annual average values. For SM, SM noise, soil temperature and evapotranspiration, the mean, median, maximum, minimum, standard deviation and coefficient of variation were computed in each point of the selected area. The temporal difference of consecutive SM and SM noise measurements was also considered, since the variation of SM is exploited in SM2RAIN to obtain rainfall. The mean, median, maximum, minimum, standard deviation and coefficient of variation were therefore calculated also over these differences, considering both the actual and the absolute values, and just the positive and negative variations, since these should be related to different mechanisms of wetting and drying of the soil (e.g., the average value of the positive variation of SM, or the maximum of the absolute variation of SM noise). For the elevation, the mean and the standard deviation within each pixel was computed; particularly the standard deviation of elevation is an indicator of topographic complexity. The latter decreases the accuracy of soil moisture retrievals, due to shadowing effects and layover (a distortion that occurs in radar imaging when the signal reflected from the top of a tall feature is received by the emitter before the one of the base, Ulaby et al., 1981).

Thirdly, the potential relationship between the descriptors and SM2RAIN parameter values was analysed through the Spearman correlation index. Spearman correlation index was calculated between each parameter and each descriptor. In order to obtain the parametric relationships, only the descriptors who showed a high absolute value of Spearman correlation (greater than 0.6) with the related parameter were considered.

SM2RAIN parametric relationships were finally obtained through a stepwise non-linear backward approach: all the possible additive and multiplicative combinations between the selected descriptors were initially used in a multilinear regression algorithm to obtain a first estimation of the relationship. An exemplary formula for the combination of *n* descriptors is here reported:

$$par = \alpha_0 + \sum_{i=1}^{n} \beta_i d_i + \sum_{i=1}^{n} \sum_{j=1}^{n-\{i\}} \gamma_i d_i d_j + \cdots + \omega \prod_{i=1}^{n} d_i \quad (3)$$

where $\alpha_i$, $\beta_i$, $\gamma_i$, and $\omega$ are the coefficients to be estimated and $d_i$ is a descriptor. The procedure was repeated iteratively by eliminating, at each step, the less significant factor, until an optimal combination of limited number of coefficients (minor or equal to 3) and good performance (drops in Spearman correlation in comparison with the previous step < 0.015) was reached. It was also verified that the Spearman correlations between each factor used in the relationship were below 0.2, in order to avoid the cross-correlation between the factors used in the parametric relationships.

## 3.3. Validation

In order to assess the goodness of the parametric relationships, the parameterized SM2RAIN rainfall product was compared with the SM2RAIN-ASCAT rainfall product obtained by calibrating SM2RAIN with the standard approach using ERA5 rainfall as reference. SM2RAIN-ASCAT was calibrated in the full available period 2013-2019, in order to compare the parameterized product with the best possible SM2RAIN version. It has to be noted that the standard calibration results presented in this paper are different from those obtained in

Brocca et al. (2019) who applied a different filtering approach and the climatology correction based on ERA5 (not considered here). Both the SM2RAIN derived rainfall products were compared with a benchmark dataset (section 3.3.1) and by using triple collocation (section 3.3.2). All the products involved in the validation were re-gridded to ASCAT grid (12.5km spacing), using the same weighted average procedure applied before (paragraph 3.2). SM2RAIN method is applicable everywhere, but the reliability of the estimated rainfall depends on the reliability of the estimated SM. This excludes all the areas with high vegetation regime, where C-band microwave measurements cannot reach the soil, coastal areas, wetlands, topographically complex areas, region characterized by subsurface scattering (Morrison et al., 2019), and frozen or snow cover terrains (Hahn et al., 2018). It was therefore defined a committed area with high confidence in the successful retrieval of surface soil moisture from MetOp ASCAT by excluding the aforementioned categories. The committed area is obtained from the EUMETSAT H SAF product validation report (Hahn et al., 2018). Two different methodologies were then used to assess rainfall products accuracy: classical performance scores and triple collocation.

### 3.3.1. Classical performance scores

Continuous metrics were applied to compare the daily rainfall estimates with the dataset taken as the 'standard', GPM-FR. In particular:

*Linear Pearson Correlation (R):* Pearson Correlation is the most common way to characterize statistical dependency between two datasets. It can be obtained from the ratio between the covariance of two dataset and the product of their standard deviation. It varies between -1 and +1, where -1 means negative linear relationship, +1 means positive linear relationship and 0 means no statistical dependency.

*Relative BIAS (BIASr):* Relative BIAS index can be calculated as the mean difference between two datasets, divided by the mean value of the reference dataset. It describes whether there is a systematic over or under-estimation with respect to the reference data. In this paper the difference is performed between the estimated and the observed rainfall. Therefore, negative BIAS values mean that the product underestimates the rainfall, while positive BIAS values indicate overestimation.

*Relative Root Mean Square Error (RMSEr):* Root mean square error (RMSE) can be calculated as the average deviation between single measurements of two dataset. It comprehends three sources of error: decorrelation, BIAS and random error. It should be noted that, since there is no "true" measure of a quantity, RMSE reliability strongly depends from the reference dataset accuracy. Relative RMSE (RMSEr) is obtained dividing RMSE by the mean value of the reference dataset.

Categorical indices were also computed to measure the performances in detecting rainfall for different precipitation classes. Five classes were selected, dividing the rainfall events in those greater than the $10^{th}$, the $30^{th}$, the $50^{th}$, the $70^{th}$ and the $90^{th}$ percentile for each point of the grid. The categorical indices were calculated for each of those classes. Naming $H$ the number of successfully predicted events, $F$ the number of falsely detected events and $M$ the number of missed events, we can define:

*False Alarm Ratio (FAR)* refers to the fraction of erroneously detected events for each class. The optimum value is 0.

$$FAR = {}^{F}/_{H+F} \tag{4}$$

*Probability Of Detection (POD)* refers to the fraction of correctly predicted events for each class. The optimum value is 1.

$$POD = {}^{H}/_{H+M} \tag{5}$$

*Threat Score (TS)* is an integrated measure of the overall performances, giving the fraction of successfully detected events over the total missed and detected events for each class. The optimum value is 1.

$$TS = H/_{H+F+M} \tag{6}$$

### 3.3.2. Triple collocation

The classical methods described above permit to assess the similarities between the analysed dataset and a reference one. Therefore, the performances reliability is dependent on the accuracy of the reference, but since no dataset has zero-error measurement (VILLARINI ET AL., 2008), not even gauges (PETERSON ET AL., 1998; KIDD ET AL., 2017) the obtained performances are subjected to error. Triple collocation (TC) method, instead, permits the assessment of uncertainties of three different products against an unknown true reference. Here a brief explanation of the theory behind the method is presented. For further information, the reader is referred to MASSARI ET AL. (2017) and STOFFELEN (1998).

Each measure related to a quantity is characterized by both a random and a systematic error:

$$X = \alpha_X + \beta_X \theta + \varepsilon_X \tag{7}$$

where $X$ is the measure, $\theta$ is the unknown truth, $\varepsilon_X$ the random error and $\alpha_X$ and $\beta_X$ are respectively the additive and multiplicative component of the systematic error. Taking into consideration three different datasets whose errors are uncorrelated, the random error of each dataset can be considered Gaussian distributed with zero mean. The error variance of each dataset can therefore be written as (MCCOLL ET AL., 2014):

$$\sigma_\varepsilon = \begin{bmatrix} \sqrt{Q_{11} - {Q_{12}Q_{13}}/{Q_{23}}} \\ \sqrt{Q_{22} - {Q_{12}Q_{23}}/{Q_{13}}} \\ \sqrt{Q_{33} - {Q_{13}Q_{23}}/{Q_{12}}} \end{bmatrix} \tag{8}$$

where $Q_{ij}$ is the covariance between the dataset $i$ and $j$. McColl underlined that, although Gaussianity ensures that the RMSE is well descripted and assists in the interpretation, Gaussian data are not required for the TC, as it is often applied to non-Gaussian data such as SM. By using the definitions of correlation and covariance, it can be derived:

$$R_{TC} = \begin{bmatrix} \sqrt{Q_{12}Q_{13}/Q_{11}Q_{23}} \\ \sqrt{Q_{12}Q_{23}/Q_{22}Q_{13}} \\ \sqrt{Q_{13}Q_{23}/Q_{33}Q_{12}} \end{bmatrix} \tag{9}$$

$R_{TC}$ is the TC correlation against the unknown truth. This measure should not be taken as an absolute measure but as a relative measure between the three datasets.

In this study, TC was used for the global analysis validation of SM2RAIN: the three products selected were therefore SM2RAIN itself (first the parameterized and then the calibrated product) and two other global rainfall datasets: GPCC and GPM_LR, chosen over GPM_FR because the latter is corrected using GPCC monthly rainfall, and therefore it does not satisfy the condition of uncorrelated error.

# 4. Results and Discussion

## 4.1. Local Analysis

The objective of this paper is to find and validate four parametric relationships to estimate the SM2RAIN algorithm parameter values from climatic and land descriptors readily available worldwide. Through these relationships, SM2RAIN can be easily applied without the need of a reference rainfall dataset. By using the 1009 points, a local analysis was performed to find the parametric relationships.

### 4.1.1. Descriptors selection

The number of potential descriptors obtained from soil data, vegetation continuous field, topography data and the statistic of time-varying quantities, exceed 50. Spearman correlation values between each of them and SM2RAIN parameters were therefore calculated, in order to reduce the number of descriptors, by selecting for each parameter the quantities that are better related to it. An example of the procedure can be found in ***Figure 2***, where three scatter density plots between the parameter $Z^*$ and three representative descriptors are shown. In the example, it can be seen how the soil water storage capacity values, obtained from the Harmonized World Soil Database, does not show significant correlation with the $Z^*$ parameter (***Figure 2c***), contrary to the expectation. Greater absolute values of Spearman correlation were obtained from the annual average daily rainfall (***Figure 2a***) and the standard deviation of the soil temperature (***Figure 2b***), with the latter showing an inverse relationship with the analysed parameter. These two descriptors were therefore selected to be used in the multilinear regression algorithm, while the soil water storage capacity was discarded (note that the standard deviation of soil temperature was discarded in the successive step). For the sake of brevity, neither the details of the descriptors selection, nor every iteration of the stepwise non-linear backward regression is described here, but the final relationships are directly shown. At the end of the procedure, most of the analysed descriptors were discarded: the descriptors who resulted more significant for SM2RAIN parameters estimation were only those related to SM, SM noise, precipitation and topography.

### 4.1.2. *T* parameter

The first obtained relationship was the one relative to the exponential filter parameter $T$. This parameter was the first to be calculated in order to obtain reduced-noise SM from satellite SM estimates. The reduced-noise SM estimates are used in equation (2) and in the calculation of the SM descriptors for the successive SM2RAIN parameters relationships. The reference values for $T$ parameter were obtained by applying the exponential filter to ASCAT SM data maximizing R between the filtered SM and the modelled SM from ERA5 (first soil layer 0-7cm). Afterward, the points with R values greater than a fixed threshold of 0.6 were retained and used as reference *T*-values to be compared with the climatic and land descriptors (see paragraph 3.2). The selection of points with correlation greater than 0.6 was done to avoid fitting the parametric relationship to not representative data. Visual inspection and Spearman correlation were used to identify which descriptors were better correlated with the reference *T*-values. A non-linear regression model was then iteratively applied to the selected descriptors in order to find the best parametric relationship. The optimal relationship can be written as:

$$T = 0.8788 + 1.7020\,\overline{SMnoise}\ std(|SM_d|) + 0.3555\ \frac{std(|SM_d|)}{\overline{P}}\ topC \tag{10}$$

where $\overline{SMnoise}$ is the temporal mean value of the SM noise relative to ASCAT estimates, $std(|SM_d|)$ is the temporal standard deviation of the absolute values of ASCAT SM temporal variations, $\overline{P}$ is the annual average of daily rainfall, and $topC$ is the topographic complexity (spatial standard deviation of elevation values within each pixel).

### 4.1.3. *b* parameter

According to Famiglietti and Wood (1994), $b$ can be considered equal to:

$$b = 3 + {}^{2}/_{\lambda} \tag{11}$$

where $\lambda$ is the pore size distribution index. A parametric relationship to estimate $\lambda$ as a function of *a* parameter was proposed by Brocca et al. (2014b):

$$\lambda = 0.085\ \log a + 0.1574 \tag{12}$$

The same relationship was adopted in this study, but the two coefficients were recalibrated using the following procedure. $T$-values from equation (10) were used to obtain filtered ASCAT SM series to which apply SM2RAIN. The three parameters of the balance equation were then calibrated against reference regional rainfall observations (1009 points). The points with R between the observed and estimated rainfall greater than the fixed threshold of 0.6 and with $T$-value less than a threshold fixed to 6, were then selected (as before to avoid fitting the parametric relationship to not representative data) and the two coefficients of equation (12) were calculated by fitting the relationship between the calibrated *a* and *b* parameter values, thus obtaining:

$$b = 3 + {}^{2}/_{(\,0.5928 * \log a + 0.3022\,)} \tag{13}$$

### 4.1.4. $Z^*$ and *a* parameter

By using equations (10) and (13), the SM2RAIN algorithm was re-applied to ASCAT SM estimates at 1009 points by only calibrating $Z^*$ and $a$ parameters. Again, the points with R between estimated and observed rainfall greater than the fixed threshold of 0.6, and with $T$-value less than 6, were selected to be compared with the climatic and land descriptors. Visual inspection and Spearman correlation were used to identify which quantities were better related with $Z^*$ parameter, then a linear regression model was applied to them in order to find the $Z^*$ parametric relationship:

$$Z^* = 10.3124 + 0.5186 \ \bar{P} \big/ \overline{|SM_d|} \tag{14}$$

The same procedure was adopted to find $a$ parametric relationship after recalibrating the SM2RAIN algorithm for only the *a* parameter and fixing the others through equations (10), (13), and (14). The obtained equations for *a* was:

$$a = -1.5748 + 13.0324 \ \ Z^* \ \overline{|SM_d|} \tag{15}$$

where $\bar{P}$ is the annual average of daily rainfall and $\overline{|SM_d|}$ is the temporal mean of the absolute values of ASCAT SM temporal variations.

### 4.1.5. Test of parametric relationships

By using equations (10), (13), (14) and (15), the four SM2RAIN parameters can be obtained from knowing the ASCAT SM timeseries and its noise, the topographic complexity and the mean annual rainfall. To avoid non-physical values for the parameters, the boundaries reported in ***Table 2*** were applied, fixing all the parameters that exceed limits to the boundary itself.

We note that the parametric relationships, obtained through a statistical regression-based approach, show physical reasoning in the expected correlation between SM2RAIN algorithm parameters and climatic and land descriptors. Indeed, equation (10) indicates that the exponential filter *T* parameter is directly proportional to the mean value of SM noise, to standard deviation of absolute SM variation, to the ratio between the latter and the annual average daily rainfall and to the topographic complexity. All these descriptors increase with either SM measurement error (i.e., SM noise and topographic complexity) or temporal SM variability (i.e., $std(|SM_d|)$ and $1/\bar{P}$); in both cases higher $T$-values are expected, since a higher value of T increases the filtering capacities. Equations (14) and (15) link the estimation of $Z^*$ and $a$ to the value of $\overline{|SM_d|}$ and $\bar{P}$. Indeed $Z^*$ increases with the ratio between $\bar{P}$ and $\overline{|SM_d|}$ because it is a measure of the amount of water stored in the soil, while $a$ is directly correlated with $\bar{P}$.

As mentioned above, to obtain the parametric relationships, SM2RAIN was applied to ASCAT SM for the 1009 points for 5 times, after and before the definition of each parameter relationship, by using the available equations and by calibrating the remaining parameters with the standard approach (i.e., minimization of RMSE). The performances of the obtained rainfall, in terms of R and RMSE are shown in ***Table 3***.

A few insights can be deduced from these results. The overall drop in performances is limited, thus demonstrating that the obtained parametric relationships are well suited to estimate the SM2RAIN parameter values. The major drop in correlation can be ascribed to the parameter $T$. This can be easily explained as the parameter $T$ is the only one related to rainfall occurrence, to which the correlation is highly sensitive, while the others parameters are more related to rainfall amount and, hence, to RMSE. A possible reason for the correlation deterioration could be due to error in modelled SM from ERA5. However, different tests with the other soil layers of ERA5 and other modelling approaches were carried out and worse results were obtained (not shown for the sake of brevity). The parametric relationship for *a* is the one that caused the greatest increase in RMSE (see ***Table 3***). Finally, we underline that soil and vegetation descriptors were found not fundamental for obtaining the parametric relationships likely due to the limited accuracy of these datasets at the considered spatial resolution, particularly for soil information, and the limited influence of vegetation on the analysed parameters, confirming the findings of Sehgal et al. (2020).

## 4.2. Global Analysis

The good results obtained at 1009 points led to the application of the parametric relationships on a global scale. The parameters maps obtained using the parametric relationships on a quasi-global scale (60° S – 60° N) are shown in ***Figure 3***. As expected, exponential filter *T* parameter is greater over desert, forest and mountain areas (***Figure 3d***), where SM quality is lower, while the distributions of $Z^*$ and *a* (***Figure 3a*** and ***3b***) reflect the known areas where the average rainfall rate is high (equatorial region). ERA5 rainfall was used to obtain the annual average daily rainfall for the parametric relationships and also to calibrate SM2RAIN ASCAT with the standard methodology to verify that the uncalibrated product performances are in line with those of the calibrated SM2RAIN. The performance of the two rainfall datasets were then assessed against the GPM-FR precipitation product, in terms of the categorical indices False Alarm Ratio (FAR), Probability of Detection (POD) and Threat Score (TS), and the continuous indices R, BIASr and RMSEr. It should be noticed that the GPM product contains both the solid and liquid fraction of the precipitation, while SM2RAIN is able to estimate only the liquid fraction. The masking of frozen condition for SM ASCAT product should be able to remove the days of solid precipitation from the comparison, but in case of failure of frozen condition detection, this issue could be a source of error, in particular over high elevation and high latitude regions. From IMERG V05B, full coverage is provided for the latitudes of 60°N-60°S, while the remaining upper and lower latitudes extending to 90° are considered "partial coverage". The current analysis was restrained to the full coverage area (60°N-60°S) to increase the accuracy of the results. From now on, the product obtained from the use of the parametric relationships will be labelled as "*parameterized*", while the one obtained using the standard calibration method will be named as "*calibrated*".

The distribution of the categorical indices is shown in ***Figure 4*** as boxplots. The indices were calculated for five rainfall classes, respectively the $10^{th}$, the $30^{th}$, the $50^{th}$, the $70^{th}$ and the $90^{th}$ percentile of the precipitation for every point. Regarding the FAR, the two products show similar performances for the first two classes, while the parameterized product has a higher percentage of false alarms for the last three classes. Different observations can be done for the POD index: the calibrated product performs slightly better than the parameterized for the first three classes, while it is true the opposite for the others two. From this information, it can be inferred that the parameterized product has greater capability in estimating the major rainfall events. However, the performances of the parameterized product are slightly worse than those of the calibration product for lower percentiles (<50$^{th}$), due to a greater number of false alarms and to a lower detection ability. These results are confirmed by the TS scores, which indicates a slightly better performance of the calibrated product for the first four classes, while for the fifth class the parameterized product performs better.

The differences in the performance are due to the different parameter values adopted by the two products. ***Figure 5a*** shows the $T$ parameter distribution for the whole area: there are clear differences between the calibrated and parameterized values, in both the median and the range of values. This is probably a consequence of the strategy used to estimate $T$, comparing the satellite SM with a modelled SM, instead of calibrating the $T$-values with respect to reference rainfall. The parameterized product tends therefore to filter the SM data less than the calibrated product (i.e., lower $T$-values), thus increasing the average SM variation, that in turn increase the overall estimated rainfall. This is the cause of the increase of both the FAR and POD indices. $Z^{*}$ and $a$ (***Figure 5b and 5c***) show instead similar behaviour between the two products. The values for the parameterized product are slightly greater than the calibrated product that is the reason of the overestimation tendency noted above. Finally, ***Figure 5d*** shows the distributions of $b$ parameter, which has similar median value but a very different variability range, due to the relationship, equation (13), used for relating $a$ and $b$ parameters.

R, BIASr and RMSEr were also calculated for the global analysis. The distribution boxplots of these performance indices are shown in ***Figure 6***. The obtained results confirm the outcomes of the categorical indices analysis: in terms of R the calibrated product is slightly better than the parameterized (0.4866 vs 0.4777 in the committed area); the range of R-values is also comparable (***Figure 6a***). BIASr for the parameterized product is around 0.2, confirming the tendency to overestimate rainfall, whereas the calibrated product has a tendency to underestimation. RMSEr values (***Figure 6c***) are very similar between the parameterized and the calibrated product, with differences lower than 5%.

The global map of R and RMSEr difference between parameterized and calibrated products are shown in ***Figure 7 and 8***, respectively. In the figures, red colour indicates that the parameterized product is better than the calibrated one, while the blue colour indicates the opposite. An overall increase in Pearson correlation is noticeable in the tropical area, while many mountainous areas show a decrement in R values. The relative error increase in forest, desert and mountainous areas, likely due to the lower filtering of SM values in the parameterized product. One possible cause of the performance deterioration in topographically complex zones could be related to the low spatial resolution of the selected DEM, ETOPO5. A different product with higher resolution will be tested in future studies. The mean Pearson correlation values shown in ***Figure 6a*** are around 0.5 for the committed area and 0.4 globally. These correlation values are likely due to the differences between the dataset used to calibrate SM2RAIN and to obtain the climatology for the parametric relationships, i.e., ERA5, and the product considered as benchmark, i.e., GPM-FR. ERA5 and GPM-FR are indeed not highly correlated by each other; the mean value of R between them equals to 0.5604 in the committed area (0.5412 globally).

Due to the difficulty to have a reliable rainfall benchmark on a global scale, a TC analysis was performed to assess the capability of SM2RAIN products in rainfall estimation against an unknown true reference. Since TC requires three different products whose errors are uncorrelated, two other global rainfall datasets were selected to be compared with the parameterized and the calibrated products, separately: GPCC and GPM-LR, chosen over GPM-FR as the latter is corrected using GPCC monthly rainfall, and therefore it does not satisfy the condition of uncorrelated error with GPCC. Since GPCC has a low spatial resolution (1-degree), it was interpolated over ASCAT grid using a weighted average, where the weights are the relative inverse of the distance between each ASCAT pixel and the four nearest GPCC points. ***Figure 9*** shows the boxplot of the obtained $R_{TC}$. SM2RAIN rainfall products have mean values comparable with those of the other products for the committed area, with a slight deterioration when the parametric relationships are used instead of the standard calibration. Moreover, the areas where the parameterized product performs better than the other two is large (red areas in ***Figure 10***): GPCC is the best performing product over most of Europe and Asia, half of North America and half of Australia, where there is a high density of gauge stations. GPM-LR performs better over forest and desert areas, where ASCAT SM has large errors (green areas in ***Figure 10***). In most of the remaining zones (red areas in ***Figure 10***), SM2RAIN ASCAT derived from the parametric relationships performs better than the other products, confirming the capacity of SM2RAIN in estimating rainfall over Africa and South America (Brocca et al., 2020; Massari et al., 2020) also when the parameterized product is considered. For further information about the mutual correlation between the products, the individual maps of TC correlation have been added in the appendix. As mentioned above, the TC correlation values should not be taken as absolute measures of accuracy but rather as relative measures between the three datasets.

# 5. Conclusions

In this paper, four parametric relationships were developed to estimate SM2RAIN parameter values from climatic and land descriptors. A local analysis was performed over a regular grid of 1009 points uniformly distributed over the areas covered by rain gauges in Australia, India, Italy and US, for which high quality observed rainfall data were available. Several climatic and land descriptor datasets were analysed to obtain an inclusive description of each point and to find the descriptors related to the four SM2RAIN parameters. The four parametric relationships were finally developed, obtaining the parameter values estimation from the knowledge of the SM timeseries and its noise, the topographic complexity and the mean annual rainfall. The major drops in correlation due the use of the parametric relationships, instead of the calibration against a reference, is caused by the $T$ parameter, while the major increase in RMSE is caused by the $a$ parameter. Possible causes of these behaviours could be attributed to the accuracy of the selected datasets for obtaining the descriptors, and these problems will be investigated in future studies.

To validate the obtained results, a global application of SM2RAIN on ASCAT SM was performed using the parametric relationships (parameterized SM2RAIN-ASCAT) and the standard calibration methodology (calibrated SM2RAIN-ASCAT). ERA5 rainfall was used to assess mean annual rainfall and as calibration dataset; while GPM-FR rainfall was used as benchmark to calculate performance indices. From the analysis of the categorical and continuous scores, an overall similar capacity in rainfall estimation between the calibrated and parameterized product is found. In particular, even if the calibrated product has slightly better performances both in terms of correlation and bias, the parameterized product resulted more capable in the detection of larger rainfall events.

Finally, a triple collocation analysis was performed by using GPM-LR, GPCC and the two SM2RAIN-ASCAT products. The analysis revealed that, even if the parameterized version of SM2RAIN-ASCAT has slightly lower correlations than the others in the committed area, there are several regions (e.g., in Africa and South America) in which its performance is better than both GPM-LR and GPCC, suggesting the utility of this product for rainfall estimation.

In future studies, the addition of new descriptors to estimate SM2RAIN parameters will be investigated (e.g. Radio Frequency Interference indicators, Land Cover, high resolution topography). These relationships could be very important for estimating rainfall from high resolution SM, since calibration data with high spatial and temporal resolution are often unavailable. Therefore, the methodology will be applied to SM timeseries from Sentinel-1 in order to assess their validity and to provide a self-calibrated high resolution (<1km) rainfall product from remote sensing.

**Acknowledgements**

The authors gratefully acknowledge support from EUMETSAT through the Global SM2RAIN project (contract no. EUM/CO/17/4600001981/BBo) and the “Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF)” CDOP 3 (grant no. EUM/C/85/16/DOC/15).

**Financial support**

This research has been supported by EUMETSAT (Global SM2RAIN project grant no. EUM/CO/17/4600001981/BBo and “Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF)” CDOP 3 grant no. EUM/C/85/16/DOC/15).

# References


Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J. C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., Martin, E. (2008). From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. Hydrology and Earth System Sciences, 12, 1323–1337. Doi: 10.5194/hess-12-1323-2008

Bauer-Marschallinger, B., Paulik, C., Hochstöger, S., Mistelbaue, T., Modanesi, S., Ciabatta, L., Massari, C., Brocca, L., Wagner, W. (2018). Soil Moisture from Fusion of Scatterometer and SAR: Closing the Scale Gap with Temporal Filtering. Remote Sensing, 10 (7), 1030. Doi: 10.3390/rs10071030

Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., Modanesi, S., Massari, C., Ciabatta, L., Brocca, L., Wolfgang, W. (2019). Towards Global Soil Moisture Monitoring with Sentinel-1: Harnessing Assets and Overcoming Obstacles. IEEE Transactions on Geoscience and Remote Sensing, 57 (1), 520–539. Doi: 10.1109/TGRS.2018.2858004.

Brocca, L., Melone, F., Moramarco, T., Wagner, W. (2013). A new method for rainfall estimation through soil moisture observations. Geophysical Research Letters, 40, 853–858. Doi: 10.1002/grl.50173

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., Levizzani, V. (2014a). Soil as a natural rain gauge: estimating global rainfall from satellite soil moisture data. Journal of Geophysical Research, 119(9), 5128–5141. Doi: 10.1002/2014JD021489

Brocca, L., Camici, S., Melone, F., Moramarco, T., Martínez-Fernández, J., Didon-Lescot, J. F., Morbidelli, R. (2014b). Improving the representation of soil moisture by using a semi-analytical infiltration model. Hydrological Processes, 28, 2103–2115. Doi: 10.1002/hyp.9766

Brocca, L., Massari, C., Ciabatta, L., Moramarco, T., Penna, D., Zucco, G., Pianezzola, L., Borga, M., Matgen, P., Martínez-Fernández, J. (2015). 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. Doi: 10.1515/johh-2015-0016

Brocca, L., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B., Wagner, W. (2019). SM2RAIN-ASCAT (2007-2018): global daily satellite rainfall from ASCAT soil moisture. Earth System Science Data, 11, 1583–1601. Doi: 10.5194/essd-11-1583-2019

Brocca, L., Massari, C., Pellarin, T., Filippucci, P., Ciabatta, L., Camici, S., Kerr, Y. H., Fernández-Prieto, D. (2020). River flow prediction in data scarce regions: soil moisture integrated satellite rainfall products outperform rain gauge observations in West Africa. Scientific Reports, 10, 12517. Doi: 10.1038/s41598-020-69343-x

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

Camici, S., Ciabatta, L., Massari, C., Brocca, L. (2018). How reliable are satellite precipitation estimates for driving hydrological models: a verification study over the Mediterranean area. Journal of Hydrology, 563, 950–961. Doi: 10.1016/j.jhydrol.2018.06.067

Cenci, L., Pulvirenti, L., Boni, G., Chini, M., Matgen, P., Gabellani, S., Squicciarino, G., Pierdicca, N. (2017). An evaluation of the potential of Sentinel 1 for improving flash flood predictions via soil moisture-data assimilation. Advances in Geosciences, 44, 89–100, 2017. Doi: 10.5194/adgeo-44-89-2017

Chen, F., Crow, W. T., Ciabatta, L., Filippucci, P., Panegrossi, G., Marra, A. C., Puca, S., Massari, C. (2020). Enhanced large-scale validation of satellite-based land rainfall products. Journal of Hydrometeorology, published online. Doi: 10.1175/JHM-D-20-0056.1

Ciabatta, L., Brocca, L., Massari, C., Moramarco, T., Puca, S., Rinollo, A., Gabellani, S., Wagner, W. (2015). Integration of satellite soil moisture and rainfall observations over the Italian territory. Journal of Hydrometeorology, 16 (3), 1341–1355. Doi: 10.1175/JHM-D-14-0108.1

Ciabatta, L., Camici, S., Brocca, L., Ponziani, F., Stelluti, M., Berni, N., Moramarco, T. (2016). Assessing the impact of climate-change scenarios on landslide occurrence in Umbria Region, Italy. Journal of Hydrology, 541A, 258-295. Doi: 10.1016/j.jhydrol.2016.02.007

Ciabatta, L., Massari, C., Brocca, L., Gruber, A., Reimer, C., Hahn, S., Paulik, C., Dorigo, W., Kidd, R., Wagner, W. (2018). SM2RAIN-CCI: a new global long-term rainfall data set derived from ESA CCI soil moisture. Earth System Science Data, 10, 267–280. Doi: 10.5194/essd-10-267-2018.

Crow, W. T., Huffman, G. F., Bindlish, R., Jackson, T. J. (2009). Improving satellite rainfall accumulation estimates using spaceborne soil moisture retrievals. Journal of Hydrometeorology, 10, 199–212. Doi: 10.1175/2008JHM986.1

Dari, J., Brocca, L., Quintana-Seguí, P., Escorihuela, M. J., Stefan, V., Morbidelli, R. (2020). Exploiting High-Resolution Remote Sensing Soil Moisture to Estimate Irrigation Water Amounts over a Mediterranean Region, Remote Sensing, 12 (16), 2593. Doi: 10.3390/rs12162593

Famiglietti, J., Wood, E. F. (1994). Multiscale modeling of spatially variable water and energy balance processes, Water Resources Research, 30, 3061–3078. Doi: 10.1029/94WR01498

Filippucci, P., Tarpanelli, A., Massari, C., Serafini, A., Strati, V., Alberi, M., Raptis, K. G. C., Mantovani, F., Brocca, L. (2020). Soil moisture as a potential variable for tracking and quantifying irrigation: a case study with proximal gamma-ray spectroscopy data. Advances in Water Resources, 136, 103502. Doi: 10.1016/j.advwatres.2019.103502

Hahn, S., Delogu, F., Gabellani, S., Brocca L. (2018). Product Validation Report (PVR) H113, Tech. rep., H SAF, Doc.No: SAF/HSAF/CDOP3/PVR/H113, available at: http://hsaf.meteoam.it/CaseStudy/GetDocumentUserDocument?fileName=H113_ASCAT_SSM_CDR_PVR_v0.3.pdf&tipo=PVR

Hannah, D. M., Demuth, S., Van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., Tallaksen, L. M. (2011). Large-scale river flow archives: importance, current status and future needs. Hydrological Processes, 25 (7), 1191-1200. Doi: 10.1002/hyp.7794

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J. N. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146, 1999– 2049. Doi: 10.1002/qj.3803

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., Iguchi, T. (2014). The global precipitation measurement mission. Bulletin of the American Meteorological Society, 95 (5), 701–722. Doi: 10.1175/BAMS-D-13-00164.1

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Stocker, E. F., Tan, J., Wolff, D. B., Xie, P. (2020) Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM) mission (IMERG). Chapter 19 in Adv. Global Change Res., Vol. 67, Satellite Precipitation Measurement, V. Levizzani, C. Kidd, D. Kirschbaum, C. Kummerow, K. Nakamura, F.J. Turk (Ed.), Springer Nature, Dordrecht, ISBN 978-3-030-24567-2 / 978-3-030-24568-9 (eBook), 343-353. Doi:/10.1007/978-3-030-24568-9_19

IPCC (2013). The Physical Science Basis, Summary for Policymakers. International Panel of Climate Change, Working Group 1 contribution, Climate Change 2013Jakeman, A. J., Hornberger, G. M., Littlewood, I. G., Whitehead, P. G., Harvey, J. W., Bencala, K. E., (1992). A systematic approach to modelling the dynamic linkage of climate, physical catchment descriptors and hydrological response components. Math. Comp. Sim. 33, 359–366. Doi: 10.1016/0378-4754(92)90122-W

Kidd, C., Levizzani, V. (2011). Status of satellite precipitation retrievals. Hydrology and Earth System Science, 15, 1109-1116. Doi: 10.5194/hess-15-1109-2011

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P.,Skofronick-Jackson, G., Kirschbaum, D. B. (2017). So, How Much of the Earth's Surface is Covered by Rain Gauges? Bulletin of the American Meteorological Society, 98, 69–78. Doi: 10.1175/BAMS-D-14-00283.1

Maggioni, V., Massari, C. (2018). On the performance of satellite precipitation products in riverine flood modeling: a review. Journal of Hydrology, 558, 214–224. Doi: 10.1016/j.jhydrol.2018.01.039

Massari, C., Crow, W., Brocca, L. (2017). An assessment of the performance of global rainfall estimates without ground-based observations. Hydrology and Earth System Sciences, 21, 4347–4361. Doi: 10.5194/hess-21-4347-2017

Massari, C., Brocca, L., Pellarin, T., Abramowitz, G., Filippucci, P., Ciabatta, L., Maggioni, V., Kerr, Y., Fernández-Prieto, D. (2020). A daily/25km short-latency rainfall product for data scarce regions based on the integration of the GPM IMERG Early Run with multiple satellite soil moisture products. Hydrology and Earth System Sciences, 24, 2687–2710. Doi: 10.5194/hess-24-2687-2020

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. Geophysical Research Letters, 41, 6229–6236. Doi: 10.1002/2014GL061322

Merlin, O., Chehbouni, A., Walker, J. P., Panciera, R., Kerr, Y.H (2008). A simple method to disaggregate passive microwave-based soil moisture. IEEE Transactions on Geoscience and Remote Sensing. Remote Sens., 46, 786–796. Doi: 10.1109/TGRS.2007.914807

Morrison, K., Wagner, W. (2020). Explaining Anomalies in SAR and Scatterometer Soil Moisture Retrievals From Dry Soils With Subsurface Scattering, IEEE Transactions on Geoscience and Remote Sensing, 58 (3), 2190–2197. Doi: 10.1109/TGRS.2019.2954771

Pai, D., Sridhar, L., Rajeevan, M., Sreejith, O., Satbhai, N., Mukhopadhyay, B. (2014). Development of a new high spatial resolution (0.25×0.25) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. Mausam, 65, 1–18. Doi: unavailable

Pellarin, T., Louvet, S., Gruhier, C., Quantin, G., Legout, C. (2013). A simple and effective method for correcting soil moisture and precipitation estimates using AMSR-E measurements. Remote Sensing of Environment, 136, 28–36. Doi: 10.1016/j.rse.2013.04.011

Peng, J., Loew, A., Merlin, O., Verhoest, N. E. (2017). A review of spatial downscaling of satellite remotely sensed soil moisture. Reviews of Geophysics, 55, 341–366. Doi: 10.1002/2016RG000543

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P.,Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gul-lett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O.,Szentimrey, T., Salinger, J., Førland, E. J., Hanssen-Bauer, I.,Alexandersson, H., Jones, P., Parker, D. (1998). Homogeneity adjustments of in situ atmospheric climate data: a review. International Journal of Climatology, 18, 1493–1517. Doi: 10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T

Pignone, F., Rebora, N., Silvestro, F., Castelli, F. (2010). GRISO – Rain, CIMA Research Foundation, Savona, Italy. Operational Agreement 778/2009 DPC-CIMA, Year-1 Activity Report 272/2010. Doi: unavailable

Ponziani, F., Pandolfo, C., Stelluti, M., Berni, N., Brocca, L., Moramarco, T. (2012). Assessment of rainfall thresholds and soil moisture modeling for operational hydrogeological risk prevention in the Umbria region (central Italy). Landslides, 9, 229-237. Doi: 10.1007/s10346-011-0287-3

Schamm, K., Ziese, M., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U., Schröder, M., Stender, P. (2014). Global gridded precipitation over land: a description of the new GPCC First Guess Daily product. Earth System Science Data, 6, 49–60. Doi: 10.5194/essd-6-49-2014

Sefton, C. E. M., Howarth, S. M. (1998). Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. J. Hydrol. 211, 1–16. Doi: 10.1016/S0022-1694(98)00163-2

Sehgal, V., Gaur, N., Mohanty, B. P. (2020). Global Surface Soil Moisture Drydown Patterns. Water Resources Research, 56, e2020WR027588. Doi: 10.1029/2020WR027588

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. Agric. Forest Meteor. 98–99, 279–293. Doi: 10.1016/s0168-1923(99)00105-7

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. Journal of Geophysical Research: Oceans, 103, 7755–7766, 1998. Doi: 10.1029/97JC03180

Tarpanelli, A., Massari, C., Ciabatta, L., Filippucci, P., Amarnath, G., Brocca, L. (2017). Exploiting a constellation of satellite soil moisture sensors for accurate rainfall estimation. Advances in Water Resources, 108, 249-255. Doi: 10.1016/j.advwatres.2017.08.010

Ulaby, F. T., Moore, R. K., Fung, A. K. (1981). Microwave Remote Sensing, Active and Passive. vol. 1, 2, 3. Artech House. Doi: unavailable

Villarini, G., Mandapaka, P. V., Krajewski, W. F., Moore, R. J. (2008). Rainfall and sampling uncertainties: A rain gauge perspective, Journal of Geophysical Research, 113, D11102. Doi: 10.1029/2007JD009214

Vörösmarty, C., Askew, A., Grabs, W., Barry, R. G., Birkett, C., Döll, P., Goodison, B., Hall, A., Jenne, R., Kitaev, L., Landwehr, J., Keeler, M., Leavesley, G., Schaake, J., Strzepek, K., Sundarvel, S. S., Takeuchi, K., Webster, F., (2001). Global water data: A newly endangered species. Eos Transactions American Geophysical Union, 82 (5), 54–58. Doi: 10.1029/01EO00031

Wagener, T., Wheater, H. S., Gupta, H. V., 2004. Rainfall-Runoff Modelling in Gauged and Ungauged Catchments. Imperial College Press, London p. 300. Doi: 10.1142/p335

Wagner, W., Lemoine, G., Rott, H. (1999). A method for estimating soil moisture from ERS scatterometer and soil data. Remote Sensing of Environment, 70, 191–207. Doi: 10.1016/S0034-4257(99)00036-X

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., Steinnocher, K. (2013). The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. Meteorologische Zeitschrift, 22, 5–33. Doi: 10.1127/0941-2948/2013/0399

Wieder, W. R., J. Boehnert, G. B. Bonan. (2014). Evaluating soil biogeochemistry parameterizations in Earth system models with observations, Global Biogeochemical Cycles, 28, 211–222. Doi: 10.1002/2013GB004665

Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., Yang S. (2007). A gauge-based analysis of daily precipitation over East Asia. Journal of Hydrometeorology, 8, 607. 626. Doi: 10.1175/JHM583.1

# Tables

Table 1: description of the dataset downloaded and processed but not selected for the derivation of SM2RAIN parametric relationships

| VARIABLE | SOURCE | TEMPORAL RESOLUTION | SPATIAL SAMPLING | ADDITIONAL INFORMATION |
|---|---|---|---|---|
| Soil Temperature (0-7 cm) | ERA5 Land | 1 h | 0.1° | https://www.ecmwf.int/en/era5-land |
| Evapotranspiration | ERA5 | 1 h | 0.25° | https://www.ecmwf.int/en/era5 |
| Rainfall | ERA5 | 1 h | 0.25° | https://www.ecmwf.int/en/era5 |
| Rainfall | IMERG Late Run | 0.5 h | 0.1° | https://gpm.nasa.gov/data/directory |
| Rainfall | IMERG Final Run | 0.5 h | 0.1° | https://gpm.nasa.gov/data/directory |
| Rainfall | GPCC | 1 d | 1° | Schamm et al. (2014) |
| Rainfall | AWAP | 1 d | 0.05° | http://www.bom.gov.au/jsp/awap/rain/index.jsp |
| Rainfall | IMD | 1 d | 0.25° | http://www.imd.gov.in/pages/service_hydromet.php |
| Rainfall | CPC | 1 d | 0.25° | https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html |
| Rainfall | ITA-DPC | 1 d | 0.1° | Ciabatta et al. (2017) |
| Soil Composition Data | HWSD | / | ~0.008° | http://www.fao.org/land-water/databases-and-software/hwsd/en/ |
| Soil Moisture | ASCAT | ~12 h | ~0.125° | Wagner et al. (2013) |
| Soil Moisture (0-7 cm) | ERA5 Land | 1 h | 0.1° | https://www.ecmwf.int/en/era5-land |
| Topography | ETOPO5 | / | ~0.08° | https://www.ngdc.noaa.gov/mgg/global/etopo5.HTML |
| Vegetation Continuous Fields | VCF5KYR | / | 0.05° | https://lpdaac.usgs.gov/products/vcf5kyrv001/ |

Table 2: Upper and Lower boundaries for SM2RAIN parameters

| Boundaries | $Z^*$ [mm] | $a$ [mm/d] | $b$ [-] | $T$ [d] |
|---|---|---|---|---|
| **Lower** | 20 | 0.1 | 1 | 0 |
| **Upper** | 800 | 200 | 50 | 8 |

Table 3: Mean value and variation of Pearson Correlation (R) and Root Mean Square Error (RMSE) for local analysis points, calculated after and before the establishment of each parametric relationship

| | Mean R [-] | ΔR [-] | Mean RMSE [mm/d] | ΔRMSE [mm/d] |
|---|---|---|---|---|
| **Calibrated SM2RAIN** | 0.5951 | | 4.4126 | |
| **T fixed** | 0.5757 | -0.0194 | 4.4909 | 0.0783 |
| **T, b fixed** | 0.5712 | -0.0045 | 4.5226 | 0.0317 |
| **T, b, Z fixed** | 0.5631 | -0.0081 | 4.6142 | 0.0916 |
| **Parameterized SM2RAIN** | 0.5567 | -0.0064 | 4.7915 | 0.1773 |

# Figures

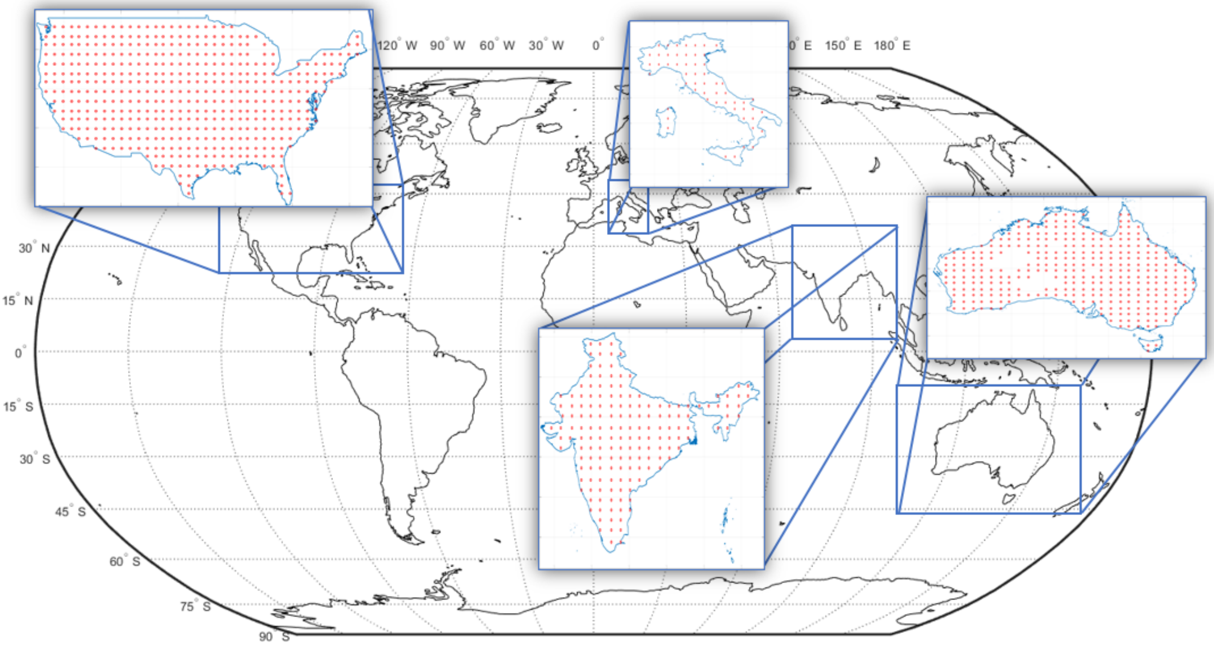



Figure 1: 1009 points grid for the local analysis, uniformly distributed over the areas covered by rain gauges in Australia, India, Italy and USA.

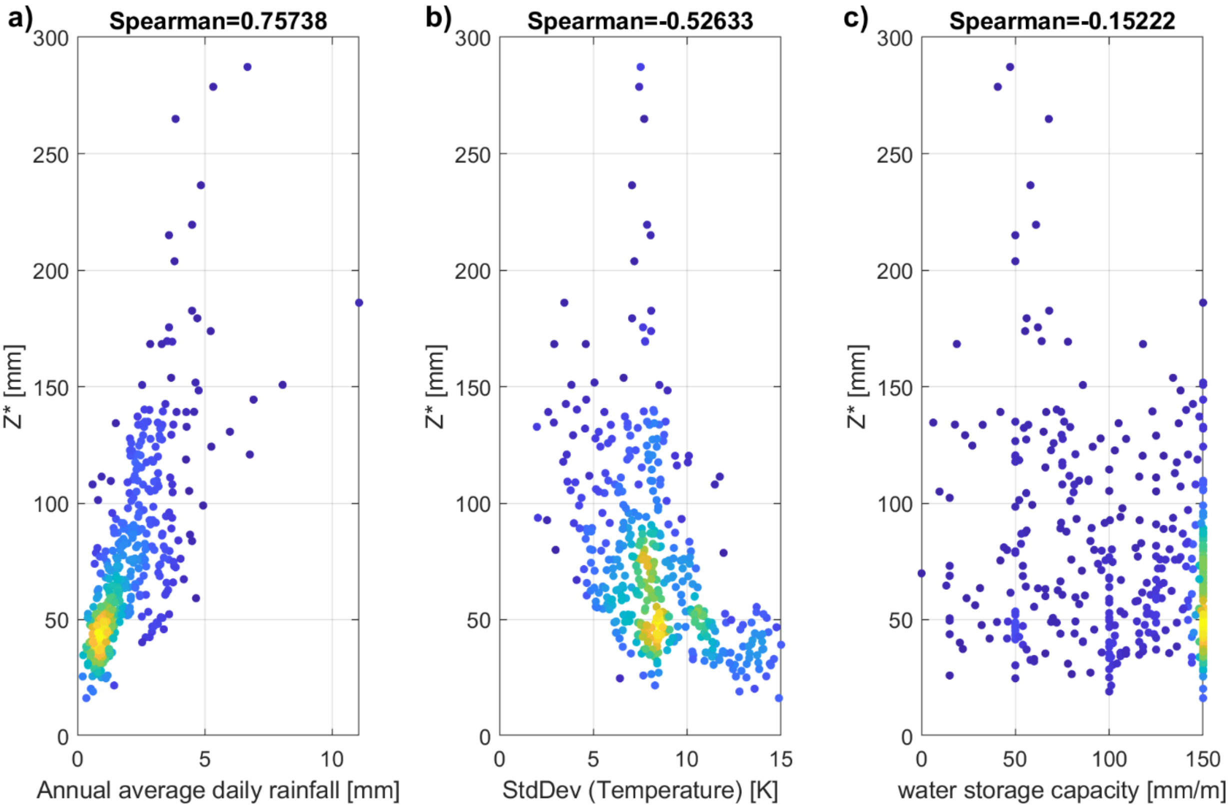



Figure 2: Example of the descriptors selection procedure. In the three panels is shown a scatter density plot of *Z** parameter distribution with respect to the annual average daily rainfall (a), the standard deviation of the soil temperature (b) and the soil water storage capacity (c) for the analysed area. Spearman correlation is shown on top of each panel.

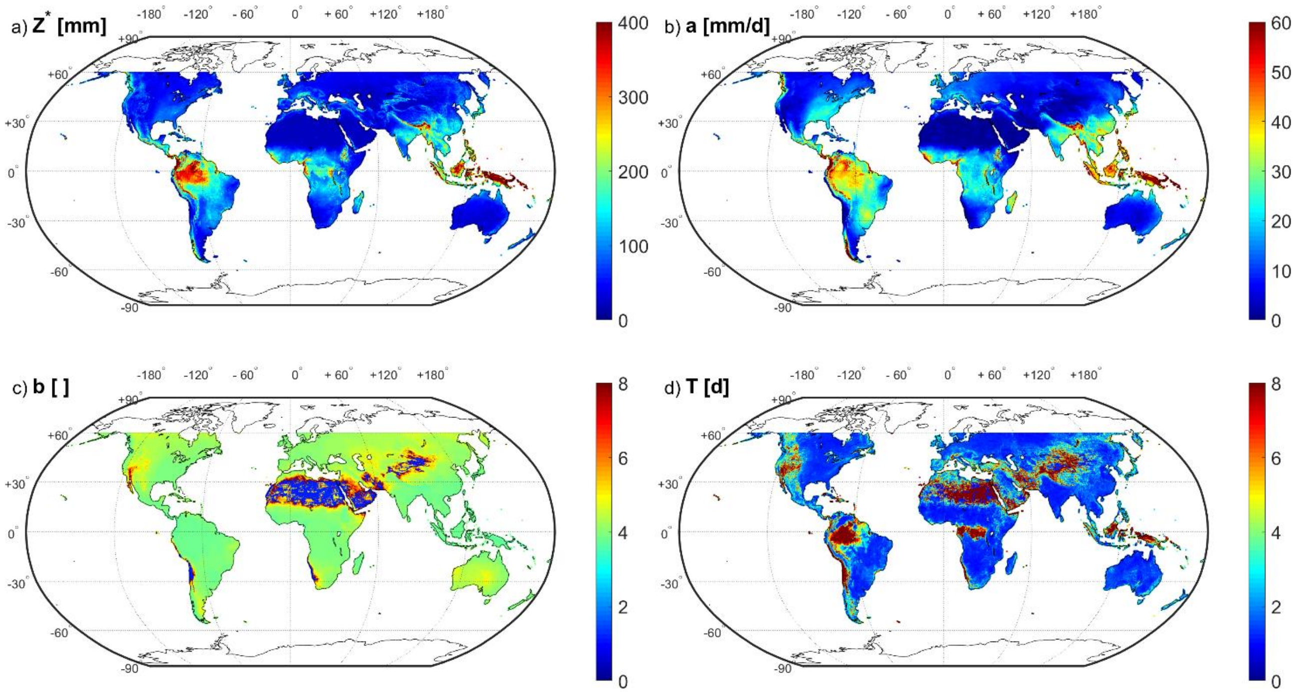
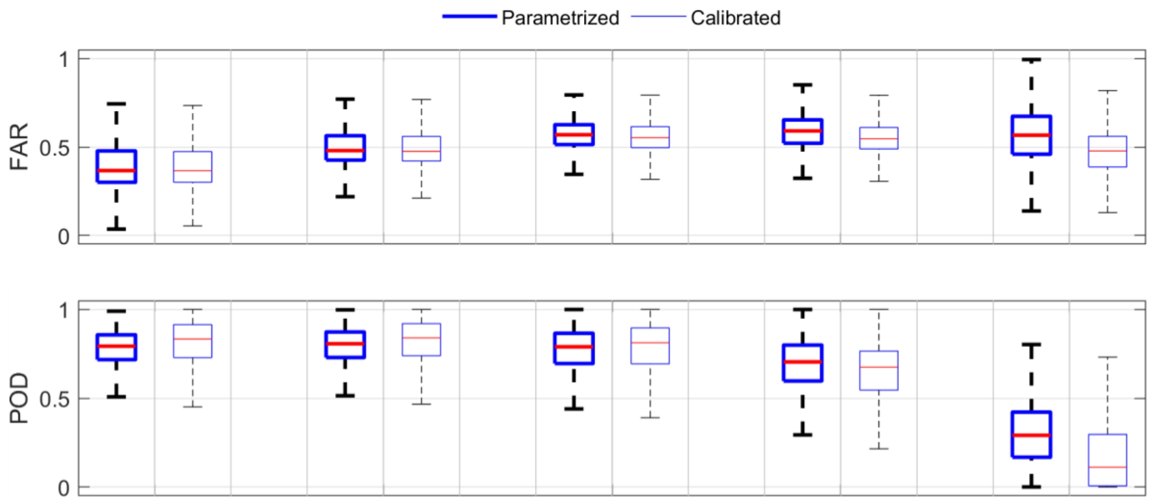



Figure 3: Global Map of SM2RAIN parameter values as obtained from the parametric relationships. Each panel shows: a) parameter $Z^*$, b) parameter *a,* c) parameter *b,* d) parameter *T.*



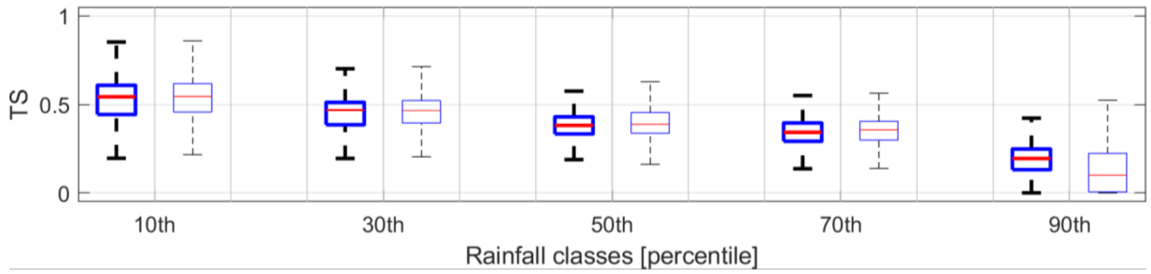



Figure 4: Distribution of False Alarm Ratio, FAR, Probability of Detection, POD and Threat Score, TS categorical indices of the parameterized and calibrated SM2RAIN rainfall products, respectively in dark and light blue, against the benchmark dataset GPM - Final Run, related to the committed area. The indices are calculated for five rainfall classes, according to the intensity of the observed rainfall events being greater than the 10th, the 30th, the 50th, the 70th and the 90th percentiles.

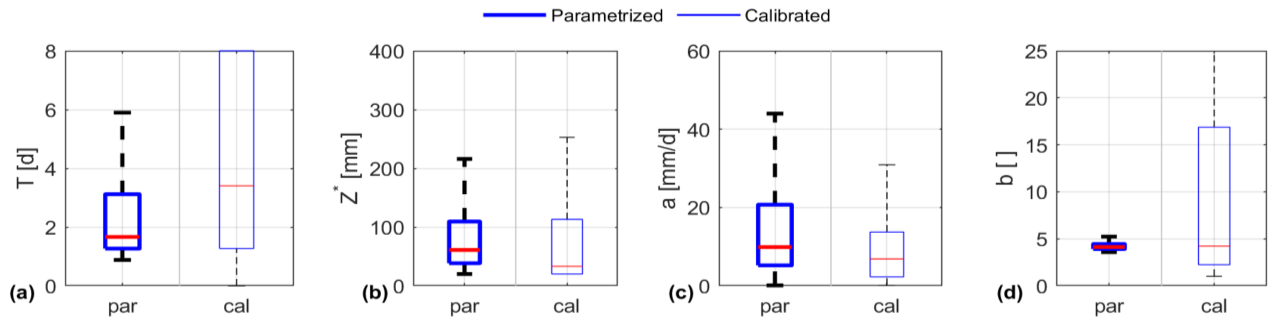



Figure 5: distribution of SM2RAIN parameters $T$, $Z^*$, $a$ and $b$ over the whole area for the parameterized, dark blue, and the calibrated, light blue, SM2RAIN rainfall products

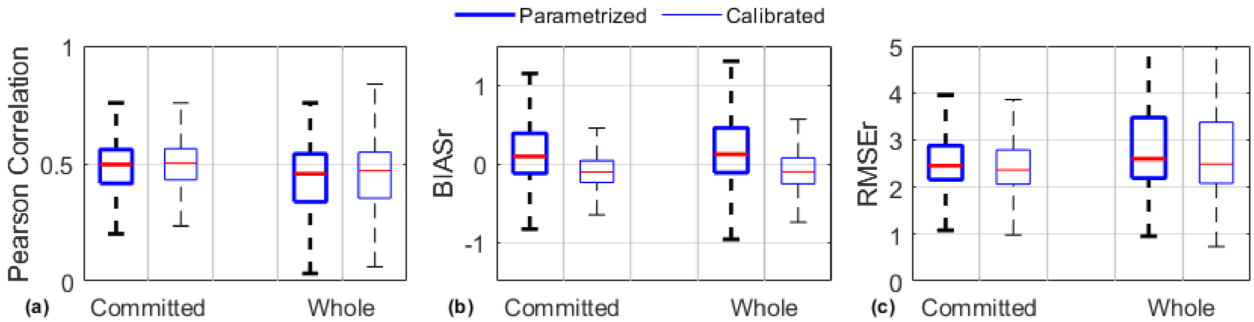



Figure 6: Distribution of Pearson Correlation, relative BIAS (BIASr) and relative Root Mean Square Error (RMSEr) indices of the parameterized and calibrated SM2RAIN rainfall products, respectively in dark and light

blue, against the benchmark dataset GPM - Final Run. In each panel, the results related to the committed area are on the left and those related to the global area are on the right.

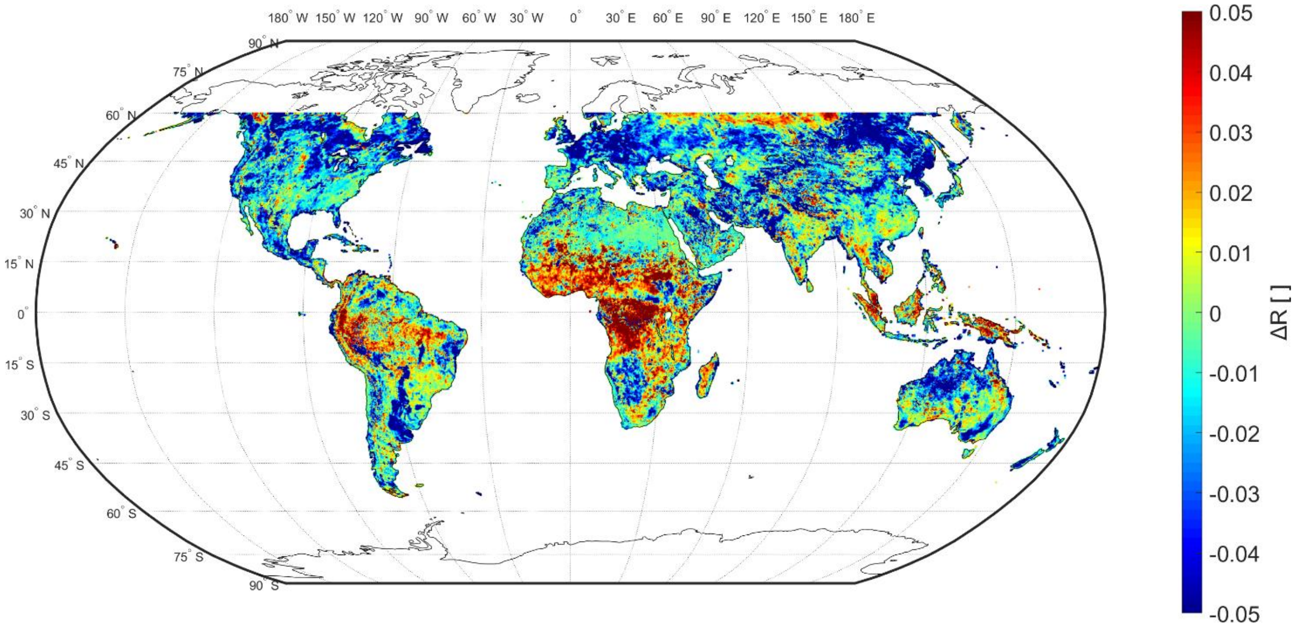



Figure 7: Global map of differences between the parameterized and calibrated SM2RAIN rainfall products for the Pearson correlation score calculated against GPM – Final Run product. Red areas mean that the parameterized product outperforms the calibrated one, the opposite for blue areas. The parameterized product shows an increase of correlation over dense forest and frozen areas.

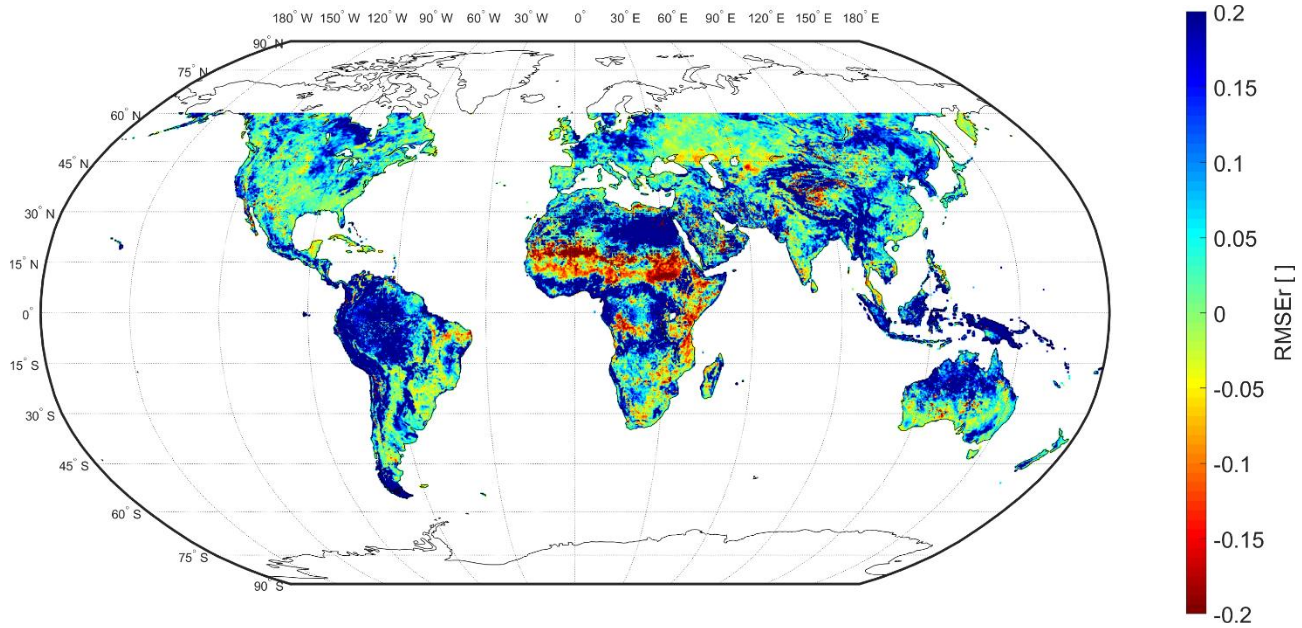



Figure 8: Global map of differences between the parameterized and calibrated SM2RAIN rainfall products for the relative Root Mean Square Error score calculated against GPM – Final Run product. Red areas mean that the parameterized product outperforms the calibrated one, the opposite for blue areas.

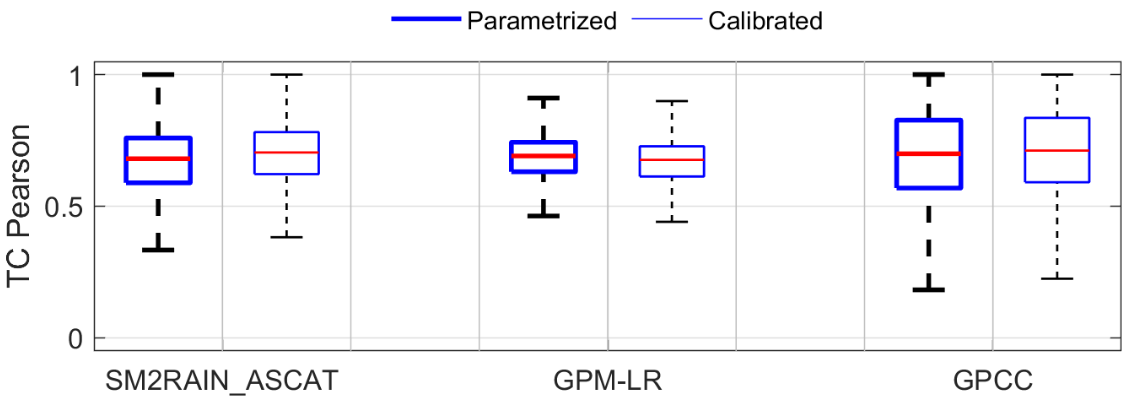



Figure 9: Distribution of the Triple Collocation correlation obtained from the rainfall products triplets composed from SM2RAIN, GPM – Late Run and GPCC, over the committed area. The results of the parameterized products are shown in dark blue, while those of the calibrated product are in light blue.

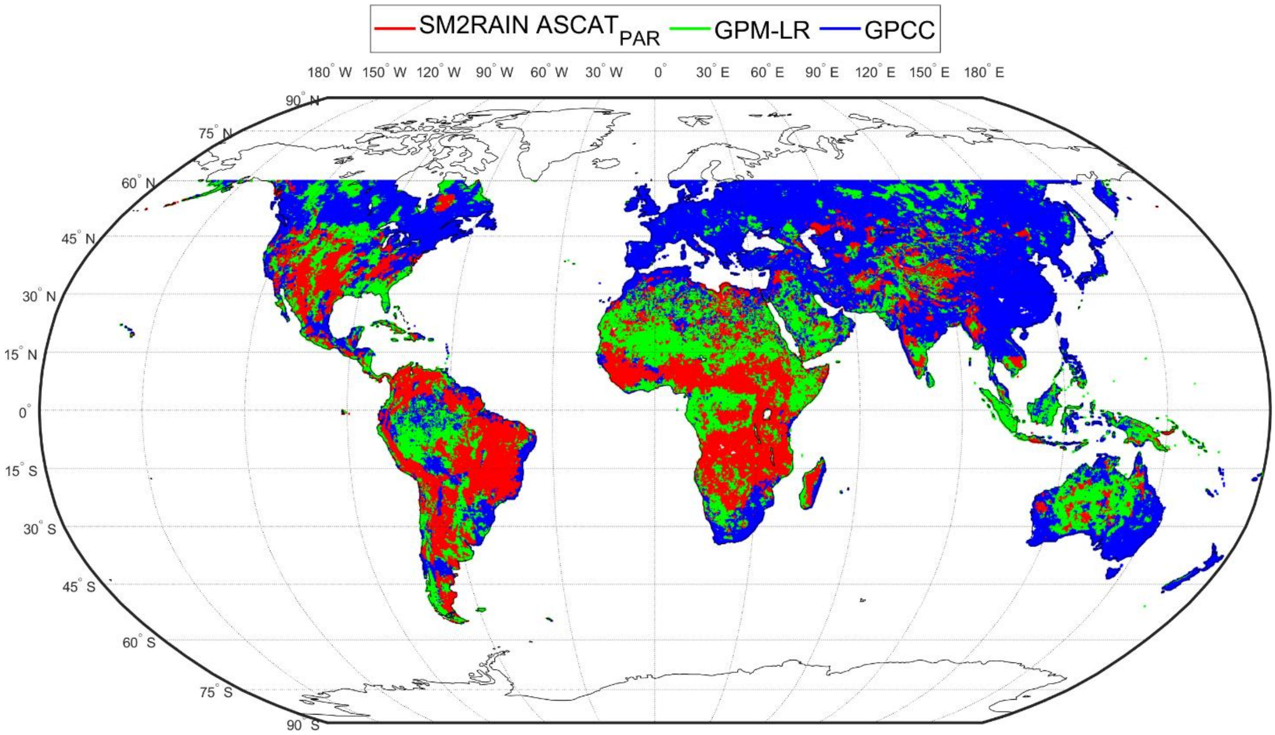



Figure 10: Map of best performing products based on the results of Triple Collocation of the rainfall products triplet SM2RAIN parameterized (red), GPM – Late Run (green) and GPCC (blue). The parameterized SM2RAIN-ASCAT product outperforms the others in those areas characterized by low density of gauge and good quality of the SM retrievals.

# Appendix

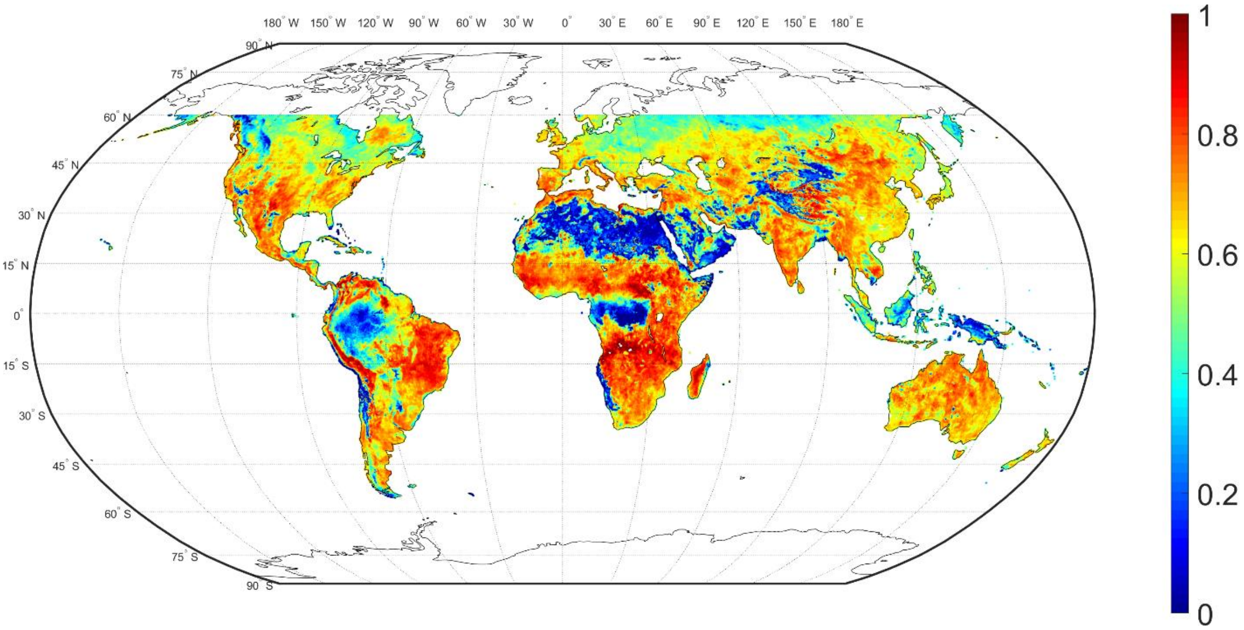



*Figure A-1: Map of TC correlation of the parameterized SM2RAIN-ASCAT based on the results of Triple Collocation of the rainfall products triplet SM2RAIN parameterized, GPM – Late Run and GPCC.*

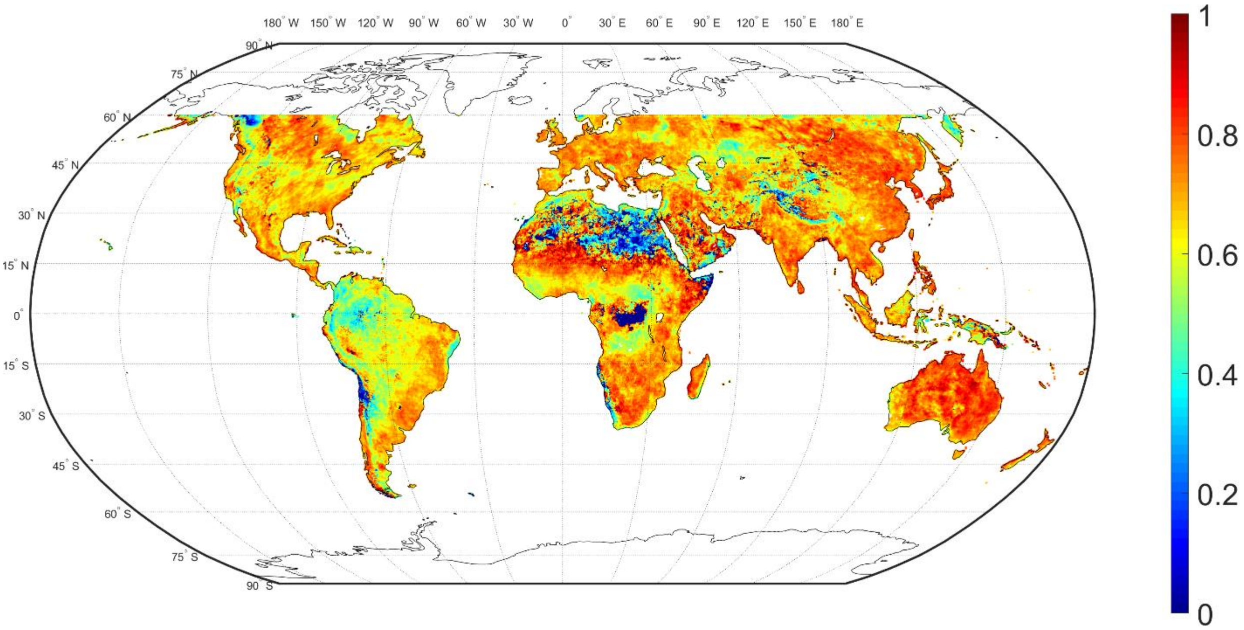



*Figure A-2: Map of TC correlation of the GPM – Late Run based on the results of Triple Collocation of the rainfall products triplet SM2RAIN parameterized, GPM – Late Run and GPCC.*

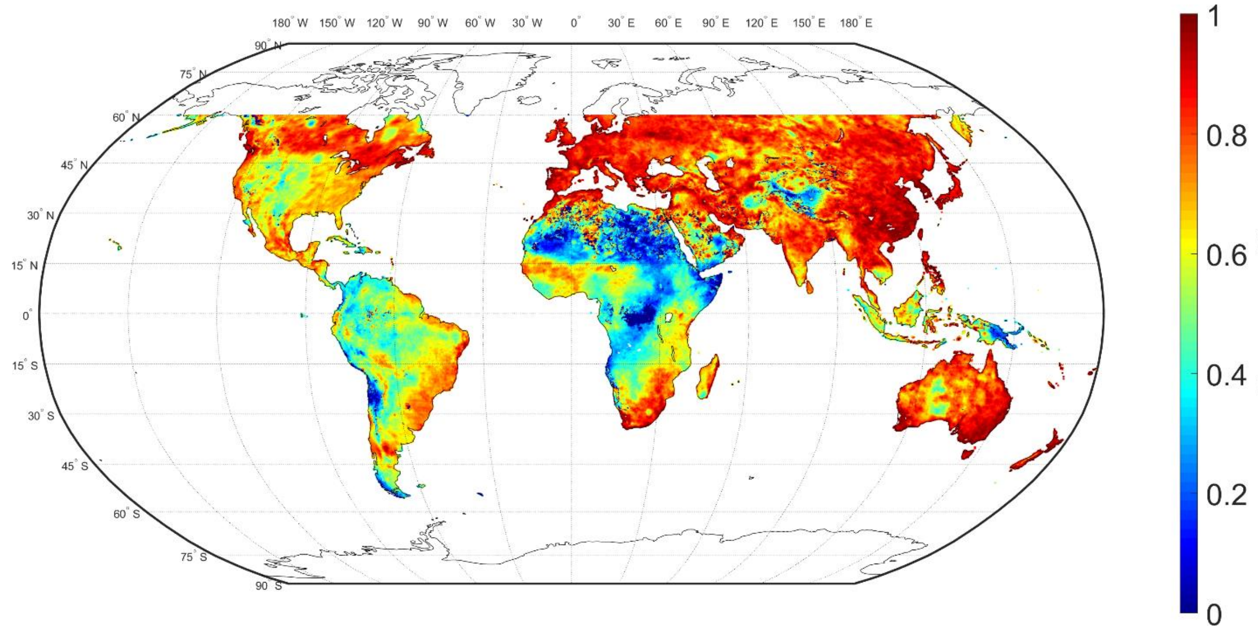



*Figure A-3: Map of TC correlation of the GPCC based on the results of Triple Collocation of the rainfall products triplet SM2RAIN parameterized, GPM – Late Run and GPCC.*