Enhanced Large-Scale Validation of Satellite-Based Land Rainfall Products

F. CHEN,[a,b] W. T. CROW,[b] L. CIABATTA,[c] P. FILIPPUCCI,[c] G. PANEGROSSI,[d] A. C. MARRA,[d] S. PUCA,[e] AND C. MASSARI[c]

[a] *SSAI, Inc., Lanham, Maryland*
[b] *Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland*
[c] *Research Institute for Geo-Hydrological Protection, Italian National Research Council, Perugia, Italy*
[d] *Institute of Atmospheric Sciences and Climate, Italian National Research Council, Rome, Italy*
[e] *National Civil Protection Department, Rome, Italy*

(Manuscript received 3 March 2020, in final form 6 November 2020)

ABSTRACT: Satellite-based precipitation estimates (SPEs) are generally validated using ground-based rain gauge or radar observations. However, in poorly instrumented regions, uncertainty in these references can lead to biased assessments of SPE accuracy. As a result, at regional or continental scales, an objective basis to evaluate SPEs is currently lacking. Here, we evaluate the potential for large-scale, spatially continuous evaluation of SPEs over land via the application of collocation-based techniques [i.e., triple collocation (TC) and quadruple collocation (QC) analyses]. Our collocation approach leverages the Soil Moisture to Rain (SM2RAIN) rainfall product, derived from the time series analysis of satellite-based soil moisture retrievals, in combination with independent rainfall datasets acquired from ground observations and climate reanalysis to validate four years of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) H23 daily rainfall product. Large-scale maps of the H23 correlation metric are generated using both TC and QC analyses. Results demonstrate that the SM2RAIN product is a uniquely valuable independent product for collocation analyses, because other available large-scale rainfall datasets are often based on overlapping data sources and algorithms. In particular, the availability of SM2RAIN facilitates the large-scale evaluation of SPE products like H23—even in areas that lack adequate ground-based observations to apply traditional validation approaches.

KEYWORDS: Rainfall; Satellite observations; Reanalysis data; Surface observations

1. Introduction

Satellite-based precipitation estimates (SPE) are increasingly being applied to important environmental applications such as numerical weather prediction, flood forecasting, and agricultural drought monitoring. A potential SPE of interest is the H23 gridded precipitation product generated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF). The H23 product is based on the temporal resampling of H-SAF's passive microwave (PMW) instantaneous precipitation rate products obtained from various satellite platforms and provides coverage of Europe, Africa, and parts of South America (60°S–75°N, 60°W–60°E) at a 0.25° spatial resolution with 1-day data latency.

Given these attributes, the H23 product is suitable for regional near-real-time monitoring and forecasting activities. However, it is a demonstrational precipitation product within H-SAF that is not currently publicly released. Prior to its release, an accurate statistical error assessment is required to ensure compliance with prespecified user accuracy and generate uncertainty estimates for downstream applications (estimations of streamflow, drought severity, inundation extent, etc.).

Over a regional or continental domain, SPEs like H23 are generally validated via comparisons against a single reference dataset (e.g., a modeling product or spatially interpolated ground observations acquired from rain gauge networks and ground-based weather radars) (e.g., Ebert et al. 2007; Sapiano and Arkin 2009; Stampoulis and Anagnostou 2012). However, such references are subject to their own uncertainties (Villarini et al. 2008; Stampoulis and Anagnostou 2012; Kidd et al. 2012; Prein and Gobiet 2017) that can bias SPE evaluation metrics.

In recent years, collocation-based mathematical solutions based on the cross-comparison of multiple independent datasets have been increasingly applied to obtain error estimates for measurement or model estimates of geophysical variables such as soil moisture (e.g., Scipal et al. 2010; Draper et al. 2013; Chen et al. 2018), leaf area index (e.g., Fang et al. 2012), and ocean wind and wave heights (e.g., Caires and Sterl 2003; Chakraborty et al. 2013; Wang et al. 2014). The application of collocation-based approaches to rainfall products is described in Mahfouf et al. (2007), Wang et al. (2018), and Massari et al. (2017).

Depending on the number of independent datasets available, either triple collocation (TC, requiring three datasets) or quadruple collocation (QC, requiring four datasets) analysis can be applied. Previous applications of TC for SPE evaluation (Roebeling et al. 2012; Alemohammad et al. 2015; Massari et al. 2017; Li et al. 2018) have demonstrated its potential value for global- and regional-scale studies that include data-scarce regions.

The chief advantage of collocation approaches is they do not require that any single available dataset be considered as an error-free reference. Instead, TC analysis requires only the availability of three datasets with mutually independent errors. To construct such a triplet, satellite- and ground-based datasets can be combined with precipitation fields from model reanalysis products. Because reanalysis datasets are typically generated retrospectively

Corresponding author: F. Chen, fan.chen@usda.gov

DOI: 10.1175/JHM-D-20-0056.1

by ingesting a broad range of satellite, atmospheric, and ground observations, their quality is relatively high. However, the breadth of data combined to generate these products can also undermine their independence with respect to other members of a rainfall product triplet. Likewise, while multiple SPE datasets are commonly available, they generally contain cross-correlated errors due to their overlapping use of common satellite instrumentation and/or retrieval algorithms (Massari et al. 2017).

An innovative and independent candidate product for the application of TC to SPEs can be generated using remotely sensed soil moisture time series via the Soil Moisture to Rain (SM2RAIN) method (Brocca et al. 2014). The SM2RAIN algorithm inverts rainfall accumulation from satellite soil moisture observations using the time difference between two successive measurements. SM2RAIN-derived rainfall products (hereinafter referred to as "SM2R") have been shown to perform relatively well at both regional and global scales and outperform other SPEs in data-scarce regions of the world, such as Africa and South America (Brocca et al. 2019). In a recent study, a short-latency rainfall product generated by combining the Integrated Multisatellite Retrievals for Global Precipitation Measurement (GPM) Early Run (IMERG-ER) product with multiple-satellite SM-based rainfall products generated using SM2RAIN was found to improve the correlation and root-mean-square error metrics of IMERG-ER by 20% and 40%, respectively (Massari et al. 2020).

Since the source of rainfall information utilized in SM2R is fundamentally different from other large-scale rainfall products, it is an appealing candidate for TC analysis. Furthermore, the availability of SM2R estimates (with the addition of a fourth rainfall dataset) enables the application of QC to back out error cross-correlation (ECC) information in cases in which error dependency between two products cannot be ruled out (Gruber et al. 2016a). Nevertheless, important limitations of SM2R must be considered, including the 1) underestimation of peak rainfall events, 2) presence of spurious low-intensity rainfall events due to high-frequency soil moisture fluctuations associated with random measurement error, and 3) limitation to only terrestrial and liquid-phase precipitation (Brocca et al. 2019).

Taken together, the newly developed SM2R product and recent advances in collocation approaches provide a novel path for validating SPE products. In this study, we explore this possibility by conducting a collocation-based, regional- and continental-scale SPE validation analysis that is enhanced relative to the common practice of relying on a single-source reference dataset. To this end, we will first perform a QC analysis to provide a robust assessment of the H23 correlation metric (with respect to unknown ground truth) when three additional independent rainfall datasets (i.e., SM2R, a gauge-based gridded rainfall dataset and a climate reanalysis product) are available. Next, we will validate an analogous TC analysis that removes the earlier dependence on ground-based observations. In this way, we will attempt to develop a credible TC analysis for assessing the H23 product that does not require the availability of high-quality, ground-based observations. Once validated, we will apply our TC analysis to the broader H23 domain (60°S–75°N, 60°W–60°E).

The remainder of the paper is organized as follows: section 2 provides information about the various rainfall datasets used in our analysis, section 3 explains our TC and QC methods and assessment strategy, section 4 describes our main findings, and section 5 summarizes and discusses our results.

2. Rainfall data products

a. H-SAF H23 product

As noted above, our primary goal is the enhanced large-scale evaluation of the H-SAF demonstrational H23 rainfall product using SM2R and collocation approaches. H-SAF demonstrational products are currently provided to users for testing and feedback without any commitment on quality or availability. The H23 is a level-3 gridded precipitation product providing daily (0000–0000 UTC) mean rainfall rate estimates in units of millimeter per hour that is based on the temporal resampling of PMW instantaneous precipitation rate estimates obtained from multiple satellites (http://hsaf.meteoam.it/description-h02b-h03b-h05b-h15b-h17-h18-h23.php) on a 0.25° × 0.25° grid over the full Meteosat Second Generation (MSG) disk (60°S–75°N, 60°W–60°E). Gridded PMW-based mean precipitation is obtained from the temporal resampling of instantaneous precipitation maps provided by H-SAF operational product H01 (Casella et al. 2013; Sanò et al. 2013; Mugnai et al. 2013a,b) and H02B (Sanò et al. 2015), for SSMIS and AMSU/MHS, respectively, by exploiting all available DMSP SSMIS and MetOp/NOAA AMSU/MHS satellite overpasses. In addition to daily-mean hourly rainfall rates, the total number of satellite-based PMW overpasses within the last 24 h that was used to derive the daily estimate is also provided on a grid-by-grid basis in the H23 product. In a future version of H23, currently under development, overpasses from all other GPM constellation radiometers will also be considered.

A prototype version of H23 described in Ciabatta et al. (2017) and provided by the H-SAF consortium for the period 2011–14 is applied here as a test-case dataset. H23 data since 2018 are available on the H-SAF FTP site (ftphsaf.meteoam.it) following registration (http://hsaf.meteoam.it/user-registration.php). For simplicity, we refer to this retrospective test dataset as "H23" since the algorithm and the input data are the same as for the contemporary H-SAF product.

b. E-OBS

The European Daily High-Resolution Observational Gridded Dataset (E-OBS) v17.0 precipitation product is based on the spatial interpolation of daily rainfall observations acquired from over 10 000 rain gauge stations in Europe on a regular 0.25° × 0.25° grid between 25° and 75°N and between 40°W and 75°E (van Engelen et al. 2008). The E-OBS data applied here were downloaded from the European Climate Assessment and Dataset project (ECA&D; https://www.ecad.eu). The E-OBS product has been designed to provide the best-available estimate of gridscale, daily (0000–0000 UTC) precipitation accumulations via a three-step interpolation process. First, monthly precipitation totals are interpolated onto a 0.25° grid using thin-plate splines. Next, daily anomalies are interpolated onto the same grid using universal kriging with an external drift factor for temperature. Last, monthly totals and daily anomalies are merged into a single daily rainfall estimate for each grid cell.

Figure 1 describes the density of rain gauges within Europe that underlie E-OBS accumulation estimates during our period

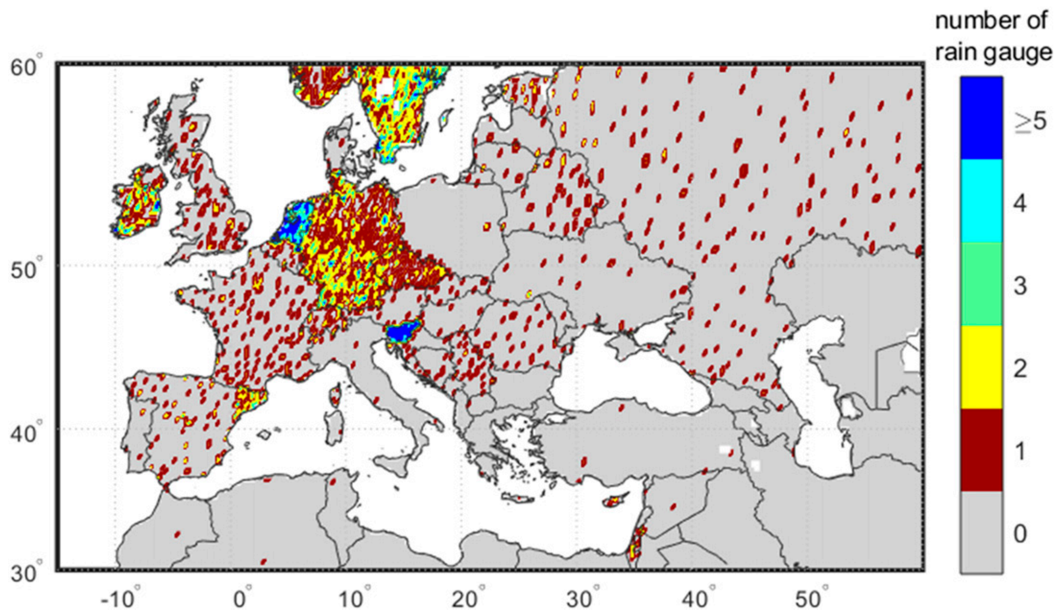

FIG. 1. Number of ground gauges per 0.25° grid cell underlying the E-OBS daily rainfall estimates (2011–14).

of interest (2011–14). Note that E-OBS v17.0 data are missing over Poland. In addition, gauge observations within Russia are missing for calendar year 2014.

c. SM2RAIN-ASCAT (SM2R)

The SM2RAIN-ASCAT (Brocca et al. 2019) daily rainfall dataset is based on the application of the SM2RAIN algorithm (Brocca et al. 2013, 2014) to soil moisture retrievals acquired from the Advanced Scatterometer (ASCAT) on board the *MetOp-A* and *MetOp-B* satellites (i.e., H-SAF product H113). The EUMETSAT H-SAF ASCAT product has a varying spatial resolution of 25–34 km from near swath to far swath and is available on a discrete global grid with a nominal spatial resolution of 25 km and a grid spacing of 12.5 km. Using *MetOp-A* (2007–present) and *MetOp-B* (2013–present) in tandem provides at least two measurements per day at midlatitudes. The ASCAT surface soil moisture retrievals are linearly interpolated into an equally spaced 12-hourly time series before application of the SM2RAIN algorithm. For additional details regarding the preprocessing of ASCAT data, see Brocca et al. (2019).

SM2RAIN is based on the inversion of the soil water balance equation to solve for surface rainfall accumulation. It assumes negligible evapotranspiration and surface runoff during rainfall events at the scale of the satellite soil moisture retrievals. Brocca et al. (2014, 2015) describes the SM2RAIN algorithm and its underlying assumptions. The SM2RAIN-ASCAT rainfall dataset (hereinafter referred to as "SM2R") provides daily rainfall accumulation with a spatial resolution of 25 ~ 34 km on a 12.5-km grid for the period 2007–18. Because of limitations on the availability of ASCAT soil moisture retrievals, SM2R only provides liquid rainfall accumulation estimates over land (Brocca et al. 2019). The SM2RAIN-ASCAT dataset (v1.0) utilized here is available for download (https://doi.org/10.5281/zenodo.2580285).

The ASCAT data used to generate SM2R were masked when their Surface State Flag indicated frozen conditions. To assess the H23 product over the widest range of surface conditions possible, additional masking of SM2R rainfall estimates (based on, e.g., soil moisture noise and/or sensitivity, vegetation density or topographic complexity) was not applied. However, the impact of vegetation density on SM2R performance, as well as the overall reliability of our TC analysis, is discussed in section 5.

The SM2R data for 2011–14 were resampled onto a regular 0.25° grid by spatially averaging 12.5-km pixels whose centers fall within each 25-km grid box. It should be noted that ERA5 rainfall (described below) is used to calibrate the parameters of SM2RAIN algorithm and also as the reference dataset to apply monthly bias corrections to SM2R data records (Brocca et al. 2019) during a postprocessing step. This raises an obvious concern about violating the error independence assumption underlying TC and QC—see section 4a for additional discussion on this point.

d. ERA5

ERA5 is the latest-generation European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis. It is available from 1950 onward and will eventually replace the ERA-Interim reanalysis product (Hersbach et al. 2018). ERA5 is based on 4D-Var data assimilation using Cycle 41r2 of the Integrated Forecasting System (IFS). Here, the global 0.25° gridded ERA5 hourly total precipitation field is used together with ERA5 snowfall estimates to obtain hourly total precipitation estimates. Daily accumulations were obtained by summing hourly accumulations between 0000 and 0000 UTC next day. All ERA5 data used here are available from the C3S Climate Data Store (C3S 2017).

Note that ERA assimilates satellite radiance observations (infrared and microwave) and ground-based radar precipitation observations (from 2009) to produce its reanalyzed precipitation field. Because of the potential cross-use of SSMIS and

AMSU/MHS microwave observations in both the ERA5 and H23 products, it is necessary to consider the possibility of H23 and ERA5 estimates containing cross-correlated error. Such error cross dependence would violate a key TC assumption—see section 4a for further discussion.

3. Method

a. Triple collocation

Triple collocation (TC) is based on the underlying assumption of a linear additive error model linking multiple independent measurement systems with the unknown truth:

$$X = \alpha + \beta T + \varepsilon, \tag{1}$$

where X is assumed to correspond to a daily rainfall product; T is the true daily rainfall accumulation; α and β are additive and multiplicative biases, respectively; and ε is mean-zero random error.

The extended triple collocation (ETC, hereinafter referred to as TC) approach (McColl et al. 2014) can be applied to estimate the correlation R of a measurement system relative to T. Given the availability of three gridded daily rainfall products (X, Y, and Z) that linearly relate to the true daily rainfall intensity as described in Eq. (1), the correlation between X and the unknown truth T can be estimated as

$$R_X = \sqrt{\frac{\sigma_{XY}\sigma_{XZ}}{\sigma_X^2\sigma_{YZ}}}, \tag{2}$$

where σ_{XY} is the temporal covariance of X and Y, for example, and σ_X^2 is the variance of X.

In addition to the linearity assumption, TC also requires 1) mutually independent error impacting X, Y, and Z; 2) errors that are uncorrelated to T (i.e., error orthogonality); and 3) the stationarity of signal and error statistics (i.e., homoscedasticity) (Gruber et al. 2016b; Draper et al. 2013; Zwieback et al. 2012).

Since TC is known to break down for very low product skill, we apply TC only at pixels where significant ($p = 0.05$) levels of mutual correlation exist between all three rainfall products. To ensure the stationarity of signal and error statistics, a common practice is to remove seasonal signals from the raw time series of estimates prior to the application of TC (e.g., Chen et al. 2017; Gruber et al. 2016b; Su and Ryu 2015). However, because of the binary nature of precipitation events, zero values are meaningful observations that should not be artificially transformed into nonzero anomaly values. Therefore, raw daily rainfall time series are used here in our TC and QC analyses.

b. Quadruple collocation

As noted above, a critical assumption in TC is that errors in each product are mutually independent. To maximize the likelihood of error independence, the rainfall products examined here are based on a wide range of measurement principles. The H23 product is a blended PWM precipitation product based on observations provided by SSMIS on board DMSP satellites and AMSU/MHS on board *MetOp-A/B* and *NOAA-18/19* satellites. Although the SM2R product is also remote sensing-based, it is derived from active scatterometer observations of the land surface acquired from the *MetOp-A/B* satellites—a completely different strategy than that employed by the H23 product. Likewise, ERA5 precipitation estimates are generated via the assimilation of a wide range of atmospheric and land surface variables into a coupled land–atmosphere modeling system. The E-OBS product is generated via spatial interpolation of ground-based rain gauge observations with no ancillary data.

However, despite our best efforts to diversify sources of rainfall information, it is difficult to eliminate the possibility of error codependence. When a fourth rainfall product is available, R can be estimated using QC based on the formulation given in Gruber et al. (2016a). The same assumptions for TC also apply for QC, but since the four datasets constitute an overconstrained system, QC can solve for one additional nonzero error covariance term via a least squares solution (Pierdicca et al. 2015). Therefore, the zero ECC assumption can be relaxed to allow nonzero ECC to exist between one, and only one, pair of data products. This provides an opportunity to verify the zero ECC assumption underlying TC.

Given four rainfall measurement systems X, Y, Z, W, and assuming that nonzero ECC exists only between X and Y, the least squares solution for the QC problem is

$$\mathbf{M} = \begin{bmatrix} \sigma_X^2 \\ \sigma_Y^2 \\ \sigma_Z^2 \\ \sigma_W^2 \\ \sigma_{XY} \\ \sigma_{XY}\sigma_{XW}/\sigma_{ZW} \\ \sigma_{YZ}\sigma_{YW}/\sigma_{ZW} \\ \sigma_{XZ}\sigma_{ZW}/\sigma_{XW} \\ \sigma_{YZ}\sigma_{ZW}/\sigma_{YW} \\ \sigma_{XW}\sigma_{ZW}/\sigma_{XZ} \\ \sigma_{YW}\sigma_{ZW}/\sigma_{YZ} \\ \sigma_{XZ}\sigma_{YW}/\sigma_{ZW} \\ \sigma_{XW}\sigma_{YZ}/\sigma_{ZW} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1&0&0&0&0&1&0&0&0&0 \\ 0&1&0&0&0&0&1&0&0&0 \\ 0&0&1&0&0&0&0&1&0&0 \\ 0&0&0&1&0&0&0&0&1&0 \\ 0&0&0&0&1&0&0&0&0&1 \\ 1&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0 \end{bmatrix}, \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} \beta_X^2\sigma_T^2 \\ \beta_Y^2\sigma_T^2 \\ \beta_Z^2\sigma_T^2 \\ \beta_W^2\sigma_T^2 \\ \beta_X\beta_Y\sigma_T^2 \\ \sigma_{\varepsilon_X}^2 \\ \sigma_{\varepsilon_Y}^2 \\ \sigma_{\varepsilon_Z}^2 \\ \sigma_{\varepsilon_W}^2 \\ \sigma_{\varepsilon_X\varepsilon_Y} \end{bmatrix}, \tag{3}$$

where σ_T^2 is the true daily rainfall variance, β is the multiplicative bias in Eq. (1), σ_ε^2 is the variance of the random error, and $\sigma_{\varepsilon_X \varepsilon_Y}$ is the error covariance between X and Y.

The least squares solution for the parameters in **S** is

$$\hat{\mathbf{S}} = (\mathbf{A}^\mathrm{T}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}\mathbf{M}. \tag{4}$$

Correlation with the unknown truth is

$$R_X = \sqrt{1 - \frac{\sigma_{\varepsilon_X}^2}{\sigma_X^2}}. \tag{5}$$

The ECC between X and Y is

$$\mathrm{ECC}_{XY} = \frac{\sigma_{\varepsilon_X \varepsilon_Y}}{\left(\sigma_{\varepsilon_X}^2 \sigma_{\varepsilon_Y}^2\right)}. \tag{6}$$

c. *Rainfall error model*

For rainfall estimates, a multiplicative error model

$$X = \alpha T^\beta e^\varepsilon \tag{7}$$

is often considered to be more appropriate than an additive error model like Eq. (1) (Hossain and Anagnostou 2006; Tian et al. 2013; Alemohammad et al. 2015). Here, e^ε is the multiplicative random error and α and β describe systematic biases. Tian et al. (2013) compared the additive and multiplicative error models applied to daily precipitation datasets across the United States and suggested that the use of a multiplicative model allows for the improved separation of random errors from systematic signals and is applicable to a wider range of daily precipitation values. In addition, applying a log transformation to a multiplicative error model converts it to a linear form that is amenable to TC (Alemohammad et al. 2015; Massari et al. 2017).

However, the challenge in applying a multiplicative error model to daily precipitation lies in the log-transformation process, which requires that all raw values be nonzero. Solutions to this problem include temporal aggregation (e.g., Alemohammad et al. 2015) to remove zeros in the time series or simply discarding zero observations (e.g., Massari et al. 2017). However, as pointed out in Massari et al. (2017), daily precipitation errors have different characteristics than multiday accumulation errors due to the frequent presence of zero daily values. Also, TC rainfall metrics are less reliable when their sampling power is reduced by the removal of zero values. To resolve this issue, Massari et al. (2017) compared the performance of multiplicative and additive error models during the application of TC to daily precipitation estimates in the United States and concluded that, despite the theoretical advantages of a multiplicative error model, an additive error model provides reasonable and more robust error results. Therefore, in this analysis, we employed an additive error model for both the TC and QC analyses presented below.

d. *Collocation analyses of R*

Correlation metric R results for H23 estimated using both TC (with a H23–SM2R–ERA5 triplet) and QC (with a H23–EOBS–SM2R–ERA5 quadruplet) will be discussed in section 4. During preprocessing, individual 0.25° grid cells are masked for the cases of 1) less than 100 complete daily data triplets (for TC) or quadruplets (for QC) or 2) the lack of significant correlation (p value < 0.05) between all cross-sampled data pairs utilized in TC or QC.

Figure 2 shows examples of the spatial coverage provided by each of the four rainfall products utilized in the QC analysis. As seen in the figure, the spatial coverage of QC results is primarily constrained by availability of the E-OBS dataset. For example, the QC analysis must be masked within Poland due to the lack of E-OBS coverage there. Due to E-OBS coverage limitations, the QC analysis was carried out over a relatively limited geographic domain (30°–60°N, 10°W–60°E) covering (most of) Europe plus a small coastal strip of northern Africa. Because it does not utilize E-OBS, TC was performed within a much larger geographic domain between 40°S and 60°N and 60°W and 60°E—comprising all of Europe and Africa as well as portions of South America.

4. Results

As discussed above, our evaluation of the H23 daily precipitation product is based on the application of QC and TC analysis. Section 4a presents a QC analysis of ECC existing between the H23, SM2R, and ERA5 rainfall products. Next, QC-estimated correlation metrics (i.e., R_{QC}) for the H23, SM2R, E-OBS, and ERA5 products are presented, and the impact of gauge density on the accuracy of E-OBS-based evaluation metrics is examined. Section 4b presents TC-estimated correlation results (i.e., R_{TC}) and discusses the impact of nonzero ECC on R_{TC} using R_{QC} as a reference.

a. *Quadruple collocation estimates of R*

1) Error cross correlation

The H23, SM2R, and ERA5 products are all based, to varying degrees, on satellite observations. As discussed earlier, there is potential overlap between the source of rainfall information contained in both the H23 and ERA5 products as well as the ERA5 and SM2R products. If present, ECC between these products will lead to biased collocation-based correlation metrics. QC provides a way to estimate ECC within product quadruplets—provided it exists between only one dataset pair (see section 3b). To investigate the presence of ECC, three separate QC scenarios were performed: (i) the "H23–ERA5" case, in which nonzero ECC is assumed to exist only between H23 and ERA5; (ii) the "H23–SM2R" case, in which nonzero ECC is assumed to exist only between H23 and SM2R; and (iii) the "ERA5–SM2R" case, in which nonzero ECC is assumed to exist only between ERA5 and SM2R.

QC-based ECC derived from (6) for these three hypothetical scenarios are presented in Fig. 3. Overall, absolute values of ECC are generally low (i.e., less than 0.20) across most of the study region. However, moderate levels of positive ECC between H23 and ERA5 are found in northwestern Russian and Finland (Fig. 3a). In contrast, moderate levels of negative ECC

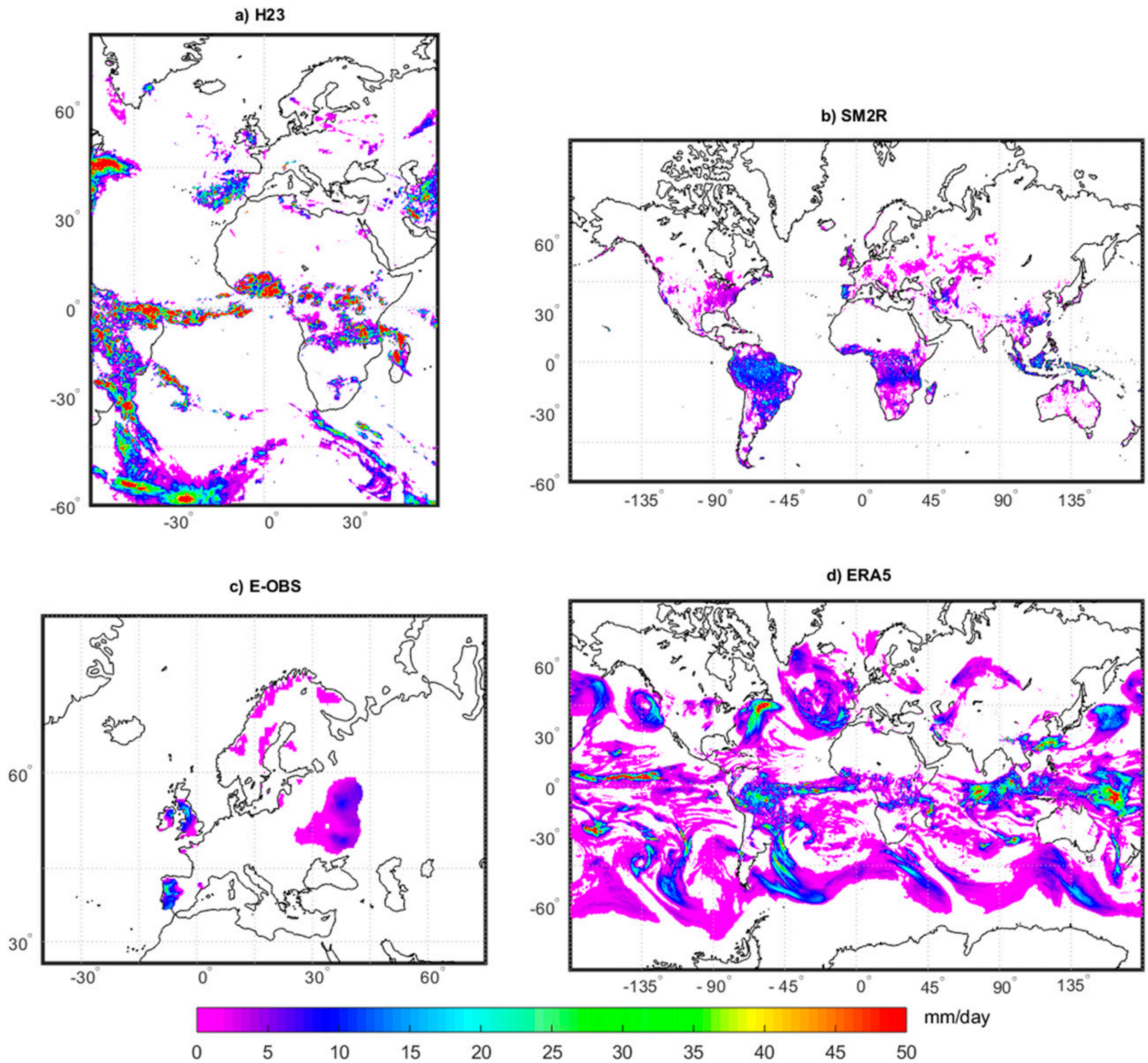

FIG. 2. Examples of 24-h (0000–2359 UTC) rainfall accumulation coverage for 31 Mar 2014 obtained by each of the four rainfall products utilized here. (a) H23 covers a geographic extent between 60°S and 75°N and between 60°W and 60°E, (c) E-OBS covers (most) European land area between 25° and 75°N and between 40°W and 75°E, (b) SM2R is functionally quasi global (with limited availability above ~60°N because of frozen soil), and (d) ERA5 is a true global product.

between H23 and ERA5 are seen around the northern coast of the Black Sea. More consistent negative ECC is found between ERA5 and SM2R across a broad swath of eastern Europe and Russia (Fig. 3c).

The weak ECC found between H23 and both SM2R (Fig. 3a) and ERA5 (Fig. 3b) is generally consistent with our earlier assumption that error in H23 is independent of both SM2R and ERA5 errors. However, the presence of negative ERA5–SM2R ECC is counterintuitive given that SM2R is scaled to match ERA5 at monthly scale (which, if anything, should induce *positive* ECC). Given our stated focus on validating H23, a complete discussion of ERA5 versus SM2R ECC is outside the scope of this analysis. Nevertheless, to mitigate the impact of nonzero ECC in the QC results, final R_{QC} values presented below are derived by averaging results from all three QC assumption scenarios. Since ECC also impacts TC analysis, it will be further examined in the context of TC results presented in section 4b.

In the above scenarios, the ECC between E-OBS and the other three rainfall products is assumed to be negligible. There are potential reasons to question this assumption. For example, both E-OBS and SM2R have known issues surrounding the detection of small rainfall events. This shared tendency could conceivably lead to nonzero ECC between the two products. To examine this issue, we performed an additional QC analysis where ECC between E-OBS and SM2R was estimated. Results

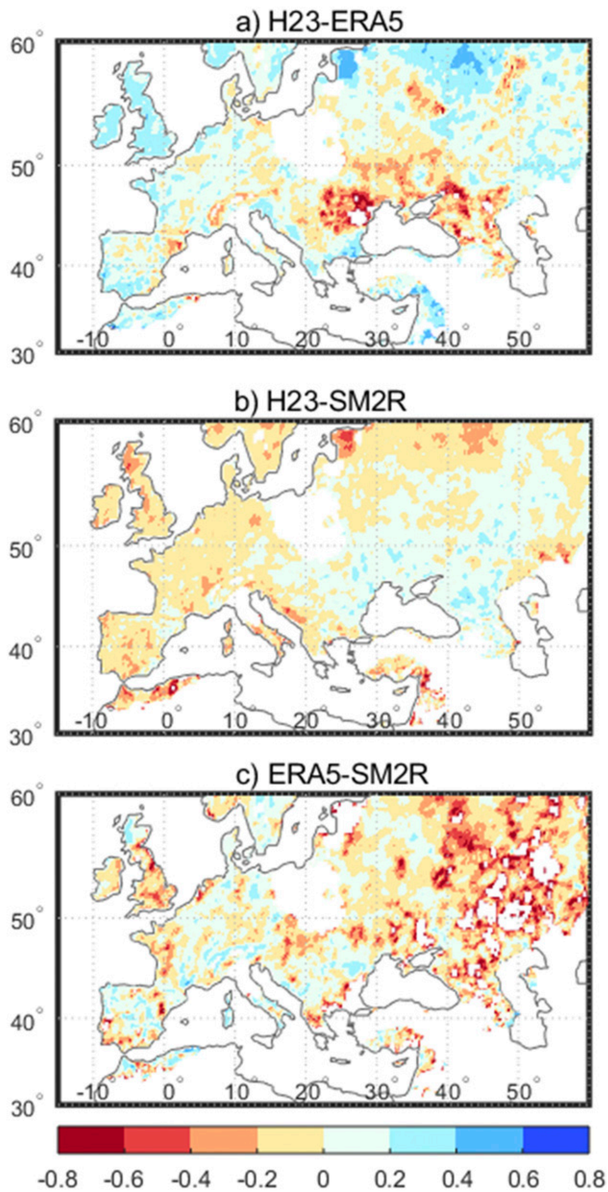

FIG. 3. The QC-derived estimates of ECC (a) between the H23 and ERA5, (b) between the H23 and SM2R, and (c) between the ERA5 and SM2R daily rainfall products.

(not shown) suggest minimal amounts of ECC between SM2R and E-OBS (i.e., 81% of pixels within ±0.20 and 98.6% within ±0.40). This suggests that ECC between E-OBS and the other rainfall products examined here is relatively small.

2) CORRELATION METRIC VIA QC

Correlation coefficients (versus unknown truth) obtained from applying quadruple collocation (R_{QC}) to the H23, E-OBS, SM2R, and ERA5 quadruplet of daily rainfall products are plotted in Fig. 4. As noted above, plotted values are obtained from averaging results from three different QC scenarios—each assuming nonzero ECC between a different pair of the H23, SM2R, and ERA5 products. Missing R_{QC} results are either due to either poor data coverage (e.g., white areas in Poland associated with missing E-OBS estimates) or nonphysical QC results (e.g., gray areas in Norway, Sweden, and Russia where SM2R struggles to accurately resolve daily rainfall). For H23, moderate to high (i.e., 0.6–0.8) R_{QC} values are generally found across the study domain. Relatively low R_{QC} values (~0.4) are found in regions characterized by complex orography (e.g., the Alps and Caucasus mountain ranges), as well as in Kazakhstan and nearby areas of Russia. The overall satisfactory performance of daily H23 rainfall estimates within a broad European domain is not unexpected, since it is derived by combining precipitation algorithms (i.e., used in H-SAF H01 and H02B products) specifically tailored for Europe and relevant areas of northern Africa. Likewise, observed difficulties in retrieving accurate precipitation estimates in complex orography are well known.

Among the four products (H23, E-OBS, SM2R, and ERA5), ERA5 demonstrates the best overall performance across the study area, with R_{QC} values consistently above 0.8. E-OBS demonstrates the highest R_{QC} scores in western Europe, consistent with its excellent rain gauge coverage there. However, the performance of E-OBS degrades considerably in areas of eastern/southeastern Europe with lower gauge densities (Fig. 1; further discussion below). Such degradation is a well-known tendency for gridded precipitation datasets based on the interpolation of rain gauge observations (e.g., Prein and Gobiet 2017; Zandler et al. 2019). Estimated R_{QC} values over Italy for SM2R (~0.7) and ERA5 (~0.8) are very close to the values reported in Brocca et al. (2019) where Pearson's correlation coefficient was calculated using reference data from high-quality regional gauge networks.

Between the two satellite-based datasets (SM2R and H23), the former has slightly better skill in southern Spain, northern Africa, the southern United Kingdom, and the Middle East. In contrast, H23 is superior in France, Germany, and the Balkans and outperforms SM2R over mountainous regions of southwestern Europe. Poorer SM2R performance over continental Europe (including Russia) can likely be traced back to the relatively low skill of ASCAT soil moisture retrievals there relative to other satellite soil moisture products (Chen et al. 2018). Notably low correlation between SM2R and other rainfall datasets was found at high latitudes (north of 50°N), which results in large areas of northeast Europe being masked in Fig. 4c. This suggests that relatively large ASCAT retrieval error over continental Europe propagates into SM2R rainfall estimates and contributes to overall lower SM2R skill.

As a product generated from quality-controlled, ground-based observations, E-OBS is a natural choice for a benchmark product to validate SPEs. However, relatively poor E-OBS performance in some areas (see Fig. 4b) indicates that E-OBS is ill suited to serve as a universal reference. For example, Fig. 5a maps the direct Pearson's correlation (R) between H23 and E-OBS (obtained without QC or TC), while Fig. 5b provides the difference between the direct R values and the H23 R_{QC} results shown in Fig. 4a as a function of E-OBS rain gauge station density. Direct H23 versus E-OBS R is lower than R_{QC} across Europe with the most severe discrepancy found in eastern Europe where E-OBS is based on very sparse ground observations (i.e., less than 1 station per 0.25° grid cell in Fig. 1).

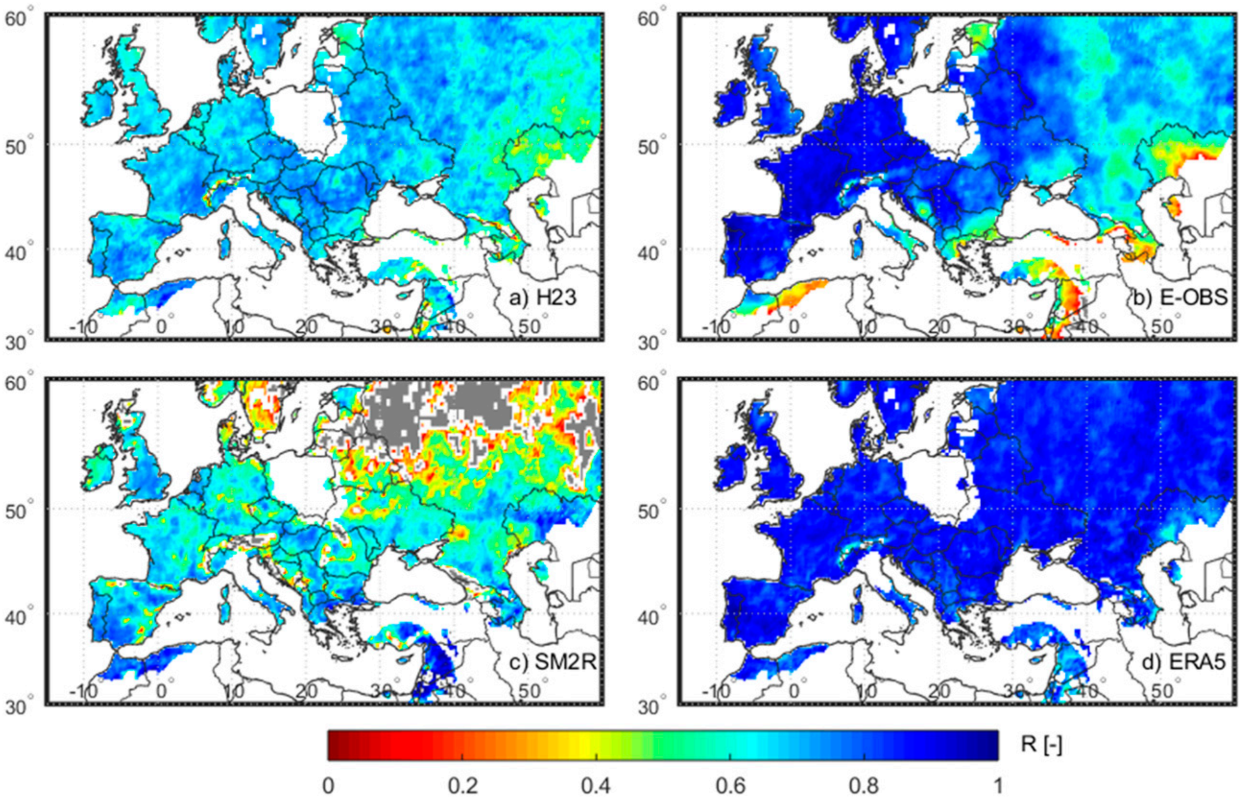

FIG. 4. QC-based estimates of correlation coefficient metrics (R_Q vs unknown truth) for the (a) H23, (b) E-OBS, (c) SM2R, and (d) ERA5 daily rainfall products. Only pixels with collocated data from all four products and passing the significant correlation test described in section 3a are shown. Gray-shaded grid cells indicate a failed QC analysis (negative or nonphysical results), whereas terrestrial grid cells in white suggest data that are missing or were masked during a prescreening process (see section 3d).

As gauge density increases from zero to one or more station(s) per grid cell, the difference between TC- and QC-based correlation estimates drops considerably and stays at a relatively constant value of around 0.1. Thus, Fig. 5 demonstrates that, due to increasing errors in E-OBS associated with relatively fewer contributing rain gauges, the validation of SPE products against E-OBS (or other ground-based dataset) will spuriously degrade correlation-based evaluation metrics—particularly in areas where observations underlying the product are very sparse.

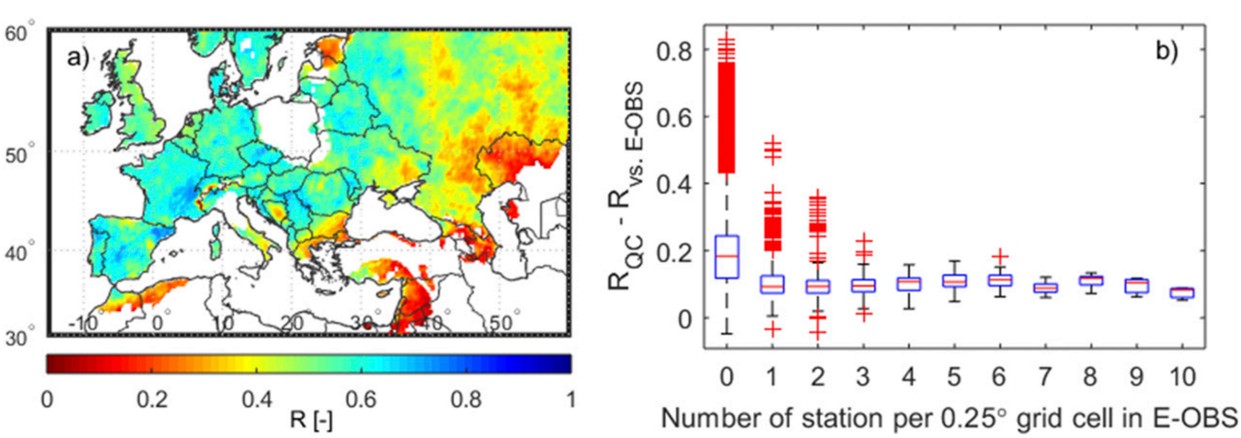

FIG. 5. (a) Map of the (directly sampled) temporal correlation between H23 and E-OBS. (b) The difference between R_{QC} and the direct correlations plotted in (a) as a function of E-OBS rain gauge station density.

b. Triple collocation estimates of correlation metrics

1) Comparison of R_{QC} and R_{TC}

Above, QC is demonstrated to provide a more robust evaluation of the H23 product over sparsely gauged areas. However, its application is still limited to areas containing a minimum level of rain gauge coverage. In the case where ground-based observations (such as E-OBS) are unavailable, TC can be applied instead. Therefore, by freeing us from the constraint of ground observations, TC allows for rainfall evaluation over a wider geographic area. Here, utilizing a H23–SM2R–ERA5 daily rainfall product triplet, R_{TC} is obtained for a larger spatial domain (relative to earlier R_{QC} results in Fig. 4) covering the entire African continent, the Arabian Peninsula, and the eastern portion of South America (Figs. 6a,c,e). Note that R_{TC} results are now also available over Poland since E-OBS availability is no longer required. Because European R_{QC} results are obtained with ground-based observations, they provide a useful point of comparison for R_{TC} estimates. To this end, Figs. 6b, 6d, and 6f compare the R_{TC} and R_{QC} results for each product over areas of Europe and North Africa where E-OBS data are available (see domain in Fig. 4).

Despite their lack of reliance on ground-based observations, H23 R_{TC} estimates in Europe and North Africa correspond closely to earlier R_{QC} results (Fig. 6a). This correspondence supports the application of non-ground-based rainfall products such as SM2R and ERA5 to obtain robust R estimates for SPE products in the absence of ground-based observations or in areas where such datasets are of low-quality due to poor spatial coverage. H23 R_{TC} results also appear to be relatively unbiased versus R_{QC} (Fig. 6b). SM2R R_{TC} estimates agree well, both in terms of spatial pattern and magnitude, with a previous TC analysis using a high-quality ground-based dataset (i.e., GPCC, see Brocca et al. 2019). However, when compared to SM2R R_{QC} results derived above, a pronounced high bias is seen in SM2R R_{TC} results at lower values of R_{QC} (<0.6). This bias suggests the violation of assumptions underlying the application of TC to SM2R and is likely linked to the observed tendency for nonzero SM2R–ERA5 ECC (Fig. 3c).

The underperformance of SM2R (as indicated by lower R_{QC} values) can be linked to the impact of vegetation density on ASCAT soil moisture retrieval accuracy. For example, we compared a 0.25° grid monthly-mean leaf area index (LAI) map for average July conditions, derived from 1981 to 2015 observations acquired from the Advanced Very High Resolution Radiometer (AVHRR; Mao and Yan, 2019), to R_{QC} results plotted in Fig. 4. A strong negative correlation ($R = -0.50$) is found between SM2R R_{QC} and climatological July LAI values and most (74%) of the low skill pixels (i.e., SM2R $R_{QC} < 0.6$) are associated with mean July LAI values above 2—a threshold commonly met in northeastern Europe where forest coverage fraction is high. This suggests that the low quality of ASCAT soil moisture retrievals in densely vegetated areas significantly reduces the quality of SM2R rainfall estimates.

However, vegetation density does not appear to affect the validity of TC results for products other than SM2R. That is, neither H23 nor ERA5's ΔR (i.e., $R_{TC} - R_{QC}$) is significantly correlated with mean July LAI. Therefore, despite the lower-quality SM2R in densely vegetated areas, the R_{TC} metric for other rainfall products (notably H23) remain largely unbiased, supporting the usefulness of SM2R in TC analysis across a wide range of surface vegetation conditions. However, caution should be used when directly comparing SM2R with other rainfall datasets without using a collocation-based analysis. In such cases, stringent masking of SM2R is advised to remove pixels where ASCAT soil moisture retrievals are less reliable due to the presence of dense vegetation, highly arid surface conditions, complex topography, and/or waterbodies.

2) Intercomparison of H23, SM2R, and ERA5 via R_{TC}

As expected, R_{TC} maps in Fig. 6 show patterns in Europe and northern Africa that echo earlier R_{QC} results in Fig. 4. However, interesting tendencies can be noted in areas outside of the E-OBS spatial domain. For example, SM2R is significantly superior to H23 over nearly all of Africa. The relatively poor performance of the H23 product in the Sahel region of Africa (i.e., the transition zone between the Sahara Desert and tropical rain forest) is not surprising given that highly variable surface emissivity conditions in the Sahel pose a unique challenge for PMW-based rainfall inversion (Alcoba and Gosset 2015). Here, the H-SAF H01 and H02B PMW rainfall products (from which H23 is derived) have been found to overestimate the frequency of rain day (i.e., false detection). This tendency is believed to be associated with the inaccurate specification of surface type and emissivity in PMW retrieval algorithms (Alcoba and Gosset 2015). On the other hand, the Sahel region is generally considered to be well suited for soil moisture remote sensing due to its relatively low levels of vegetation biomass (Chen et al. 2018).

South of the Sahel region of Africa, SM2R continues to show the highest overall skill, followed by ERA5 and H23. In particular, the performance of H23 and ERA5 is significantly worse in the equatorial zone between 15°S and 15°N where the seasonal movement of the intertropical convergence zone leads to strong rainfall seasonality. In this region, satellite soil moisture retrievals (utilized in SM2R) appear to provide a better proxy to infer rainfall amounts than PMW-based estimates. However, this conclusion should be treated with caution given the possibility that SM2R R_{TC} values are positively biased due to high vegetation density [see Fig. 6f and discussion in section 4b(1)]. Note that extremely low SM2R–ERA5 correlations were found in the same area by Brocca et al. (2019).

In South America, H23 achieves higher R_{TC} values (~0.8 and above) over Atlantic coastal regions and moderate (~0.5–0.6) R_{TC} values inland. SM2R, on the other hand, shows excellent skill (~0.8) in the northern part of the Brazilian Highlands—a drier region east of the Amazon basin. Brocca et al. (2019) obtained similar R_{TC} values for SM2R in this area. Again, note that SM2R R_{TC} values are likely incorrectly inflated in regions covered by tropical rain forests. Finally, relatively high H23 R_{TC} values are found in areas of the Middle East (e.g., Saudi Arabia) that fall within the H23 domain (but are not shown in the overlapping E-OBS–H23 domain in Fig. 4). Finally, note that missing R_{TC} values in Fig. 6 are due to a combination of poor data coverage (e.g., SM2R in the Sahara Desert), lack of temporal correlation between SM2R and the other two products (i.e., H23 and ERA5) or a failed TC analysis (e.g., equatorial Congo, indicated via gray shading).

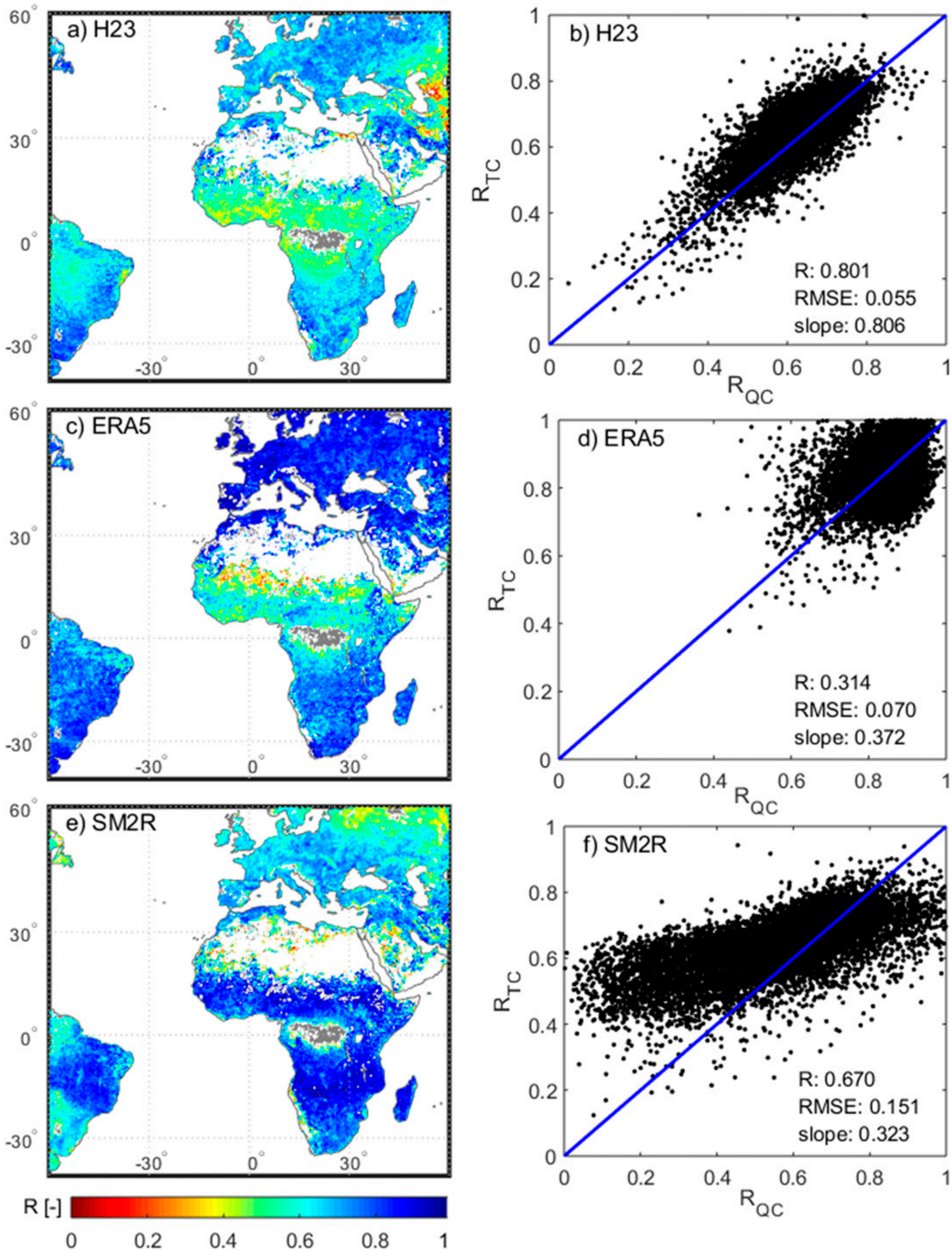

FIG. 6. Estimates of correlation coefficient obtained from a TC analysis (R_{TC}) for (a) H23, (c) ERA5, and (e) SM2R, along with (b),(d),(f) respective scatterplots of R_{TC} vs R_{QC} for each product for areas of Europe and North Africa where E-OBS estimates are available (see the domain in Fig. 4). Gray-shade grid cells indicate a failed TC analysis (negative or nonphysical results), whereas terrestrial grid cells in white indicate data that are missing or were masked during a prescreening process (see section 3d).

Figure 7 presents an R_{TC}-based pairwise comparison between H23, SM2R, and ERA5 that indicates areas where one product is significantly superior (at 95% confidence) to the other. Terrestrial pixels shaded white indicate either: missing results due to data availability, masking during preprocessing or cases where neither product is significantly better than the other. White pixels in Europe (Fig. 7a) indicates that H23 and SM2R have comparable temporal correlation (versus unknown truth)

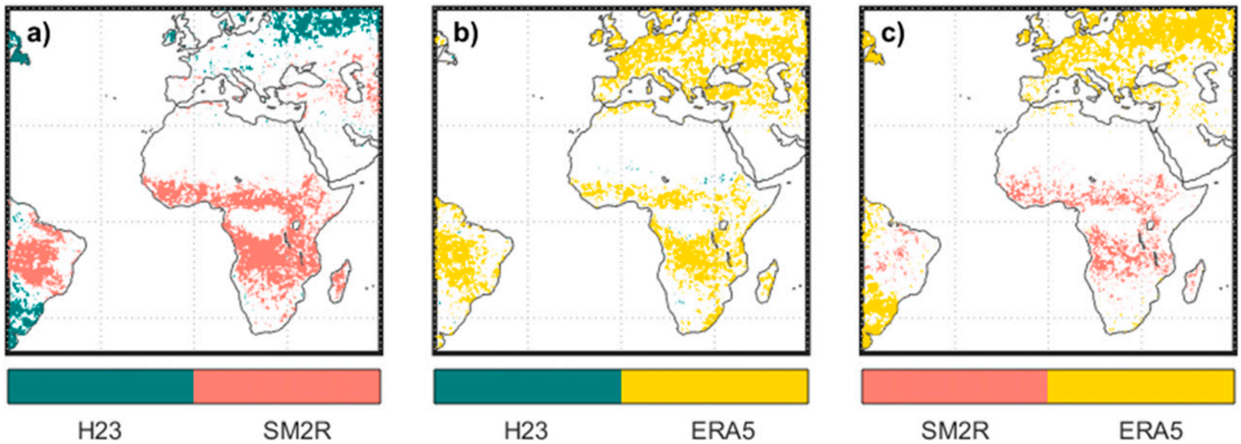

FIG. 7. Pairwise comparison of R_{TC} between the (a) H23 and SM2R, (b) H23 and ERA5, and (c) SM2R and ERA5 products. Grid cells in color indicate corresponding product having higher R_{TC} while the difference between the two products is statistically significant with 95% confidence (on the basis of a 100-member bootstrapped distribution). Terrestrial pixels in white reflect areas with data that are missing or were masked during preprocessing or areas where observed R_{TC} differences are nonsignificant.

there. Likewise, large masked areas of southern Africa (south of ~25°S) in Figs. 7a–c reflect that R differences between these products are largely nonsignificant. In other areas, the significant advantages of each product are relatively well defined geographically: H23 outperforms SM2R in areas of northern Europe, southern Brazil, and Uruguay. Likewise, SM2R outperforms both H23 and ERA5 in the Sahel, southern Africa, and eastern Brazil and ERA5 is consistently superior to H23. However, as discussed above, the superior performance of SM2R outside of Europe should be viewed with caution given the potential for SM2R R_{TC} values to be biased high (Fig. 6f).

5. Summary

This study presents a validation strategy for satellite precipitation products based on the application of triple (TC) and quadruple (QC) collocation techniques and the newly available SM2RAIN-ASCAT product (i.e., SM2R). The proposed method is applied to evaluate the H-SAF H23 daily precipitation product relative to three other large-scale precipitation products. Specifically, estimates of correlation versus unknown truth are generated via both a QC analysis (based on an H23–E-OBS–SM2R–ER5 quadruplet) and a TC analysis (based on a H23–SM2R–ERA5 triplet).

On the basis of collocation results, H23 correlation versus true daily rainfall accumulations is moderate to high (~0.6–0.8) across most of western Europe, South America, and Africa and relatively poor (~0.2–0.4) in areas of equatorial Africa and continental Europe. The performance of SM2R is comparable to H23 in the Mediterranean region but poorer in Scandinavia and over continental Europe. SM2R outperforms ERA5 and H23 in Africa south of the Sahel and in northeastern Brazil—with the caveat that SM2R R_{TC} is potentially overestimated in areas of (genuinely) low skill. Across all products, ERA5 demonstrates the best overall performance and exhibits poor R results (<0.4) only in equatorial areas of Africa.

The evaluation of collocation results indicates that SM2R is a useful independent rainfall dataset that enables the application of QC and TC approaches for the evaluation of large-scale rainfall products. Despite instances of poor SM2R performance and error correlation with H23 or ERA5, no noticeable bias is found in H23's R_{TC} results (generated using SM2R) relative to R_{QC} (generated using ground-based rainfall observations). This suggests that SM2R, together with collocation-based analyses, can be applied to reliably evaluate H23, even in areas where ground observations are scarce. Our analysis also suggests that, in areas where SM2R accuracy is negatively affected by suboptimal surface conditions for ASCAT soil moisture retrieval, QC/TC results are still generally robust for other participating datasets in the collocation analyses.

Although the ability of SM2R to validate SPEs via the application of QC and TC is encouraging, several caveats should be noted. First, TC analysis provides only a correlation evaluation metric and does not provide absolute estimates of other commonly applied rainfall metrics like additive and multiplicative bias, as well as categorical statistics such as false alarm ratio and the probability of detection. The recent development and application of Categorical Triple Collocation approaches (McColl et al. 2016; Dong et al. 2020) suggests the potential application of TC to obtain categorical metrics describing the rainfall detection performance of SPEs; however, these approaches have yet to be widely applied.

Our collocation analyses are sensitive to the presence of ECC between products (see section 3d)—particularly since we have not removed seasonal signals from each product prior to collocation analysis. The presence and impact of potential ECC between H23, SM2R, and ERA5 is discussed in section 4. In addition, SM2R's tendency to underestimate extreme rainfall events can potentially introduce error correlation with respect to other products suffering from a similar systematic bias. It should be stressed that our QC analysis largely failed to find evidence of such error cross correlation—at least within the European domain over which it was applied. Nevertheless, careful assessment of cross-correlated error impacts is necessary before collocation results can be used with confidence. In particular, while SM2R enables a relatively accurate representation of ERA5 and SM2R skill in data-poor regions (via collocation analysis), the R_{TC} of SM2R itself may be biased. Eventually, the reliability of QC will be maximized by its application to the highest quality precipitation product. Therefore, the use of new gauge analysis datasets, such as recently released Global Precipitation Climatology Centre (GPCC) daily products (Ziese et al. 2018) is recommended for future study.

At present, the reliable, comprehensive validation of SPE products remains focused on the use of interpolated rain gauge data (e.g., Puca et al. 2014). However, this approach requires sufficient rain gauge density to accurately reproduce spatial patterns at the satellite's resolution, which can only be satisfied over a small fraction of Earth's land surface. It is estimated that between 60°S and 60°N, where a majority of the world's rain gauges operate, only 6.5% of land area is located within 10 km of a gauge (Kidd et al. 2017). Therefore, collocation-based SPE validation approaches are a valuable tool for vast land regions that are ungauged or data-scarce. Given the difficulty of obtaining wholly independent estimates of rainfall at the global scale, innovative rainfall products like SM2R provide a unique resource for efforts to improve the large-scale validation of SPEs.

Acknowledgments. This study is supported by the EUMESAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) project (H_AVS18_04). We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). The authors also gratefully acknowledge the Italian Civil Protection Department and thank Drs. Luca Brocca, Daniel Casella, and Paolo Sanò for their extremely valuable suggestions.

Data availability statement. H23 is requested from http://hsaf.meteoam.it/. ERA5 is downloaded from http://doi.org/10.24381/cds.e2161bac. E-OBS is downloaded from: http://www.ecad.eu. SM2RAIN-ASCAT is downloaded from: http://doi.org/10.5281/zenodo.2591215.

REFERENCES

Alcoba, M., and M. Gosset, 2015: Verification study over West Africa of PMW precipitation products using X-pol radar observations and rain gauges. H-SAF Final Rep. HSAF_AS15_03, 45 pp., https://hsaf.meteoam.it/documents/visiting-scientist/HSAF_AS15_03_Final_Report.pdf.

Alemohammad, S. H., K. A. McColl, A. G. Konings, D. Entekhabi, and A. Stoffelen, 2015: Characterization of precipitation product errors across the United States using multiplicative triple collocation. *Hydrol. Earth Syst. Sci.*, **19**, 3489–3503, https://doi.org/10.5194/hess-19-3489-2015.

Brocca, L., F. Melone, T. Moramarco, and W. Wagner, 2013: A new method for rainfall estimation through soil moisture observations. *Geophys. Res. Lett.*, **40**, 853–858, https://doi.org/10.1002/grl.50173.

——, and Coauthors, 2014: Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *J. Geophys. Res. Atmos.*, **119**, 5128–5141, https://doi.org/10.1002/2014JD021489.

——, and Coauthors, 2015: Rainfall estimation from in situ soil moisture observations at several sites in Europe: An evaluation of SM2RAIN algorithm, *J. Hydrol. Hydromech.*, **63**, 201–209, https://doi.org/10.1515/johh-2015-0016.

——, and Coauthors, 2019: SM2RAIN–ASCAT (2007–2018): Global daily satellite rainfall data from ASCAT soil moisture observations. *Earth Syst. Sci. Data*, **11**, 1583–1601, https://doi.org/10.5194/essd-11-1583-2019.

C3S, 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), accessed 7 February 2019, https://doi.org/10.24381/cds.e2161bac.

Caires, S., and A. Sterl, 2003: Validation of ocean wind and wave data using triple collocation. *J. Geophys. Res.*, **108**, 3098, https://doi.org/10.1029/2002JC001491.

Casella, D., and Coauthors, 2013: Transitioning from CRD to CDRD in Bayesian retrieval of rainfall from satellite passive microwave measurements: Part 2. Overcoming database profile selection ambiguity by consideration of meteorological control on microphysics. *IEEE Trans. Geosci. Remote Sens.*, **51**, 4650–4671, https://doi.org/10.1109/TGRS.2013.2258161.

Chakraborty, A., R. Kumar, and A. Stoffelen, 2013: Validation of ocean surface winds from the OCEANSAT-2 scatterometer using triple collocation. *Remote Sens. Lett.*, **4**, 84–93, https://doi.org/10.1080/2150704X.2012.693967.

Chen, F., and Coauthors, 2017: Application of triple collocation in ground-based validation of Soil Moisture Active/Passive (SMAP) Level 2 data products. *IEEE J. Sel. Topics Appl. Earth Obs.*, **10**, 489–502, https://doi.org/10.1109/JSTARS.2016.2569998.

——, W. T. Crow, R. Bindlish, A. Colliander, M. Burgin, J. Asanuma, and K. Aida, 2018: Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote Sens. Environ.*, **214C**, 1–13, https://doi.org/10.1016/j.rse.2018.05.008.

Ciabatta, L., A. C. Marra, G. Panegrossi, D. Casella, P. Sanò, S. Dietrich, C. Massari, and L. Brocca, 2017: Daily precipitation estimation through different microwave sensors: Verification study over Italy. *J. Hydrol.*, **545**, 436–450, https://doi.org/10.1016/j.jhydrol.2016.12.057.

Dong, J., W. T. Crow, and R. H. Reichle, 2020: Improving rain/no-rain detection skill by merging precipitation estimates from different sources. *J. Hydrometeor.*, **21**, 2419–2429, https://doi.org/10.1175/JHM-D-20-0097.1.

Draper, C., R. Reichle, R. de Jeu, V. Naeimi, R. Parinuss, and W. Wagner, 2013: Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sens. Environ.*, **137**, 288–298, https://doi.org/10.1016/j.rse.2013.06.013.

Ebert, E. E., J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, **88**, 47–64, https://doi.org/10.1175/BAMS-88-1-47.

Fang, H., S. Wei, C. Jiang, and K. Scipal, 2012: Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method. *Remote Sens. Environ.*, **124**, 610–621, https://doi.org/10.1016/j.rse.2012.06.013.

Gruber, A., C.-H. Su, W. T. Crow, S. Zwieback, W. A. Dorigo, and W. Wagner, 2016a: Estimating error cross-correlation in soil moisture data sets using extended collocation analysis. *J. Geophys. Res. Atmos.*, **121**, 1208–1219, https://doi.org/10.1002/2015JD024027.

——, ——, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner, 2016b: Recent advances in (soil moisture) triple collocation analysis. *Int. J. Appl. Earth Obs. Geoinf.*, **45**, 200–211, https://doi.org/10.1016/j.jag.2015.09.002.

Hersbach, H., and Coauthors, 2018: Operational global reanalysis: Progress, future directions and synergies with NWP. ECMWF ERA Rep. Series 27, 63 pp., https://doi.org/10.21957/tkic6g3wm.

Hossain, F., and E. N. Anagnostou, 2006: A two-dimensional satellite rainfall error model. *IEEE Trans. Geosci. Remote Sens.*, **44**, 1511–1522, https://doi.org/10.1109/TGRS.2005.863866.

Kidd, C., P. Bauer, J. Turk, G. Huffman, R. Joyce, K.-L. Hsu, and D. Braithwaite, 2012: Intercomparison of high-resolution precipitation products over Northwest Europe. *J. Hydrometeor.*, **13**, 67–83, https://doi.org/10.1175/JHM-D-11-042.1.

——, A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, 2017: So, how much of the Earth's surface is covered by rain gauges? *Bull. Amer. Meteor. Soc.*, **98**, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1.

Li, C., G. Tang, and Y. Hong, 2018: Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the triple collocation method across mainland China. *J. Hydrol.*, **562**, 71–83, https://doi.org/10.1016/j.jhydrol.2018.04.039.

Mahfouf, J.-F., B. Brasnett, and S. Gagnon, 2007: A Canadian Precipitation Analysis (CaPA) Project: Description and preliminary results. *Atmos.–Ocean*, **45**, 1–17, https://doi.org/10.3137/ao.v450101.

Mao, J., and B. Yan, 2019: Global monthly mean leaf area index climatology, 1981-2015. ORNL DAAC, accessed 15 September 2020, https://doi.org/10.3334/ORNLDAAC/1653.

Massari, C., W. Crow, and L. Brocca, 2017: An assessment of the performance of global rainfall estimates without ground-based observations. *Hydrol. Earth Syst. Sci.*, **21**, 4347–4361, https://doi.org/10.5194/hess-21-4347-2017.

——, and Coauthors, 2020: A daily 25 km short-latency rainfall product for data-scarce regions based on the integration of the Global Precipitation Measurement mission rainfall and multiple-satellite soil moisture products. *Hydrol. Earth Syst. Sci.*, **24**, 2687–2710, https://doi.org/10.5194/hess-24-2687-2020.

McColl, K. A., J. Vogelzang, A. G. Konings, D. Entekhabi, M. Piles, and A. Stoffelen, 2014: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophys. Res. Lett.*, **41**, 6229–6236, https://doi.org/10.1002/2014GL061322.

——, A. Roy, C. Derksen, A. G. Konings, S. H. Alemohammed, and D. Entekhabi, 2016: Triple collocation for binary and categorical variables: Application to validating landscape freeze/thaw retrievals. *Remote Sens. Environ.*, **176**, 31–42, https://doi.org/10.1016/j.rse.2016.01.010.

Mugnai, A., and Coauthors, 2013a: Precipitation products from the hydrology SAF. *Nat. Hazards Earth Syst. Sci.*, **13**, 1959–1981, https://doi.org/10.5194/nhess-13-1959-2013.

——, and Coauthors, 2013b: CDRD and PNPR satellite passive microwave precipitation retrieval algorithms: EuroTRMM/EURAINSAT origins and H-SAF operations. *Nat. Hazards Earth Syst. Sci.*, **13**, 887–912, https://doi.org/10.5194/nhess-13-887-2013.

Pierdicca, N., F. Fascetti, L. Pulvirenti, R. Crapolicchio, and J. Muñoz-Sabater, 2015: Quadruple collocation analysis for soil moisture product assessment. *IEEE Geosci. Remote Sens. Lett.*, **12**, 1595–1599, https://doi.org/10.1109/LGRS.2015.2414654.

Prein, A. F., and A. Gobiet, 2017: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *Int. J. Climatol.*, **37**, 305–327, https://doi.org/10.1002/joc.4706.

Puca, S., and Coauthors, 2014: The validation service of the hydrological SAF geostationary and polar satellite precipitation products. *Nat. Hazards Earth Syst. Sci.*, **14**, 871–889, https://doi.org/10.5194/nhess-14-871-2014.

Roebeling, R. A., E. L. Wolters, J. F. Meirink, and H. Leijnse, 2012: Triple collocation of summer precipitation retrievals from SEVIRI over Europe with gridded rain gauge and weather radar data. *J. Hydrometeor.*, **13**, 1552–1566, https://doi.org/10.1175/JHM-D-11-089.1.

Sanò, P., D. Casella, A. Mugnai, G. Schiavon, E. A. Smith, and G. J. Tripoli, 2013: Transitioning from CRD to CDRD in Bayesian retrieval of rainfall from satellite passive microwave measurements: Part 1. Algorithm description and testing. *IEEE Trans. Geosci. Remote Sens.*, **51**, 4119–4143, https://doi.org/10.1109/TGRS.2012.2227332.

——, G. Panegrossi, D. Casella, F. Di Paola, L. Milani, A. Mugnai, M. Petracca, and S. Dietrich, 2015: The Passive microwave Neural network Precipitation Retrieval (PNPR) algorithm for AMSU/MHS observations: description and application to European case studies. *Atmos. Meas. Tech.*, **8**, 837–857, https://doi.org/10.5194/amt-8-837-2015.

Sapiano, M. R. P., and P. A. Arkin, 2009: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *J. Hydrometeor.*, **10**, 149–166, https://doi.org/10.1175/2008JHM1052.1.

Scipal, K., W. Dorigo, and R. de Jeu, 2010: Triple collocation – A new tool to determine the error structure of global soil moisture products. *2010 IEEE Int. Geoscience and Remote Sensing Symp.*, Honolulu, HI, IEEE, 4426–4429, https://doi.org/10.1109/IGARSS.2010.5652128.

Stampoulis, D., and E. N. Anagnostou, 2012: Evaluation of global satellite rainfall products over continental Europe. *J. Hydrometeor.*, **13**, 588–603, https://doi.org/10.1175/JHM-D-11-086.1.

Su, C.-H., and D. Ryu, 2015: Multi-scale analysis of bias correction of soil moisture. *Hydrol. Earth Syst. Sci.*, **19**, 17–31, https://doi.org/10.5194/hess-19-17-2015.

Tian, Y., G. J. Huffman, R. F. Adler, L. Tang, M. Sapiano, V. Maggioni, and H. Wu, 2013: Modeling errors in daily precipitation measurements: Additive or multiplicative? *Geophys. Res. Lett.*, **40**, 2060–2065, https://doi.org/10.1002/grl.50320.

van Engelen, A., A. Klein Tank, G. van de Schrier, and L. Klok, 2008: Towards an operational system for assessing observed changes in climate extremes. Royal Netherlands Meteorological Institute European Climate Assessment & Dataset Rep., 68 pp., https://knmi-ecad-assets-prd.s3.amazonaws.com/documents/ECAD_report_2008.pdf.

Villarini, G., P. V. Mandapaka, W. F. Krajewski, and R. J. Moore, 2008: Rainfall and sampling uncertainties: A rain gauge perspective. *J. Geophys. Res.*, **113**, D11102, https://doi.org/10.1029/2007JD009214.

Wang, H., J. Zhu, and J. Yang, 2014: Error analysis on ESA's Envisat ASAR wave mode significant wave height retrievals using triple collocation model. *Remote Sens.*, **6**, 12 217–12 233, https://doi.org/10.3390/rs61212217.

Wang, Q., Y. Zeng, C. M. Mannaerts, and V. R. Golroudbary, 2018: Determining relative errors of satellite precipitation data over the Netherlands. *Proc. Second Int. Electronic Conference on Remote Sensing (ECRS 2018)*, Online, MDPI, 7 pp., https://doi.org/10.3390/ecrs-2-05139.

Zandler, H., I. Haag, and C. Samimi, 2019: Evaluation needs and temporal performance differences of gridded precipitation products in peripheral mountain regions. *Sci. Rep.*, **9**, 15118, https://doi.org/10.1038/s41598-019-51666-z.

Ziese, M., A. Rauthe-Schöch, A. Becker, P. Finger, A. Meyer-Christoffer, and U. Schneider, 2018: GPCC full data daily version.2018 at 1.0°: Daily land-surface precipitation from rain-gauges built on gts-based and historic data. DWD, accessed 1 September 2020, https://doi.org/10.5676/DWD_GPCC/FD_D_V2018_100.

Zwieback, S., K. Scipal, W. Dorigo, and W. Wagner, 2012: Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Processes Geophys.*, **19**, 69–80, https://doi.org/10.5194/npg-19-69-2012.