H SAF Associated Scientist activity

H_AS18_04

TITLE

Leveraging coincident soil moisture and precipitation products for improved global validation of satellite-based rainfall products

FINAL REPORT

18 October 2019

Fan Chen[1,2] and Wade T. Crow[2]

[1] Science Systems and Applications Inc., Greenbelt, Maryland, USA

[2] USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland,

Table of Contents

List of Figures

p. 28 - Figure 14. Same as Figure 13 but for H05B.

Abstract

Satellite-based rainfall products are generally validated against ground-based observations from gauge or radar networks. However, uncertainty in such reference datasets, especially in regions where instruments are spatially sparse, can leads to biased assessment of the target product, and, at regional or continental scales, an objective basis to evaluate satellite-based rainfall products is currently unavailable. Here, we seek to enhance the continental-scale validation practice via utilizing of both direct and indirect application of collocation-based techniques (i.e. triple collocation (TC) and quadruple collocation (QC) analyses) while leveraging a newly developed precipitation product derived from remote sensing soil moisture retrievals (SM2R) in addition to traditional validation datasets such as ground observations and climate reanalysis. The proposed procedure is applied to validate four years (2011-2014) of the H23 and 2.5 years (2015-2017) of the H05B daily rainfall products developed by the EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF). The H23 correlation metric is first estimated from TC and QC separately, and a QC analysis of the error cross correlation among the input datasets confirms that the error-independence assumption underlying TC is generally satisfied. Then, a high-quality benchmark rainfall product is generated by merging the SM2R, reanalysis and ground-based datasets using optimal scaling parameters derived via TC. The merged rainfall product is then used to perform a more complete and robust assessment of the H23 and H05B products by providing continuous and categorical performance metrics, as well as seasonal error behaviors. The SM2R product is demonstrated to be a useful, independent dataset to apply collocation analyses to the evaluation of rainfall products, given that many of the large-scale products share considerable overlapping data sources and algorithms. With the availability of a large-scale rainfall product that optimally combines the strengths of different existing rainfall datasets, validation of satellite rainfall products is not only more robust but also possible in areas lacking of ground observations.

1. Introduction

Satellite-based precipitation estimates (SPE) are increasingly being applied to critical applications including numerical weather predictions, hydrological modeling and agricultural drought monitoring. For instance, the H23 gridded daily precipitation product generated by the EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites)'s Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) has a one-day data latency and is therefore

applicable to near-real-time monitoring and forecasting activities in Europe, Africa and parts of South America. An accurate statistical assessment of errors in SPE products is required to ensure that certain pre-specified user accuracy requirements are met and to provide uncertainty estimates for subsequent hydrologic model predictions (e.g., stream flow and/or inundation extent) derived from these products.

Over a regional or continental scale, SPE products are generally validated by comparisons against a single reference dataset (e.g., a modelling product or spatially interpolated ground observations acquired from dense rain gauge networks and ground-based weather radars) (e.g. Ebert et al., 2007; Sapiano and Arkin, 2009; Stampoulis and Anagnostou, 2011). However, such datasets are subject to their own uncertainties – originating from various sources such as instrument failures, measuring mechanisms, sampling density, interpolation etc. (Villarini et al., 2008; Stampoulis and Anagnostou, 2011; Kidd et al., 2017; Prein and Gobiet, 2017) – that can result in biased evaluation metrics for the satellite-based products.

Recently, collocation-based mathematical solutions have been increasingly applied to calculate unbiased estimates of product error variance and correlation coefficient in the absence of a high-quality benchmark dataset. Depending on the number of independent datasets available, a triple collocation (TC, requiring three datasets) or quadruple collocation (QC, requiring four datasets) analysis can be performed. Various applications of triple collocation (TC) for rainfall product evaluations (Roebeling et al., 2012; Alemohammad et al., 2015; Massari et al. 2017; Li et al. 2018) have demonstrated its value for evaluating satellite-based rainfall products – particularly for global- and regional-scale studies that include data-scarce areas. TC analysis requires the availability of three datasets with mutually independent errors. To construct such a triplet, satellite and ground-based datasets can be combined with modeled precipitation fields from reanalysis products. Because the reanalysis datasets are generated retrospectively by ingesting a wealth of satellite, atmospheric and ground observations, their quality is relatively high. However, the mixture of data used to generate these products can potentially undermine their independence with respect to other members of the triplets. Likewise, while multiple satellite-based rainfall products 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 candidate product for the application of rainfall TC can be generated using remotely sensed soil moisture via the SM2RAIN method (Brocca et al. 2014). The SM2RAIN algorithm inverts rainfall accumulation from satellite soil moisture observations from the difference between two successive measurements. As a valuable and independent source of information, the rainfall estimates provided as SM2RAIN-ASCAT (Brocca et al., 2019), based on the application of SM2RAIN to ASCAT soil

moisture product are here used. The SM2RAIN-ASCAT product, hereinafter called "SM2R", are therefore largely independent of the other types of rainfall estimates and thus a suitable TC input. In addition, so-called quadruple collocation (QC) approaches can be applied to back out error correlation information in cases where the error-dependency between two data products is uncertain (Gruber et al., 2016).

TC and QC approaches cannot be directly applied to obtain categorical metrics (e.g., probability of detection or false alarm ratio) commonly used in a rainfall product assessment. However, one of the most important applications of error variance information gleaned from TC and QC is the estimation of optimal weighting factors when merging various product via mutual averaging (Yilmaz et al., 2012; Crow et al., 2015; Gruber et al., 2017). Thus, a high-quality reference dataset generated by optimally merging multiple rainfall products can be obtained and used to provide a full range of validation metrics and comparisons.

In this report, TC and QC analyses are applied to utilize several rainfall products, including one generated from satellite soil moisture retrievals, to provide an enhanced regional- and continental-scale validation analysis of H-SAF H23 and H05B rainfall products, compared to the traditional practice of using a single benchmark rainfall product for validation. Analyses and discussions are primarily focused on the four years (2011-2014) of the H23 daily rainfall product. In addition, results for 2.5 years (2015-2017) of the H05B daily rainfall product are also presented. In the first part of our analysis TC is applied to obtain the correlation (versus an unknown truth) metric for H23 along with SM2R and a reanalysis precipitation product. In the second part, by adding another gridded daily precipitation dataset generated from ground observations, a QC analysis is performed to provide a more robust estimate of the H23 correlation metric and a verification of the error-independence assumptions underlying the TC analysis. Finally, a high-quality benchmark rainfall product is generated by linearly merging the SM2R, reanalysis and ground-based datasets using optimal parameters derived via TC. This benchmark dataset allows a thorough assessment of the H23 error structure in terms of continuous and categorical performance metrics, as well as seasonal error behaviors.

This report is organized as follows: Section 2 provides information about the rainfall datasets used in the analysis; Section 3 explains the TC and QC methodology and assessment strategy; Section 4 describes the main findings and Section 5 summarizes and discusses the results.

2. Rainfall Data Products

2.1 H-SAF H23, H05B

As noted in the introduction, our primary motivation here is an enhanced evaluation of the H-SAF demonstrational H23 rainfall product. The H23 is a Level 3 gridded precipitation product providing daily mean rainfall rate in mm/hour based on the temporal resampling of passive microwave (PMW) instantaneous precipitation rate estimates (http://hsaf.meteoam.it/description-h02b-h03b-H05B-h15b-h17-h18-h23.php). This off-line product is provided on a 0.25° x 0.25° grid over the whole Meteosat Second Generation (MSG) full disk (60°S - 75°N, 60°W -60°E). The gridded PMW mean precipitation is obtained from the H-SAF instantaneous precipitation rate operational products H01 (Casella et al., 2013; Sanò et al., 2013, Mugnai et al., 2013) and H02B (Sanò et al., 2015), for SSMIS and AMSU/MHS, respectively, by exploiting all the DMSP SSMIS and MetOp/NOAA AMSU/MHS satellite overpasses available throughout a day. Daily rainfall accumulation values between Jan 01, 2011 and Dec 31, 2014 used in this study are calculated from mean hourly rates.

The H05B product is based on frequent precipitation measurements as retrieved by blending PWM precipitation rate measurements (i.e. the H-SAF H01B, H02B products) and infrared imagery (http://hsaf.meteoam.it/description-H05B.php). H05B rainfall accumulation covers the MSG full disk area and is generated and distributed every 3 hours at 00, 03, 06, 09, 12, 15, 18 and 21 UTC in a space view perspective or orthographic grid with an average resolution of approximately 8 km over Europe. Daily rainfall accumulation values between Jun 03, 2015 and Nov 10, 2017 used in this study are resampled to a 0.25° geographic grid using bilinear interpolation.

2.2 E-OBS

The European Daily High-Resolution Observational Gridded Dataset (E-OBS) v17.0 daily precipitation dataset is based on a spatial interpolation of observations acquired from more than 10,000 rain gauge stations in Europe on a regular 0.25° x 0.25° grid between 25° N and 75° N latitudes and 40° W and 75° E longitudes (van Engelen et al., 2008). E-OBS data is provided by and downloaded from the European Climate Assessment & Dataset project (ECA&D; https://www.ecad.eu). It has been designed to provide the best available estimate of grid-scale, daily (0 to 24 UTC) precipitation accumulations. The interpolation process is based on three steps. First, monthly precipitation totals are interpolated onto the 0.25° grid using thin-plate splines. Next, daily anomalies are interpolated onto the same grid using universal kriging with an external drift for temperature. Finally, monthly totals and daily anomalies are merged into a single daily rainfall estimates for each grid cell. Note that E-OBS data is completely missing over Poland due to rain gauge observations being unavailable during the study period (2011-2014), and moreover, no gauge observations are available within Russia during 2014.

2.3 SM2RAIN-ASCAT

SM2RAIN-ASCAT (Brocca et al., 2019) is a newly released rainfall dataset based on the application of the SM2RAIN algorithm (Brocca et al., 2013, 2014) to satellite soil moisture observations derived from the Advanced SCATterometer (ASCAT), provided as Climate Data Record product H113 (2007-2017) and H114 (2018) within the H-SAF project.

SM2RAIN is based on the inversion of the soil water balance equation solved for rainfall and assumes 1) negligible evapotranspiration during rainfall and 2) null "surface" runoff component at the scale of the satellite pixel (for detailed description of the algorithm and its assumptions, refer to (Brocca et al., 2014, 2015). SM2RAIN-ASCAT rainfall dataset (referred to as "SM2R" for short here) covers the period 2007-2018 and is characterized by a spatial/temporal sampling of 12.5 km/1-day. The dataset is available through https://doi.org/10.5281/zenodo.2580285.

ERA5 rainfall is used as the reference dataset to calibrate SM2RAIN for the generation of SM2RAIN-ASCAT rainfall (Brocca et al., 2019), which raises concern about violating the error-independency assumption of TC and QC. However, as our QC-based analysis (see Section 3.5) reveals low level of error correlation between the ERA5-SM2R pair, the use of both products in the collocation analyses is justified.

2.4 ERA5

ERA5 is the latest generation of ECMWF's atmospheric reanalyses of global climate which will be available from 1950 to eventually replace the current ERA-interim (Hersbach et al., 2018). ERA5 is based on 4D-Var data assimilation using Cycle 41r2 of the Integrated Forecasting System (IFS) and is produced since 2016. In this work, the global 0.25□ gridded ERA5 hourly total precipitation field is used together with ERA5 snowfall estimates to obtain rainfall. The daily accumulation is obtained by summing up all the hourly estimates within each day. The data were downloaded from the C3S Climate Data Store (CDS, https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview).

ERA5 assimilates satellite radiance observations (infrared and microwave) and ground-based radar precipitation observations (from 2009) to produce its reanalysis precipitation field. Because of the potential sharing of microwave (SSMIS and AMSU/MHS) observations used in both the ERA5 and H23 products, it is necessary to consider the possibility of H23-ERA5 cross-correlated error which violates a key TC assumption. (see Section 3.5 below).

3. Methodology

3.1 Triple Collocation (TC)

Underlying the triple collocation technique is the linear additive error model between the measurement systems and the unknown truth:

$$X = \alpha + \beta T + \varepsilon \quad (1)$$

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

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

$$R_X = \sqrt{\frac{\sigma_{XY}\sigma_{XZ}}{\sigma_X^2\sigma_{YZ}}} \quad (2)$$

where σ_{XY} is the covariance of X and Y, etc., and σ_X^2 is the variance of X. For further details please refer to McColl et al. (2014). 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*. 2016a; Draper *et al*. 2013; Zwieback *et al*. 2012).

Due to the linearity assumption expressed in (1), we applied TC only at pixels where significant (p = 0.05) levels of mutual correlation exist between the members of the triplet. To ensure the stationarity of signal and error statistics (see above), a common practice is to remove seasonal signals from the raw time-series prior to application of TC (e.g., Chen et al., 2017; Gruber *et al*. 2016a; Su and Ryu, 2015). However, because of the intermittent nature of precipitation, no-rain days are an important category of observation and should not be artificially altered into non-zero anomaly values. Therefore, raw daily rainfall time-series were used here in the TC and QC analyses.

3.2 Quadruple Collocation (QC)

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 different measurement principles. The H23 is a blended PWM precipitation product, based on precipitation retrieved from radiometric observations provided by SSMIS onboard DMSP satellites and AMSU/MHS onboard Metop-A/B and NOAA18/19 satellites. The H05B is a blended PWM and IR precipitation product. Although the SM2RAIN product is also remote sensing-based, the

dataset used here 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 and H05B products). The ERA5 precipitation field is generated via assimilation of a wide range of atmospheric and land surface variables into a coupled land/atmosphere modelling system. Finally, the E-OBS data are generated via spatial interpolation of ground-based rain gauge observations. However, despite our best efforts to diversify the sources of information, it is difficult to eliminate the possibility of error co-dependence. For example, both the H23/H05B and ERA5 products rely on PMW observations from the atmosphere. In such cases, quadruple collocation (QC) provides an opportunity to verify this assumption by calculating the error cross-correlation (ECC) between different product pairs.

When a fourth rainfall product is available, R can be estimated using QC based on the formulation in Gruber *et al*. (2016b). The same assumptions for TC also apply for QC, but since the four datasets constitute an over-constrained system, it allows the designation of one non-zero error covariance term which can be estimated with a least squares solution (Pierdicca *et al*. 2015). Therefore, the zero ECC assumption can be relaxed to allow non-zero ECC to exist between one, and only one, pair of data products.

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

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

(3)

where σ_T^2 is the true daily rainfall variance, β is the multiplicative bias in (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.

And the least-squares solution for the parameters in S is given as

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

Based on Draper et al. (2013) and McColl et al. (2014), correlation with the unknown truth is then calculated from

$$fRMSE_X = \frac{\sigma^2_{\varepsilon_X}}{\sigma^2_X}, \quad R_X = \sqrt{1 - fRMSE_X}, \quad (5)$$

in which fRMSE is the fractional-RMSE introduced in Draper et al. (2013). Finally, ECC between *X* and *Y* can then be obtained from

$$ECC_{XY} = \sigma_{\varepsilon_X \varepsilon_Y} / (\sigma^2_{\varepsilon_X} \sigma^2_{\varepsilon_Y}) . \quad (6)$$

3.3 Rainfall error model

For rainfall measurements, a multiplicative error model is often considered more appropriate than an additive error model (Anagnostou, 2006; Tian et al., 2013; Alemohammad et al., 2015), *i.e.*

$$X = \alpha T^{\beta} e^{\varepsilon} \quad (7)$$

where e^{ε} is the multiplicative random error and α and β describe the systematic errors. Tian et al. (2013) compared the additive and multiplicative error models applied to daily precipitation datasets across the US and suggested that the use of a multiplicative model allows for the improved separation of random errors from systematic signals and is applicable to 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 to be non-zero. Solutions to this problem include temporal aggregation (e.g. Alemohammad et al., 2015) to remove the zeros in the time-series, or simply discarding the zero observations (e.g. Massari et al. 2017). As Massari et al. (2017) pointed out, daily precipitation error has different characteristics than multi-day accumulation errors due to the frequent presence of zeros values. Also, TC results are less reliable when its sampling power is reduced by the removal of zero-observations. Massari et al. (2017) compared the performance of multiplicative and additive error models during the application of TC to daily precipitation in the U.S. and concluded that, despite the theoretical advantages of a multiplicative errors model, an additive error model provides reasonable and robust error results and little practical advantage can be attributed to the use of a multiplicative error model. Therefore, in this analysis we employed the additive error model for both the TC and QC analyses presented below.

3.4 Merging of rainfall datasets using TC

TC has also been applied to the optimal merging of geophysical data products obtained from multiple sources (Yilmaz et al., 2012; Crow et al., 2015; Gruber et al., 2017). Here, we apply TC to merge the ERA5, E-OBS and SM2R daily rainfall products into a benchmark dataset for H23 and H05B assessments. As discussed earlier, zero values in daily precipitation are meaningful observations and should not be altered by turning into anomalies. Therefore, the analyses are based on raw, not anomaly time-series. TC-based merging is performed at 0.25□ grid cells where significant positive correlations ($p<0.05$) between ERA, E-OBS and SM2R are found. Prior to merging, one of the three products is chosen as the reference dataset and the other two products are scaled to its climatology using the optimal scaling method described in Yilmaz et al. (2013). The reference dataset, i.e., its climatology in which the merged dataset will be generated, should be carefully chosen to minimize the (complicated) impact on the satellite product validation results. For example, metrics describing error magnitude such as bias and root-mean-square difference are directly affected by the dynamic range of the reference dataset. Statistics concerning mutual trend (e.g. correlation) or event detection are less affected as all three datasets contribute (though unevenly) to these aspects. Here, ERA5 is chosen as the reference dataset based on its superior spatial-temporal coverage, prediction skill (see QC results in section 4.1) as well as continuous availability through ECMWF (ERA5 will be maintained as an operational product to at least the mid-2020s and to be replaced with a version with shorter latency). Note that, the choice of the reference dataset does not change the scaling weights assigned to each product (from (12) below) after the scaling steps in (9). Nor does it impact our H23/H05B correlation metrics obtained from TC or QC.

Following Yilmaz et al. (2013), optimal multiplicative scaling factors for transforming the E-OBS (Y) and SM2Rain (Z) products into the ERA5 (X) climatology are given by:

$$\beta_Y = \frac{\sigma_{XZ}}{\sigma_{YZ}}\ , \beta_Z = \frac{\sigma_{XY}}{\sigma_{YZ}}, \tag{8}$$

where $\sigma_{XZ}, \sigma_{YZ}, \sigma_{XY}$ are covariances between the three datasets. Scaled E-OBS (Y^*) and SM2Rain (Z^*) are then obtained as

$$Y^* = \beta_Y(Y - \bar{Y}) + \bar{X} \text{ and } Z^* = \beta_Z(Z - \bar{Z}) + \bar{X}\ , \tag{9}$$

respectively, where the overbar indicates temporal averaging. The error variances for the X, Y* and Z* can then be derived from TC as:

$$\sigma^2_{\varepsilon_X} = \sigma^2_X - \frac{\sigma_{XY^*}\sigma_{XZ^*}}{\sigma_{Y^*Z^*}}\ , \sigma^2_{\varepsilon_{Y^*}} = \sigma^2_{Y^*} - \frac{\sigma_{XY^*}\sigma_{Y^*Z^*}}{\sigma_{XZ^*}}\ , \sigma^2_{\varepsilon_{Z^*}} = \sigma^2_{Z^*} - \frac{\sigma_{XZ^*}\sigma_{Y^*Z^*}}{\sigma_{XY^*}}\ . \tag{10}$$

These three products are linearly combined into a merged product (*M*) as:

$$M = w_X \cdot X + w_Y \cdot Y^* + w_Z \cdot Z^*, \quad (11)$$

where w_X, w_Y, w_Z are weights with a sum of 1 and derived from

$$w_X = \frac{\sigma^2_{\varepsilon_X}}{\sigma^2_{\varepsilon_X}+\sigma^2_{\varepsilon_{Y^*}}+\sigma^2_{\varepsilon_{Z^*}}}, w_Y = \frac{\sigma^2_{\varepsilon_{Y^*}}}{\sigma^2_{\varepsilon_X}+\sigma^2_{\varepsilon_{Y^*}}+\sigma^2_{\varepsilon_{Z^*}}}, w_Z = \frac{\sigma^2_{\varepsilon_{Z^*}}}{\sigma^2_{\varepsilon_X}+\sigma^2_{\varepsilon_{Y^*}}+\sigma^2_{\varepsilon_{Z^*}}} \quad (12)$$

that lead to a maximum reduction of random error variance in the merged time-series (Gruber et al., 2017). Any negative value obtained from (9) is preserved and applied in subsequent calculations (10-12). Finally, negative values of *M* obtained from (11) are removed and replaced by zero. As for H23, values of *M* less than 1 mm/day are also treated as zero when comparing with each other. Hereinafter, *M* is referred to as the "MESMO" (Merged ERA5-SM2Rain-E-OBS) product.

3.5 Error statistics

In addition to the correlation coefficient (*R*) metric obtained from TC and QC analyses, several other performance metrics – obtained from direct comparison with MESMO – are also calculated to provide a more complete understanding of error in the H-SAF H23 and H05B products. Two general types of metrics are evaluated in this analysis, some of which are defined below: 1) continuous error statistics including R, root mean square error (RMSE), root mean square error percent (RMSE%), mean bias (Bias) and *mean* multiplicative bias (MB); 2) categorical error statistics such as Probability Of Detection (POD), False Alarm Rate (FAR), Critical Success Index (CSI), and fraction correct accuracy (ACC). Daily accumulated rainfall values of less than 1 mm are treated as zero in both products.

MB and RMSE% are calculated as:

$$MB = \overline{obs}/\overline{T}, \quad (13)$$

$$RMSE\% = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{(obs_i - T_i)^2}{T_i^2}} * 100, \quad (14)$$

in which *obs* is the satellite observation (H23 or H05B) and MESMO serves as *T* ("truth"). The categorical error statistics are derived based on the following contingency table:

Table 1. Contingency table of rainfall observations.

		Truth	
		rain	no rain
Satellite	rain	**hits**	**false alarms**

observation	no rain	**misses**	**correct negatives**

and calculated as follows:

$$POD = \frac{hits}{hits+misses} \quad (15)$$

$$FAR = \frac{false\ alarms}{hits+false\ alarms} \quad (16)$$

$$CSI = \frac{hits}{hits+misses+false\ alarms} \quad (17)$$

$$ACC = \frac{hits+correct\ negatives}{hits+misses+false\ alarms+correct\ negatives} \quad (18)$$

Therefore, POD represents the fraction of successfully detected events among all rainfall events. FAR is the fraction of false events among all events detected by the satellite product and CSI (also referred to as the "Threat Score") combines POD and FAR and represents the success rate of rainfall event prediction over all predicted and missed events. ACC captures the fraction of total correct predictions, which – besides correctly predicted rainfall event predictions – includes correct no-rain predictions that are not considered in POD, FAR or CSI. All of POD, CSI and ACC are all bounded in the range [0, 1] and have a perfect score of 1. FAR also ranges between 0 and 1 but has a perfect score of 0.

Parallel assessments for H23 and H05B are conducted using the same methods described earlier. The correlation metric (R) of H23 (as well as H05B, which is omitted in the rest of this section for brevity) is estimated using both TC (with a H23-SM2R-ERA5 triplet) and QC (with a H23-EOBS-SM2R-ERA5 quadruplet). These results will be discussed in Section 4.1. Other error statistics described above, as well as seasonal error behavior of H23, are obtained by comparison against MESMO (the E-OBS-SM2R-ERA5 merged product) and are discussed in Section 4.2. In the preprocessing procedure, the 0.25°-grid cells are masked for the cases of 1) less than 100 data triplets (for TC) or quadruplets (for QC); or 2) the lack of significant correlation (p-value < 0.05) between all the data pairs utilized for TC or QC. The spatial coverage of QC results is primarily constrained by the E-OBS dataset. For example, analysis was masked for most of the Poland territory due to the lack of rain gauges across the country during the study period. As a result, the QC analysis is carried out in a geographic region approximately between 30° N-60° N and 10° W-60° E across Europe plus a small coastal strip of northern Africa. TC is performed in a larger geographic domain, between 15° S-60° N and 60° W-60° E including Africa and parts of South America.

The H23, SM2R and ERA5 (the latter through assimilation) products are all derived from satellite observations. As discussed earlier, some overlapping between H23 and ERA5 data sources may exist (see Section 3.2). Also, ERA5 rainfall was used to calibrate

the SM2RAIN algorithm that generated SM2R (see Section 2.3). Potential error cross-correlation (ECC) between these products may lead to biased QC and TC correlation estimates. Therefore, three QC scenarios are performed, *i.e.* the "H23-ERA5" case, in which non-zero ECC is assumed to be exist only between H23 and ERA5; and analogously, the "H23-SM2R" and "ERA5-SM2R" cases. Results (not shown) suggested that relatively low levels of ECC (between ±0.20) are found across the study region, indicating that these datasets can be reasonably assumed to contain independent errors and the impact of violation of the zero-ECC assumption in TC analysis (using a H23-SM2R-ERA5 triplet) should be small. Similar results were also obtained for H05B.

4. Results

Our evaluation of the H23/H05B daily precipitation products is based on a two-fold application of the collocation analyses. First, we will provide unbiased estimates of R for H23/H05B using both TC (i.e., R_{TC}) and QC (i.e., R_{QC}) and collocated SM2R, E-OBS and ERA5 data (Section 4.1). Next, in order to obtain a more complete set of evaluation metrics, we will apply TC to optimally merge the SM2R, E-OBS and ERA5 products into the MESMO product, and then directly compare H23/H05B to MESMO (Section 4.2).

4.1 Triple and quadruple collocation analyses

4.1.1 Quadruple collocation estimates of R (R_{QC})

Correlation coefficient obtained from quadruple collocation (i.e., R_{QC}) applied to the H23, E-OBS, SM2R and ERA5 daily rainfall products are plotted in Figure 1. Plotted values are obtained from averaging the results from the three different QC scenarios, each assuming non-zero ECC between a different pair of products (see Section 3.5). For H23, moderate to high (0.6 – 0.8 [-], same below) R_{QC} scores are found over much of the study area. Relatively low R_{QC} values (~0.4) are found in regions characterized by a complex orography (such as Alps and Caucasus), Kazakhstan and bordering area of Russia. The overall satisfactory H23 performance is somehow expected, since this product is derived by combining precipitation products (H01 and H02B) derived using algorithms specifically tailored for the European (and African) areas. At the same time, difficulties in retrieving accurate precipitation estimates in regions characterized by complex orography are well known. Among the four rainfall products, E-OBS demonstrates the highest R_Q scores among the four products in Western Europe, which coincides with regions of high rain gauge density. On the other hand, E-OBS' performance declines considerably in Eastern/Southeastern Europe. Among the four products, ERA5 demonstrates the best overall performance across the study area, with a homogeneous distribution of R_Q values consistently near or greater-than 0.80.

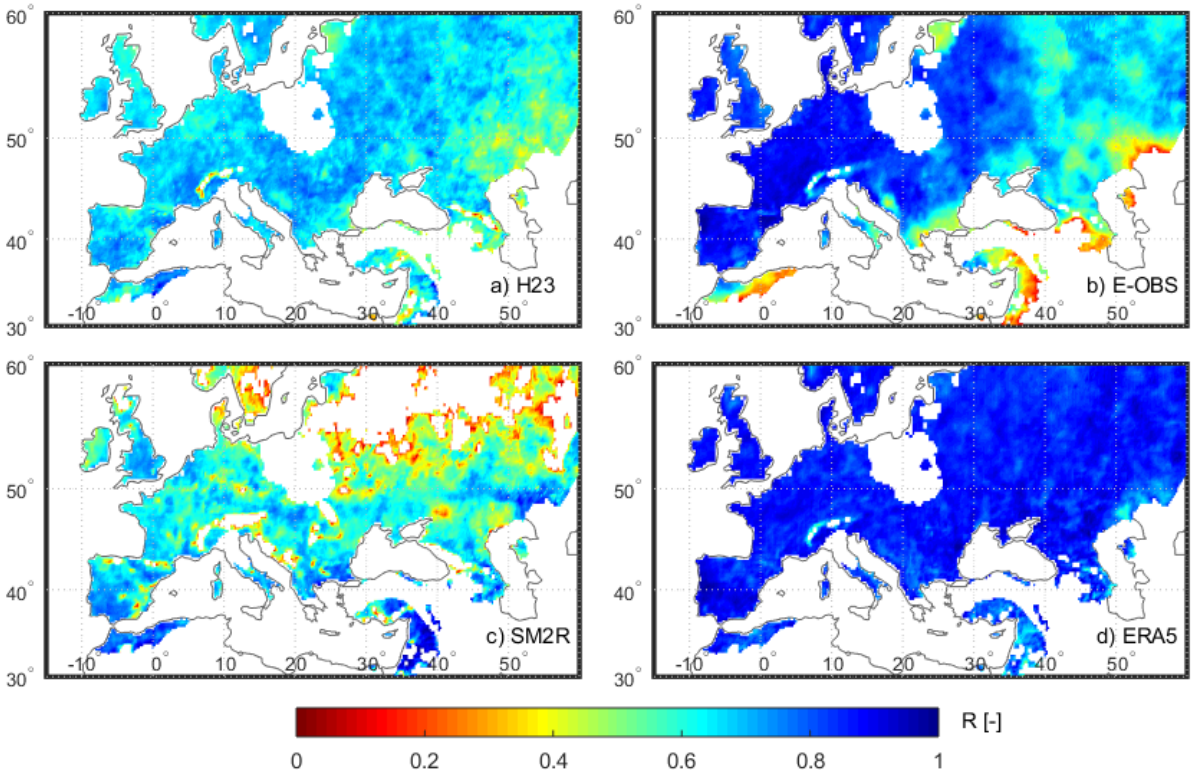

Figure 1. QC-based estimates of correlation coefficient (R_Q) for the a) H23, b) E-OBS, c) SM2R and d) ERA5 daily rainfall products for the period 2011 – 2014.

The performance of SM2R is comparable with H23 in the Mediterranean region. SM2R is slightly better than H23 in southern Spain and northern Africa, south UK and Middle East/Israel, while H23 is better in France, Germany and the Balkan Peninsula. Low SM2R R_Q values are in the Scandinavia region are likely due to the low ASCAT soil moisture retrieval quality and availability over frozen surfaces. Poor SM2R performance over the continental Europe (including Russia) is consistent with the finding that ASCAT soil moisture has lower skill compared to SMAP and SMOS in this region (while having better skill than SMOS in western European) from a global TC analysis of the three satellite surface soil moisture products (Chen et al., 2018). Our preprocessing analysis (not shown) also indicates relatively low correlation between SM2R and other datasets at higher latitudes. Both findings suggest that relatively large errors in the ASCAT soil moisture over continental Europe would propagate into the SM2R rainfall estimates, which explains the relatively low R_Q found for SM2R compared to other products. In addition, low skill in SM2R estimates also contributes to the widespread masking of Northeast Europe in Fig. 1c – reflecting areas where QC yields an unrealistic R_Q value (i.e., fRMSE>1) estimated by (5).

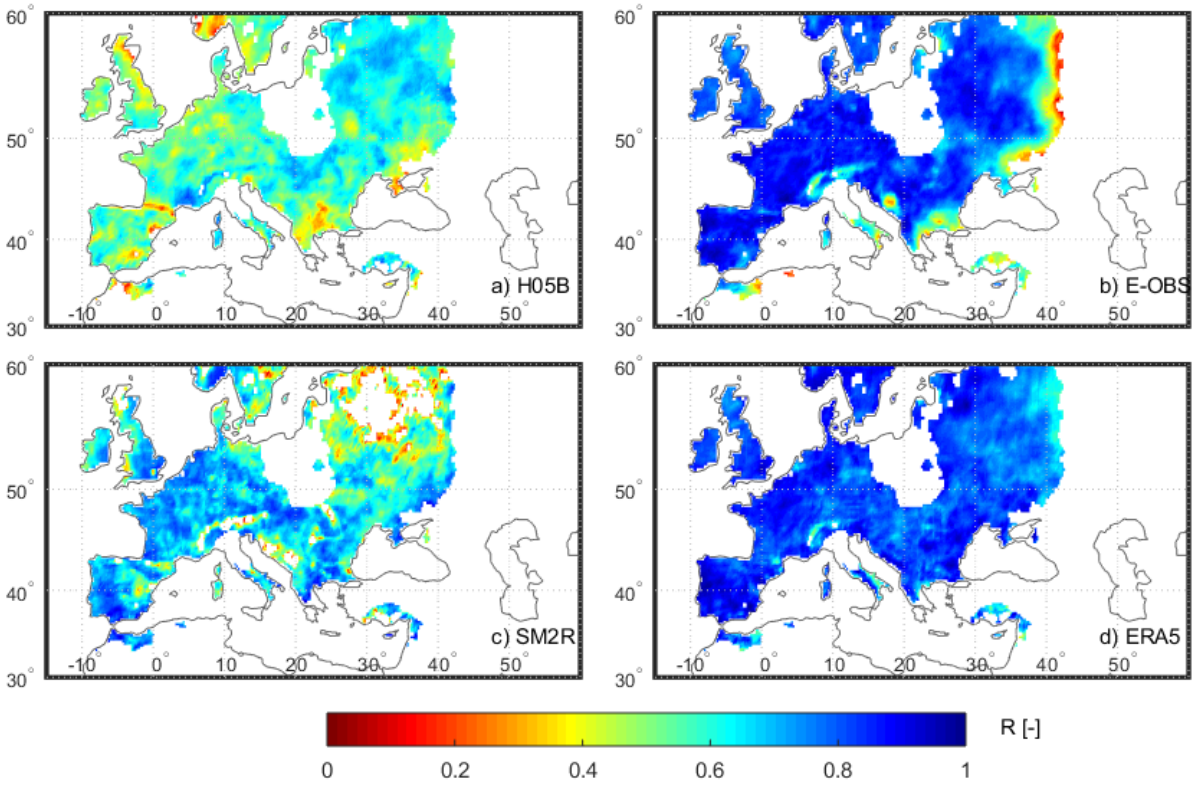

Figure 2. Same as Figure 1 for H05B for the period June 2015 – November 2017.

QC results for H05B and other rainfall products (Figure 2) show similar spatial patterns to results for H23. However, lower overall QC_R values (i.e. domain-average of 0.56) for H05B are observed compared to H23. Figure 2a also suggests quite poor H05B skills over the mountain ranges of the Alps, the Pyreness, the Balkans and the Scandinavia. Note the missing results in eastern Europe are due to lack of E-OBS data for the H05B evaluation period (June 2015 – November 2017).

As a product generated from quality-controlled, ground-based observations, E-OBS is a natural choice for benchmark dataset to validate airborne rainfall products. However, relatively poor E-OBS performance in some areas (see Figure 1b) indicates that E-OBS is ill-suited to serve as a universal ground truth and direct validation of H23 against E-OBS is likely to be biased in areas of low E-OBS quality. Figure 3a maps the direct correlation of H23 with E-OBS (obtained without QC or TC), while Fig. 3b shows its difference with respect to H23 R_{QC} results (Fig. 1a), as a function of the rain gauge density that E-OBS is based on. Direct H23 versus E-OBS correlation 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. The impact of E-OBS station density on the H23-E-OBS correlation is evident in Figure 3b, which shows sharp increase in the difference between the two

estimates (ΔR) from one rain gauge per grid cell to none. As gauge density increases the spread of ΔR shrinks and is kept at a relatively stable level of ~0.1 with three or more gauges per grid cell. However, the underestimation of H23 R using E-OBS alone is not eliminated by densities as high as 10 gauges per grid cell. Thus Figure 3 demonstrates that the direct comparison of H23 against E-OBS inflates statistical representations of H23 error virtually everywhere – but especially in areas where ground observations (underlying the E-OBS product) are spatially sparse.

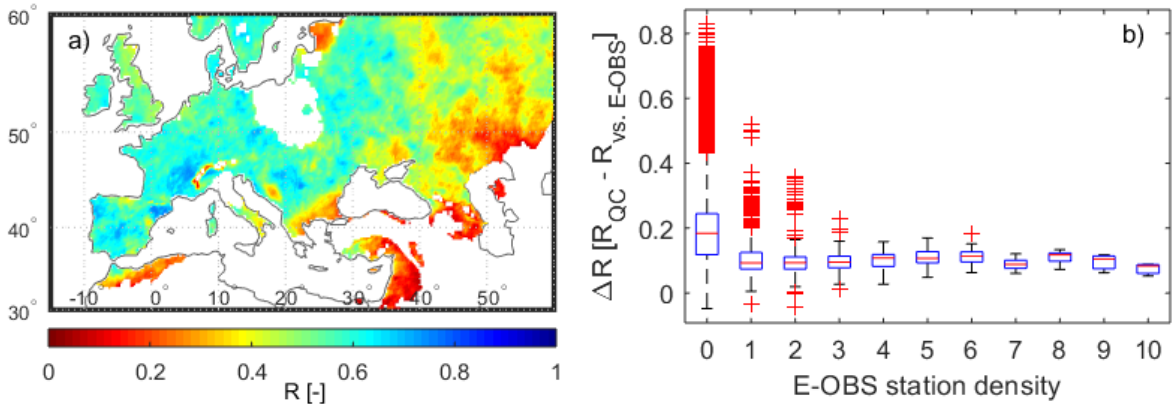

Figure 3. a) Map of the (directly sampled) temporal correlation between H23 and E-OBS. b) The difference between R_{QC} and the direct correlation plotted in a) as a function of E-OBS station density.

4.1.2 Triple collocation estimates of R (R_{TC})

Above, QC is demonstrated to improve H23 evaluation over sparsely gauged areas. In cases where ground-based observations (i.e., E-OBS) are completely unavailable, however, TC can be applied instead. Utilizing an SM2R-ERA5-H23 triplet, R_{TC} for these rainfall products can be obtained for a larger domain covering the entire African continent, the Arabian Peninsula, and eastern part of South America (Fig. 3a-c). Note that

R_{TC} results are also available over Poland now given that E-OBS is not considered.

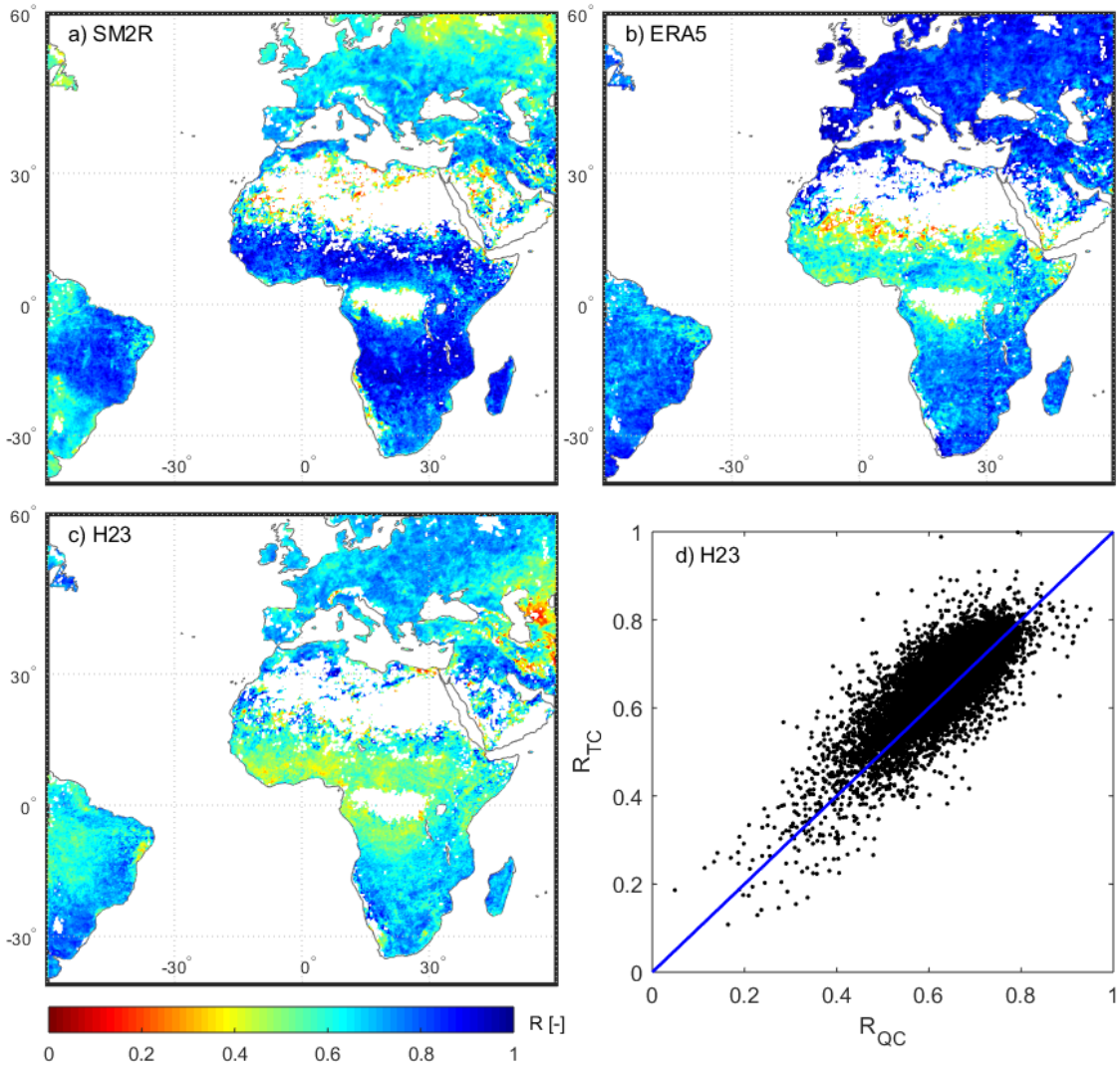

Figure 4. Estimates of correlation coefficient (R) from TC analysis for a) SM2R, b) ERA5 and c) H23. Scatterplot of H23 R estimates via TC (R_{TC}) versus QC (R_{QC}) is shown in d).

R_{TC} maps show patterns in Europe and northern Africa similar to the R_{QC} results (Fig. 1). Strong H23 correlations (with true precipitation accumulations) are found in western Europe and parts of Saudi Arabia. In African regions south of the Sahel, SM2R shows the highest skill, followed by ERA5 and H23. Both of H23 and ERA5's performance is worse in the equatorial zones between approximately 15□ S – 15□ N, i.e. where seasonal movements of the ITCZ brings shift of wet/dry episodes to the land. Here,

satellite soil moisture retrieval appears to be a better proxy (i.e. used in SM2R) to retrieve rainfall amounts than PMW-based estimates. Missing results over the Sahara Desert are due to the low occurrence of rainfall as well as missing SM2R data (due to a lack of ASCAT soil moisture retrievals in the region). In South America, H23 obtains higher R_{TC} values (~0.8 and above) over the Atlantic coastal regions and moderate (~0.5 – 0.6) R_{TC} values inland. SM2R, on the other hand, shows excellent skill in the northern part of the Brazilian Highlands, a drier region east of the Amazon Basin.

Despite the complete absence of E-OBS, TC yields robust R estimates (R_{TC}) for H23 that are highly comparable to R_{QC} results (obtained earlier over Europe considering E-OBS). Comparison of collocated H23 R_{TC} and R_{QC} (Fig. 4d) values shows a high degree of agreement with a correlation of 0.81 and root-mean square difference of only 0.05. This finding supports the use of non-ground-based rainfall products such as SM2R and ERA5 to robustly derive performance metrics for H23 in the absence of ground-based observations. R_{TC} results for H05B, SM2R and ERA5 are shown in Figure 5.

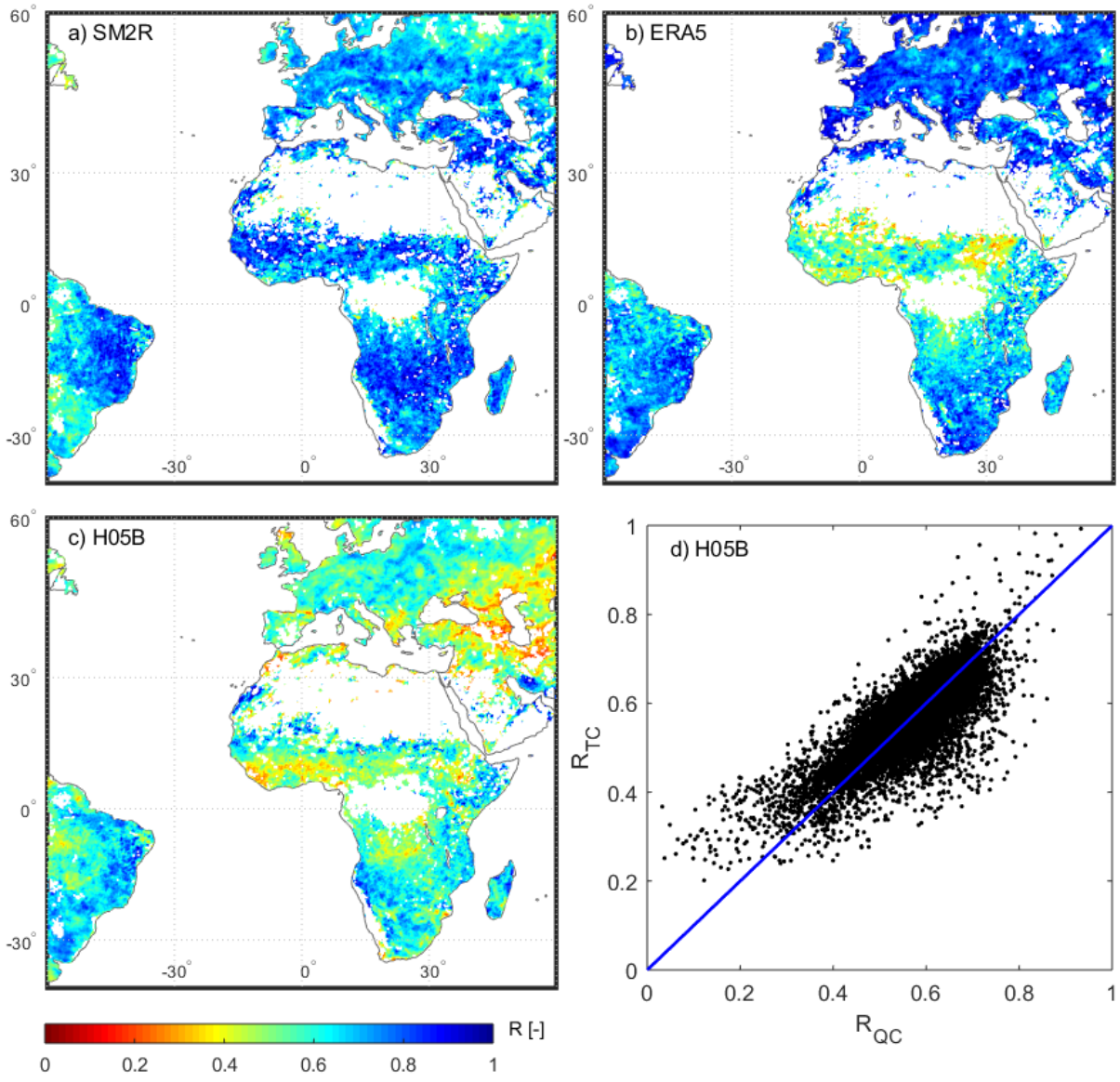

Figure 5. Estimates of correlation coefficient (R) from TC analysis for a) SM2R, b) ERA5 and c) H05B. Scatterplot of H05B R estimates via TC (R_{TC}) versus QC (R_{QC}) is shown in d).

4.2 Merged daily rainfall product

QC and TC are useful tools to provide (relatively) unbiased, direct estimates of R. However, a variety of other error metrics must be considered, in order to complete a comprehensive validation of a rainfall product. To this end, we applied the TC-based optimal merging scheme (Section 3.4) to derive a benchmark daily rainfall product (i.e.

"MESMO") via TC-based weighted averaging of the E-OBS, SM2R and ERA5 datasets. The statistical evaluation of H23 based on direct comparisons against MESMO are presented and discussed in the following subsections.

4.2.1 Generation of merged daily rainfall product (MESMO) via TC

A reference daily rainfall dataset (MESMO) is generated by optimally merging the E-OBS, SM2R and ERA5 daily rainfall products using a TC-based procedure. MESMO is generated in the ERA5 climatology based on consideration of data availability as well as spatially consistent high skills (Figure 1). Scaling factors, or weights, assigned to each of the three products derived from TC for the 2011-2014 H23 validation period are shown in Figure 6. Scaling factors for the H05B validation period are calculated separately (not shown). As expected the weighting pattern closely reflects the rainfall gauge density behind the E-OBS dataset. In particular, E-OBS is weighted heavier in areas where the 0.25° grid cells are directly gauged, which also reflects the skill of E-OBS determined by collocation analyses (e.g., Figure 1b). Over Western Europe, where higher gauge density and R scores of E-OBS are found, the TC-merging scheme determines that ~50% of MESMO should be contributed by E-OBS. ERA5 and SM2R share most of the weight for ungauged areas, with ERA5 generally weighting more than SM2R.

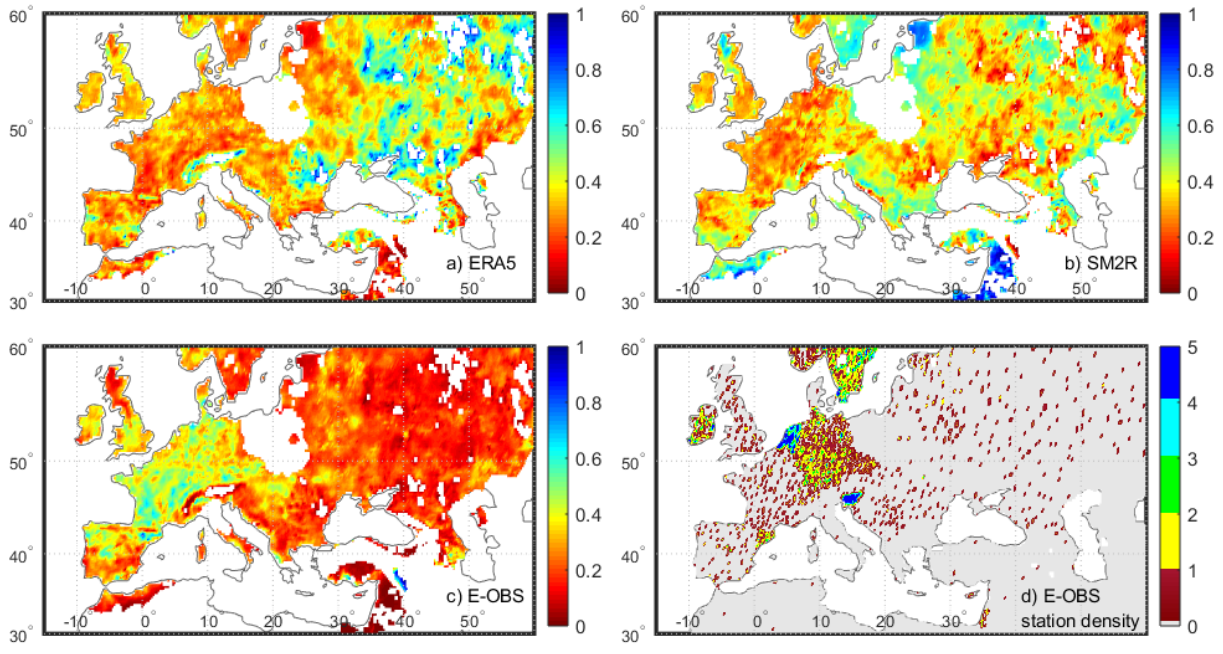

Figure 6. TC-based relative weights assigned to the a) ERA5, b) SM2R and c) E-OBS products for merging into MESMO. Rain gauge density used to produce the E-OBS daily rainfall during the 2011-2014 period is shown in d).

4.2.2 Evaluation of H23 against MESMO

A non-TC based estimate of H23 correlation coefficient can be calculated directly against MESMO. Although it is generated by optimally merging E-OBS, SM2R and ERA5, residual random errors will remain in MESMO. As a result, direct H23-MESMO correlation values are somewhat lower than R_{QC} values that have been explicitly corrected for this effect (Figure 7). The two sets of R estimates are strongly correlated (R = 0.89) and a mean bias of -0.07 is relatively small compared to sampled correlation values. Therefore, results support the general robustness of MESMO-based H23 errors metrics.

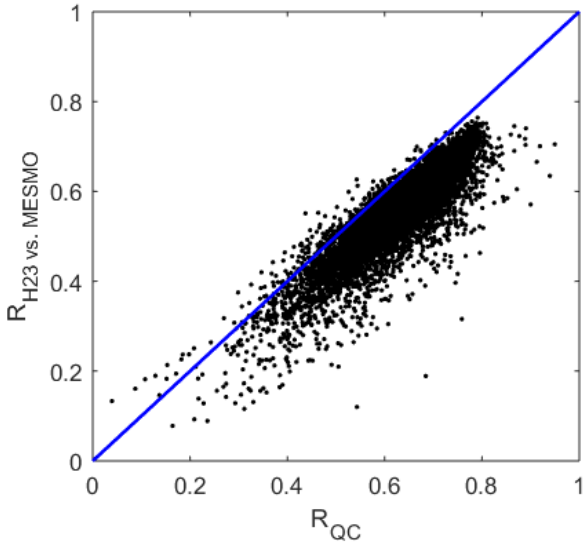

Figure 7. Scatterplot of the relationship between sampled correlation of H23 versus MESMO and the H23 correlation obtained from QC (R_{QC}).

In addition to correlation, H23 validation metrics calculated against MESMO also include RMSE, RMSE%, bias and multiplicative bias (MB), shown in Figure 8. H23 RMSE (Fig. 8a) ranges from 2 to 6 mm/day in most of the study area, with a mean of 4.6 mm/day and standard deviation of 1.5 mm/day. Caution should be used to interpret the large RMSE values (and other metrics as well) found along the coast of Mediterranean Sea and Caspian Sea, where uncertainty of MESMO is potentially high due to the relatively large errors in ERA5, E-OBS and SM2R presented (see Fig. 1). In other words, because of the difficulty of various products behind MESMO to accurately represent rainfall events in this coastal area, the estimated H23 error metrics calculated against MESMO are of considerable uncertainty there. H23 RMSE% across the domain (Figure 8b) ranges between 81 and 3894, with an average of 209 and standard deviation of 133. For H23 RMSE% values are less than 100 in only ~0.5% of the grid cells, most of which located in the UK, Norway and Sweden.

H23 bias map (Figure 8c) indicates underestimated rainfall in most Western European countries and Northwestern Russia and overestimation elsewhere. Averaged over the 4-year period, mean daily bias of the study area is only 0.07 mm/day with a standard deviation of 0.7 mm/day. There is a corresponding northwest-to-southeast gradient of increasing multiplicative bias (MB ([-]), Figure 8d) from less than one to greater than 2. Large MB (>3) are observed in Kazakhstan and the most severe overestimation occurs in areas of Syria and Northern Egypt. Note that MESMO is completely missing for Russia and Kazakhstan during 2014 (due to the lack of E-OBS data) and the metrics for the region are based on only three years of data (2011-2013).

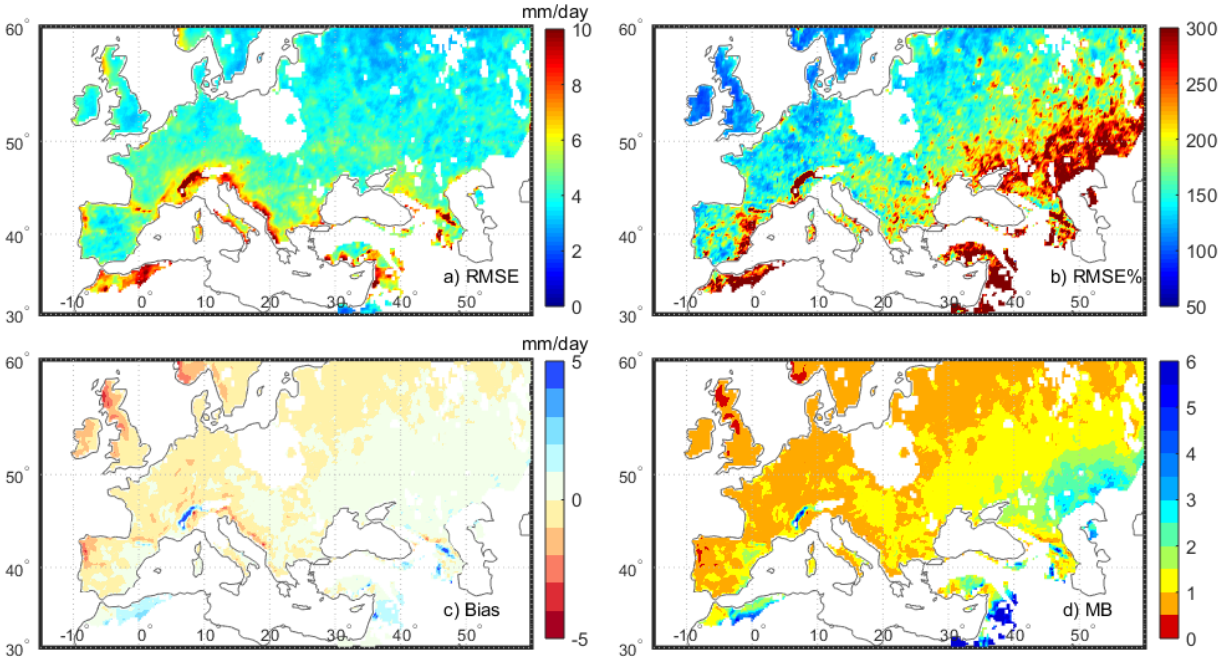

Figure 8. H23 performance metrics calculated against MESMO between 01/01/2011 and 12/31/2014: a) RMSE, b) RMSE%, c) additive bias and d) multiplicative bias.

As noted earlier, MESMO time-series for the H05B evaluation period is generated separately as to the one for the H23 evaluation. H05B performance metrics calculated against MESMO (Figure 9) reveal similar spatial pattern as in Fig. 8 for H23, with significantly larger errors in the southern regions indicated by RMSE and RMSE% (Fig. 9 a, b). A prominent contrast of wet bias in the south-southeast versus dry bias in the north-northwest is present (Fig. 9c).

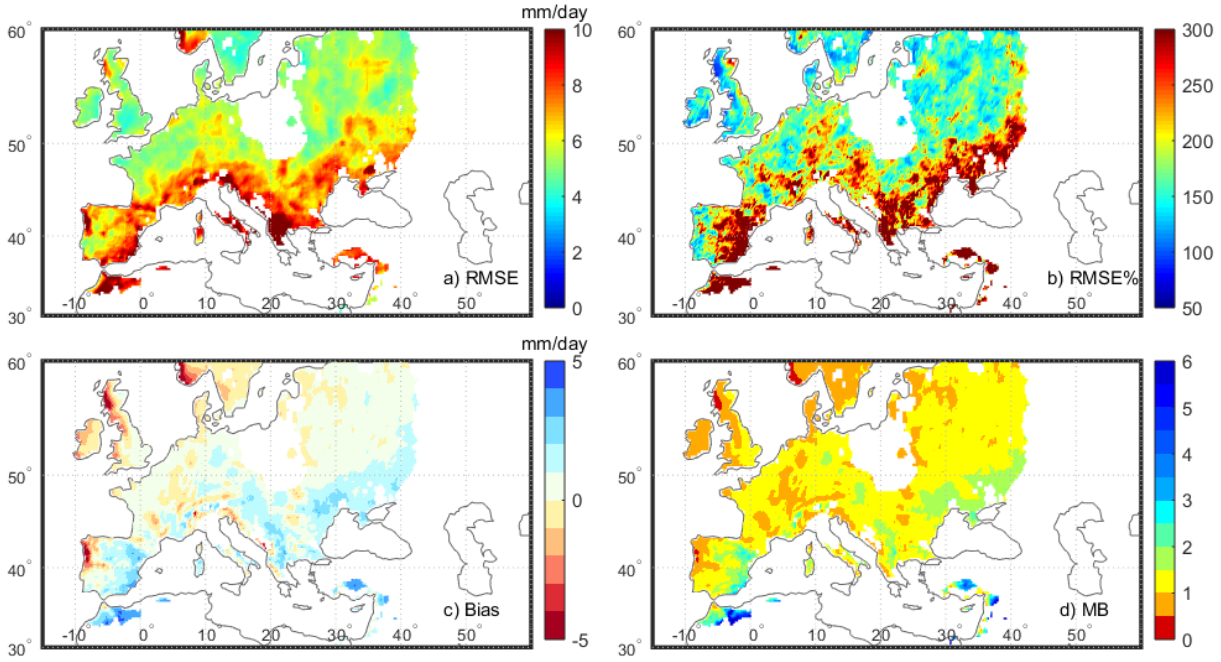

Figure 9. H05B performance metrics calculated against MESMO between 01/01/2011 and 12/31/2014: a) RMSE, b) RMSE%, c) additive bias and d) multiplicative bias.

Relative to MESMO, H23 is overall slightly wetter when averaged over four years (Fig. 8c). Seasonal variabilities of H23 error within the study area are examined by monthly mean biases (Fig. 10). H23 manifests a persistent low bias during the winter months (November – March) and a high bias in the summer months (May – August). H23 has a greater dynamic range (~ 0.5 – 3 mm/day, averaged monthly), while MESMO lacks a strong seasonal cycle of spatially-averaged rainfall amounts. Note the H23 biases show a similar level of spatial variability (standard deviation in Fig. 10b) throughout the seasonal cycle.

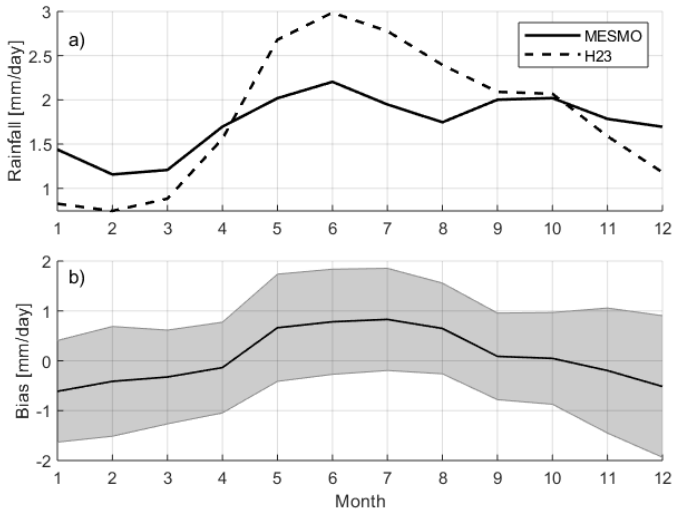

Figure 10. a) Domain-average monthly mean daily rainfall of MESMO and H23; b) H23 monthly-mean daily bias (solid line) with one standard deviation (shaded area) across the study area.

Spatial patterns of the H23 bias in the months of January and July are shown in Fig. 11. In January, H23 significantly underestimates rainfall in areas of the Northern UK, Northern Portugal, Western Spain, most of Germany, and the western coast of Adriatic Sea. However, a significant wet bias in the Italian Po Valley stands out in January amid the dry biases. In contrast, a wet bias in July is primarily found in Kazakhstan, Southwestern Russia, Ukraine and Northeastern Spain. As such, H23 seasonal errors are characterized by a general dry bias in the winter in Western Europe and a summertime wet bias in Southern and Eastern Europe.

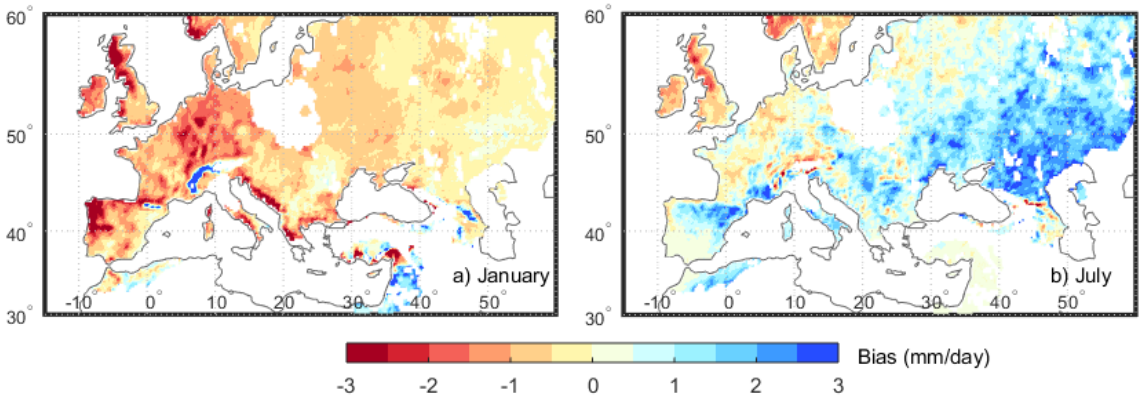

Figure 11. Mean daily bias of H23 versus MESMO for a) January and b) July.

The H05B daily rainfall estimates also show an overall dry bias in January, but areas of eastern Spain, the Po Valley and Balkans are slightly wetter than MESMO (Figure

12a). In July, H05B shows larger areas of wet bias than H23 with more severe overestimation over Spain.

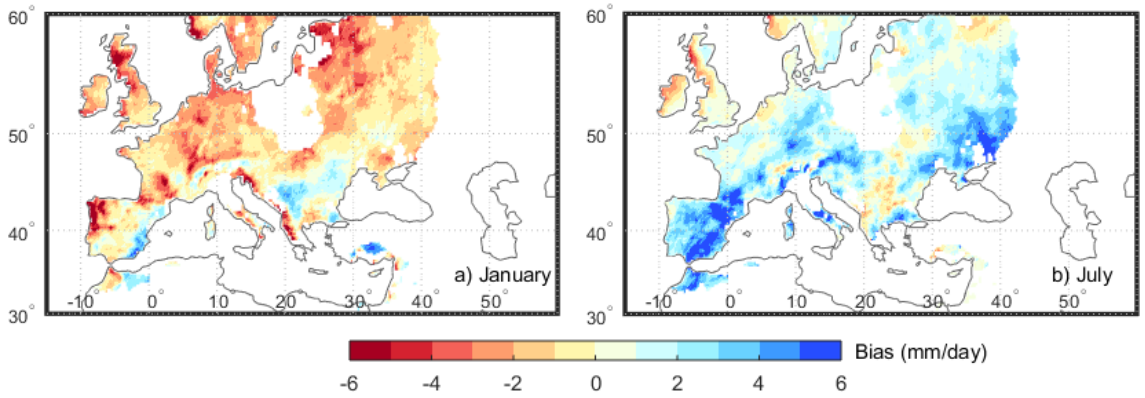

Figure 12. Same as Figure 11 for H05B.

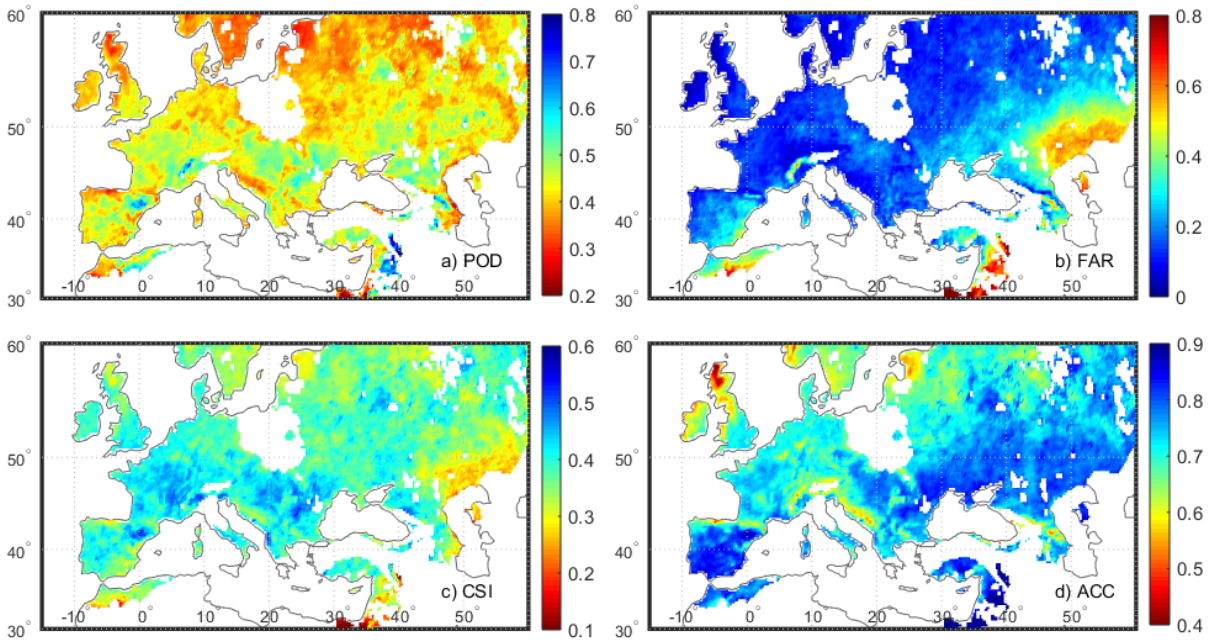

Figure 13. Multi-categorical error statistics of H23 compared to MESMO: a) POD, b) FAR, c) CSI and d) ACC.

Categorical performance scores of H23 against MESMO are shown in Figure 13. The POD ranges from 0.4 to 0.5 [-] in most of Europe and is generally poorer over the northern part of the domain, which suggests that H23 misses a large fraction of rainfall events in Northern Europe. The FAR is generally low (~0.20 [-]) throughout the study area except Southeastern Spain, parts of Syria and areas north of the Caspian Sea (mainly Kazakhstan). The strong gradient in FAR north of the Caspian Sea generally corresponds

to the regional increasing annual precipitation trend towards higher latitudes. A large proportion of H23-predicted rainfall events appears to be false-positive in this relatively dry area (just north of the Caspian Sea).

The CSI suggests that H23 has relatively good success rainfall event prediction rate in parts of France and the Balkan region (~0.50 [-]). Relatively low CSI values in Kazakhstan and Southern Russia are associated with both high FAR and low POD. The overall correct prediction rate for H23 including both rain and no-rain events indicated by ACC is generally above 0.6 [-] across Europe. ACC is higher in the drier regions of southern Europe, where H23 identifies the frequent occurrence of no-rain days relatively well, in addition to correctly predicting actual events.

Categorical metrics for H05B daily rainfall are shown in Figure 14. Compared to H23, H05B has a relatively high POD of mostly above 0.5. POD is high (0.6~0.8) over Western Europe (except the British Isles) and is relatively low in Eastern Europe (below 0.65). The FAR is generally below 0.2, with higher values in areas of Spain and north of Black Sea which are associated with the wet bias (Fig. 9). The CSI suggests that H05B has a correct rate of above 0.5 in predicted rainfall events.

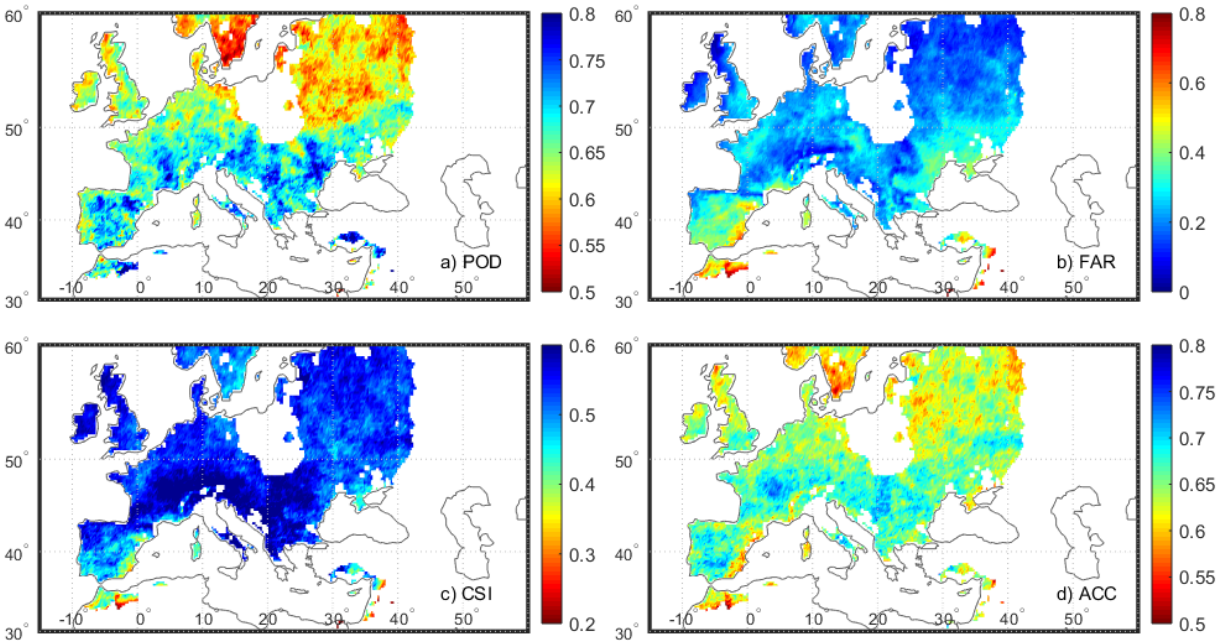

Figure 14. Same as Figure 13 but for H05B.

5. Summary

This report presents an innovative validation strategy of satellite precipitation products based on both direct and indirect applications of the collocation techniques. The proposed methodology is applied to validate the H-SAF H23 and H05B daily precipitation products. TC and QC analyses are carried out utilizing a ground-based observation dataset (E-OBS), a satellite soil moisture-converted rainfall product (SM2R) and a reanalysis dataset (ERA5). SM2R proves be a useful, largely independent dataset to apply QC and TC analyses in the evaluation of rainfall products. Despite SM2R's relatively poor performance compared to other types of rainfall observations and reanalysis products over Europe, no noticeable degradation is found in H23's R_{TC} results compared to R_{QC}.

Correlation metric of H23 is directly estimated by triple collocation using SM2R and ERA5 and quadruple collocation using E-OBS, SM2R and ERA5. H23 correlation values versus true daily rainfall accumulations are moderate-to-high (~ 0.6 – 0.8) across most of the study area in Western Europe, South America and Africa, and relatively poor in areas of equatorial Africa and continental Europe. The performance of SM2R is comparable with H23 in the Mediterranean region but poor in the Scandinavia region and over continental Europe. SM2R also outperforms ERA5 and H23 in regions south of the Sahel and northeastern Brazil. ERA5 demonstrates the best overall performance across the study area and shows poor R metric only in the equatorial Africa region.

In addition to correlation, a suite of continuous and categorical error statistics for H23 are obtained by comparing directly against MESMO, a new benchmark daily rainfall dataset generated in the climatology of ERA5 by TC-based optimal merging of E-OBS, SM2R and ERA5. Across Europe, H23 has an average RMSE of 4.7 mm/day and an average RMSE% of 200. A notable increasing northwest – southeast gradient in multiplicative bias indicates the tendency of H23 to overestimate rainfall rates in the drier areas (i.e. towards areas north of the Caspian Sea), which is also where severe False Alarm Ratios (FAR) are found. On the other hand, high ACC (fraction correct accuracy) values in the drier regions of Southern Europe suggest that H23 accurately recognizes a large fraction of the no-rain days. In the northern regions, however, H23 tends to miss a large fraction of rainfall events (i.e. poor POD). The overall correct prediction rate for H23 including both rain and no-rain events indicated by ACC is generally above 0.6 across Europe. Analysis of seasonal error behavior also suggests that H23 has stronger seasonal dynamics than the reference dataset (MESMO in ERA5 climatology) and tends to underestimate rainfall during winter (primarily in Western Europe) and overestimate during summer (mainly in Southern and Eastern Europe).

A high-quality, optimally merged rainfall product that combines the strengths of different data sources provides an opportunity to thoroughly investigate the spatio-temporal error characteristics as well as to quantify more aspects of the error statistics not covered in this study. Future work should include examination of the uncertainty in the merged dataset, as well as its impacts on the validation results in different aspects of rainfall estimation (e.g. intensity, temporal accumulation, temporal variability, etc.).

Overall, this report provides a viable strategy for the validation of accumulated precipitation also in data scarce regions and thus can be considered the basis for the future validation of the H-SAF accumulated precipitation products. The approaches developed here are also potentially applicable to H-SAF instantaneous precipitation rate products as well – an extension that should be developed in future research.

6. Acknowledgements

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 gratefully acknowledge EUMETSAT through the "Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF)" (http://hsaf.meteoam.it/) and the Italian Civil Protection Department. The authors would like to thank also Christian Massari, Luca Ciabatta, Luca Brocca, Giulia Panegrossi, Anna Cinzia Marra, Paolo Filippucci and Silvia Puca for their help and support on this project.

7. References

Alemohammad, S. H., McColl, K. A., Konings, A. G., Entekhabi, D., and Stoffelen, A. 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., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B., and Wagner, W. (2019). SM2RAIN-ASCAT (2007–2018): global daily satellite rainfall from ASCAT soil moisture. *Earth Syst. Sci. Data Discuss.*, https://doi.org/10.5194/essd-2019-48, in review.

Ebert, E. E., Janowiak, J. E., and Kidd, C. 2007: Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models, B. Am. Meteorol. Soc., 88, 47–64, https://doi.org/10.1175/BAMS-88-1-47.

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., and Kirschbaum, D. B. 2017: So, How Much of the Earth's Surface IS Covered by Rain Gauges?, *B. Am. Meteorol. Soc*., 98, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1.

Massari, C., Crow, W., and Brocca, L. 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.

Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), https://cds.climate.copernicus.eu/cdsapp#!/home

Crow, W. T., Su, C.-H., Ryu, D., and Yilmaz, M. T. 2015: Optimal averaging of soil moisture predictions from ensemble land surface model simulations, *Water Resour. Res*., 51, 9273– 9289, doi:10.1002/2015WR016944.

Gruber, A., W. A. Dorigo, W. Crow and W. Wagner, 2017: Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals, *IEEE Transactions on Geoscience and Remote Sensing*, 55(12), 6780-6792, doi: 10.1109/TGRS.2017.2734070.

Gruber, A., Su, C.-H., Crow, W.T., Zwieback, S., Dorigo, W.A., Wagner, W., 2016. 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.

Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones and M. New. 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. *J. Geophys. Res. Atmos.*, 113, D20119, doi:10.1029/2008JD10201.

Hersbach, H., P. de Rosnay, B. Bell, D. Schepers, A. Simmons, C. Soci, S. Abdalla, M. Alonso Balmaseda, G. Balsamo, P. Bechtold, P. Berrisford, J. Bidlot, E. de Boisséson, M. Bonavita, P. Browne, R. Buizza, P. Dahlgren, D. Dee, R. Dragani, M. Diamantakis, J. Flemming, R. Forbes, A. Geer, T. Haiden, E. Hólm, L. Haimberger, R. Hogan, A. Horányi, M. Janisková, P. Laloyaux, P. Lopez, J. Muñoz-Sabater, C. Peubey, R. Radu, D. Richardson, J.-N. Thépaut, F. Vitart, X. Yang, E. Zsótér & H. Zuo, 2018: Operational global reanalysis: progress, future directions and synergies with NWP, *ECMWF ERA Report Series* 27.

Li, C., Tang, G., and Hong, Y. 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.

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

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

Sapiano, M. R. P. and Arkin, P. A. 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.

Stampoulis, D. and Anagnostou, E. N. 2012: Evaluation of Global Satellite Rainfall Products over Continental Europe, *J. Hydrometeorol*., 13, 588–603, https://doi.org/10.1175/JHM-D-11-086.1.

Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., and Wu, H. 2013: Modeling errors in daily precipitation measurements: Additive or multiplicative*? Geophys. Res. Lett*., 40, 2060–2065, doi: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. *European Climate Assessment & Dataset Rep*., 68pp.

Villarini, G., Mandapaka, P. V., Krajewski, W. F., and Moore, R.J. 2008: Rainfall and Sampling Uncertainties: A Rain Gauge Perspective, *J. Geophys. Res*., 11, D11102, https://doi:10.1029/2007JD009214.

Yilmaz, M. T., Crow, W. T., Anderson, M. C., and Hain, C. 2012: An objective methodology for merging satellite and model based soil moisture products, *Water Resour. Res*., 48, W11502, doi:10.1029/2011WR011682.