https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Open Access
Discussions

WHICH RAINFALL METRIC IS MORE INFORMATIVE ABOUT THE FLOOD SIMULATION PERFORMANCE? A COMPREHENSIVE ASSESSMENT ON 1318 BASINS OVER EUROPE

Stefania Camici [(1)], Christian Massari [(1)], Luca Ciabatta [(1)], Ivan Marchesini [(1)], Luca Brocca [(1)]

(1) National Research Council, Research Institute for Geo-Hydrological Protection, Perugia, Italy (s.camici@irpi.cnr.it)

* Correspondence to: Ph.D. Stefania Camici, Research Institute for Geo-Hydrological Protection, National Research Council, Via della Madonna Alta 126, 06128 Perugia, Italy. Tel: +39 0755014419 Fax: +39 0755014420 E-mail: stefania.camici@irpi.cnr.it.

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions

ABSTRACT

The global availability of satellite rainfall products (SRPs) at an increasingly high temporal/spatial resolution has made possible their exploitation in hydrological applications, especially over in-situ data scarce regions. In this context, understand how uncertainties transfer from SRPs into flood simulation, through the hydrological model, is a main research question.

SRPs accuracy is normally characterized by comparing them with ground observations via the calculation of categorical (e.g., threat score, false alarm ratio, probability of detection) and/or continuous (e.g., bias, root mean square error, Nash-Sutcliffe index, Kling-Gupta efficiency index, correlation coefficient) metrics. However, whether these metrics are informative about the associated performance in flood simulations (when the SRP is used as input to an hydrological model) is an underdiscussed research topic.

This study aims to relate the accuracy of different SRPs both in terms of rainfall and in terms of flood simulation. That is, the following research question are addressed: is (are) there appropriate performance metric (s) to drive the choice of the best performing rainfall product for flood simulation? To answer this question three SRPs, namely the Tropical Rainfall Measurement Mission Multi-satellite Precipitation Analysis, TMPA; the Climate Prediction Center Morphing algorithm, CMORPH, and the SM2RAIN algorithm applied to the ASCAT (Advanced SCATterometer) soil moisture product, SM2RAIN-ASCAT, have been used as input into a lumped hydrologic model (MISDc, "Modello Idrologico Semi-Distribuito in continuo") on 1318 basins over Europe with different physiographic characteristics.

Results have suggested that, among the continuous metrics, correlation coefficient and Kling-Gupta efficiency index are not reliable scores to select rainfall product performing best for hydrological modelling whereas bias and root mean square error seem more appropriate. In particular, by constraining the relative bias to values lower than 0.2 and the relative root mean square error to values lower than 2, good hydrological performances (Kling-Gupta efficiency index on discharge

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

greater than 0.5) are ensured for almost 75% of the basins fulfilling these criteria. Conversely, the categorical scores have not provided suitable information to address the SRPs selection for hydrological modelling.

Key words: satellite rainfall products, hydrological validation, rainfall-runoff modelling, Europe.

1. INTRODUCTION

Accurate rainfall estimate is essential in many fields spanning from climate change research, weather prediction and hydrologic applications (Tapiador et al., 2017, Ricciardelli et al., 2018, Lu et al., 2018). In particular, the delivery of real time rainfall observations is one of the most challenging task in operational flood forecasting both for technical reasons related to the need of a prompt release of the observations and for scientific motives linked to the necessity of ensuring sufficient accuracy to provide a reliable forecasting. Generally rainfall observations are obtained through real time ground monitoring networks (e.g., Artan et al., 2007), meteorological and numerical weather prediction models (e.g, Montani et al., 2011; Zappa et al., 2008) and, more recently, by satellite observations (Mugnai et al., 2013) that, albeit with some difficulties (Maggioni and Massari, 2018) are gaining ground with respect to the classical rainfall monitoring methods.

The global availability of near real time satellite rainfall products (SRPs) has boost their use for hydrological applications, specifically for river discharge estimation via rainfall-runoff models (Casse et al., 2015; Elgamal et al., 2017; Camici et al., 2018; Beck et al., 2017, see Maggioni and Massari, 2018 and Jiang and Wang, 2019 for a more complete review). In particular, in the past decade a special attention has been paid on the propagation of the satellite rainfall error on flood simulations (Hong et al., 2006; Hossain, and Anagnostou, 2006; Pan et al., 2010; Maggioni et al. 2013; Thiemig et al. 2013; Bhuiyan et al., 2019) and two approaches, one probabilistic and one statistical, can be recognized (Quintero et al., 2016). In the probabilistic approach a statistical model is first used to produce an ensemble of possible rainfall realizations. Then, each rainfall realization

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions

is used to simulate discharge time series through an hydrologic model and the difference between simulated and in situ discharge data is used to assess how rainfall accuracy transfers to the flood simulation (e.g., Hong et al., 2006; Hossain, and Anagnostou, 2006; Demaria et al. 2014; Maggioni et al. 2013, 2011). In the deterministic approach, SRPs are first compared with a reference dataset to assess the accuracy in terms of rainfall estimate. Then, SRPs are used as input in rainfall-runoff models to estimate river discharge that is then compared with in situ discharge observations. Eventually, the existence and the shape of the relationship between the SPR accuracy and the associated discharge score is analysed (e.g, Serpetzoglou et al. 2010; Pan et al., 2010; Thiemig et al. 2013; Chintalapudi et al. 2014; Pakoksung and Takagi, 2016; Shah and Mishra, 2016; Qi et al. 2016; Ren et al., 2018; Bhuiyan et al., 2019).

In both approaches, several continuous (e.g., bias, root mean square error, RMSE, correlation coefficient, R, Nash-Sutcliffe efficiency index, NSE, Kling-Gupta efficiency index, KGE) and categorical (e.g., probability of detection, POD, false alarm ratio, FAR, threat score, TS) performance scores are used to characterize the accuracy in terms of rainfall and river discharge. Generally this comparison has been carried out for few basins (e.g., Hong et al., 2006; Pan et al., 2010; Demaria et al., 2014; Chintalapudi et al., 2014; Qi et al. 2016; Ren et al., 2018; Thiemig et al. 2013), rarely at regional scale (e.g., Bhuiyan et al., 2019), whereas no studies investigated the hydrological propagation of SRP error at a continental scale. In Beck et al (2017), the authors carried out an evaluation of multiple (22) global daily rainfall datasets both in terms of rainfall and discharge for many (+9000) basins over the globe, however, the relationship between the accuracy in terms of rainfall and discharge was not investigated in detail.

From both the probabilistic and the statistical approach, arises that the hydrological performances of SRPs depend on a complex interaction among the characteristics of the input data (i.e., precipitation type, seasonality, data resolution or time window considered, see e.g., Ebert et al., 2007; Vergara et al., 2014; Satgé et al., 2019), the hydrological model formulation (i.e. parameter estimation and modelled processes, Quintero et al., 2016; Mei et al., 2017; Bhuiyan et al., 2019), the characteristics of the basin (e.g., area and initial soil moisture conditions, land use and land cover Yong et al., 2010; Yilmaz et al., 2005; Nikolopoulos et al., 2010; Mei et al., 2016; Shah and Mishra, 2016; Gebregiorgis et al., 2012) and observations (i.e., streamflow data, see e.g., Nikolopoulos et al., 2012). In this context, it is not trivial to draw general guidelines about which SRPs should be favoured or which error metric(s) should be used to identify the best performing rainfall product for flood forecasting (Qi et al., 2016; Hossain and Huffman, 2008). The only largely accepted suggestion is about SRP bias, recognized as a major issue for a reliable flood forecast across several basins around the world (Maggioni et al., 2013; Thiemig et al. 2013; Shah and Mishra 2016; Jiang and Wang, 2019). Based on that, bias correction methods have shown to significantly reduce streamflow errors (e. g, Yilmaz et al. 2005; Bitew et al., 2012; Valdes-Pined et al., 2016). For instance, by using MIKE SHE model on a small and mountainous basin in the Blue Nile basin, Bitew et al. (2012) stated that large biases in satellite rainfall directly translate into bias in one or more of the hydrology simulation components. Zhu et al. (2016) found that for two humid basins in China, the accuracy on flood simulations is related to the mean error and to bias in the rainfall estimates as also found by Yilmaz et al. (2005). Besides bias, it is difficult to found literature studies advising on rainfall error metrics able to indicate flood simulation performances. The work of Bisselink et al. (2016), even if conducted over only 4 basins in south Africa, is an exception. The authors, by using different SRPs as input to LISFLOOD model, proved that a high correlation between monthly rainfall and observed streamflow is a needed prerequisite for obtaining good hydrological performances, as long as the rainfall variability in time is not too high.

Based on that, there is a need to investigate metrics that can more effectively advance the use of SRPs for hydrological applications, and specifically for flood modelling at regional scales. This paper aims to explore the link between satellite rainfall accuracy of different products and their flood modelling performance. The following research questions are addressed: which is the most appropriate performance metric to be used to select the best performing satellite rainfall product for

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

flood modelling? Are R and RMSE, generally used to characterize the rainfall accuracy, informative about the hydrological modelling performance?

In pursuing this goal, three different near real time SRPs are considered, i.e., Tropical Rainfall Measurement Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) real time product (TMPA 3B42RT, Huffman et al., 2010), the Climate Prediction Center (CPC) morphing technique (CMORPH, Joyce et al., 2004) and SM2RAIN-ASCAT rainfall product (Brocca et al., 2019) obtained by applying the SM2RAIN algorithm (Brocca et al., 2014) to the ASCAT satellite soil moisture product are used to force a lumped hydrological model, MISDc (Brocca et al., 2011) over 1318 basins spread out over Europe. An intercomparison of SRPs with respect to a benchmark rainfall dataset, i.e., E-OBS (Haylock et al., 2008), is carried out. This step, along with the reliability assessment of the different SRPs for flood modelling over Europe, constitutes only an intermediate output of the work. The ultimate aim of the paper is to investigate how SRPs accuracy propagates through the river discharge simulations, as to help in the selection of the rainfall metrics more informative of better hydrological performances. As the intent of the paper is to analyse the performances of near-real time satellite rainfall products, gauge-corrected satellite or reanalysis rainfall products are not considered in this work.

2. STUDY AREA

The study area is composed by 1318 basins belonging to 23 different countries and spread over the whole Europe, over longitude varying from -10° to 25° and latitude from 35° to 70° (Figure 1). European continent is characterized by a complex topography ranging, from south to north, from huge mountains towards hilly plateaus to a large plain. The Alpine mountain chain, crossing the continent from west to east represents the highest and more extensive mountain range system in Europe. Hilly plateaus gently slops towards the Great European Plain, a low flat region, extending from the Atlantic coast of France to the Urals, crossed by many rivers and with densely populated cities.

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences Discussions
Open Access

The climate is humid continental with cold summers in central and eastern Europe. Mean annual rainfall across Europe ranges between 300 mm year^{-1} and 4000 mm year^{-1}, depending on the location. The area east, west and north of the Alps generally is interested by higher rainfall amount, while along the edges of the Mediterranean Sea and in northern Scandinavia, lighter rainfall is common. In terms of floods, their occurrence range from spring to summer moving from northeastern Europe towards the Alps, whereas Mediterranean region and western Europe are prevailing subject to winter floods (Berghuijs et al., 2019).

The main features of the study basins, clustered according the latitude of the outlet section, are summarized in Table 1: among the 1318 basins, more than half (889) have the outlet section located below the 50° latitude and about 11% of them the outlet section is placed above 60° latitude. Basin areas range in size from 200 to 136'000 km^2 and the median area of the basins located below 50° is lower than the one of basins located in northern part of Europe (above 50° latitude). By considering these features, the selected set of basins can be considered a comprehensive sample of the European basin characteristics, definitely.

3. DATASETS

The datasets used in this study include both ground observations and satellite rainfall products (Table 2).

3.1 Ground observations

Ground observations comprise rainfall, air temperature and river discharge data. Rainfall and air temperature are extracted from the European high-resolution gridded data sets version 16.0 (E-OBS, Haylock et al., 2008), currently maintained by the Copernicus Climate Change Service. The E-OBS dataset is built using data from nearly 2316 stations (i.e., equivalent on average to a density of 1 stations every 4000 km^2) but the station density significantly varies across Europe (see Haylock et al., 2008; Cornes et al., 2018). However, as this product is composed by time series thoroughly checked both in terms of quality and homogeneity (Klok and Tank, 2009) and it is continuously available from 1950 up to now at daily time step, it can be considered a good benchmark for the analysis.

Daily river discharge data are obtained through an european daily dataset, compiled by the authors merging stations from 5 different databases: the Global Runoff Data Base (GRDC, https://www.bafg.de/GRDC/EN/Home/homepage_node.html), the European Water Archive (EWA, https://www.bafg.de/GRDC/EN/04_spcldtbss/42_EWA/ewa.html?nn=201574), the Italian ISPRA HIS national database (http://www.hiscentral.isprambiente.gov.it/hiscentral/default.aspx); the Portuguese national database (http://snirh.pt/) and the Spanish national database (http://ceh-flumen64.cedex.es/anuarioaforos/default.asp). From the resulting European dataset, composed by 3913 quality checked stations covering the period 1900-2016, 1318 stations with available observations after 2007 (according the availability of SRPs, see paragraph 3.2) have been extracted. The authors, using the EU-DEM digital elevation model (Mouratidis and Ampatzidis, 2019) resampled at 100m ground resolution, developed an automatic and rapid procedure to delineate the drainage watersheds located upstream of each discharge measurement location (outlet section). The study basins and the related observation period length after 2007 is shown in Figure 1a: more than 50% of the basins have an observation period longer than 7 years; Spanish, Italian and Northern European basins have a nearly complete observation period (10 years), whereas for Central Europe some stations end the monitoring period on 2012 and the median length of discharge observations is about 6/7 years (see Table 1).

3.2 Satellite rainfall products

Three different SRPs have been used in this study: TMPA 3B42RT, CMORPH and SM2RAIN-ASCAT satellite products. As these products have been largely used in literature, only a brief product description is reported in the following whereas for major details the reader is referred to Huffman et al. (2010); Joyce et al. (2004) and Brocca et al. (2019) for TMPA 3B42RT, CMORPH and SM2RAIN-ASCAT, respectively.

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

TMPA 3B42RT, provided by NASA (National Aeronautics and Space Administration, http://disc.sci.gsfc.nasa.gov/) covers ±50° north-south latitude band with a spatial sampling of 0.25° and a temporal resolution of 3 h from 1997 onward.

CMORPH is provided by CPC (Climate Prediction Center, ftp://ftp.cpc.ncep.noaa.gov) for the +60°/-60° latitude band from March 2000 up to now. In this study, the CMORPH raw version is extracted with a spatial/temporal resolution of 0.25°/3 hours.

In addition to these state-of-the-art SRPs, we used the SM2RAIN-ASCAT rainfall product (Brocca et al., 2019) obtained through the application of the SM2RAIN algorithm (Brocca et al., 2014) to the ASCAT satellite soil moisture product (Wagner et al., 2013). SM2RAIN-ASCAT, downloadable at https://zenodo.org/record/3405563, is available for the period 2007-2019, with a 12.5 km spatial sampling and a daily temporal aggregation.

For sake of simplicity, the TMPA 3B42RT, CMORPH and SM2RAIN-ASCAT satellite datasets are indicated in the following as TMPA, CMOR and $SM2R_{ASCAT}$, respectively. By considering the spatial/temporal availability of both ground-based and satellite observations (see Table 2 for a summary), the analysis has been carried out to cover the maximum common observation period, i.e., from 2007 to 2016 at daily time scale (TMPA and CMOR are aggregated at daily scale), with three different areal masks cut: 1) at the original spatial coverage of each SRP, i.e., until 50°, 60° and 70° latitude for TMPA, CMOR and $SM2R_{ASCAT}$, respectively; 2) below the TMPA area (latitude <50°); 3) above TMPA area (latitude >50°).

4. METHOD

4.1 Hydrological model

The model applied to carried out the flood simulation is MISDc ("Modello Idrologico Semi-Distribuito in continuo" Brocca et al. 2011), a two-layer continuous hydrological model composed by a component simulating the temporal pattern of soil moisture and by a rainfall-runoff model simulating flood. By using as input daily rainfall and air temperature data, MISDc simulates the

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions

most important processes involved in the rainfall-runoff transformation (e.g., infiltration, evapotranspiration, saturation excess and percolation). The geomorphological Instantaneous Unit Hydrograph (IUH) is used to transfer surface and subsurface runoff to the outlet of the catchment. The model (downloadable at: http://hydrology.irpi.cnr.it/download-area/midsc-code/) uses 9 parameters calibrated by maximizing the Kling-Gupta efficiency index (KGE, Gupta et al., 2009; Kling et al., 2012, see paragraph 4.5 for more details) between observed and simulated river discharge.

The successful results obtained through MISDc model for flood simulation in many different basins (in Italy, see e.g., Brocca et al., 2011; 2013a, Massari et al. 2015; Masseroni et al. 2016; Cislaghi et al. 2019, and in Europe, see e.g., Brocca et al., 2013b; Massari et al. 2018; Camici et al., 2018) and for different applications (e.g., climate change impact studies, see Camici et al., 2014) allow to consider the model suitable for the analysis purpose.

4.2 Experimental design

The first step of the analysis concerned on the quality assessment of the SRPs in terms of rainfall. For that, each SRP has been compared with the daily E-OBS data used as reference. Then, discharge simulations have been performed by running the lumped version of MISDc model with E-OBS dataset (river discharge reference) and with each SRP as input. Specifically, the two following steps have been performed:

1) MISDc model has been calibrated over the entire 2007-2016 period by using as input the mean areal E-OBS rainfall and air temperature data for each basin; these simulated discharge data, $Q_{E\text{-}OBS}$, has been used as benchmark to estimate the accuracy of the selected SRPs for river discharge simulation.
2) MISDc has been run for each basin by using as input the mean areal SRPs and E-OBS air temperature data. In accordance with literature studies (e.g, Thiemig et al., 2013), in these runs the model parameters are calibrated separately for each SRP. The period 2007-2012 is

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions
Open Access

used for the parameter values calibration, whereas the remaining 2013-2016 period is used for the validation; $Q_{E\text{-}OBS}$ is used as reference for parameter values calibration.

The use of $Q_{E\text{-}OBS}$ as benchmark presents three advantages as it allows: 1) to consider a common and extended analysis period for all basins, 2) to consider a common benchmark in evaluating the SRP accuracy both in terms of rainfall and in terms of discharge and, more important, 3) to neglect the uncertainty due to the hydrological model structure in the SRPs comparison.

4.5 Performance metrics

The quality assessment of the different SRPs has been calculated by four continuous dimensionless metrics and three categorical scores. Among the continuous scores, the Pearson correlation coefficient, R, the relative BIAS, rBIAS, the relative root mean square error, RRMSE and the KGE, an index increasingly used in hydrology to measure the goodness-of-fit between simulated and observed data, have been computed between the daily E-OBS and the satellite rainfall data averaged over the area of each basin as follows:

$$R = \frac{\mathrm{Cov}(SRP, P_{ref})}{\sigma_{SPR} \times \sigma_{P_{ref}}} \quad (1)$$

$$rBIAS = \frac{\frac{1}{n}\sum_{i=1}^{n}(SRP_i - P_{ref_i})^2}{\frac{1}{n}\sum_{i=1}^{n}(P_{ref_i})} \quad (2)$$

$$RRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(SRP_i - P_{ref_i})^2}}{\frac{1}{n}\sum_{i=1}^{n}(P_{ref_i})} \quad (3)$$

$$KGE = 1 - \sqrt{(R-1)^2 + \left(\frac{\frac{1}{n}\sum_{i=1}^{n}(SRP_i)}{\frac{1}{n}\sum_{i=1}^{n}(P_{ref_i})} - 1\right)^2 + \left(\frac{\frac{1}{n}\sum_{i=1}^{n}(P_{ref_i}) * \sigma_{SPR}}{\frac{1}{n}\sum_{i=1}^{n}(SRP_i)\ \sigma_{P_{ref}}} - 1\right)^2} \quad (4)$$

where SRP and P_{ref} represent the SRPs and E-OBS rainfall time series; Cov and σ are the covariance and the standard deviation operator, respectively; n corresponds to the length of the time series. R values range from -1 to 1; rBIAS ranges from –∞ to +∞; RRMSE is bounded from 0 to +∞ while KGE varies between –∞ to 1. More R, rBIAs, RRMSE and KGE values goes toward 1, 0, 0,

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences Discussions

1 respectively, higher is the agreement between E-OBS and SRPs. In particular, for KGE, model performance in the range -0.41 < KGE <= 1 indicate that the model outperforms the mean of the observations (Knoben et al., 2019). In addition, for each SRP and for different rainfall thresholds three categorical metrics are evaluated (Chen et al., 2012, Brocca et al., 2014): probability of detection (POD), false alarm ratio (FAR) and threat score (TS) POD reports on the capability of SRP to correctly detect rainfall events, FAR counts the fraction of rainfall events that are actually non-events and TS takes into account the correctly detected, missed rainfall events and false alarms.,. These categorical metrics range from 0 to 1: higher POD and TS along with lower FAR values indicate a better capability of SRPs to detect rainfall events.

To evaluate the suitability of rainfall products for flood modelling, the KGE index between the observed and simulated discharge data has been computed. Specifically, KGE index has been evaluated both between the observed and simulated $Q_{E\text{-}OBS}$ discharge and between $Q_{E\text{-}OBS}$ and the simulated discharge data obtained by using SRPs as input, in order to establish the hydrological performances of E-OBS and SRPs, respectively. Discharge simulations characterized by KGE values greater than 0.5 have been considered good with respect to their ability to reproduce observed discharge time series (Thiemig et al., 2013).

5. RESULTS

The findings of this work for the three SRPs are presented below. The SRP quality has been evaluated first in terms of rainfall and then in terms of discharge. The propagation of the rainfall error into the discharge simulation has been finally investigated.

5.1 Rainfall assessment

The performances of the three SRPs against the E-OBS datasets are illustrated in Figure 2. For sake of brevity, the SRPs performances are presented only for the validation period (2013-2016), but similar findings are obtained in the calibration period (see Table 3). Specifically, rBIAS, R, RRMSE and KGE values are illustrated in the rows of Figure 2 for each study basin, for the three

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

products TMPA, CMOR and SM2R$_{ASCAT}$ in each column. At the top of each plot, the median score value is reported by considering the original spatial coverage of each SRP whereas in Table 3 the performances of the basins whose outlet section is located below/above 50° latitude, i.e. below/above the TMPA coverage, are listed. Already at first glance, it is possible to note that the three products show similar patterns in terms of R and RRMSE whereas the same does not hold for the rBIAS and KGE. By focusing the analysis over the TMPA area, median R (RRMSE) values are equal to 0.626 (1.968), 0.551 (1.969), 0.609 (1.781) for TMPA, CMOR and SM2R$_{ASCAT}$, respectively. Higher/lower R/RRMSE values are obtained in Central Europe; the opposite is observed in the Mediterranean area. The rBIAS is low for TMPA and SM2R$_{ASCAT}$, with median values equal to -0.127 and 0.081, respectively, whereas CMOR show a clear underestimation of the daily rainfall data over the entire European area. In terms of KGE, TMPA presents higher values with respect to the other two products above all over the basins whose outlet section is located between 40° and 50° latitude. Median KGE value for TMPA is equal to 0.516; this value reduces of about 24% and 42% for SM2R$_{ASCAT}$ and CMOR, respectively.

Outside the TMPA area and until 60° latitude, CMOR and SM2R$_{ASCAT}$ show quite similar performances in terms of R and RRMSE, while SM2R$_{ASCAT}$ outperforms CMOR in terms of rBIAS and KGE. Likely due to soil freezing and snow presence, the performances of SM2R$_{ASCAT}$ decrease in terms of R, rBIAS and KGE moving toward northern Europe (Brocca et al., 2019).

Results in terms of categorical metrics are summarized in Figure S1, where POD (first row), FAR (second row) and TS (third row) have been computed for the validation period for three rainfall thresholds (0.5, 5, and 10 mm/day) in order to assess the capability of SRPs to detect low to high rainfall events. Numbers at the top of each plot represent the median score value obtained by considering the original spatial coverage of each product. For all the three metrics and for moderate to heavy rainfall events, TMPA presents the highest values of POD (median values equal to 0.500/0.415 for moderate/high events) and TS (median values equal to 0.368/0.288 for moderate/high events), overperforming the other two products. Conversely, SM2R$_{ASCAT}$ show higher ability to detect small and moderate rainfall events with performances in terms of TS slightly lower than the ones of TMPA product.

5.2 Discharge assessment

Prior to assess the hydrological performances of the satellite rainfall data, MISDc model has been run with the E-OBS rainfall data as input to obtain $Q_{E\text{-}OBS}$, the benchmark river discharge data. The results of this calibration, carried out for the entire observation period (2007-2016), are good as illustrated in Figure 1b: the median KGE value obtained for the European area is equal to 0.768 (0.770 over the TMPA area). This ensures the good quality of $Q_{E\text{-}OBS}$ data that are assumed as benchmark for the successive analysis. Hereinafter, the hydrological performance has been assessed in terms of KGE with respect to $Q_{E\text{-}OBS}$, with values higher than 0.5 considered as good.

Depending on the product, SRPs show different hydrological performances as illustrated in Figure 3 for the validation period and in Table 4 for both the calibration and the validation periods. At the top of each plot in Figure 3, the median KGE value, averaged over the spatial coverage of each product, is reported whereas in Table 4 the performances of the basins whose outlet section is located below/above 50° latitude are listed. In addition, in Table 4 the percentage of basins showing KGE values higher than 0.5 is computed.

By averaging the performances over the spatial coverage of each product, median KGE values range from 0.279 to 0.722 for CMOR and $SM2R_{ASCAT}$, respectively, in the calibration period and from -0.090 to 0.569 for the same products in the validation period (Figure 3). The percentage of the basins showing KGE values higher than 0.5, is 88% and 18% for CMOR and $SM2R_{ASCAT}$, respectively, whereas the same percentage drop in the validation period up to about 39% and 3% for the same products. TMPA is in the middle between the two products in terms of performances; the percentage of basins with good hydrological performances is similar to the one of $SM2R_{ASCAT}$.

Similar findings hold if the comparison is carried out below the TMPA area (see Table 4): poor results are obtained by CMOR during the validation period (median KGE<0; only 2.6% show KGE higher than 0.5), whereas $SM2R_{ASCAT}$ outperforms TMPA in both periods. In particular, during the

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

validation, period a median KGE value equal to 0.580 is obtained for $SM2R_{ASCAT}$ against a value equal to 0.428 for TMPA. Moreover, by comparing $SM2R_{ASCAT}$ against TMPA in terms of basins with KGE greater than 0.5, the ratio is two to one, i.e., 64% of basins show good hydrological performances when forced with $SM2R_{ASCAT}$ with respect to 39% for TMPA. The lowest performances for both products are obtained over southern Spain and northern Italy. Conversely, the basins located over northern Spain and central Europe show a better agreement with respect to $Q_{E\text{-}OBS}$ benchmark data, above all when $SM2R_{ASCAT}$ is used as rainfall input. The performances of $SM2R_{ASCAT}$ remain good also when the analysis is extended above the TMPA area, with a median KGE higher than 0.5 (Table 4). This is the first notably result of the paper, i.e., among the SRPs available in near real time, there are some products that can be reasonably used to force a hydrological model in order to obtain reliable discharge data over Europe. However, a question remains: why do some SRPs perform better than others? Is it possible to find a rainfall score to select a priori the best SRP to obtain reliable discharge simulations? Is there any link between rainfall and discharge performances? The answer to these questions is given in the next paragraph where the rainfall performances are compared with the discharge performances.

5.3 Rainfall vs discharge performances

By comparing the patterns of Figure 2 against the patterns of Figure 3, some insights about the link between the rainfall accuracy and the hydrological performance can be noted: the basins with the highest RRMSE (e.g., in the Mediterranean area and in particular in southern Spain and northern Italy) correspond to basins with poorer hydrological performances (KGE<0.4). In addition, as occurs for CMOR product, high rBIAS values (both negative or positive) produce negative KGE values. Interestingly, R and KGE rainfall scores seem to be weakly linked to the hydrological performances. Finally, no clear link can be highlighted between KGE of discharge and the rainfall categorical scores as for instance, the high/low values of $SM2R_{ASCAT}$ in terms of FAR/TS do not explain the higher performances of this products in terms of discharge (see Figure 3 against Figure S1).

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

To better investigate these relationships, the scatterplots of Figure 4 and Figure S2 (in the supplementary material) have been constructed for the continuous and categorical scores, respectively. For each basin and for each SRP, the rainfall scores (x-axis) are plotted against the KGE values (y-axis), resulting in a large ensemble of points spread out in the full range of rainfall/discharge scores without any apparent relationship. The unique remark from Figure 4 is that CMOR shows higher rBIAS and lower KGE values with respect to the other two products; rBIAS of $SM2R_{ASCAT}$ varies near zero and, in terms RRMSE, $SM2R_{ASCAT}$ is characterized by a reduced range of variability, (i.e., most of the $SM2R_{ASCAT}$ data are characterized by RRMSE ranging from 1.5 and 2.5) with respect to the other two products. By looking at the categorical scores (Figure S2), the three products show a similar variability range for moderate to high rainfall events whereas some differences are evident for low rainfall events, that however should have a minor impact on flood modelling. In particular, $SM2R_{ASCAT}$ tend to have higher POD values for rainfall threshold equal to 0.5, due to the tendency of the product to overestimate the rainfall occurrence (Brocca et al., 2019).

To extract useful information from Figure 4 and Figure S2, the scores obtained separately for each product have been grouped and the KGE data points have been binned into uniform ranges (with step 0.1) of rainfall scores. The median KGE, and the 25th and 75th percentiles of KGE values, have been computed for each rainfall score within each bin. The white dots in Figure 4 and Figure S2 represent, for each bin of each rainfall score, the median KGE value, the two ends of the black lines in the same figure represent the 25th and 75th percentile of the KGE data points. By looking at the boxplots so obtained, some insights already anticipated by inspecting Figure 2 versus Figure 3 for the continuous scores can be confirmed: SRP hydrological performances strongly decrease by increasing the absolute value of rBIAS, |rBIAS|, and the RRMSE values (Figure 4a and b) whereas KGE of discharge slightly increase with R and KGE of rainfall (Figure 4c and d). If these relationships have reflected the expectations, the same did not occur for the categorical scores. Indeed, except for the rainfall threshold equal to 10 mm/day, the relationships between KGE of

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

discharge and the categorical scores of small and moderate rainfall are different (and sometimes inverse) from what can be expected. This could be due to the lowest impact of small/moderate rainfall events on flood generation. Then, focusing the attention only on high rainfall events, seems that KGE of discharge slightly increase with POD whereas a stronger link can be noted between KGE of discharge and FAR/TS.

The findings obtained so far become even more interesting if the following question is posed: for which values of rainfall scores is it possible to obtain good results in terms of river discharge simulation (i.e., KGE>0.5 evaluated on the discharge data)? The straight grey line in Figure 4 (and Figure S2), drawn for a threshold value of KGE equal to 0.5, helps us to answer the question suggesting that good hydrological performances can be obtained for SRPs characterized by rBIAS values close to 0 and small RRMSE scores. Conversely, R and KGE of rainfall seem to have a small impact on KGE of discharge as for a large range of R and KGE values (from 0.5 to 0.8 and from 0.4 to 0.8, respectively), it is possible to obtain high KGE values. Similar conclusions hold for the categorical scores evaluated for heavy rainfall events: the higher capability of SRPs to detect rainfall events does not affect the hydrological performances, i.e., it is possible to obtain KGE of discharge higher than 0.5 for a large range of POD, FAR and TS values. A further question remains: how small/large should be the rainfall scores to obtain good hydrological performances, i.e., KGE greater than 0.5? In particular, what about rBIAS and RRMSE that seem to have a stronger link with the hydrological performances?

The boxplot of Figure 5a shows the hydrological performances that have been obtained during the validation period by the three SRPs without any constrain on the rainfall scores. In order to consider always the same number of basins for all the products, the area of analysis is cut below the TMPA area and a median KGE value equal to 0.342 is obtained for the 889 basins. According to Table 4, nearly 35% of the basins show KGE greater than 0.5. If the absolute value of rBIAS, |rBIAS|, is constrained to values lower than 0.2 (Figure 5b), the median KGE value over the 400 basins that fulfil the criteria is equal to 0.525. As shown in Figure 5c, a constrain on RRSME lower than 2 is

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions

not enough to assure good hydrological performances (median KGE lower than 0.5) whereas if a combination of the two rainfall scores is considered, the threshold on KGE>0.5 is exceeded by nearly 75% of the basins fulfilling the criteria (see first boxplot of Figure 5d). In other words, it means that nearly less than 25% of the basins fulfilling the criteria shows low performance (first boxplot of Figure 5d). Alternatively, less than 25% of basins not fulfilling the rainfall constrains shows good hydrological performances (see second boxplot of Figure 5d).

For completeness of the work, a figure similar to Figure 5 has been added in the Supplementary material (Figure S3) for the other rainfall scores (R, KGE, POD, FAR and TS and relative combinations), but no one of the shown rainfall constrain can be considered satisfactory for the analysis purpose. Indeed, no one of the rainfall constrain in Figure S3 allows a clear separation between basins fulfilling/not fulfilling the criteria with a corresponding increase of KGE on discharge.

6. DISCUSSION

The findings of Figure 4 and Figure 5 draw some interesting conclusions about the main research question of the paper, i.e., for which rainfall metric it is possible to obtain good results in terms of river discharge simulation. rBIAS along with RRMSE seem to be the most appropriate error metrics to be used in conjunction to select the best performing SRP for flood modelling. With respect to bias, the finding is in line with literature studies. For instance, Maggioni et al., (2013) showed that bias can double from rainfall to runoff consistently from small to large basins. Conversely, no suggestions can be found with respect to RRMSE or R metrics to characterize the SRPs potentiality in terms of flood modelling. In the scientific literature, we have found thresholds on metric scores to express the quality of SRPs in terms of rainfall. In particular, some authors considered an R value equal or greater than 0.7 (Condom et al., 2011), a normalized RMSE values less than or equal to 0.5 (Adeyewa and Nakamura, 2003, Condom et al., 2011; Satgé et al., 2016; Shrestha et al., 2017) and

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

bias ranging from $-10\% \leq$ bias $\leq 10\%$ (Brown, 2006, Yang and Luo, 2014) to be associated with good satellite rainfall performances, but without a reference to justify these numbers.

Specifically, in this study we have found that constraining |rBIAS| to values lower than 0.2 and RRMSE to values lower than 2, good hydrological performances are assured for nearly 75% of the basins fulfilling the criteria. The remaining percentage of basins for which the rainfall/discharge performance relationship is not satisfied highlights that it is not straightforward to find such kind of relationships as errors on rainfall and discharge data used as benchmark as well as the hydrological model recalibration could influence the analysis. These findings corroborate those obtained by Qi et al. (2016), stating that a good discharge simulation is a results of a good combination between a rainfall product and an hydrological model, and the selection of the most accurate rainfall product alone does not guarantee the most accurate hydrological performances.

7. CONCLUSIONS

This study represents the most comprehensive European-scale evaluation to date of satellite rainfall products (SRPs). Three different near real time SRPs are used to force a lumped hydrological model over 1318 basins throughout Europe. The results can be summarized as follows:

1. In terms of rainfall accuracy, the three SRPs show similar patterns in terms of R and RRMSE whereas the same does not hold for the rBIAS. For the three products, higher/lower R/RRMSE values are obtained in Central Europe; the opposite, is observed in the Mediterranean area. The rBIAS is low for TMPA and SM2R$_{ASCAT}$, whereas CMOR shows a clear underestimation of the daily rainfall data over the entire European area.
2. Among the SRPs available in near real time, there are some SRPs that can be reasonably used to force a hydrological model in order to obtain reliable discharge data over Europe. In particular, SM2R$_{ASCAT}$ is the best performing product for flood simulation across Europe (even at high latitudes).

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions

3. There is a link between rainfall accuracy and discharge performance. In particular, by constraining |rBIAS| to values lower than 0.2 and RRMSE to values lower than 2, good hydrological performances are assured for almost 75% of the basins fulfilling these criteria.

Overall, we believe the results obtained from this study provide very useful information about the application of SRPs to simulate river discharge at basin scale. In particular, for the first time, this work has addressed the topic of providing quantitative guidelines in the use of SRPs for near real time hydrological applications.

Nevertheless, some limitations can be recognized in the analysis. One of the main limitations lies in the use of only one hydrological model for flood simulation. In this respect, further analysis with multiple hydrological models will be carried out to better investigate the link between rainfall, hydrological model and discharge performances. In addition, in future researches the rainfall metrics ranges here defined will be checked also with the use of different satellite rainfall products (e.g., the Global Precipitation Measurement, GPM, Huffmann et al., 2018) and in different regions worldwide. In particular, the extension of the analysis over different regions in the world could allow to explore the connection between rainfall accuracy and discharge performances as a function of additional criteria such as climate type, soil characteristics and terrain features (topography).

Another limitation of the study relies in having considered only one performance score for the discharge. Indeed, as the main purpose of this study has been to reproduce the entire discharge time series, any special attention to high/low flows was not paid. In a further analysis, a more comprehensive study could consider a larger set of discharge metrics to better address the SRP selection. Finally, the results of this study are likely sensitive to the quality of data taken as "reference", i.e., the E-OBS datasets, used as benchmark to evaluate the performances of SRPs both in terms of rainfall and, through the hydrological model, in terms streamflow.

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Despite the aforementioned limitations this study, contributing in the purpose of better understand the propagation of the satellite rainfall error to streamflow simulations, could be very helpful for data users facing the selection of the best satellite rainfall for hydrological applications.

Author contribution

S.C. collected discharge data, performed the analysis and wrote the manuscript. L.C. collected satellite rainfall data; I.M. performed the basins delineation; C.M. and L.B. contributed on the supervision of the work. All authors discussed the results and contributed to the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

The authors wish to thank the Global Runoff Data Centre (GRDC) for providing most of the streamflow data throughout Europe and the E-OBS dataset from the EU-FP6 project UERRA (http://www.uerra.eu) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://www.ecad.eu). The authors gratefully acknowledge support from EUMETSAT through the "Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF)" CDOP3 (EUM/C/85/16/DOC/15) and the Global SM2RAIN project (contract n° EUM/CO/17/4600001981/BBo), and from ESA through SMOS+rainfall project (contract n° 4000114738/15/I-SBo).

REFERENCE

Adeyewa, Z. D., & Nakamura, K. (2003). Validation of TRMM radar rainfall data over major climatic regions in Africa. Journal of Applied Meteorology, 42(2), 331-347.

Artan, G., Gadain, H., Smith, J. L., Asante, K., Bandaragoda, C. J., & Verdin, J. P. (2007). Adequacy of satellite derived rainfall data for stream flow modeling. Natural Hazards, 43(2), 167-185.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I., Weedon, G. P., ... & Wood, E. F. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrology and Earth System Sciences, 21(12), 6201-6217.

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The relative importance of different flood-generating mechanisms across Europe. Water Resources Research.

Bhuiyan, E., Abul, M., Nikolopoulos, E. I., Anagnostou, E. N., Polcher, J., Albergel, C., ... & Munier, S. (2019). Assessment of precipitation error propagation in multi-model global water resource reanalysis. Hydrology and Earth System Sciences, 23(4), 1973-1994.

Bisselink, B., Zambrano-Bigiarini, M., Burek, P., & De Roo, A. (2016). Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions. Journal of Hydrology: Regional Studies, 8, 112-129.

Bitew, M. M., & Gebremichael, M. (2011). Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. Water Resources Research, 47(6).

Brocca, L., Melone, F., Moramarco, T. (2011). Distributed rainfall-runoff modelling for flood frequency estimation and flood forecasting. Hydrological Processes, 25 (18), 2801-2813, doi:10.1002/hyp.8042.

Brocca, L., Liersch, S., Melone, F., Moramarco, T., Volk, M. (2013a). Application of a model-based rainfall-runoff database as efficient tool for flood risk management. Hydrology and Earth System Sciences, 17, 3159-3169, doi:10.5194/hess-17-3159-2013.

Brocca, L., Moramarco, T., Dorigo, W., & Wagner, W. (2013b). Assimilation of satellite soil moisture data into rainfall-runoff modelling for several catchments worldwide. In 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS (pp. 2281-2284). IEEE.

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

Brocca, L., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B., Wagner, W. (2019). SM2RAIN-ASCAT (2007-2018): global daily satellite rainfall from ASCAT soil moisture. Earth System Science Data Discussion, under review, doi:10.5194/essd-2019-48. https://doi.org/10.5194/essd-2019-48.

Brown, J. E. (2006). An analysis of the performance of hybrid infrared and microwave satellite precipitation algorithms over India and adjacent regions. Remote Sensing of Environment, 101(1), 63-81.

Camici, S., Brocca, L., Melone, F., Moramarco, T. (2014). Impact of climate change on flood frequency using different climate models and downscaling approaches. Journal of Hydrologic Engineering, 19(8), 04014002, doi:10.1061/(ASCE)HE.1943-5584.0000959. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000959

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

Casse, C., Gosset, M., Peugeot, C., Pedinotti, V., Boone, A., Tanimoun, B. A., & Decharme, B. (2015). Potential of satellite rainfall products to predict Niger River flood events in Niamey. Atmospheric Research, 163, 162-176.

Chen, F., Crow, W., & Holmes, T. R. (2012). Improving long-term, retrospective precipitation datasets using satellite surface soil moisture retrievals and the soil moisture analysis rainfall tool. Journal of Applied Remote Sensing, 6(1), 063604.

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Chintalapudi, S., Sharif, H., & Xie, H. (2014). Sensitivity of distributed hydrologic simulations to ground and satellite based rainfall products. Water, 6(5), 1221-1245.

Cislaghi, A., Masseroni, D., Massari, C., Camici, S., Brocca, L. (2019). Combining rainfall-runoff model and regionalization approach for flood and water resource assessment in the western Po-Valley (Italy). Hydrological Science Journal, in press, doi:10.1080/02626667.2019.1690656.

Condom, T., Rau, P., & Espinoza, J. C. (2011). Correction of TRMM 3B43 monthly precipitation data over the mountainous areas of Peru during the period 1998–2007. Hydrological Processes, 25(12), 1924-1933.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J., & Jones, P. D. (2018). An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. Journal of Geophysical Research: Atmospheres, 123(17), 9391-9409.

Demaria, E. M., Nijssen, B., Valdés, J. B., Rodriguez, D. A., & Su, F. (2014). Satellite precipitation in southeastern South America: how do sampling errors impact high flow simulations?. International journal of river basin management, 12(1), 1-13.

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

Elgamal, A., Reggiani, P., & Jonoski, A. (2017). Impact analysis of satellite rainfall products on flow simulations in the Magdalena River Basin, Colombia. Journal of Hydrology: Regional Studies, 9, 85-103.

Gebregiorgis, A. S., Tian, Y., Peters-Lidard, C. D., & Hossain, F. (2012). Tracing hydrologic model simulation error as a function of satellite rainfall estimation bias components and land use and land cover conditions. Water Resources Research, 48(11).

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), 80-91.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., & New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. Journal of Geophysical Research: Atmospheres, 113(D20).

Hong, Y., Hsu, K. L., Moradkhani, H., & Sorooshian, S. (2006). Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. Water resources research, 42(8).

Hossain, F., & Anagnostou, E. N. (2006). A two-dimensional satellite rainfall error model. IEEE transactions on geoscience and remote sensing, 44(6), 1511-1522.

Hossain, F., & Huffman, G. J. (2008). Investigating error metrics for satellite rainfall data at hydrologically relevant scales. Journal of Hydrometeorology, 9(3), 563-575.

Huffman, G. J., Adler, R. F., Bolvin, D. T., & Nelkin, E. J. (2010). The TRMM multi-satellite precipitation analysis (TMPA). In Satellite rainfall applications for surface hydrology (pp. 3-22). Springer, Dordrecht.

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E., and Xie, P. (2018). Algorithm Theoretical Basis Document (ATBD) Version 4.5. NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) NASA.

Jiang, D., & Wang, K. (2019). The Role of Satellite Remote Sensing in Improving Simulated Streamflow: A Review. Water, 11(8), 1615.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. Journal of Hydrometeorology, 5(3), 487-503.

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of Hydrology, 424, 264-277.

Klok, E. J., and Tank, A. K. (2009). Updated and extended European dataset of daily climate observations. International Journal of Climatology, 29(8), 1182-1191.

Knoben, W. J. M., Freer, J. E., and Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-327, in review.

Lu, D.; Yong, B. (2018) Evaluation and Hydrological Utility of the Latest GPM IMERG V5 and GSMaP V7 Precipitation Products over the Tibetan Plateau. Remote Sens., 10, 2022.

Maggioni, V., Reichle, R. H., & Anagnostou, E. N. (2011). The effect of satellite rainfall error modeling on soil moisture prediction uncertainty. Journal of Hydrometeorology, 12(3), 413-428.

Maggioni, V., Vergara, H. J., Anagnostou, E. N., Gourley, J. J., Hong, Y., & Stampoulis, D. (2013). Investigating the applicability of error correction ensembles of satellite rainfall products in river flow simulations. Journal of Hydrometeorology, 14(4), 1194-1211.

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

Massari, C., Brocca, L., Ciabatta, L., Moramarco, T., Gabellani, S., Albergel, C., de Rosnay, P., Puca, S., Wagner, W. (2015). The use of H-SAF soil moisture products for operational hydrology: flood modelling over Italy. Hydrology, 2(1), 2-22, doi:10.3390/hydrology2010002. http://dx.doi.org/10.3390/hydrology2010002

Massari, C., Camici, S., Ciabatta, L., Brocca, L. (2018). Exploiting satellite surface soil moisture for flood forecasting in the Mediterranean area: state update versus rainfall correction. Remote Sensing, 10(2), 292, doi: 10.3390/rs10020292.

Masseroni, D., Cislaghi, A., Camici, S., Massari, C., & Brocca, L. (2016). A reliable rainfall–runoff model for flood forecasting: review and application to a semi-urbanized watershed at high flood risk in Italy. Hydrology Research, 48(3), 726-740.

Mei, Y., Nikolopoulos, E., Anagnostou, E., Zoccatelli, D., & Borga, M. (2016). Error analysis of satellite precipitation-driven modeling of flood events in complex alpine terrain. Remote Sensing, 8(4), 293.

Mei, Y., Anagnostou, E. N., Shen, X., & Nikolopoulos, E. I. (2017). Decomposing the satellite precipitation error propagation through the rainfall-runoff processes. Advances in water resources, 109, 253-266.

Montani, A., Cesari, D., Marsigli, C., & Paccagnella, T. (2011). Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. Tellus A: Dynamic Meteorology and Oceanography, 63(3), 605-624.

Mouratidis, A., & Ampatzidis, D. (2019). European Digital Elevation Model Validation against Extensive Global Navigation Satellite Systems Data and Comparison with SRTM DEM and ASTER GDEM in Central Macedonia (Greece). ISPRS International Journal of Geo-Information, 8(3), 108. https://doi.org/10.3390/ijgi8030108

Mugnai, A., Casella, D., Cattani, E., Dietrich, S., Laviola, S., Levizzani, V., ... & Biron, D. (2013). Precipitation products from the hydrology SAF. Natural Hazards and Earth System Sciences, 13(8), 1959-1981.

Nikolopoulos, E. I., Anagnostou, E. N., Hossain, F., Gebremichael, M., & Borga, M. (2010). Understanding the scale relationships of uncertainty propagation of satellite rainfall through a distributed hydrologic model. Journal of Hydrometeorology, 11(2), 520-532.

Nikolopoulos, E. I., E. N. Anagnostou, and M. Borga (2012). Using high-resolution satellite rainfall products to simulate a major flash flood event in northern Italy. J. Hydrometeor., 14, 171–185.

Pakoksung, K., and Takagi, M. (2016). Effect of satellite based rainfall products on river basin responses of runoff simulation on flood event. Modeling Earth Systems and Environment, 2(3), 143.

Pan, M., Li, H., & Wood, E. (2010). Assessing the skill of satellite-based precipitation estimates in hydrologic applications. Water Resources Research, 46(9).

Qi, W., Zhang, C., Fu, G., Sweetapple, C., & Zhou, H. (2016). Evaluation of global fine-resolution precipitation products and their uncertainty quantification in ensemble discharge simulations. Hydrology and Earth System Sciences, 20(2), 903-920.

Quintero, F., Krajewski, W. F., Mantilla, R., Small, S., & Seo, B. C. (2016). A spatial–dynamical framework for evaluation of satellite rainfall products for flood prediction. Journal of Hydrometeorology, 17(8), 2137-2154.

Ren, P., Li, J., Feng, P., Guo, Y., & Ma, Q. (2018). Evaluation of multiple satellite precipitation products and their use in hydrological modelling over the Luanhe River basin, China. Water, 10(6), 677.

Ricciardelli, E.; Di Paola, F.; Gentile, S.; Cersosimo, A.; Cimini, D.; Gallucci, D.; Geraldi, E.; Larosa, S.; Nilo, S.T.; Ripepi, E.; et al. (2018). Analysis of Livorno Heavy Rainfall Event: Examples of Satellite Observation Techniques in Support of Numerical Weather Prediction. Remote Sens., 10, 1549.

Satgé, F., Bonnet, M. P., Gosset, M., Molina, J., Lima, W. H. Y., Zolá, R. P., ... & Garnier, J. (2016). Assessment of satellite rainfall products over the Andean plateau. Atmospheric research, 167, 1-14.

Satgé, F., Ruelland, D., Bonnet, M. P., Molina, J., & Pillco, R. (2019). Consistency of satellite precipitation products in space and over time compared with gauge observations and snow-hydrological modelling in the Lake Titicaca region. Hydrology and Earth System Sciences, 23(1), 595-619.

Serpetzoglou, E., Anagnostou, E. N., Papadopoulos, A., Nikolopoulos, E. I., & Maggioni, V. (2010). Error propagation of remote sensing rainfall estimates in soil moisture prediction from a land surface model. Journal of Hydrometeorology, 11(3), 705-720.

Shah, H. L., & Mishra, V. (2016). Uncertainty and bias in satellite-based precipitation estimates over Indian subcontinental basins: Implications for real-time streamflow simulation and flood prediction. Journal of Hydrometeorology, 17, 615–636.

Shrestha, N. K., Qamer, F. M., Pedreros, D., Murthy, M. S. R., Wahid, S. M., & Shrestha, M. (2017). Evaluating the accuracy of Climate Hazard Group (CHG) satellite rainfall estimates for precipitation based drought monitoring in Koshi basin, Nepal. Journal of Hydrology: Regional Studies, 13, 138-151.

Tapiador, F.J.; Navarro, A.; Levizzani, V.; García-Ortega, E.; Huffman, G.J.; Kidd, C.; Kucera, P.A.; Kummerow, C.D.; Masunaga, H.; Petersen, W.A.; et al. (2017). Global precipitation measurements for validating climate models. Atmos. Res., 197, 1–20.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., & De Roo, A. (2013). Hydrological evaluation of satellite rainfall estimates over the Volta and Baro-Akobo Basin. Journal of Hydrology, 499, 324-338.

Valdés-Pineda, R., Demaría, E. M., Valdés, J. B., Wi, S., & Serrat-Capdevila, A. (2016). Bias correction of daily satellite rainfall estimates for hydrologic forecasting in the Upper Zambezi, Africa. Hydrol. Earth Syst. Sci, 1, 28.

Vergara, H., Hong, Y., Gourley, J. J., Anagnostou, E. N., Maggioni, V., Stampoulis, D., & Kirstetter, P. E. (2014). Effects of resolution of satellite rainfall estimates on hydrologic modeling skill at different scales. Journal of Hydrometeorology, 15(2), 593-613.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., ... & Komma, J. (2013). The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. Meteorologische Zeitschrift, 22(1), 5-33.

Yang, Y., Luo, Y. (2014). Evaluating the performance of remote sensing precipitation products CMORPH, PERSIANN, and TMPA, in the arid region of northwest China. Theoretical and applied climatology, 118(3), 429-445.

Yilmaz, K. K., Hogue, T. S., Hsu, K. L., Sorooshian, S., Gupta, H. V., & Wagener, T. (2005). Intercomparison of rain gauge, radar, and satellite precipitation estimates with emphasis on hydrologic forecasting. Journal of Hydrometeorology, 6(4), 497-517.

Yong, B., Ren, L. L., Hong, Y., Wang, J. H., Gourley, J. J., Jiang, S. H., ... & Wang, W. (2010). Hydrologic evaluation of Multisatellite Precipitation Analysis standard precipitation products in basins beyond its inclined latitude band: A case study in Laohahe basin, China. Water Resources Research, 46(7).

Zappa, M., Rotach, M. W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., ... & Jaun, S. (2008). MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. Atmospheric Science Letters, 9(2), 80-87.

Zhu, Q., Xuan, W., Liu, L., & Xu, Y. P. (2016). Evaluation and hydrological application of precipitation estimates derived from PERSIANN-CDR, TRMM 3B42V7, and NCEP-CFSR over humid regions in China. Hydrological processes, 30(17), 3061-3083.

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions

Table 1. Main characteristics of the study basins clustered according to the latitude coordinate of the outlet section.

#	latitude	Number of basins	Median Area (km^2)	Median length of available discharge data after 2007 (years)
1	35°- 50°	889	800	8
2	50°- 60°	288	960	7
3	> 60°	141	2484	8

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions

Table 2. Main characteristics of the datasets used in this study.

#	Satellite-only rainfall datasets	Spatial/ temporal resolution	Spatial coverage	Time period
1	TMPA RT (3B42RT V7)	0.25° / 3-hour	±50° north-south latitude band	2000 – 2018
2	CMORPH	0.25° /3-hour	±60° north-south latitude band	1998 – 2018
3	$SM2R_{ASCAT}$	0.25° / 24-hour	global, over land	2007 – 2018
#	**Large scale gauge based rainfall dataset**	**Spatial/ temporal resolution**	**Coverage**	**Time period**
1	E-OBS	0.22° / 24-hour	Europe	1950 – 2018
#	**Gauge based discharge dataset**	**Spatial/ temporal resolution**	**Coverage**	**Time period**
1	European daily dataset	1318 sites/daily	Europe	1900 - 2016

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Table 3. Performance scores for rainfall (in terms of rBIAS, R and RRMSE) time series computed during the calibration (in italic) and the validation periods. Rainfall performances are evaluated with respect to E-OBS rainfall data and distinguished between basins whose outlet section is below or above 50° latitude.

Rainfall performances

Score / **Product**	**rBIAS**	**R**	**RRMSE**	**KGE**	**rBIAS**	**R**	**RRMSE**	**KGE**
	below TMPA area (latitude <50°)				above TMPA area (latitude >=50°)			
TMPA	-0.127 *(-0.095)*	0.626 *(0.619)*	1.968 *(1.978)*	0.516 *(0.533)*	---	---	---	---
CMOR	-0.462 *(-0.406)*	0.551 *(0.576)*	1.969 *(1.974)*	0.299 *(0.375)*	-0.635 *(-0.618)*	0.544 *(0.562)*	1.607 *(1.621)*	0.114 *(0.147)*
$SM2R_{ASCAT}$	0.081 *(0.084)*	0.609 *(0.595)*	1.781 *(1.805)*	0.393 *(0.436)*	-0.086 *(-0.080)*	0.572 *(0.548)*	1.477 *(1.514)*	0.331 *(0.372)*

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Table 4. Median KGE index computed by comparing $Q_{E\text{-}OBS}$ simulated data against simulated discharge data obtained by forcing MISDc hydrological model with satellite (TMPA, CMOR, $SM2R_{ASCAT}$) rainfall data. Percentage of the basins showing KGE values higher than 0.5 is also listed. Performances and percentages are averaged over different spatial windows: the original spatial coverage of the product and below/above the TMPA area (latitude ±50°).

	KGE					
	Spatial coverage of the product		below TMPA area (latitude <50°)		above TMPA area (latitude >=50°)	
Score / **Product**	**cal**	**val**	**cal**	**val**	**cal**	**val**
TMPA	0.692	0.428	0.692	0.428	---	---
CMOR	0.279	-0.090	0.324	-0.014	0.201	-0.248
$SM2R_{ASCAT}$	0.722	0.569	0.751	0.580	0.670	0.539
	% of basins with KGE>0.5					
TMPA	87.9	38.6	87.9	38.6	---	---
CMOR	17.5	2.40	21.6	2.60	4.90	1.80
$SM2R_{ASCAT}$	87.6	61.7	92.6	64.0	77.2	56.9
Average	**64.4**	**34.2**	**67.4**	**35.1**	**41.1**	**29.4**

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

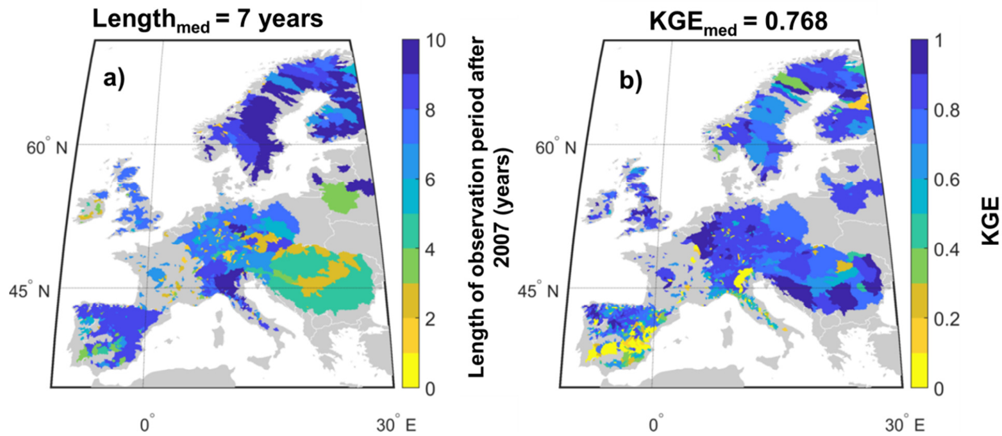
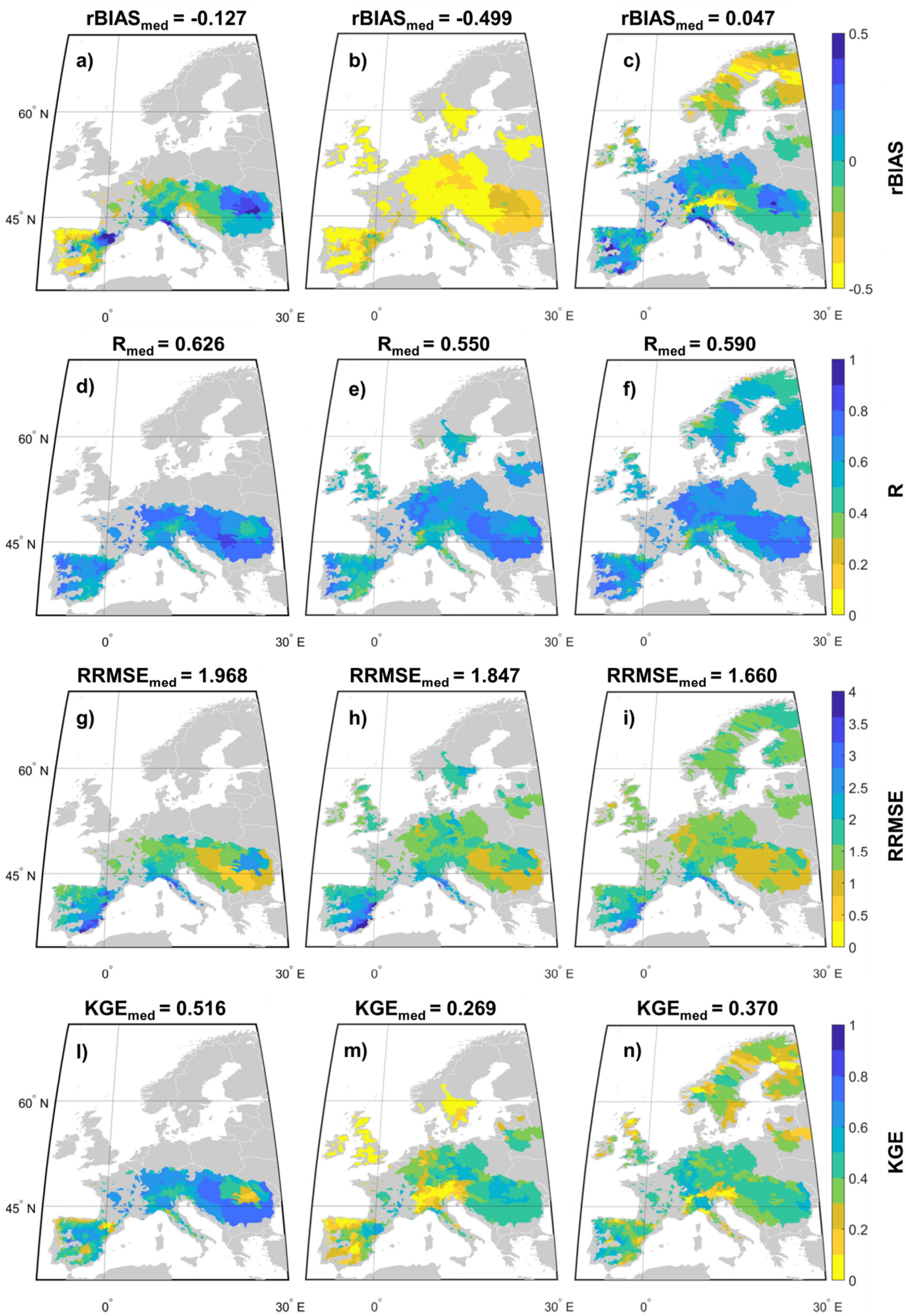


Figure 1. Location of study basins and a) length of discharge observation period after 2007; b) KGE index obtained by comparing observed against modelled discharge data over the period 2007-2016. Modelled data have been obtained by using E-OBS rainfall dataset as input to MISDc model.

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions



725

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Discussions

Figure 2. Performances of satellite rainfall during the validation period in terms of rBIAS (first row), R (second row), RRMSE (third row), KGE (fourth row) over the study basins, for the three products TMPA (first column), CMOR (second column) and $SM2R_{ASCAT}$ (third column). Numbers in each plot represent the median score value obtained by considering the original spatial coverage of each product.

https://doi.org/10.5194/hess-2020-31
Preprint. Discussion started: 19 February 2020

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

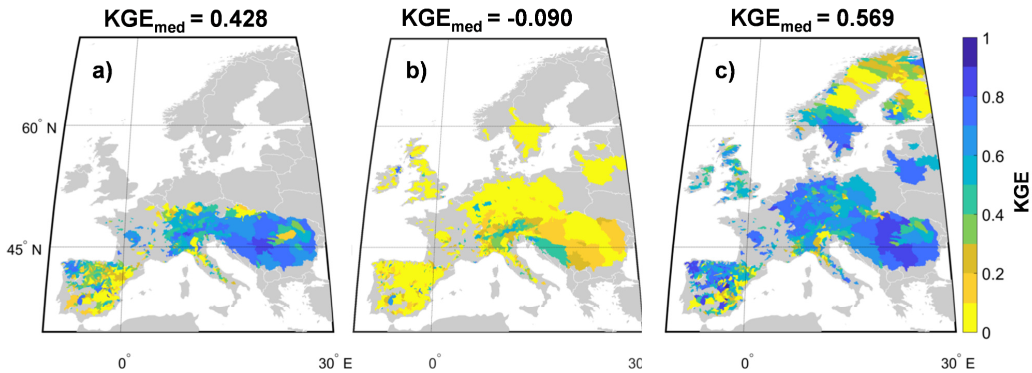


Figure 3. Maps of KGE index obtained by considering a, d) TMPA, b, e) CMORPH and c, f) $SM2R_{ASCAT}$ rainfall dataset in the validation period. In a), b) and c) plots, the median KGE value averaged over the original product coverage is reported.

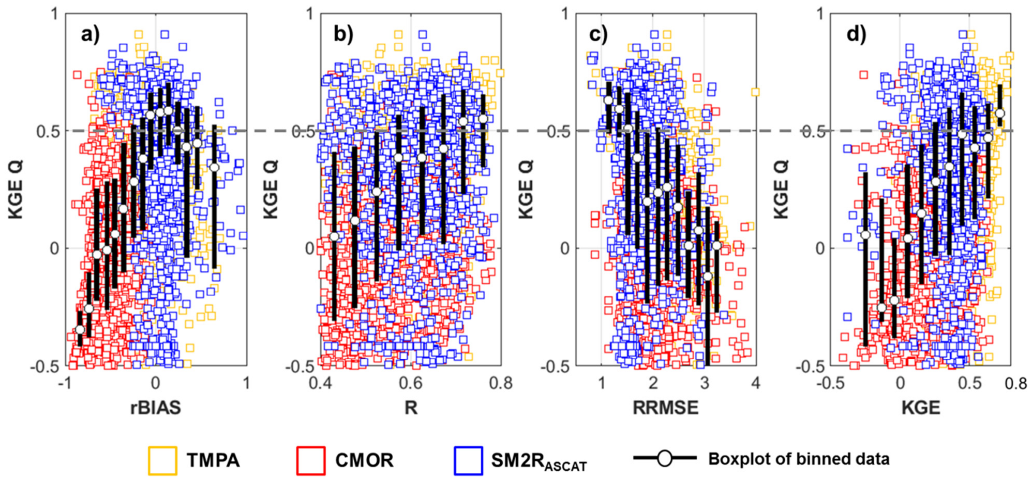


Figure 4. Performances of discharge in terms of KGE against a) relative rainfall bias, rBIAS; b) rainfall correlation, R; c) relative root mean square error of rainfall, RRMSE, d) KGE of rainfall. The scores are evaluated for the validation period (2013-2016) for all the 1318 basins.

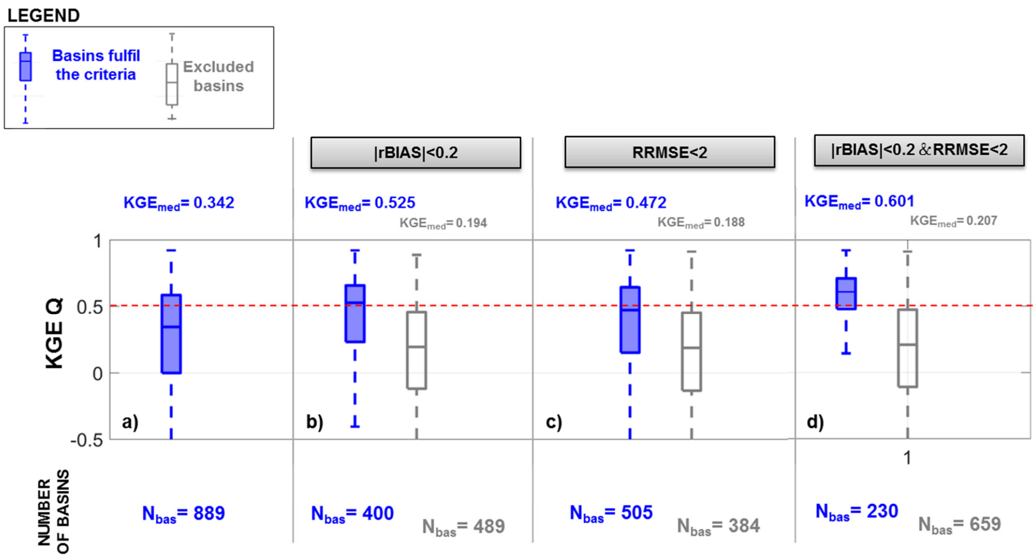


Figure 5. Hydrological performances in terms of KGE values that can be obtained during the validation period by the three satellite rainfall products for all the basins whose outlet section is located below the TMPA area (889), a) without any constrain on the rainfall scores; b) constraining the module of rBIAS to values lower than 0.2; c) constraining RRMSE to values lower than 2; d) constraining the module of rBIAS to values lower than 0.2 and RRMSE to values lower than 2.