Referee 1

General comments

The paper addresses the relevant scientific question of what are the most important metrics to assess the goodness of a SRPs product for hydrological applications. The question as well as the motivation of this work are stated clearly in the context of a comprehensive literature review. The methodology is appropriate to answer the question and the extensive analysis over 1318 basins across Europe defines the main novelty of this paper. Substantial conclusions about the most relevant indexes for assessing the quality of SRPs product for hydrological applications are reached, so overall this is good contribution for the scientific community. However, there are a number of issues that the authors need to address before the paper is accepted for publication.

R: We thank the reviewer for his/her supportive review. In the revised version of the manuscript the following changes have been implemented:

- the title has been changed in “Which rainfall score is more informative about the performance in river discharge simulation? A comprehensive assessment on 1318 basins over Europe”.
- Any reference to flood has been removed and modified with river discharge to highlight that the purpose of the study is to investigate the performances between rainfall and river discharge time series (without specific focus on high and/or low flows);
- a discussion about the quality of the E-OBS rainfall data and the impact of its density network on river discharge simulation has been added;
- to avoid misunderstanding between KGE of rainfall and discharge throughout the manuscript (and in the figures) KGE has been replaced by KGE-P and KGE-Q, to indicate the KGE index referred to precipitation and the one referred to river discharge respectively;
- more information about why only the KGE-Q index has been selected for the analysis has been added to the revised manuscript;
- Tables and figures have been modified according to the reviewer’s suggestions.

Specific comments

1. Line 158-169: the E-OBS dataset is built on a station network with an average station density of 1 in 4000 km^2 and the basin areas range from 200 to 136’000 km^2. Is the E-OBS dataset a reliable benchmark for the smaller basins? Maybe it is worth to discuss this in your discussion section.

R: In the submitted version of the paper, an error occurred in the definition of station used within the E-OBS datasets. Indeed, 2316 stations (i.e., equivalent on average to a density of 1 station every 4000 km^2) is referred to the first versions of E-OBS whereas in the version 17 (used in the manuscript) the number of stations increased up to 9618 (equivalent on average to a density of 1 station every 1000 km^2). However, as correctly raised by the reviewer even the E-OBS density network referred to version 17 could be too low to correctly represent the rainfall spatial variability over small basins. This, in turn could affect the river discharge simulation. To consider this aspect, it has been verified that 1) for all the analysed basins the KGE-Q values obtained by the calibration of the model by using E-OBS dataset as input were greater than -0.41, i.e., the model improves upon the mean flow benchmark 2) no relationship between basin area and KGE-Q exists (see Figure 1, below). As these conditions were satisfied and as the purpose of the study was to investigate the performances between rainfall and discharge time series (without specific focus on high and/or low flows), the limitations about the E-OBS station density can be assumed to have a negligible impact for the analysis purpose.

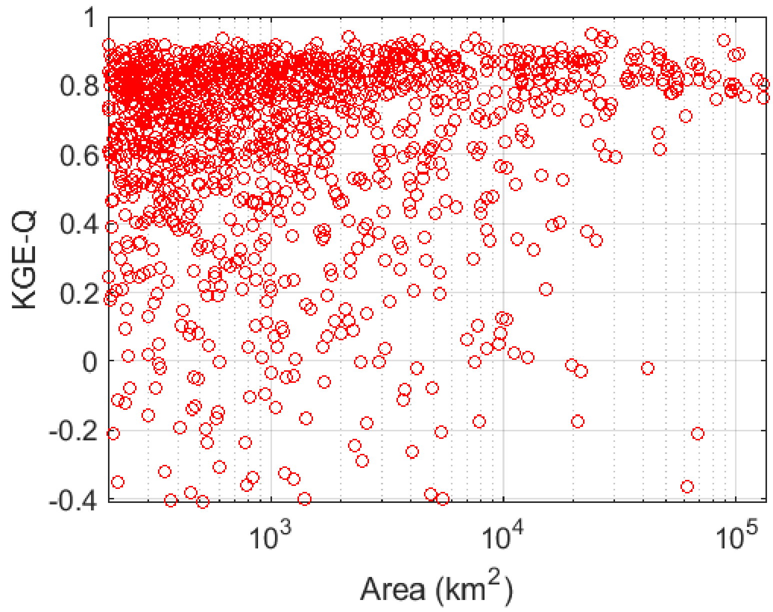


Figure 1. Relationship between basin area and KGE-Q for the analysed catchments.

Accordingly, two sentences have been added in the revised version of the manuscript (see Lines 361-371, section 5.2):

"The results of this calibration, carried out for the entire observation period (2007-2016), are good as illustrated in Figure 3a: for all the analysed basins the KGE-Q values are greater than -0.41, i.e., the model improves upon the mean flow benchmark and the median KGE-Q value obtained for the European area is equal to 0.768 (0.770 over the TMPA area). In addition, to explore the impact of the density of E-OBS rainfall on smaller basins (area<1'000 km2), the relationship between basin area and KGE-Q has been investigated (not shown). As no relationship was found, and considering that the purpose of the study is to investigate the performances between rainfall and discharge time series (without specific focus on high and/or low flows), the limitations about the E-OBS station density can be assumed to have a negligible impact on the analysis results and QE-OBS data can be assumed as a good benchmark for the successive analysis."

2. Line 331: for the discharge assessment you used only one performance score, the KGE. Can you provide more information about why you selected this score?

R: We selected only the KGE score to evaluate the hydrological model performances for three main reasons:

1) due to inherent limitations recognized for NSE (e.g., Schaefli and Gupta 2007; Gupta et al., 2009), KGE is today the criterion most commonly recommended and applied to evaluate the performance of hydrological models and therefore its use allows meaningful comparisons with other studies.

2) the purpose of the of analysis was to investigate the relationship between rainfall score and discharge simulation, without specific focus on high and/or low flows. In this respect, it is known that KGE assign a relatively more importance to discharge variability with respect to other scores (e.g., NSE or RMSE) generally found to be highly sensitive to high discharge values (Gupta et al., 2009);

3) for a practical reason, i.e., it was a decision of the author to limit the number of investigated performance scores to communicate in the most efficient way the results of the work. However, as stated in the conclusion section, in the future a more comprehensive study could consider a larger set of discharge scores metrics to better address the SRP selection.

The reasons of why we selected the KGE score have been added in the revised manuscript in the section "performance scores" as in the following (see Lines 303-314):

"To evaluate the suitability of rainfall products for river discharge modelling, the KGE index between observed and simulated river discharge data has been computed. In particular, we selected only this score for three main reasons: 1) due to inherent limitations recognized for other indices (e.g., Nash-Sutcliffe Efficiency index, Schaefli and Gupta 2007; Gupta et al., 2009), KGE is today the criterion most commonly recommended and applied to evaluate the performance of hydrological models and therefore its use allows meaningful comparisons with other studies; 2) the purpose of the analysis was to investigate the relationship between rainfall score and river discharge simulation, without specific focus on high and/or low flows. In this respect, it is known that KGE assigns a relatively more importance to discharge variability with respect to other scores (e.g., NSE or RMSE) generally found to be highly sensitive to high discharge values (Gupta et al., 2009); 3) for a practical reason, i.e., it was a decision of the author to limit the number of investigated performance scores to communicate in the most efficient way the results of the work."

3. Line 397: Can you explain better why KGE of rainfall is not relevant? From figure 4 the increasing trends of KGE-Q with rBIAS and KGE of rainfall look quite similar.

R: The authors verified the increasing trend both for KGE-Q vs rBIAS and KGE-Q vs KGE-P. Although a difference in the magnitude and correlation of the relationship between KGE-Q vs rBIAS and KGE-Q vs KGE-P can be noted, i.e., the slope coefficient is equal to 1.07 (R^2= 0.98) and 0.80 (R^2= 0.81) for KGE-Q vs rBIAS and for KGE-Q vs KGE-P, respectively, the sentence in the revised manuscript has been smoothed as (see Lines 435-438):

"*SRP hydrological performances decrease by increasing the absolute value of rBIAS, |rBIAS|, and the RRMSE values (higher |rBIAS| and RRMSE values indicate lower rainfall performances, Figure 4a and c) whereas KGE-Q increases with R and KGE-P (higher R and KGE-P values indicate higher rainfall performances, Figure 4b and d).*"

4. Line 411: How do assess that R and KGE ranges are large?

R: In Line 411 it has been observed that "*R and KGE-P seem to have a small impact on KGE-Q as for a large range of R and KGE-P values (from 0.5 to 0.8 and from 0.4 to 0.8, respectively), it is possible to obtain high KGE-Q values.*" The assessment about the "large ranges" for R and KGE-P values has been carried out by considering that, even if the two scores potentially range from -1 to 1 and from $-\infty$ to 1, respectively, meaningful range of R (KGE-P) is between 0 and 1 (-0.41 and 1). Therefore, a range of 0.3 and 0.4 can be considered "large" with respect to the variability range for which the rainfall scores suggest reliable rainfall data.

To better explain this aspect in the revised manuscript a sentence has been added in the discussion section as follows (see Lines 482-486):

"In particular, it has been noted that R and KGE-P rainfall scores have a small impact on KGE-Q as for R ranging from 0.5 to 0.8 and for KGE-P ranging from 0.4 to 0.8, it is possible to obtain high (>0.5) KGE-Q values. As the meaningful range of R (KGE-P) is between 0 and 1 (-0.41 and 1), we

can conclude that R and KGE-P are not suitable scores to define a criterion able to discern between good/bad hydrological simulations."

Technical corrections

1. Line 124: State all the questions here. I can see that you have more questions later (e.g. lines 416, 417, 418)

R: According to the reviewer suggestion, all the questions have been moved at the end of the introduction (see Lines 123-128).

2. Line 167: add spatial resolution of the product in the text.

R: The resolution of the E-OBS dataset has been added to the revised manuscript (see Line 170).

3. Line 215: it is a bit confusing when you say below TMPA area because I guess you mean the TMPA area. Change accordingly also in the other paragraphs and tables.

R: The reviewer is right. With "below TMPA area" the authors were referring to the TMPA area. The sentence has been modified with "TMPA area" throughout the manuscript.

4. Line 262-263: swap the two lines because in the plots you present first rBIAS

R: According to the review suggestion, the two lines have been swapped in the new version of the manuscript.

5. Line 262: remove "x"

R: Accordingly, the "x" has been removed in the formula.

6. Line 263: I think the numerator shouldn't be squared

R: The reviewer is right; the numerator shouldn't be squared. In the revised version of the manuscript the rBIAS formula has been modified, accordingly.

7. Line 265: in the second bracket under the square root I think there is a mistake (see Gupta et al., 2009). The ratio in the bracket should be just between standard deviation of the SRP and of the E-OBS.

R: The reviewer is right. The KGE formula has been modified in the revised version of the manuscript.

8. Line 300-Figure 2: you are talking about "patterns" so I assume you are referring to Figure 2, but then the values at line 302 are the ones reported in Table 3, so for the TMPA area. It is a bit complicated to follow, maybe you can just condense the most relevant information in figure 2 and put table 3 in supplementary material, since it doesn't provide much more information.

R: The reviewer is right; this part is difficult to follow. Therefore, in the revised version of the manuscript it has been modified as (see Lines 348-358): ". *Already at first glance of Figure 2, it is possible to note that the three products show similar patterns in terms of R (Figure 2d-f) and RRMSE (Figure 2g-i) whereas the same does not hold for the rBIAS (Figure 2a-c) and KGE-P (Figure 2l-n). The rBIAS is small for TMPA and SM2RASCAT, with median values equal to -0.127 and 0.047, respectively, whereas CMOR show a clear underestimation of the daily rainfall data over the entire European area. Higher/lower R/RRMSE values are obtained in Central Europe; the opposite is observed in the Mediterranean area. In terms of KGE-P, 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-P value for TMPA is equal to 0.516; this value reduces of about 24% and 42% for SM2RASCAT and CMOR, respectively. The median rBIAS, R, RRMSE and KGE-P rainfall score values for the three products remain approximately the same if the analysis is focused over the TMPA area (see Table 2).*"

However, to be consistent with Table 3, Table 2 has not been removed from the main manuscript.

9. Line 328: I would put Figure 1.b which belong to result section in a separate figure from Fig.1.a which belong to the dataset section

R: Accordingly, Figure 1b has been merged in Figure 3.

10. Line 342: the 39% is not for CMOR but for TMPA

R: We thank the reviewer. The values in the manuscript have been modified in accordance with Table 4 (see Lines 391-392).

11. Line 379: higher absolute values of rBIAS

R: We thank the reviewer. The sentence has been modified accordingly (see Line 419).

12. Line 389: maybe name the KGE as KGE-Q otherwise it can be confused with KGE of rainfall

R: We thank the reviewer for this suggestion. The KGE has been modified as KGE-P and KGE-Q to refer to KGE of rainfall and discharge, respectively. A sentence to clarify this distinction has been added to the revised manuscript (section 4.5, Lines 315-316):

"To distinguish between the KGE of rainfall and discharge, hereinafter, the symbols KGE-P and KGE-Q will be used."

13. Figure 2-3: you can add the name of each SRP product at the top of each column.

R: Figure 2 and 3 have been modified according to the reviewer suggestion.

14. Line 734: there are no figures d), e), f). Change CMORPH to CMOR to be consistent.

R: The caption of figure 3 has been modified accordingly.

References

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Referee 2

General comments

The authors propose the evaluation of satellite rainfall products with different metrics and compare the results to the performance of a hydrological discharge model. The aim is to determine which rainfall accuracy metrics are suitable in describing satellite rainfall accuracy in regard to flood simulation performance. The authors compare the performance of a hydrological model forced with a benchmark rainfall dataset with the performance of the same model forced with three different satellite rainfall products. In my opinion the work described is novel and worthy of publication. The results presented support the conclusion reached. The findings of the study will be very relevant for future research. However, there are several minor issues that need clarification, which are outlined below.

R: We thank the reviewer for his/her supportive review. In the revised version of the manuscript the following changes have been implemented:

- the title has been changed in "Which rainfall score is more informative about the performance in river discharge simulation? A comprehensive assessment on 1318 basins over Europe".
- Any reference to flood has been removed and modified with river discharge to highlight that the purpose of the study is to investigate the performances between rainfall and river discharge time series (without specific focus on high and/or low flows);
- a discussion about the quality of the E-OBS rainfall data and the impact of its density network on river discharge simulation has been added;
- to avoid misunderstanding between KGE of rainfall and discharge throughout the manuscript (and in the figures) KGE has been replaced by KGE-P and KGE-Q, to indicate the KGE index referred to precipitation and the one referred to river discharge respectively;
- more information about why only the KGE-Q index has been selected for the analysis has been added to the revised manuscript;
- Tables and figures have been modified according to the reviewer's suggestions.

Specific comments

1. The authors claim in the title, introduction and methodology that satellite rainfall performance is evaluate in regard to flood modelling. However, no high flow specific analyses are performed and in their conclusions the authors them-selves state that the focus was on the entire discharge time series. I therefore propose to remove any reference to flood simulation and instead refer to runoff or discharge simulation.

R: According to the reviewer suggestion, the title and the manuscript have been changed to outline that the analysis is not specifically oriented to floods but it is related to the simulation of the entire river discharge time series (see Lines 308-310).

2. The title can be perceived as misleading, since "rainfall metric" generally refers to rainfall statistics (e.g. spatial and temporal distribution, amount, seasonality, ... etc.). Please change the title

(and mentions in the paper, for example L135) in a way that reflects to focus on satellite rainfall product performance metrics.

R: We thank the reviewer for raising this issue. Accordingly, the terms "rainfall metric" have been changed as "rainfall performance score" throughout the revised manuscript. In particular, the title, by considering also the comment n#1 has been modified as: "Which rainfall score is more informative about the performance in river discharge simulation? A comprehensive assessment on 1318 basins over Europe".

3. Please discuss what effect the station density of the benchmark dataset (E-OBS) has on the results. Similarly, please mention if any quality checks have been performed on the discharge data.

R: A discussion on the station density has been added to the revised manuscript (see Lines 173-182) to highlight that *"E-OBS station density significantly varies across Europe (see Haylock et al., 2008; Cornes et al., 2018): for some regions, the station density is sufficiently low to expect a strong tendency for interpolated daily rainfall and temperature values to be underestimated with respect to the "true" area-average stations (Hofstra et al., 2009; Hofstra et al., 2010; Kyselý and Plavcová, 2010). As the smoothing is greatest for higher percentiles, an underestimation of peak floods is expected if E-OBS rainfall data are used for rainfall-runoff modelling above all for basins with area lower than 1000 km2 (Hofstra et al., 2010). 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 of long rainfall time series."*
Similarly, some sentences have been added to the revised manuscript to describe the quality checks performed on the discharge data. In particular, it has been specified that (Lines 192-198): *"To ensure quality on discharge observations the following steps have been followed: 1) visual hydrograph inspection, which is probably the most thorough method (Crochemore et al., 2020); 2) check on data availability; 3) check the presence of outliers; 4) check the presence of inhomogeneities. Only stations with less than 20% of missing data in one year, showing no inhomogeneities in the time series were retained in the compiled European dataset. The time series were checked also against the presence of anomalous values (i.e., values greater than five times the standard deviation), flagged as outliers."*

4. In Line 185 the authors very briefly mention that they developed and used a catchment delineation algorithm. Since this is a new approach, please elaborate on the methodology and quality checks used. Do et al. (2018) applied a catchment delineation procedure to a global river dataset and found only 68 % of catchments to have a "high" quality result. An evaluation of the quality of the delineated catchments is therefore imperative.

R: We thank the reviewer for the very relevant comment. In the manuscript we have added details about the adopted procedure and associated quality checks (see Lines 201-210): *"The procedure is based on the following steps: (i) we select cells having contributing area larger or equal to 4 km2 over the entire study area, (ii) we move the discharge measurement locations from the coordinates reported in the original metadata to the closest cells of the river network, (iii) we delineate the catchments. Adopting the method used by Do et al. (2018), we evaluated the quality of the products* *comparing the area of the delineated catchment (Ad) with that available from the original metadata (Am). The absolute percentage difference (Dp) was calculated according to the following formula Dp =(Ad - Am)/ Ad *100 |. Median and 75th percentile of the distribution of the Dp values were, respectively, 2.67% and 22.07%. We excluded from the following hydrological simulation, catchments having Dp values larger than 50% (less than the 20% of the total number of catchments)."*

5. L62: Please clarify what do you mean with "gaining ground"? Are satellite rainfall observations used more often? Do they improve in accuracy?

R: Yes, the meaning of the sentence was "*satellite rainfall observations are are becoming potential alternative to the classical rainfall monitoring methods, thanks to their global availability and increasing accuracy*". The manuscript has been modified, accordingly (see Lines 62-64).

6. L122: Please specify that "best performing" in this context is meant in regard to hydrological model performance and not in regard to rainfall accuracy in comparison to a benchmark rainfall product.

R: In the revised version of the manuscript this part has been modified as, according to the suggestion of reviewer 1, all the questions raised throughout the manuscript have been moved here. For that this sentence has been modified as follows (see Lines 123-128):
"*The following research questions are addressed: is there any performance score that can be used to select the best performing rainfall product for river discharge simulation? Are multiple scores needed? And, which are these scores? Are R and RMSE, generally used to characterize the rainfall accuracy, informative about the hydrological modelling performance? How small/large should be these rainfall scores to obtain good performances in river discharge simulations, i.e., KGE on discharge greater than 0.5?*

We hope that in this way the misunderstanding highlighted by the reviewer has been solved.

7. L150-152: Please improve your description of the rainfall distribution over Europe. The Alps receive high rainfall amounts (not just the surrounding areas), as does the coast of Croatia (which is at the edge of the Mediterranean Sea).

R: Accordingly, the description of the rainfall distribution over Europe has been improved. The sentence has been modified as follows (see Lines 154-156):
"The north Atlantic coast of Spain, the Alps and Balkan Mediterranean countries generally receive higher rainfall amounts while along the west edges of the Mediterranean Sea, in northern Europe and in northern Scandinavia, lighter rainfall is common."

8. L204-208: Since the SM2RAIN-ASCAT product is relatively can add a brief explanation of the SM2RAIN algorithm.

R: Accordingly, in the revised version of the manuscript a brief description of the SM2RAIN algorithm has been given. Specifically, it has been added to the manuscript that (see Lines 230-232): *"SM2RAIN is an algorithm based on the concept that the soil acts as a "natural rain gauge": by inverting the soil water balance equation, the algorithm allows to estimate the accumulated rainfall from soil moisture observations."*

9. L249: Can you elaborate what you mean with "QE-OBS is used as reference for parameter values calibration"

R: The sentence has been modified to better explain that (see Line 273): "*QE-OBS is used as benchmark to calibrate the parameters of MISDc model.*"

10. L312/313: Please move this thought to the discussion.

R: The authors would prefer to leave this sentence in section 5.1, focused on rainfall assessment, instead of to move it in the discussion section, focused on the relationship between rainfall and discharge performances.

11. L341-343: Please check these values. They do not match the values reported in Table 4.

R: We thank the reviewer. The values in the manuscript have been modified in accordance with Table 4.

12. L394/395: Please check the reference to the Figure. RRMSE is Figure 4c) and R is Figure 4b).

R: The text has been changed accordingly.

13. L394/395: Please elaborate how you differentiate between strong increase/decrease and how the individual increase/decrease might be related to the definition of the individual metric.

R: To differentiate between strong increase/decrease, the authors verified the increasing trend between both KGE-Q with rBIAS and KGE-Q and KGE-P. Although a difference in the magnitude and correlation of the relationship between KGE-Q vs rBIAS and KGE-Q vs KGE-P can be noted, i.e., the slope coefficient is equal to 1.07 (R^2= 0.98) and 0.80 (R^2= 0.81) for KGE-Q vs rBIAS and for KGE-Q vs KGE-P, respectively, the sentence in the revised manuscript has been smoothed.

Concerning how the individual increase/decrease might be related to the definition of the individual metric, the meaning of each score will be recalled in the revised manuscript. Specifically, the sentence in line 394/395 has been modified as (see Lines 435-438): "*SRP hydrological performances decrease by increasing the absolute value of rBIAS, |rBIAS|, and the RRMSE values (higher |rBIAS| and RRMSE values indicate lower rainfall performances, Figure 4a and c) whereas KGE-Q increases with R and KGE-P (higher R and KGE-P values indicate higher rainfall performances, Figure 4b and d).*."

14. L400/401: Can you explain how the categorical values are different than expected? Higher (= better) POD and TS scores lead to better performance. Only FAR behaves differently than expected (but only for rainfall >0.5 and rainfall >5 mm). It even seems like not very high values of TS and POD are necessary to still get high KGE.

R: The reviewer is right, only FAR behaves differently than expected and only for small and moderate rainfall events. This sentence has been modified according to the reviewer suggestion (see Lines 440-442):

"*Indeed, it has been found that higher (= better) POD and TS scores lead to better performance whereas the relationships between KGE-Q and the FAR for small and moderate rainfall are different (i. e, inverse) from what can be expected.*"

15. L410-418: Could the difference in which range reaches high KGE performance be due to differences in how these metrics are calculated? This will impact what is considered a "large range" of values. This makes the interpretation of the plot slightly subjective. Please elaborate how you came to the conclusion that "rBIAS and RRMSE [...] seem to have a stronger link with the hydrological performance". In regard to this, please also see the comment on Figure 4 below.

R: According to the authors this result could be linked to the hydrological model structure and to the parameters calibrated into the model. Indeed, it has been largely demonstrated in the scientific literature (e.g., Zeng et al., 2018) that the impact of imperfect precipitation estimates on model efficiency can be reduced to some extent through the adjustment of model parameters. In this case, it is clear that the hydrological model calibration step is able to correct the rainfall time shift, allowing to obtain good hydrological performances (KGE-Q) for a large range of R values. A similar consideration holds for KGE-P, largely influenced by the correlation coefficient.

A sentence highlighting this aspect has been added in the revised manuscript (in the discussion section, Lines 482-493):

"In particular, it has been noted that R and KGE-P rainfall scores have a small impact on KGE-Q as for R ranging from 0.5 to 0.8 and for KGE-P ranging from 0.4 to 0.8, it is possible to obtain high (>0.5) KGE-Q values. As the meaningful range of R (KGE-P) is between 0 and 1 (-0.41 and 1), we can conclude that R and KGE-P are not suitable scores to define a criterion able to discern between good/bad hydrological simulations. This result could be linked to the hydrological model structure and to the parameters calibrated into the model. Indeed, it has been largely demonstrated in the scientific literature (e.g., Zeng et al., 2018) that the impact of imperfect precipitation estimates on model efficiency can be reduced to some extent through the adjustment of model parameters. In this case, it is clear that the hydrological model calibration step is able to correct the rainfall time shift, allowing to obtain good hydrological performances (KGE-Q) for a large range of R values. A similar consideration holds for KGE-P, largely influenced by the correlation coefficient."

16. Tables/Figures: If possible, the authors might want to consider including the supplement figures in the main text.

R: Thanks for this suggestion, but we think that to make the manuscript more readable it would be better to not increase the number of figures and related comments.

17. Table 1: This table is not necessary. Instead a plot showing catchment area distribution might be more useful.

R: In the revised version of the manuscript, the table has been removed and, according the reviewer suggestion, in Figure 1 a plot showing catchment area distribution has been added.

18. Table 3: KGE is missing in the list of metrics mentioned in the caption. For better readability, can you add the information from this sentence "The more R, rBIAs, RRMSE and KGE values goes to-ward 1, 0, 0, 1 respectively, the higher is the agreement between E-OBS and SRPs. "to the table caption?

R: The caption of the table has been improved according to the reviewer suggestion, specifying also that "*the more R, rBIAs, RRMSE and KGE-P values goes to-ward 1, 0, 0, 1 respectively, the higher is the agreement between E-OBS and SRPs.*"

19. Figure 2: Please add column headings. Although using the same colour scale is aesthetically pleasing, it makes it difficult to compare the different metrics, since "best" value varies. E.g. for rBIAS a diverging colour scale would be more appropriate.

R: Figure 2 has been modified in the revised version of the manuscript, adding a column heading for each product and modifying the colour scale for rBIAS and RRMSE. Specifically, as suggested by the reviewer a diverging colour scale has been used for rBIAS whereas for RRMSE an inverse colorbar with respect to R and KGE has been considered.

20. Figure 3: Plots d, e and f are mentioned in the caption but are not part of the Figure.

R: The caption of the figure has been modified deleting any reference to plots d, e, f that are not part of Figure 3.

21. Figure 4: There is a high density of points. Can you use empty (e.g. transparent) filling of the points, so that the points do not cover each other. Otherwise the distribution, particularly of the TMPA points, is not visible. (Same for Figure S2). Also, can you clarify if the boxplots are for all products together? S1: Please add which values are considered better (e.g. higher for POD and TS and lower for FAR) to the figure caption.

R: Figure4, Figure S2 and caption of Figure S1 have been modified according to the reviewer suggestion.

Concerning the boxplot, they are evaluated for all products together. This aspect is clarified in the manuscript at lines 388-390: "to extract useful information from Figure 4 and Figure S2, the scores obtained separately for each product have been grouped and the KGE-Q data points have been binned into uniform ranges (with step 0.1) of rainfall scores".

Technical corrections

I want to compliment the authors for communicating a complex topic very well, however the manuscript would benefit from a thorough grammar and spell check to improve understanding.

22. L25: "understanding how uncertainties[...]"
R: In the correct version of the manuscript the sentence has been correct.

23. L42: "Results suggest that, among [...] are not reliable scores to select the best performing rainfall product for hydrological modelling[...]"
R: In the correct version of the manuscript the sentence has been correct.

24. L58: "Generally, rainfall observations [...]"
R: In the correct version of the manuscript the sentence has been correct.

25. L63: "[...] (SRPs) has boosted their use [...]"

R: In the correct version of the manuscript the sentence has been correct.

26. L72: "is used to simulate a discharge time series [...]"

R: In the correct version of the manuscript the sentence has been correct.

27. L86: "Generally, this comparison [...]"
R: In the correct version of the manuscript the sentence has been correct.

28. L98/99: Missing bracket
R: In the correct version of the manuscript bracket has been added.

29. L108: "by using the MIKE SHE model[...]"
R: In the correct version of the manuscript the sentence has been corrected.

30. L112: "it is difficult to find literature [...]"
R: In the correct version of the manuscript the sentence has been corrected.

31. L140/141: "is composed of 1318 basins[...] over the whole of Europe [...]"
R: In the correct version of the manuscript the sentence has been corrected.

32. L142: "The European continent [...]"
R: In the correct version of the manuscript the sentence has been corrected.

33. L145: “gently slopes towards [...]”
R: In the correct version of the manuscript the sentence has been corrected.

34. L150: “the Alps generally has higher rainfall amounts [...]”
R: In the correct version of the manuscript the sentence has been corrected.

35. L154: “prevailingly subject to [...]”
R: In the correct version of the manuscript the sentence has been corrected.

36. L155: “according to the latitude[...]”
R: In the correct version of the manuscript the sentence will be corrected.

37. L157: “and for about11% [...]”
R: In the correct version of the manuscript the sentence has been corrected

38. L161: “basin characteristics.”
R: In the correct version of the manuscript the sentence has been corrected

39. L175: “an European daily dataset [...]”
R: In the correct version of the manuscript the sentence the sentence has been corrected.

40. L190: “period in 2012[...]”
R: In the correct version of the manuscript the sentence has been corrected

41. L201:” provided by the CPC [...]”
R: In the correct version of the manuscript the sentence has been corrected.

42. L219-221: “applied to carry out[...] model composed of a component [...] of soil moisture and a rainfall-runoff model[...]”
R: In the correct version of the manuscript the sentence has been corrected.

43. L233/234: “allow us to consider the model suitable for the purpose of this analysis.”

44. L236: “analysis regards the quality assessment [...]”
R: done

45. L270/271: “The more R [...], respectively, the higher is [...]”
R: done

46. L275: “(TS). POD reports [...]”
R: done

47. L278: too many dots
R: removed.

48. L356: “This is the first notable result [...]”
R: done

49. L368: “for the CMOR product[...]”
R: done

L371/404. As amusing as it is, it might be better to refer to FAR and TS as TS/FAR instead of the other way around.
R: done

50. L372: “of this product in terms [...]”
R: done

51. L380: “in terms of RRMSE[...]”
R: done

52. L454-457: This sentence would benefit from commas.
R: done

53. L456: “errors in rainfall [...]”
R: done

54. L498: “limitation, this study contributes to the better understanding of the propagation of [...] simulations. This could be very [...]”
R: done

References

Do, H.X., Gudmundsson, L., Leonard, M. and Westra, S., 2018. The GlobalStreamflow Indices and Metadata Archive (GSIM)-Part 1: The production of a dailystreamflow archive and metadata. Earth System Science Data, 10(2), pp.765-785.C4
Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-31, 2020

WHICH RAINFALL ~~METRIC~~ ~~PERFORMANCE~~ SCORE IS MORE INFORMATIVE ABOUT THE PERFORMANCE IN ~~FLOOD~~ RIVER DISCHARGE 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)

February 2020

Submitted to:

* 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.

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, understanding how uncertainties transfer from SRPs into ~~flood~~ river discharge 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) performance scores. However, whether these scores are informative about the associated performance in ~~flood~~ river discharge simulations (when the SRP is used as input to a~~n~~ 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~~ river discharge simulation. That is, the following research questions are~~are~~ addressed: is ~~(are)~~ there any ~~appropriate~~ performance score that can be used to ~~(s)~~ select~~to drive the choice of~~ the best performing rainfall product for ~~flood~~ river discharge simulation? Are multiple scores needed? And, which are these scores? ~~?~~ To answer these~~is~~ questions 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~~ suggest~~ed~~ that, among the continuous scores, correlation coefficient and Kling-Gupta efficiency index are not reliable indices to select the best ~~rainfall product~~ performing rainfall product ~~best~~ for hydrological modelling whereas bias and root mean square error seem more appropriate. In particular, by constraining the relative bias to absolute 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 river discharge 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 becoming potential alternative~~gaining ground with respect~~ to the classical rainfall monitoring methods, thanks to their global availability and increasing accuracy.

The global availability of near real time satellite rainfall products (SRPs) has boosted 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

Codice campo modificato

Codice campo modificato

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 is used to simulate a river discharge time series through a~~n~~ hydrological model and the difference between simulated and observed 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 river discharge for many (+9000) basins over the globe, however, the relationship between the accuracy in terms of rainfall and river discharge was not investigated in detail.

From the analysis of both the probabilistic and the statistical approaches 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~~performance score(s) should be used to identify the best performing rainfall product for ~~flood forecasting~~river discharge estimation (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 the 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 fi~~ou~~nd literature studies advising on rainfall error metrics able to indicate ~~flood~~ river discharge 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~~ river discharge modelling at regional scales. This paper aims to explore the link between satellite rainfall accuracy of different products and their ~~flood~~ river discharge modelling performance. The following research questions are addressed: is there

Codice campo modificato
Codice campo modificato
Codice campo modificato
Codice campo modificato

any performance score that can be used to select the best performing rainfall product for river discharge simulation? Are multiple scores needed? And, which are these scores? ~~which is the most appropriate performance metric to be used to select the best performing satellite rainfall product for flood modelling?~~ Are R and RMSE, generally used to characterize the rainfall accuracy, informative about the hydrological modelling performance? How small/large should be these rainfall scores to obtain good performances in river discharge simulations, i.e., KGE on discharge greater than 0.5?

In pursuing this goal, three different near real time SRPs, 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 ~~basins spread out over~~across 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 performance scores ~~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~~ of 1318 basins, with area ranging in size from 200 to 136’000 km^2 belonging to 23 different countries and spread over the whole of Europe, over longitude varying from -10° to 25° and latitude from 35° to 70° (Figure 1a). The 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 slopes 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.

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 north Atlantic coast of Spain, the Alps and Balkan Mediterranean countries generally receive higher rainfall amounts ~~area east, west and north of the Alps generally is interested by higher rainfall amount,~~ while along the west edges of the Mediterranean Sea, in northern Europe 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 prevailingly ~~prevailing~~ subject to winter floods (Berghuijs et al., 2019).

The main features of the study basins, clustered according to the latitude of the outlet section, are ~~summarized~~ illustrated in Figure 1b and c~~Table 1~~: among the 1318 basins, more than half (889) have the outlet section located below the 50° latitude and for 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 t~~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~~1).

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 0.22°x0.22° gridded data sets version

17~~6~~.0 (E-OBS, Haylock et al., 2008), currently maintained by the Copernicus Climate Change Service. The E-OBS dataset is built by using data from nearly ~~2316~~ 9618 stations (i.e., equivalent on average to a density of 1 stations every ~~4000~~ 1000 km^2) but the station density significantly varies across Europe (see Haylock et al., 2008; Cornes et al., 2018): for some regions, the station density is sufficiently low to expect a strong tendency for interpolated daily rainfall and temperature values to be underestimated with respect to the “true” area-average stations (Hofstra et al., 2009; Hofstra et al., 2010; Kyselý and Plavcová, 2010). As the smoothing is greatest for higher percentiles, an underestimation of peak floods is expected if E-OBS rainfall data are used for rainfall-runoff modelling above all for basins with area lower than 1000 km^2 (Hofstra et al., 2010). 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 of long rainfall time series.

~~.~~

Daily river discharge data are obtained through an E~~e~~uropean 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. To ensure quality on discharge observations the following steps have been followed: 1) visual hydrograph inspection, which is probably the most thorough method (Crochemore et al., 2020); 2) check on data availability; 3) check the presence of outliers; 4) check the presence of inhomogeneities. Only stations with less than 20% of missing data in one year, showing no inhomogeneities in the time

Codice campo modificato

Codice campo modificato

Codice campo modificato

Codice campo modificato

Codice campo modificato

series were retained in the compiled European dataset. The time series were checked also against the presence of anomalous values (i.e., values greater than five times the standard deviation), flagged as outliers.

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 procedure is based on the following steps: (i) we select cells having contributing area larger or equal to 4 km2 over the entire study area, (ii) we move the discharge measurement locations from the coordinates reported in the original metadata to the closest cells of the river network, (iii) we delineate the catchments. Adopting the method used by Do et al. (2018), we evaluated the quality of the products comparing the area of the delineated catchment (Ad) with that available from the original metadata (Am). The absolute percentage difference (Dp) was calculated according to the following formula Dp =(Ad - Am)/ Ad *100 |. Median and 75th percentile of the distribution of the Dp values were, respectively, 2.67% and 22.07%. We excluded from the following hydrological simulation, catchments having Dp values larger than 50% (less than the 20% of the total number of catchments).

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 ~~ono~~in 2012 and the median length of discharge observations is about 6/7 years (see ~~Table~~ Figure 1a).

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.

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.

Codice campo modificato

CMORPH is provided by the 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 is an algorithm based on the concept that the soil acts as a "natural rain gauge": by inverting the soil water balance equation, the algorithm allows to estimate the accumulated rainfall from soil moisture observations. SM2RAIN-ASCAT, downloadable at https://zenodo.org/record/3635932~~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~~ 1 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~~ over the TMPA area (latitude <50°); 3) above TMPA area (latitude >50°).

4. METHOD

4.1 Hydrological model

~~The model applied to carryied out the flood discharge simulation is~~ MISDc ("Modello Idrologico Semi-Distribuito in continuo" Brocca et al. 2011) is~~,~~ a two-layer continuous hydrological model

~~composed~~ characterized by~~by of~~ a component simulating the temporal pattern of soil moisture and ~~by~~ a rainfall-runoff transformation component for simulating river ~~model simulating flood~~discharge time series. By using ~~as input~~ daily rainfall and air temperature data, MISDc simulates the 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~~ discharge 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 us to consider the model suitable for the purpose of this analysis~~for the analysis purpose~~.

4.2 Experimental design

The first step of the analysis ~~concerned~~ ~~regards on the~~is 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, river discharge simulations have been ~~performed~~ obtained 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.

Codice campo modificato

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 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~~ benchmark to calibrate the parameters of MISDc~~for parameter values calibration~~model.

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 river discharge and, more important, 3) to neglect the uncertainty due to the hydrological model structure in the SRPs comparison.

4.5 Performance ~~metrics~~scores

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 Pearson correlation coefficient, R, 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:

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

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

$$R = \frac{Cov(SRP, P_{ref})}{\sigma_{SPR}\ \sigma_{P_{ref}}} \qquad (2)$$

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

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

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 +∞; R values range from -1 to 1; RRMSE is bounded from 0 to +∞ while KGE varies between –∞ to 1. More ~~R,~~ rBIAs, R, RRMSE and KGE values goes toward ~~1,~~ 0, 1, 0, 1 respectively, higher is the agreement between E-OBS and SRPs. In particular, for KGE, ~~model performance~~values in the range -0.41 < KGE <= 1 indicate that satellite rainfall data~~the model~~ outperform~~s~~ the mean of the E-OBS 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 river ~~flood~~ discharge modelling, ~~the KGE index between the~~ the KGE index between observed and simulated river discharge data ~~has~~has been computed. In particular, we selected only this score for three main reasons: 1) due to inherent limitations recognized for other indices (e.g., Nash-Sutcliffe Efficiency index, Schaefli and Gupta 2007; Gupta et al., 2009), KGE is today the criterion most commonly recommended and applied to evaluate the performance of hydrological models and therefore its use allows meaningful comparisons with other studies; 2) the purpose of the analysis was to investigate the relationship between rainfall score and river discharge simulation, without specific focus on high and/or low flows. In this respect, it is known that KGE assigns a relatively more importance to discharge

Codice campo modificato

variability with respect to other scores (e.g., NSE or RMSE) generally found to be highly sensitive to high discharge values (Gupta et al., 2009); 3) for a practical reason, i.e., it was a decision of the author to limit the number of investigated performance scores to communicate in the most efficient way the results of the work.

To distinguish between the KGE of rainfall and discharge, hereinafter, the symbols KGE-P and KGE-Q will be used. Specifically, KGE-Q 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. River d~~D~~ischarge simulations characterized by KGE-Q values in the range -0.41 and 1 can be assumed as reliable; KGE-Q values greater than 0.5 have been considered good with respect to their ability to reproduce benchmark river 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 river discharge. The propagation of the rainfall error into the river 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~~2). Specifically, rBIAS, R, RRMSE and KGE-P values are illustrated in the rows of Figure 2 for each study basin, for the three 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~~ 2 the performances of the basins whose outlet section is located below/above 50° latitude, i.e. ~~below~~over/above the TMPA coverage, are listed. Already at first glance of Figure 2, it is possible to note that the three products show similar patterns in terms of R (Figure 2d-f) and RRMSE (Figure

2g-i) whereas the same does not hold for the rBIAS (Figure 2a-c) and KGE-P (Figure 2l-n). The rBIAS is ~~low~~small for TMPA and SM2R$_{ASCAT}$, with median values equal to -0.127 and 0.08~~1~~47, respectively, whereas CMOR show a clear underestimation of the daily rainfall data over the entire European area. Higher/lower R/RRMSE values are obtained in Central Europe; the opposite is observed in the Mediterranean area. In terms of KGE-P, 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-P value for TMPA is equal to 0.516; this value reduces of about 24% and 42% for SM2R$_{ASCAT}$ and CMOR, respectively. The median rBIAS, R, RRMSE and KGE-P rainfall score values for the three products remain approximately the same if the ~~By focusing the~~ analysis is focused over the TMPA area (see Table 2).~~, 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-P. ~~Likely d~~Due to soil freezing and snow presence, the performances of SM2R$_{ASCAT}$ decrease in terms of R, rBIAS and KGE-P 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. The n~~N~~umbers 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~~ outperforming the other two products. Conversely, $SM2R_{ASCAT}$ shows a 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 3~~1b~~a: for all the analysed basins the KGE-Q values are greater than -0.41, i.e., the model improves upon the mean flow benchmark and the median KGE-Q value obtained for the European area is equal to 0.768 (0.770 over the TMPA area). In addition, to explore ~~take into account that due to~~ the impact of the density of ~~network~~ E-OBS rainfall on ~~data could be not reliable for~~ smaller basins (area<1’000 km^2), the relationship between basin area and KGE-Q has been investigated (not shown). As no relationship was found, and ~~as~~ considering that the purpose of the study is to investigate the performances between rainfall and discharge time series (without specific focus on high and/or low flows), the limitations about the E-OBS station density can be assumed to have a negligible impact on the analysis results and ~~This ensures the good quality of~~ $Q_{E\text{-}OBS}$ data ~~that~~ can be ~~are~~ assumed as a good benchmark for the successive analysis. Hereinafter, the hydrological performance has been assessed in terms of KGE-Q 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 3b-d for the validation period and in Table ~~4~~ 3 for both the calibration and the validation periods. At the top of each plot in Figure 3, the median KGE-Q value, averaged over the spatial coverage of each product, is reported whereas in Table ~~4~~ 3 the performances of the basins whose outlet section is located below/above 50° latitude are listed. In addition, in Table ~~4~~ 3 the percentage of basins showing KGE-Q values higher than 0.5 is computed.

By averaging the performances over the spatial coverage of each product, median KGE-Q 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 3b-d). The percentage of the basins showing KGE-Q values higher than 0.5, is ~~88~~18% and ~~18~~88% for CMOR and SM2R$_{ASCAT}$, respectively, whereas the same percentage drop in the validation period up to about ~~39~~2% and ~~3~~62% 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~~ over the TMPA area (see Table ~~4~~3): poor results are obtained by CMOR during the validation period (median KGE-Q<0; only 2.6% show KGE-Q higher than 0.5), whereas SM2R$_{ASCAT}$ outperforms TMPA in both periods. In particular, during the validation period a median KGE-Q 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-Q greater than 0.5, the ratio is nearly 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$_{E-OBS}$ 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-Q higher than 0.5 (Table ~~4~~3). This is the first notable~~y~~ result of the paper, i.e., among the SRPs available in near real time, there are some products that can be ~~reasonably~~ profitably used to force a hydrological model for obtaining ~~in order to obtain~~ reliable river discharge data over Europe. However, ~~a~~ some questions raised in the introduction are still unsolved~~question remains:~~, i.~~-~~e., if there ~~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? I~~is any link between rainfall and river discharge performances and if it is possible to find a rainfall ~~performance~~ score to select a priori the best SRP to obtain reliable river discharge simulations.~~?~~ The answer to these questions is given in the next paragraph where the rainfall performances are compared with the river discharge performances.

~~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?~~

5.3 Rainfall vs river discharge performances: is there any link between them?

By comparing the patterns of Figure 2 against the patterns of Figure 3b-d, 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-Q<0.4). In addition, as occurs for the CMOR product, high rBIAS values (both negative or positive) produce negative KGE-Q values. Interestingly, R and KGE-P rainfall scores seem to be weakly linked to the hydrological performances. Finally, no clear link can be highlighted between KGE-Q ~~of discharge~~ and the rainfall categorical scores as for instance, the low/high~~/low~~ values of $\mathrm{SM2R}_{\mathrm{ASCAT}}$ in terms of ~~FAR/~~TS/ FAR do not explain the higher performances of this product~~s~~ in terms of discharge (see Figure 3 against Figure S1).

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-Q 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 absolute values of rBIAS and lower KGE-P values with respect to the other two products; rBIAS of $\mathrm{SM2R}_{\mathrm{ASCAT}}$ varies near zero and, in terms of RRMSE, $\mathrm{SM2R}_{\mathrm{ASCAT}}$ is characterized by a reduced range of variability, (i.e., most of the $\mathrm{SM2R}_{\mathrm{ASCAT}}$ 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~~ river discharge 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-Q data points have been binned into uniform ranges (with step 0.1) of rainfall scores. The median KGE-Q, and the 25th and 75th percentiles of KGE-Q 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-Q value, the two ends of the black lines in the same figure represent the 25th and 75th percentile of the KGE-Q 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 decrease by increasing the absolute value of rBIAS, |rBIAS|, and the RRMSE values (higher |rBIAS| and RRMSE values indicate lower rainfall performances, Figure 4a and c) whereas KGE-Q increases with R and KGE-P (higher R and KGE-P values indicate higher rainfall performances, Figure 4b and d).~~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 all the categorical scores and the rainfall events here investigated. Indeed, it has been found that higher (= better) POD and TS scores lead to better performance whereas ~~except for the rainfall threshold equal to 10 mm/day,~~ the relationships between KGE-Q ~~of discharge~~ and the ~~categorical scores~~FAR ~~of~~ for small and moderate rainfall are different (i. e, ~~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-Q ~~of discharge~~ slightly increase with POD whereas a stronger link can be noted between KGE-Q and TS/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-Q>0.5~~ evaluated on the discharge data~~)? The straight grey line in Figure 4 (and Figure S2), drawn for a threshold value of KGE-Q 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, i. e. for good rainfall data. Conversely, R and KGE-P~~ of rainfall~~ seem to have a small impact on KGE-Q~~ of discharge~~ as for a large range of R and KGE-P values (from 0.5 to 0.8 and from 0.4 to 0.8, respectively), it is possible to obtain high KGE-Q values. Similar conclusions hold for the categorical scores evaluated for heavy rainfall events: it can be noted that the higher capability of SRPs to detect rainfall events does not affect the hydrological performances, i.e., it is possible to obtain KGE-Q ~~of discharge~~ higher than 0.5 for a large range of POD, FAR and TS values. Finally, a last point has to be addressed to fulfil the purpose of the manuscript, i.e., it has to be investigated ~~A further question remains:~~ how small/large should be the rainfall scores to obtain good hydrological performances, i.e., KGE-Q greater than 0.5. ~~?~~ In particular, should be defined a range of variability for ~~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 constraint 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~~ over the TMPA area and a median KGE-Q value equal to 0.342 is obtained for the 889 basins. According to Table ~~4~~3, nearly 35% of the basins show KGE-Q greater than 0.5. If the absolute value of rBIAS (i.e., ~~,~~ |rBIAS|)~~,~~ is constrained to values lower than 0.2 (Figure 5b), the median KGE-Q value over the 400 basins that fulfils the criteria is equal to 0.525. As shown in Figure 5c, a constraint on RRSME lower than 2 is not enough to ~~assure~~ ensure good hydrological performances (median KGE-Q lower than 0.5) whereas if a combination of the two rainfall scores is considered, the threshold on KGE-Q>0.5

is exceeded by nearly 75% of the basins fulfilling the criteria (see first boxplot of Figure 5d). In other words, this~~it~~ means that nearly less than 25% of the basins fulfilling the criteria show~~s~~ low performance (first boxplot of Figure 5d). Alternatively, less than 25% of basins not fulfilling the rainfall constraints~~s~~ show~~s~~ good hydrological performances (see second boxplot of Figure 5d).

For the sake of 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-P, POD, FAR and TS and relative combinations), but no one of the shown rainfall constraint can be considered satisfactory for the purpose of the analysis~~ purpose~~. Indeed, no one of the rainfall constraint in Figure S3 allows a clear separation between basins fulfilling/not fulfilling the criteria with a corresponding increase of KGE-Q~~ 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 ~~performance score(s) can be used to select the best performing rainfall product for river discharge simulation~~it is possible to obtain good results in terms of river discharge simulation~~. In particular, it has been noted that R and KGE-P rainfall scores ~~seem to ~~have a small impact on KGE-Q as for R ranging from 0.5 to 0.8 and for KGE-P ranging from 0.4 to 0.8, it is possible to obtain high (>0.5) KGE-Q values. As the meaningful range of R (KGE-P) is between 0 and 1 (-0.41 and 1), we can conclude that R and KGE-P are not suitable scores to define a criterion able to discern between good/bad hydrological simulations. This result could be linked to the hydrological model structure and to the parameters calibrated into the model. Indeed, it has been largely demonstrated in the scientific literature (e.g., Zeng et al., 2018) that the impact of imperfect precipitation estimates on model efficiency can be reduced to some extent through the adjustment of model parameters. In this case, it is clear that the hydrological model calibration step is able to correct the rainfall time shift, allowing to obtain good hydrological performances (KGE-Q) for a large range of R values. A similar consideration holds for KGE-P, largely influenced by the correlation coefficient. Conversely, 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~~river ~~—~~discharge ~~modelling~~simulation. 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 river discharge simulation~~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 bias ranging from $-10\% \leq \text{bias} \leq 10\%$ (Brown, 2006, Yang and Luo, 2014) to be associated with good satellite rainfall performances, but without a reference to justify these numbers.

Codice campo modificato

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 i~~o~~n rainfall and river 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 river discharge simulation is a result~~s~~ ~~of~~ from 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 river discharge ~~data~~ simulations over Europe. In particular, $SM2R_{ASCAT}$ is the best performing product for river ~~flood~~ discharge simulation across Europe (even at high latitudes).
3. There is a link between rainfall accuracy and river 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~~ addresses~~d~~ 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~~ river discharge 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 ranges of rainfall ~~metrics~~ performance scores ~~ranges here~~ defined here 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 river 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 river discharge. Indeed, as the main purpose of this study has been to reproduce the entire river discharge time series, any special attention to high/low flows was not paid. ~~In a further analysis,~~ A~~a~~ more comprehensive study ~~could~~ should consider a larger set of river 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.

Despite the aforementioned ~~limitations~~ limitation, this study ~~,~~contribut~~ing~~es ~~in the purpose of~~to a better understanding of the propagation of the satellite rainfall error to streamflow simulations~~,~~. This 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

Codice campo modificato

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). Finally, the authors wish to thank the two anonymous reviewer who with their valuable comments help to improve the manuscript.

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.

Codice campo modificato

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.

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.

Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A., & Arheimer, B. (2020). Lessons learnt from checking the quality of openly accessible river flow data worldwide. Hydrological Sciences Journal, 65(5), 699-711.

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.

Do, H. X., Gudmundsson, L., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. Earth System Science Data, 10(2), 765–785. https://doi.org/10.5194/essd-10-765-2018.

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.

Hofstra, N., Haylock, M., New, M., & Jones, P. D. (2009). Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *Journal of Geophysical Research: Atmospheres*, *114*(D21).

Hofstra, N., New, M., & McSweeney, C. (2010). The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate dynamics*, *35*(5), 841-858.

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.

Kyselý, J., & Plavcová, E. (2010). A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations. Journal of Geophysical Research: Atmospheres, 115(D23).

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.

Schaefli, B., and H. V. Gupta (2007), Do Nash values have value? Hydrol. Processes, 21, 2075–2080, doi:10.1002/hyp.6825.

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.

Zeng, Q., Chen, H., Xu, C. Y., Jie, M. X., Chen, J., Guo, S. L., & Liu, J. (2018). The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach. Journal of hydrology, 563, 106-122.

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.

~~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~~

Table ~~2~~1. 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	SM2RASCAT	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

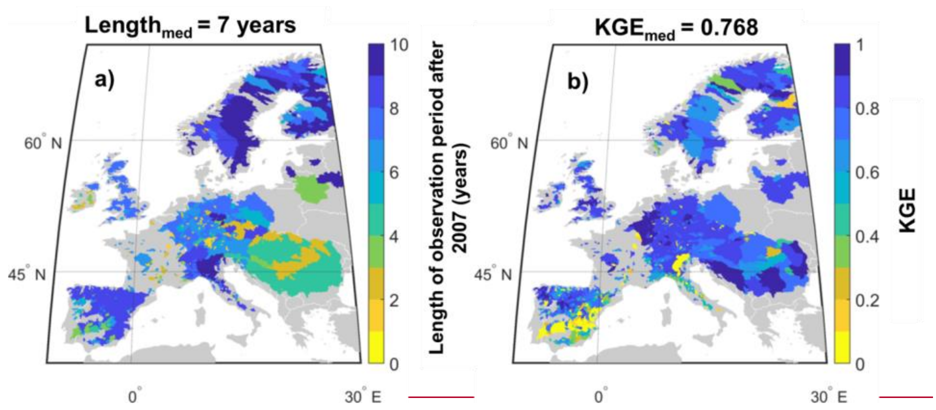
Table ~~3~~2. Performance scores for rainfall (in terms of rBIAS, R ~~and~~ RRMSE and KGE-P) time series computed during the calibration (in italic) and the validation periods. Rainfall performances are evaluated with respect to E-OBS rainfall data and distinguished between basins whose outlet section is below or above 50° latitude. It has to be noted that the more rBIAs, R, RRMSE and KGE-P values goes toward 0, 1, 0, 1 respectively, the higher is the agreement between E-OBS and SRPs.

Rainfall performances

Score / **Product**	**rBIAS**	**R**	**RRMSE**	**KGE-P**	**rBIAS**	**R**	**RRMSE**	**KGE-P**
	~~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)*

Table ~~4~~3. Median KGE-Q 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-Q 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~~over/above the TMPA area (latitude ±50°).

Score / Product	Spatial coverage of the product		~~below~~ TMPA area (latitude <50°)		above TMPA area (latitude >=50°)	
	cal	**val**	**cal**	**val**	**cal**	**val**
KGE-Q						
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**



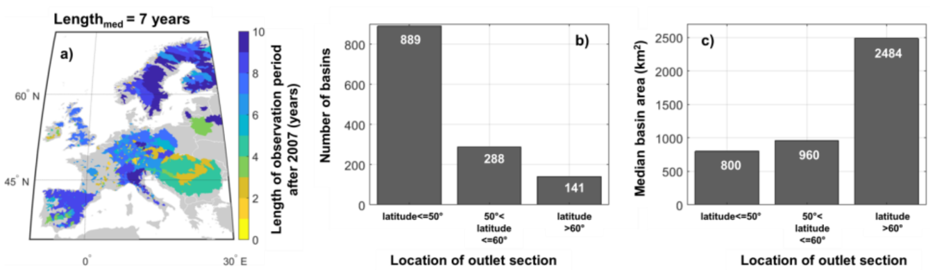
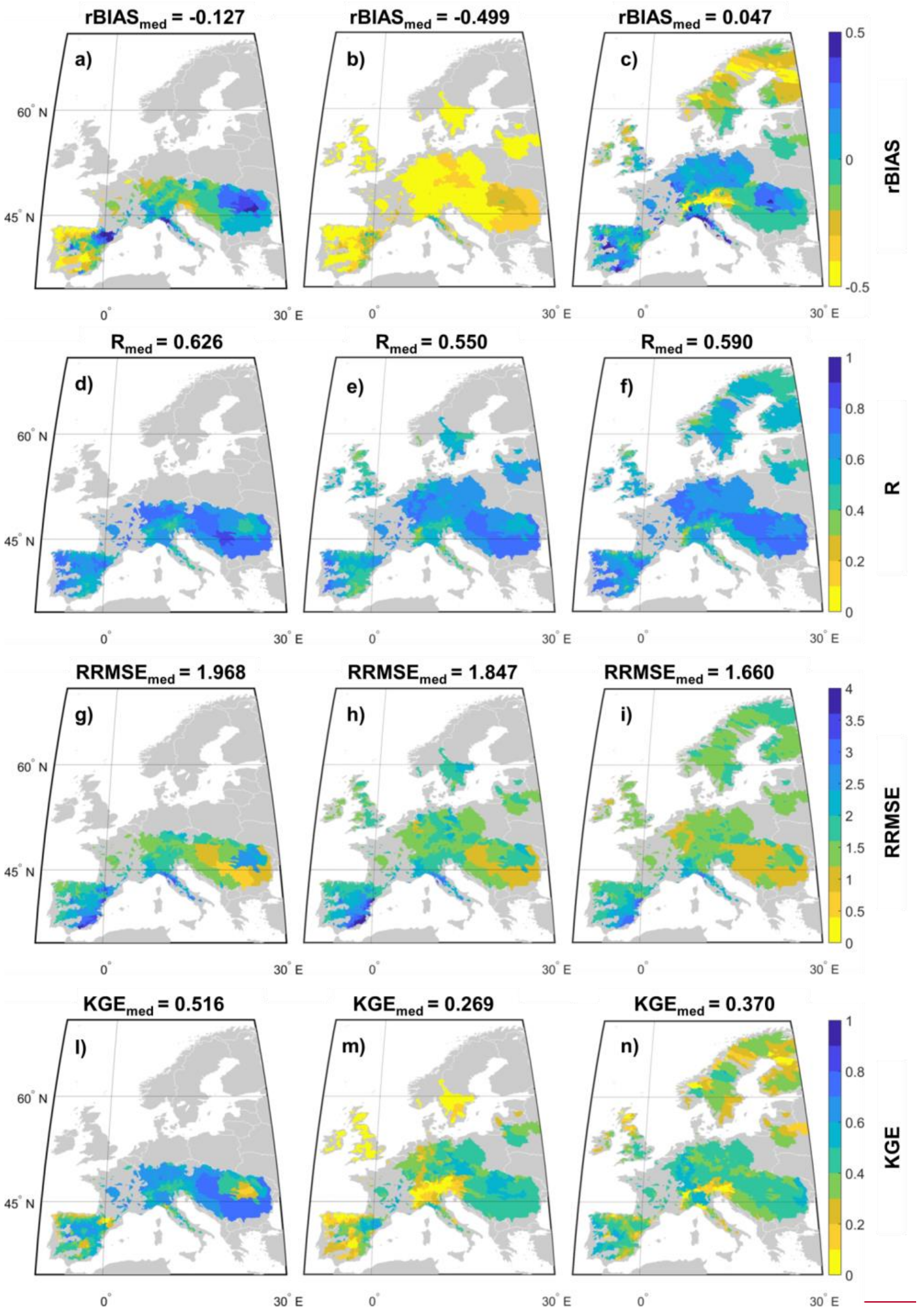


Figure 1. Location of study basins and ~~a)~~ length of discharge observation period after 2007 (a); ~~b)~~ number of basins (b) and median basin area (c) clustered according to the latitude coordinate of the outlet section of the basins.

~~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.~~


rBIAS_med = -0.127
rBIAS_med = -0.499
rBIAS_med = 0.047
a)
b)
c)
rBIAS
R_med = 0.626
R_med = 0.550
R_med = 0.590
d)
e)
f)
R
RRMSE_med = 1.968
RRMSE_med = 1.847
RRMSE_med = 1.660
g)
h)
i)
RRMSE
KGE_med = 0.516
KGE_med = 0.269
KGE_med = 0.370
l)
m)
n)
KGE
60° N
45° N
0°
30° E

838

839

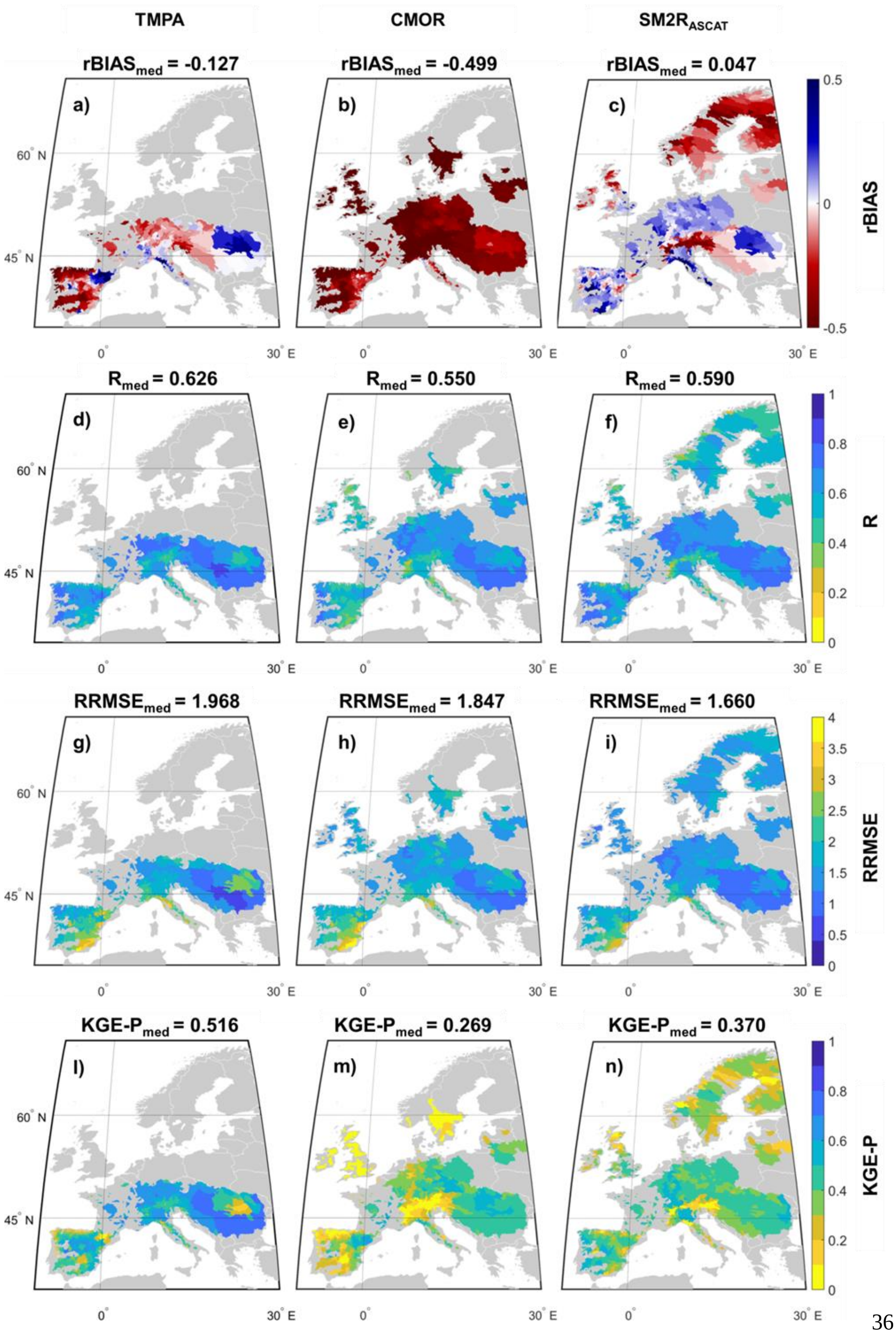
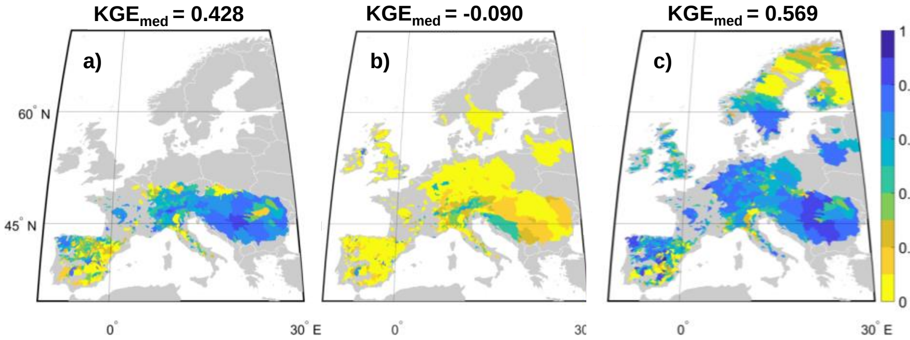

TMPA
CMOR
SM2R_ASCAT
rBIAS_med = -0.127
rBIAS_med = -0.499
rBIAS_med = 0.047
R_med = 0.626
R_med = 0.550
R_med = 0.590
RRMSE_med = 1.968
RRMSE_med = 1.847
RRMSE_med = 1.660
KGE-P_med = 0.516
KGE-P_med = 0.269
KGE-P_med = 0.370
a)
b)
c)
d)
e)
f)
g)
h)
i)
l)
m)
n)
60° N
45° N
0°
30° E
rBIAS
R
RRMSE
KGE-P

Figure 2. Performances of satellite rainfall during the validation period in terms of rBIAS (~~first row~~a, b, c), R (~~second row~~d, e, f), RRMSE (~~third row~~g, h, i), KGE-P (~~fourth row~~l, m, n) 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.


KGEmed = 0.428
KGEmed = -0.090
KGEmed = 0.569
a)
b)
c)
60° N
45° N
0°
30° E
KGE
1
0.8
0.6
0.4
0.2
0

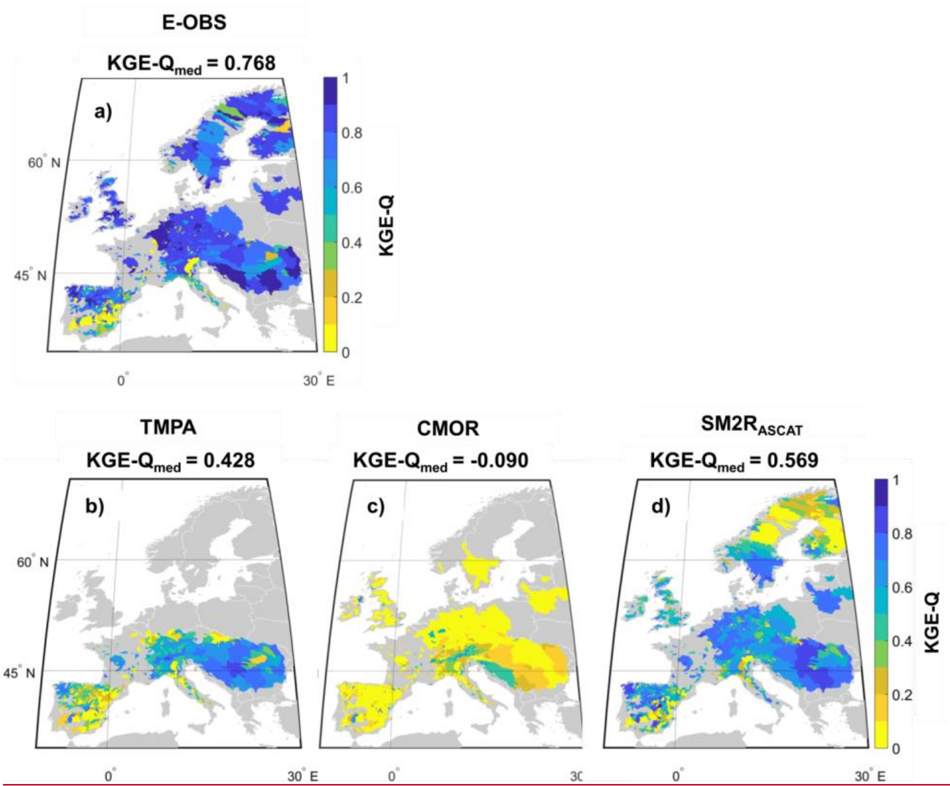

E-OBS
KGE-Qmed = 0.768
a)
TMPA
KGE-Qmed = 0.428
b)
CMOR
KGE-Qmed = -0.090
c)
SM2RASCAT
KGE-Qmed = 0.569
d)
60° N
45° N
0°
30° E
KGE-Q
1
0.8
0.6
0.4
0.2
0

Figure 3. Maps of KGE-Q index obtained by considering a~~, d~~) E-OBS, b)TMPA, ~~b~~c~~, e~~) CMOR~~PH~~ and ~~c~~d~~, f~~) SM2R$_{ASCAT}$ rainfall datasets. For E-OBS, KGE-Q index has obtained by comparing observed against modelled discharge data over the period 2007-2016. Modelled discharge data have been obtained by using E-OBS rainfall dataset as input to MISDc model. For the satellite data, KGE-Q refer to~~ in~~ the validation period (2013-2016). In a), -b)~~ and~~, c) and d) plots, the median KGE value averaged over the original product coverage is reported.

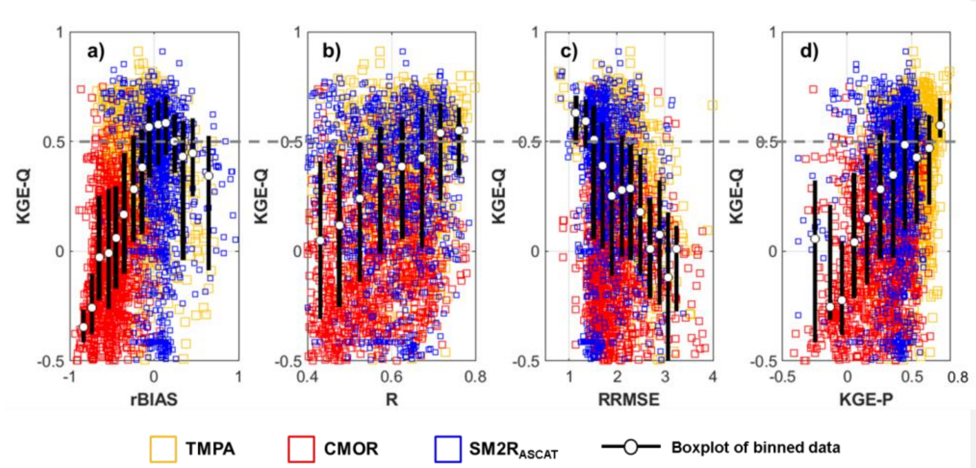


Figure 4. Performances of discharge in terms of KGE (KGE-Q) against a) relative rainfall bias, rBIAS; b) rainfall correlation, R; c) relative root mean square error of rainfall, RRMSE, d) KGE-P. The scores are evaluated for the validation period (2013-2016) for all the 1318 basins.

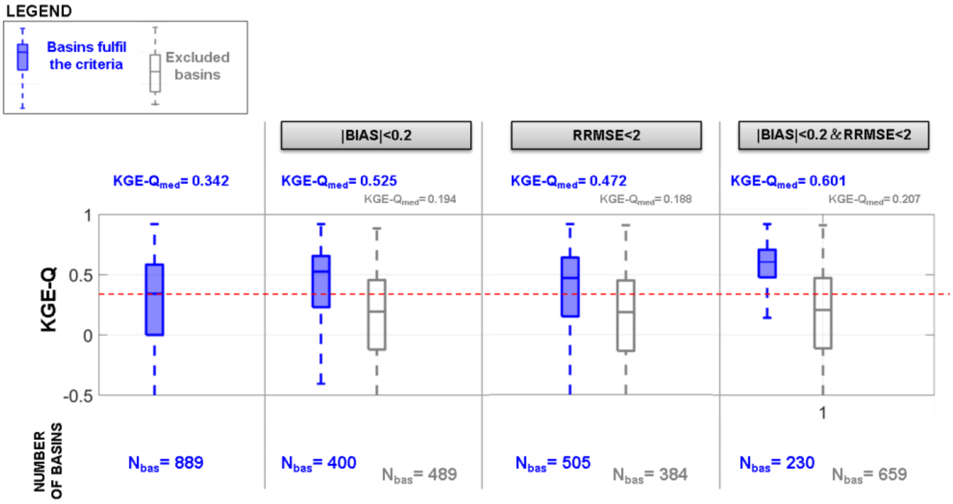


Figure 5. Hydrological performances in terms of KGE values ~~that can be~~ obtained during the validation period by the three satellite rainfall products for all the basins whose outlet section is located ~~below~~ over the TMPA area (889), a) without any constrain on the rainfall scores; b) constraining 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.