

## ***Interactive comment on “Which rainfall metric is more informative about the flood simulation performance? A comprehensive assessment on 1318 basins over Europe” by Stefania Camici et al.***

### **Anonymous Referee #2**

Received and published: 12 May 2020

**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.

C1

**Specific comments** – The authors claim in the title, introduction and methodology that satellite rainfall performance is evaluated in regard to flood modelling. However, no high flow specific analyses are performed and in their conclusions the authors themselves 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. – 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. – 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. – 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.

L62: Please clarify what do you mean with “gaining ground”? Are satellite rainfall observations used more often? Do they improve in accuracy? 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. 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. L204-208: Since the SM2RAIN-ASCAT product is relatively new, can you add a brief explanation of the SM2RAIN algorithm. L249: Can you elaborate what you mean with “QE-OBS is used as reference for parameter values calibration” L312/313: Please move this thought to the discussion. L341-343: Please check these values. They do not match the values reported in Table 4. L394/395: Please check the reference to the Figure. RRMSE

C2

is Figure 4c) and R is Figure 4b). 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. 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. 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.

Tables/Figures: If possible, the authors might want to consider including the supplement figures in the main text. Table 1: This table is not necessary. Instead a plot showing catchment area distribution might be more useful. 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 toward 1, 0, 0, 1 respectively, the higher is the agreement between E-OBS and SRPs." to the table caption? 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. Figure 3: Plots d, e and f are mentioned in the caption but are not part of the Figure. 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.

C3

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.

L25: "understanding how uncertainties[...]" L42: "Results suggest that, among [...] are not reliable scores to select the best performing rainfall product for hydrological modelling[...]" L58: "Generally, rainfall observations [...]" L63: "[...] (SRPs) has boosted their use [...]" L72: "is used to simulate a discharge time series [...]" L86: "Generally, this comparison [...]" L98/99: Missing bracket L108: "by using the MIKE SHE model [...]" L112: "it is difficult to find literature [...]" L140/141: "is composed of 1318 basins [...] over the whole of Europe [...]" L142: "The European continent [...]" L145: "gently slopes towards [...]" L150: "the Alps generally has higher rainfall amounts [...]" L154: "prevailingly subject to [...]" L155: "according to the latitude[...]" L157: "and for about 11% [...]" L161: "basin characteristics." L175: "an European daily dataset [...]" L190: "period in 2012[...]" L201: "provided by the CPC [...]" L219-221: "applied to carry out [...] model composed of a component [...] of soil moisture and a rainfall-runoff model [...]" L233/234: "allow us to consider the model suitable for the purpose of this analysis." L236: "analysis regards the quality assessment [...]" L270/271: "The more R [...] , respectively, the higher is [...]" L275: "(TS). POD reports [...]" L278: too many dots L356: "This is the first notable result [...]" L368: "for the CMOR product[...]" 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. L372: "of this product in terms [...]" L380: "in terms of RRMSE [...]" L454-457: This sentence would benefit from commas. L456: "errors in rainfall [...]" L498: "limitation, this study contributes to the better understanding of the propagation of [...] simulations. This could be very [...]"

References Do, H.X., Gudmundsson, L., Leonard, M. and 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), pp.765-785.

C4

