

This manuscript greatly improved after the first round of revision and now illustrates the capability of different rainfall datasets to represent high-impact dry and wet spells rather nicely. Particularly the discussion of good and bad performing datasets in view of their data sources is very informative and this paper importantly illustrates the difficulties in deriving information on drought and rainfall extremes, even from the wealth of rainfall datasets available nowadays. It is therefore an important addition to literature, particularly for West Africa where analyses of extreme weather metrics from available rainfall observations become more and more important but reliability of datasets for such applications is too rarely questioned.

Unfortunately and in spite of the authors stating that the manuscript was properly proof-read and even checked by a native speaker, non-cosmetic corrections are still necessary in high density. I would urge the authors to read the entirety of this manuscript again – to actually do it and to add all the –s and cross the ts. It is rather tedious for a reviewer to do this job and was very much not enjoyable. We all know that if only half of those errors make it through to publication it reduces the trustworthiness and readability (and therefore the impact) of this study, which is completely avoidable.

I have a couple of major comments regarding the dry spell results that need to be clarified and the minor comments need to be addressed before I can recommend this manuscript for publication.

Major comments:

- I have methodological concerns (1) regarding the definition of dry spells causing large spreads at the end of the year, inconsistent with early in the year (Fig5) and (2) whether the inclusion of the dry season into metrics skews the aggregated results for DSC10, DSC20 and DSx1 (Fig 4,6,7) away from anything that is relevant regarding “drought hazard” or “high impact event” if they are dominated by the dry season signal. Please clarify the behaviour of (1) and illustrate that the dry season sensitivity is not dominating Fig 4,6,7 results (I.e. I’d like to see Fig4,6, for ‘rainy’ months only, say April-Nov or similar).

General comments:

- Different from the author’s answer, there were in fact no additional maps of any metrics added to the supplementary material that would illustrate the spatial pattern of discussed metrics. Could you please add those as mentioned in the response?
- Please add letters to figures and refer to panels in plots consistently throughout the manuscript.
- Where possible provide some key quantification of discussed discrepancies and biases. Currently the analysis remains predominantly qualitative.

Minor comments (those corrections are unfortunately not exhaustive!):

P1 line 6 remove first “same”

P1 line 10 “while,” -> meanwhile

P1, Line 10: why does this suggest an “early cessation” of precipitation – what would be a normal cessation? Doesn’t this rather say that rainfall is more intermittent during the cessation of precipitation (“false onset” expression captures that)

P1, LL 10-11: remove “while,” , the strongest contrasts between the data products [...] observed _for_ the amplitude

P1, Ll 12-13: remove “quite similar”, better to say “show a comparable/similar frequency of wet sequences”

P1, L 18 thE Sahel

P2, l 7: “will increase significantly [...] during the lean season of 2018” what does this mean? Is it projected to increase in the future or did the number increase in 2018?

P2 ll14-15: “these hybrid rainy seasons, illustrating a rainfall regime intensification , **ARE** part of”

P2 l34 remove “although”, it’s “however”

P3, l7; l28; p4 l3: techniques (same for all other cases of “technic”)

P4 l3: remove “is” from “is uses a moving neighborhood”

P4 l5: centered **AT** the block mean

P4 l7: “has a minimum variance of estimation error” - does that mean it minimises the variance of the estimation error? In any case, please clarify

P4 ll 9-10: This should be changed to “we use the kriging RMSE to mask out areas [...]” etc. The part with “can be possibly masked out” is not fitting since Fig1 indeed shows vast masked out areas, so it is applied rather than “can potentially be applied”

P 4, l32: resampled to 0.25 (remove <), same p5, line 1

P5 l16: what kind of sensitivity is meant here and how is the analysis of those datasets related to future impacts?

P5, l19 used

P5, l20 precipitation IS lower

P5, l23 descripted -> shown

P5, l23-24 during a specific period(s) - remove s - and **IS** called

P5 26 when the rainfall IS not sufficient ... and therefore DOES not provide..

P5, l28 The results presented in THIS study

P5, l29 from the other durationS

P6, 15 “and because of the synoptic systems associated..” this should be better explained and the characteristics of MCSs (e.g. propagating and therefore extremes rarely stationary etc) mentioned. It’s otherwise not clear for people who don’t work in the region

P6, 110-11 “these periods are defined according to the different synoptic components that drive the rainfall variability” again, this is very vague. If there is related reasoning it should be stated explicitly, and those factors at least mentioned (preferably with a reference about the importance of that factor)

P6. L21: remove “the” from a south-north gradient

P6 122 in termS of

P6 122-23 It would be good to state in the text that this is rainfall values for June-October, possibly right in the introductory sentence of the “seasonal rainfall” section

P6 123 closeD -> remove D

P6 124 kriged observed precipitation **datasets** (better: in-situ datasets or rain gauge datasets etc)

P6 125-26 Our results from CMORPH (..) - there are several language problems in this sentence, please correct (confirmS, “which” showed, “these” precipitation)

What is the result here for CMORPH? I assume this refers to lower seasonal precipitation compared to the gauge-based products but it’s not stated. It would be useful to quantify “the results are close” or “underestimating” by giving a percentage range for the rainfall differences between those datasets, or correspondence in pattern correlation or anything that underpins the qualitative statements in this section.

P6 132-33 “When looking over smaller areas differences are more important and any of the products is able to get this structure even if their bias stay low” Please correct (language problem) and clarify this sentence, which region this refers to and where biases stay low. I’d assume this means something like “Regional-scale patterns in rainfall are of particular importance. All products seem to approximately agree on the magnitude of spatial rainfall variation. Such variation is particularly pronounced across the peanut basin, for which the bias between rainfall products is low” - again, can “low” be quantified? It seems difficult to assess those statements by just eyeballing the maps and no indication of what the authors refer to.

P7, 18 closest to BK in intensity - can “closest” be quantified, just to give the reader some idea what the magnitudes here are in terms of biases, agreement etc.

P7, 114 Does this paragraph now refer to Fig3? Reference missing

P7, 119 accurate productS – remove S

P7 20-21 On P6, 25-26 it says that CMORPH misses local convective rainfall between scans, resulting in somewhat lower seasonal total rain, and here it says it tends to overestimate small (low-intensity?) precipitation, where the authors say “which would explain why the difference appears here but not when looking at the cumulated rainfall”. This can be confusing and would

be worth clarifying. I think this says that high-intensity rain dominates the wet season, of which CMORPH misses events in-between scans, but low-intensity events during the dry season are overestimated. But please state this more clearly.

P7 l23: in terms of, better would simply be “This is also visible for the cumulated rainfall”

P7 l27 finalLy

P7 l29: it is a difficult – remove “a”, better than “to find the reasons” would be “to suggest an explanation for”

P8 l7 different types, depending on their

P8 l8-10 “In the main document”.. and reference to supplementary can be shortened to “We focus on..” with (see Table 2 for the definitions, further results in supplementary material). “Nevertheless [..]” can be dropped.

P8 l13: dry days is Fig 3, not Fig4

P8 l13 “This is in agreement with the previous result” - as the authors show later on, this is not in agreement regarding the CMORPH / TRMM behaviour, which so not agree for the dry days.

Fig 4 caption: “Boxplots of **the** average number [..] the **left and right** edges of the box” this should be bottom and top edges. What does “extreme values” for the whisker position mean. Min and max? Is this really per year or again from June-October like Fig2? If it is per year, wouldn't the dry season performance shown in Fig2 predominantly affect those extreme dry spell indices?

P8 l16 “**than** TAMSAT and CHIRPS” replace with “as”

P8 l19 cloud top temperature

P8 l20 MO was already introduced in l15

P8 ll 21-22 This can explained ...compared to the observations -> This may explain the relative good performance [..] compared to the gauge observations

Fig5: how is this frequency defined? Description in caption and text just says “seasonal cycle of dry spells” without further specification. Also, why are there such inconsistencies moving from Dec to Jan? Particularly visible for DCS10, 20 and DSx1. Is there in problem in how the dry spells are identified at the end of the year? Must be a methodological issue that the spread is large in Dec and gone in Jan. Does this affect the aggregated metrics in the other plots?

P8 l25: It is a very important point that those dry spell metrics are so strongly affected by the dry season and should be pointed out much earlier in the manuscript. While the behaviour of the datasets during the dry season is interesting (and sufficiently shown in the seasonal cycle plots), the importance of dry spells depends on whether they appear during the wet or dry season. For example, Fig4 shows that DSC20 is around 1 or below per year, questioning the usefulness of this metric in the hazard context. It suggests that this metric reaches “1 occurrence per year”,

which likely reflects the dry season - this is not very interesting and not reflecting an “extreme event”. On the other hand, it would be an important information if this event occurred once a year during the monsoon season. How much are the dry spell results skewed towards rainfall dataset dry season skill (affected by low-intensity precipitation breaks rather than MCSs)? Why weren’t the non-seasonal cycle plots restricted to June-Oct (or at least months outside the dry season)?

P8, 131 in agreement FOR the observations

P8 31-34 “The evolution of DSI **is also interesting by focusing** on relative **mild droughts with specific durations that are sensitive to dry spells** during the onset and retreat phases of the monsoon. This detection is, **by far, the more variable** from one product to another. For this **specific drought** it is difficult to **distinguish specific behavior** of a group of products. Each possesses a **specific time evolution** [..]”

Please improve wording. [..] is also interesting as it represents/characterises relatively mild droughts with a fixed duration. This metric is most sensitive to dry spells during [..], and is by far the most variable [..]. For this dry spell metric, it is difficult to distinguish any specific behavior [..]. Each possesses an individual time evolution [..]

P9 113-5: **gauge** observations (the difference to satellite observations is otherwise not clear). Indeed, the difference between the interpolated gauges is remarkable and, if ignoring ERA5, almost as large as the spread between the satellite observations. Again, it would be worth to quantify this uncertainty in the text. Looking at DSI at the high at the rainy season between Aug-Sep, the frequency difference between BK and OK is around 20%. The dry day frequency increases by more than 100% just changing from BK to OK, based on the same set of stations. Please be more explicit in numbers about statements rather than to rely on handwaving only

P9 110: spatial datasets -> gridded datasets

P9 111: are providing in -> are provided in THE

Fig6 caption: it should be BK which is mentioned as reference dataset here. What is the x-axis? If it is standard deviation too the ticks should be similar to the y-axis.

Is it correct that this diagram was calculated from the spatial maps (like Fig2) of those metrics and e.g. spatially correlated? Which leads me to the question why no metric map was added to the supplementaries (contrary to what was stated in the reviewer response)?

Again this relatively good aggregated agreement may be artificially boosted by including the long dry season. What would this look like for the rainy period only (or say April-Nov?). I think it doesn’t reflect well what was shown based on the seasonal plots and distracts from the fact that discrepancies are large when it’s most important.

P9 l12-13 “For the DSC10 and DSC20 and the DSI there is no clear difference amongst the datasets. However, DSC10 is more sensitive to the datasets.”

Is this supposed to refer to DSC20, DSI and DSxl, which all sit in the area of low standard deviation? The spatial correlation for those metrics seems rather low compared to DSC10

P9, l22 similitude -> similarity

P9 l23-24 “Finally, DSI displays a specific time evolution.” -> displays a time evolution that seems distinct from the other metrics?

P9 l32 observations -> in-situ / gauge observations

P10 l3 I would suggest the authors add lettering to their plots and refer to Figx a,b etc throughout the manuscript. That would make it much easier to follow which panel is being discussed without having to check and recheck the acronyms.

P10 l7 I think that should read “WS1 99P” instead of WSI

P10 l10 “This distribution shows to see tipping points on daily rainfall.” language problem, please rephrase

P10, l21 except to the -> except FOR the WSM

P10 l24-25 contributes bias correction -> allows for such biases to be taken into account

P11 l3 in-situ observations

P11 l4-5 are more likely to be compared with -> are more likely to be comparable to gridded [..]

P11 l25 the monitoring [..] are compared -> the monitoring [..] is tested OR the representation [..] is compared

P11 l26 3 products BASED on raingauges

P11 l27 by upgrading or -> by area averaging, interpolation or[..]

P11 l29 THE large-scale climatology

P11 l33 for an average rainfall like most of -> remove “an”, like FOR most of

L33 this good agreement start to dissipate -> startS

P12, l2 “It turned out that each of the kriging methods were positioned in these groups.” -> Interestingly, from the kriging methods each falls into one of these groups.

P12 l10 “However, there is less agreement between the different data products for dry spells than for the wet spells.” Shouldn’t this be “there is MORE agreement for dry spells than for wet spells” ?

P12 114 “record the rainiest days but minimize these high rainfall events.” -> record highest rainfall intensities but show lowest rainfall frequencies (?), otherwise please clarify what that means.