**Reviewer #1**

**General comments**

This paper analyses the occurrence of wet and dry spells in Senegal estimated from a set of precipitation products (in situ, satellite and reanalysis). Although the purpose of the paper is valuable and results potentially significant, the overall quality of the analysis and presentation is below acceptable standard.

We first thank the reviewer for its valuable contribution and evaluation of the document. We have modified and corrected the document as suggested to reach the standard of the journal. Significant improvements have been done and clarifications have been included. Please find the responses of the general and specific comments in red.

1) Language needs substantial revision and improvement, by a native speaker if possible;

we have checked the document with a native english speaker.

2) The presentation of the method (description of kriging in Section 1.1 and definition of Wet and Dry spells in Section 1.3) needs improvement, the description of the methods should be expanded and the presentation clarifyed;

These sub section have been modified and additional informations have been provided.

3) The presentation of the results is often unclear, and results themselves are not discussed enough: Section 2 is basically a description of the figures, and the keypoint of the paper, i.e. the comparison of the precipitation datasets, is never actually addressed;

First, the figures have been modified. We have also modified the section 2 with more discussions on the results and less descriptions. That implies more discussions about the comparison of precipitation. Nevertheless, the reviewer mentions the keypoint of the paper is the comparison of the precipitation datasets. This is not correct. The main objective is the comparison of the detection of wet and dry spells, that are more specific that precipitation comparisons. We have clarified this point in the objective of this study.

4) Figure captions should be improved;

Modified as suggested.

5) Conclusion section is not conclusive at all, it is just a summary of the paper.

We have improved the conclusion section by providing the keys messages of this inter-comparison. That results are really relevant for users and researchers working over these regions or with theses datasets and provide the uncertainties that are related to the monitoring of extreme wet and dry spells.

For the above reasons I cannot recommend the publication of the paper in NHESS. I suggest the authors to undertake a substantial revision of the paper, and resubmit it for new consideration. I list below my specific comments that I hope can help the authors in improving the paper.

We thank again the reviewer for providing these useful comments . We took into account all of them to substantially improved the quality of this study. These changes increase significantly the quality of the paper and, we hope, they allow to reach the standard of the journal.

**Specific comments**

Page 1, line 6: “more variability”, do you refer to some intra-dataset variability? If this is the case, this expression is not suitable, please rephrase.

Indeed, the term "more variability" is used to illustrate the differences between products.The sentence has been rephrased to clarify.

Page 1, lines 9-10: these lines are unclear.

All the paragraph has been rewritten to correct an to clarify the text.

Page 1, line 15: add the reference to Taylor et al. 2017, https://www.nature.com/articles/nature22069.

Modified as suggested

Page 1, lines 15-16: this sentence should be moved at line 19, before “Recently extreme events: : :”

Modified as suggested

Page 2, lines 25-27: sentence unclear.

According to the reviewer comments, all the paragraph has been rewritten.

Page 2, line 30: when stating the objective of the paper, it is not clear that you compare satellite/reanalysis products to rain gauges, which you consider as reference. Please clarify this point.

We thank the reviewer for this comment. Indeed, this important point need to be clarified. This paper aims to make an inter-comparison between several products from observations, satellite data and models. More specifically, the objective is to compare them on their ability to detect potentially high-impact dry and wet events occurring in Senegal. However, according to literature and the data characteristics, there is no ideal product, especially when studying strong to extreme events. Indeed, the network of surface observations is too coarse, the products derived from satellite datasets are sometimes unprecise either because of approximations (due to some proxy used) or because of orbital characteristics (revisit period, resolutions), and finally the models are well known to have uncertainties, especially over the tropical areas. Thus, it seems difficult to say which product is the reference even if ground observations are expected to be an ideal candidate. However, the uncertainties associated with krigging technics, highlighted in our study, show that even these products should be used with caution.

Page 2, line 33: are wet and dry spells “extreme hazards”?

Dry and wet spells are defined with different intensities and durations. Thus, events can be considered moderate to extreme for the highest intensities and/or the longest durations. This work is a first step in identifying potential high impact events. It is therefore important of having a large sample of different events. However, most of the results presented in the study focus on events that have a return period of about ten years. We have clarified this point in the text.

Page 3, line 1: DS, DSC, WS and WSC are not defined in the text.

Sorry for this mistake. Every wet and dry spells is now well defined in the text.

Section 1.1: do rain-gauge time series have passed any objective homogeneity check?

The rain-gauges used in this study have passed two levels of quality control. The first one is a manual check for suspicious records was carried out and then additional checks were carried out, including verification of station locations, identification of repeated data, identification of outliers, comparative tests using neighbouring stations and search for suspicious zero values (missing data or zero precipitation). This information has been added in the text.

Section 1.1: BK is poorly described.

A better description of the technical procedure to generate block krigging is now added. This description gives sufficient information for the reader to understand the technic. The details of this technic are described in supported litteratures cited in the new version.

how the lambda weight are assigned?

The basis of kriging is the variogram which is a function of the variance and distances between stations. This function gives two properties, the nugget effect which corresponds to the limit of the variogram in zero. The nugget effect represents the variation between two very close measurements and the range that corresponds to reaching a plateau indicating that there is no longer any spatial dependence between the data. Depending on these two parameters we have the screen effect: the nearest points receive the most important weights (lambda). This screen effect varies according to the configuration and the variogram model used for kriging. The greater the nugget effect, the less screen effect there is.

Page 3, line 18: why kriging reduces high values and increase low values? Please explain. Equation 2: what is xq?

Kriging tends to greatly smooth intensities. This effect is well known and tends to reduce extreme wet and dry events and converge to the mean values, especially when the network of rain-gauges is coarse. In consequences, the intensities are reduced as well as the occurrence of zero rain. The effect is similar to a smoothing when applying mean values.

Equation 2 is a bias correction using the quantile mapping method. This method corrects the underestimation of high rainfall intensities by kriging.

xq are the kriged data values; xo are the observed data values. The purpose of quantile mapping is to bring the two Cumulative Distribution Function (CDF) together. By correcting the distribution of the kriging outputs, percentiles and high intensities are corrected as well.

Page 3, line 22: square root of OK variance is used for what?

Square root of variance is a derived product for the kriging technic. It is also commonly called kriging error. It was used with a threshold of 0.5 to identify regions with low rain gauge density. Looking at the results in Figure 2, it can be seen that the eastern central part of the country is ignored because the error on kriging is less than 0.5. This choice is essential for the reliability of the data collected.

Section 1.2: there are discrepancies in dataset resolutions in the text and Table 1, please check.

Corrected

Page 5, line 9: where did you define wet and dry spells as extremes? “A maximum number of definition”, this sentence is unclear.

We agree that these sentences were not clear. The entire paragraph has been rewritten.

Page 5, line 16: on which statistical basis do you define “DS extreme long”? You should use the “extreme” word carefully, and only after an analysis of distributions.

We agree with the reviewer, and the term “extreme” is now used with more caution. Nevertheless, it is worth noting that the selection of dry and wet spells is based on the PDF of the precipitation from 1991 to 2010 to select some of the most extreme events. It is, obviously, depending to the criterias but most of the events have a 10-y return period.

Page 5, line 20: what do you mean by “DSC duration is known”? Please clarify.

We have clarified that DSC have fixed durations, from 5 to 20 days.

Page 6, line 12: why BK is more comparable to averaged values and more in agreement with satellites?

According to the literature (Lloyd and Atkinson, 2001; Maidment et al., 2013;panthou et al., 2018), the two methods of kriging are different: Ordinary Kriging (OK) estimates rain as a punctual value while Block Kriging (BK) estimates rain on space blocks with an average value rainfall. Rainfall satellite estimates are closer to an average value over a space block than a punctual value. For this reason, the BK should be closer to satellite datasets.

Page 6, line 14: please explain why radar on board of TRMM explain TRMM performance.

The TRMM satellite is the first satellite with an active radar instrument onboard. It is a powerful added value since it provides a profil of the rainfall activity. This is especially important over tropical region where unsateurated downdraft and evaporation of the rainfall is important. It is also important since the rainfall estimates is not based on a proxy of the top of the convective cells. This estimation is quite common and derived from passive radiometer but shows some bias (overestimation) during the collapsing period of the convective cells. This point is now clarified in the document.

Page 6, line 15: please improve the description of the Peanut Basin or highlight it on the map, to facilitate the reader to locate the region.

Modified as suggested

Page 6, lines 20-21: This actually means that during dry season date are not collected,

isn't it? In this case you cannot consider these days as dry.
This is correct. We have clarified the document as suggested.

Page 6, line 26: does "variability" refer to the datasets? I suggest to find another expression in this case. How can I compare seasonal and intra-seasonal variability in Fig. 3?
These sentences have been clarified. The term has been changed as suggested to clarify this point.

Page 6, line 29: datasets are actually four.
Corrected as mentionned

Page 8, line 4: "depth of wet day"?
Modified by "rain amount of day"

Page 8, lines 22-24: here again you discuss TRMM performance, but discussion should be expanded.
It is obviously difficult to conclude that the radar explains all the results. But it is worth to note that active instrument provide a more accurate datasets. This is especially true for the wet events and not for the dry ones.

Page 9, line 8: sentence unclear.
We agree that this sentence were unclear and the entire paragraph has been modified.

Page 9, line 19: how do you define "extreme rainfall deficits"?
This was used as a synonym of the dry spells but we agree that can generate some confusions. The sentence has been rewritten.

Page 9, line 22: according to Table 1, not all the products are upgraded (BTW this is just regridding, so another word should be used).
We define upgridding when the regridding is done to a coarser resolution. For most of the datasets, the regridding is actually an upgrading. Nevertheless, we have clarified that some datasets are just regridded.

Page 10, line 5: as in the Abstract, please clarify this sentence.
We agree that these sentences were not clear. The sentence has been modified.

Page 10, line 14: same sentence on TRMM performance, but no discussion.

According to the comments of the reviewer. We have modified and clarified all the sentences related to the radar of TRMM. See previous comments.

Figure 4: what is depicted here? Yearly averages?

This is the average number of dry spells (DS and DSC) per year collected on all grid points. Thus the boxplots illustrate the spatial variability of these indicators. The caption has been clarified.

Figure 5: what is displayed here? What does the y-axis refer to?

This figure illustrates the seasonal cycle of occurrences of dry spells (DS and DSC) from 1998 to 2010 over all grid points. This occurrence was previously indicated in term of number (depending the number of years and grid cells used). In the new version of the document, the frequency of the event, in relation to the total number of event, has been provided. This indicator, in percent, is more understandable.

Figure 8: what do the y-axis refer to? Are monthly values represented?

Same answer as with Figure 5. Again, the new figure and caption, on the frequencies of these events, facilitate the comprehension of these results.

Table 1: check table header.

Modified as suggested

Technical corrections

Language is below acceptable standard, and the paper requires a substantial revision in this sense, therefore I omit here to indicate individual issues.

The entire document has been revised and clarified. A native English speaker has corrected the English.

**Reviewer #2**

**General comments**

- The language of this manuscript is imprecise and altogether not of acceptable quality, which makes it difficult to review in the first place

We understand the reviewer's opinion. To correct these errors we had the entire article proofread by a native English speaker. We have thus made significant changes to the text to ensure that it is in line with the journal's requests.

- Interpretation and putting into context of the presented work / comparison with existing literature on dry and wet spells in the region is essentially entirely missing

We partially agree with the reviewer. Significant efforts have been made to better contextualize this work in relation to previous work. However, the bibliography on this topic in this region is very limited and we have focused on the articles that are relevant to this study. Through the review, no specific study is proposed. We hope that the integrated articles will be satisfactory for reviewer.

- The goal of evaluating the performance of different datasets for dry spell and wet spell identification was not achieved as there is hardly any information that goes beyond a pure description of given plots, leaving it to the reader to come up with a conclusion

Here again, we partially agree with the reviewer. Indications on the quality of detection of wet and dry spell have been made in this study and, in the conclusions, recommendations are made. However, we agree with both reviewers that it is necessary to do a better synthesis work to avoid too much descriptions and to discuss in more detail the conclusions and origins for these results. This is why a large effort has been made in section 3 (results, simplified in the new version) and section 4 (discussion, where the consequences and possible reasons for these results are discussed). The conclusion section has also been modified to highlight the most significant results.

- Figures need to be revised (axis, overall readability, caption descriptions, annotations and plots cut off)

All figures have undergone quality control to meet the requirements of the journal.

- Conclusion only summarises plots all over again

As previously mentioned, the conclusions has been entirely rewrote to focus on the key results of this study.

**Specific comments, which are not exhaustive:**

p5 Methodological approach: This section is too short – were the metrics calculated on the entire time series / per month / per pixel etc? Please also say something about the usefulness of those metrics. For example, WS99P seems to be an unnecessary metric as, per definition, the "number of extreme days" is will be 1% of the number of identified >=1mm wet days.

We agree, this section needs to be developed, particularly on the indicators developed. Indeed, all DS and WS indicators are calculated for each pixel from the daily data.

The usefulness and the objectives of having a large spectre of indices (from mild to the most extreme definitions) is now better justified. WS99Ps are the most extreme events we could detect in our analysis. Because of their potentially extreme impact we have to include them in our analysis even if they are rare.

- Could you state the rationale behind looking at 90 - 99th centile 'wet spells' only, rather than including lower thresholds that are agriculturally relevant?

That is a relevant question. The purpose of this paper is to provide a wide range of potentially high-impact event indicators. Thus, we focused on potentially impacting wet (WS) and dry (DS) events. DS are studied because their presence and duration can generate a rain deficit and droughts that can impact yields. WSs are able to destroy seedlings and crops through heavy rainfall or subsequent flooding. The detection of DS is based on previous studies that define these periods as periods of non-precipitation.

- Please show the number of rainy days per dataset as that's what the other metrics are based on

This information is already provided by the distribution of dry days (Figure 3) and the distribution of cumulative rainy days (Figure 10). The text has been revised to clarify this point.

p5 ll 25-26: "The duration categories of wet spells are chosen to correspond to the different synoptic systems causing rain in West Africa (Froidurot and Diedhiou, 2017)." What does that mean? How does it address the "different synoptic systems"?

In this sentence, we justify the durations of the wet spells according to previous studies that highlighted synoptic origins of the rainfall. One of the most important drivers of the mesoscale convective systems is the African Easterly Jet. This perturbation generate a signal at around 3 to 5-d period. Nevertheless. We have completely modified this paragraph to clarify this point.

p6 ll13-14: why does the TRMM radar explain this?

The TRMM satellite is the first satellite with an active radar instrument onboard. It is a powerful added value since it provides a profil of the rainfall activity. This is especially important over tropical

region where unsateurated downdraft and evaporation of the rainfall is important. It is also important since the rainfall estimates is not based on a proxy of the top of the convective cells. This estimation is quite common and derived from passive radiometer but shows some bias (overestimation) during the collapsing period of the convective cells. This point is now clarified in the document.

p6 p15: Please indicate on the map where the Peanut Basin is
Indicated in the map as suggested

Figure2: While larger patterns are reflected in the datasets, local differences in the transition zone can be large. Please consider to show all datasets as a difference from OK (since OK is used as reference dataset later on) to help the reader spot biases more easily
Modified as suggested

p6 17-19: how was this % of dry days calculated? Is it this per day and pixel or is a "dry day" when the entirety of Senegal is < 1mm?
Thank you for the question, this percentage of dry days is the number of dry days compared to the total number in each month averaged from 1998 to 2010 and on all pixels in Senegal. Note that a dry day is defined as any rain <=1 mm. This has been clarified in the caption of the figure and in the text.

- Generally, please consider adding 1-2 further maps that can take into account the extreme rainfall gradient and illustrate regional differences in those metrics
As suggested by the reviewer, additional figures have been added in the supplementary material to address this point. Nevertheless all types of uncertainties and case cannot be discussed and plotted in the main document.

p 6 l24: why is this a paradox? Also, what is the take-away message for the reader from this section?
The paradox is illustrated by the fact that the products with the lowest number of dry days in our study (namely TAMSAT and CHIRPS) tend to underestimate seasonal accumulation. This is particularly true in the South. The take away message, related to the uncertainties of the detection of wet and dry spells, and about the quality of each product, is clearly provided in conclusion.

p6 l26: " Fig. 3 also illustrates a higher variability from the seasonal scale to intraseasonal scale." please explain more clearly

The idea is to show that despite a fairly good agreement on the spatial distribution of seasonal accumulation, the intra-seasonal distribution of dry days shows larger disparities between products. The sentences have been rewritten to clarify this point.

Figure 4 has boxes cut off – please replot

Corrected as suggested

p7 ll11-12: “The seasonal cycle of dry spell shows slight differences between products which confirm that this events characterize false start and early cessation of season in Senegal. “ Please explain more clearly

In this sentence, we relate the detection and the anomaly of dry spells and some other characteristics of the rainy season, namely the false start (when a dry spells occurs just after the first rainfall event of the season) and cessation. As mentioned previously, this paragraph has been rewritten to clarify this point.

p7 ll13-14: what does illustrate the severity of DSC and DS? What is this severity?

These rainy breaks (dry spells) can be detrimental to yields, especially when they occur at the beginning of the season, causing farmers to lose seedlings. In addition, they can occur in the middle of the crop maturation season to generate water stress by lowering the WS (Water satisfaction) of the plants. The severity of dry spells defines their potential to be dangerous. Due to the nature of the precipitation values, it is not possible, unlikely wet spells, to assess this severity by using the intensity of the rainfall during dry spells. So they are only estimated by the duration of events. In this study, we have selected the longest and therefore the most severe DSC and DS.

P7 l16 It’s rather within the seasonal cycle than at a given date

Corrected

Figure 6: what is the x axis? Where does the “0” belong? Please provide complete descriptions in your figure captions

Figure 6 is a typical Taylor diagram. Radius expresses the standard deviation, the angle the correlation and the distance from the bottom right point, the RMSD. The caption has been clarified as suggested.

Figure 10: why does the daily rainfall only start at about 15mm per day?

Sorry for the missing information. We started the analysis with precipitation above 10mm. This is now mentioned in the text and in the caption.

P9 l5: worth to note that WSM 99P is only the rarest because it is the only defined wet spell metric that does not have a predefined number of occurrences per definition of percentile thresholds

This is correct. We have mentioned that in the modified version.

p9 ll 8-9: please explain the part with the fraction more clearly

The fraction related to the total number of data available has been clarified. Due to the comment of the first reviewer, this paragraph has been deeply modified.

Generally, this is a section when having a map would be interesting in order to see in which region the dry spells are particularly hard to catch for certain datasets

Corrected

Figure 12: It's very difficult to make out the different datasets in this plot (please improve)

We agree there is a lot of information in that figure. Nevertheless, we have modified it.

and the last two years of OK look rather questionable - can you comment on that? Please explain where you see a clear increase in this.

As suggested, we have discussed the issues of the recent years using OK that could be due to the missing datasets.