

Review of the manuscript Egusphere-2022-458 (Lam et al.)

I appreciate the authors for taking into account my comments and conducting additional analysis. The scientific representation of the manuscript has improved compared to the previous version, and the results appear to be robust.

However, I would like to suggest some minor corrections to enhance the clarity of the manuscript. Some of the comments are related to sentence structures and word usage. I'd recommend that the authors review the manuscript again to check wordy sentences and correct any misuse of adverbs, which I may not have been able to identify all here.

Comments:

Line 10) *"Longer timescales (≥ 12 months) and the Standardized Streamflow Index were mostly present, "*

The sentence is not clear. "Present" in where?

Line 15) *"In doing so, however, spatial differences in aridity and water scarcity conditions have to be taken into account."*

I'd suggest removing this sentence or moving it to line 9, after "malnutrition.". As it is now, it is not coherent with the previous sentence. And remove "however".

Line 18) "Impactful" instead of *"Impactfull"*

Line 75) *"Strong gradients characterize the country..."*

I'd suggest "The country is characterized by strong gradients in precipitation ...".

Line 108) "because of" instead of *"according to"*

Line 123) I'd suggest removing *"and also for consistency purposes."*

Line 128) "reanalysis-based" instead of *"reanalysis"*

Line 184) The sentence *"When we compared the SPI calculated with a statistical distribution and SPI with ranking, the results were similar."* can be removed. The previous sentence is sufficient to explain the SPI (and other indices) methodology.

Line 235) About the Jaccard similarity equation: What are the A and B? When does the coefficient indicate that the two data series are statistically similar and when not?

Also, the significance test the authors applied (Chung et al. (2019)) should be briefly explained here.

Line 238) Suggest replacing *"namely"* with "known as".

Line 239) *"per region with the same aridity level"* with "for regions with the same aridity level".

Line 249) Suggest replacing *"per drought impact category"* with "for each drought impact category".

Line 253) *"after Kuhn (2008)." to "(Kuhn, 2008)".* Just citing Kuhn is enough.

Line 253) What are the typical values for variable importance? Is the link between two variables more significant when the variable importance is higher?

Line 264) Similar question to the previous comment. What are the good values for OBB, F1-score, and AUC? Are higher or lower values better? Should they be close to 1 or 0?

Line 266) I'd suggest replacing *"method"* with "the point-biserial correlation coefficient".

Line 275) Wrong sentence structure. *"Noticeable"* is an adjective. It should be "It is noticeable that ..."

Line 282) Suggest adding here that Masarabit is the arid region and Nyeri is the sub-humid region.

Line 283) *"Noticable"* to "It is noticeable that" or something else.

Figure 2) I wonder whether it would be possible to enlarge the labels on the x-axis (years) for the SPEI indices. The font is too small and not well visible when the manuscript is printed.

Table 3) Not the "*statistical significance*" is marked with asteriks. "The coefficients that are statistically significant (with a p-value less than 0.05)" are marked.

Line 293) Suggest changing "*SPEI-12 values lower than 0*" to "negative SPEI-12". Also, remove "*which is prevalent*" just leaving "between April 2016 and March 2018".

Line 294) Again, suggest changing "*SPEI-12 values lower than 0*" to "negative SPEI-12".

Line 295) "*drought impact occurrence does not happen simultaneously with the drought time period of the other accumulation periods, except for SPEI-24*".

So, does this means that except SPEI-24, other drought indices with different time scales (or accumulated periods) do not occur simultaneously with drought impacts, right? The sentence is confusing. I'd suggest something like:

"In general, drought impact occurrence happens simultaneously with SPEI-24 but not with the drought indices with other accumulation periods."

Line 296) "*In particular*" does not fit here well. The connection is not very smooth from the previous sentence. Do you mean "moreover", "Also" or "In addition"?

Line 301) Suggest replacing "*have*" with "shows"

Line 302) Suggest replacing "*other relations*" with "other impact categories", and "*have*" to "show".

Line 305) "*This*" is confusing. Suggest replacing "*This is done...*" with "the analysis is performed (or done) for the period of ..."

Line 317) The sentence "*The RF models have been trained on 75% of the data and tested on 25% of the data.*" can be removed as it is already in the method section.

Line 320) Same as my question for line 264. So, close to 1 means better performance?

Figure 5) Mention that the values on the x-axis are different for each category.

Figure 4-6) Change "*importance*" to "variable importance".

Table 4) Include in the caption what "*Discussed*" indicates. (That the authors are selecting only those indices and categories that are also statistically significant in the point-biserial correlation analysis).

Line 344) "*namely*" does not fit here. Suggest something like "which are".

Line 354) Remove "*namely*".

Line 356) "*Although the slightly underestimation of the MSWEP data compared to CHIRPS over East Africa*" to "Although there is a slight underestimation of precipitation in MSWEP compared to CHIRPS, ..."

Line 359) Remove "*namely*".

Line 359) "*different per county*," to something like "depending on the county".

Line 361) "*months*" to "period".

Line 362) "*reliable list*" to "reliability"

Line 363) Remove either "data" or "information". "*Data information*" is wordy.

Line 364) "*Europa*" to "Europe".

Line 364) "*assess*" means "estimate, measure". I don't think this verb fits there. Change to something else.

Line 371) Mention the point biserial technique too as this method was also used. Suggest something like "Random Forest technique in combination with the point biserial correlation analysis... "

Line 375) "*This study confirms this,*". What does the second "this" mean? And I'd replace first "this" with "our".

Line 379-381) *"deviations in the link between drought impacts and indices among the counties, making the model perform worse when the counties are combined together (i.e., the models in relation to the semi-arid regions)."*

So does it mean that aggregation can affect the result because the links between drought indices and impacts may differ for each county? But how can the authors make this claim if the analysis in finding the link is only performed by using aggregated regions? (or was the analysis also performed for each region individually from the beginning?)

Also, This part of the sentence is confusing *"deviations in the link between drought impacts and indices among the counties"*. If this means that each country may have a different link, please be more explicit.

Line 383) *"in accordance with"* to *"related to"*.

Line 395) Throughout the text, the authors sometimes use British English spelling and at other times U.S. English spelling (analyze vs analyse). Be consistent.