

Clim. Past Discuss., author comment AC3
<https://doi.org/10.5194/cp-2021-167-AC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

George C. D. Adamson et al.

Author comment on "Quantifying and reducing researcher subjectivity in the generation of climate indices from documentary sources" by George C. D. Adamson et al., Clim. Past Discuss., <https://doi.org/10.5194/cp-2021-167-AC3>, 2022

The article presents interesting results of a remarkable experiment between students and historical climatologists. The possibility of repeating the experiment dealing with a production of annual rainfall index series by two groups of raters with different experience is an indisputable advantage that can bring new findings. The article itself can serve as a good inspiration for other researchers and institutions occupied with historical climatology and having access to daily/monthly weather observations. The future possible experiments can be focused on different variable, area or time scale as the authors stated at the end of the discussion.

From the general point of view, the paper is written comprehensibly and no serious shortages haven't been noticed. The applied methods seem to be suitable and the results are clearly presented. Therefore, the study should be accepted and only several suggestions and questions are given for consideration.

Response: Many thanks for your positive comments.

Page 6: It would be better if the ICC values were expressed via interval, i.e. "ICC values of 0.5–0.74 are taken to represent moderate reliability, values between 0.75 and 0.89 to represent good reliability, and values ≥ 0.9 to represent excellent reliability."

Response: We have made this change as suggested.

Page 7, line 188: "IRR" abbreviation should be explained.

Response: We have changed this to 'inter-rater reliability' (it was a legacy from a previous draft).

Why did the authors decide to apply target reliability just of 0.9? Even though the reliability of 0.9 is commonly used in many studies did the author's decision coincide with ICC values ≥ 0.9 representing excellent reliability? If did this fact should be emphasized at least by one sentence.

Response: We have clarified this as follows:

The results suggest that a target ICC of 0.9 – considered 'excellent' inter-rater reliability

by Portney and Watkins (2007) – can be achieved for a group of 4 raters who are highly trained in the climatology of the region.

I miss a picture of Lesotho in the paper because not every reader can be aware of its precise localization within Africa.

Response: We have inserted a map as Figure 1.

I suggest joining a picture of a used documentary source to freshen up the paper.

Response: We have inserted an image of an example source as Figure 2.

Did the authors find out what was the average time to process your task by one student/historical climatologist and if time differed significantly? It could be also an interesting point in your study.

Response: This is a good point. Students were allowed 2 hours to complete the reconstruction. Historical climatologists were not given a time limit but were informed in advance it should take no more than 1-2 hours. At the time no one informed us differently, but we are currently consulting with our participants and will add the response to the final edit.