

The manuscript entitled “Comparing CFSR and conventional weather data for discharge and soil loss modelling with SWAT in small catchments in the Ethiopian Highlands” submitted by Roth and Lemann has been greatly improved according to the reviewers comments.

However, some of my comments were skipped (maybe my comments were not clear enough, sorry for it) and I also regret some of the changes mentioned in the reply letter were not included in the submitted revised version (e.g. P11057 L19, P11058 L3-4 and P11058 L5-6, P11059 L16 and P11059 L19, see reply letter).

Here I list some points that could, but also some other points that should, be improved before final acceptance in HESS. The lines refer to the manuscript version without track change.

L35 “A first evaluation, carried out by our research group, of CFSR-modelled rainfall data with that measured by the Water and Land Resource Centre (WLRC, formerly the Soil Conservation Research Programme [SCRIP]) in Ethiopia has shown substantial differences in daily, monthly, and annual rainfall.” I see the authors did not understand my previous comment (P11055 L12-14). In their paper, the authors also show “substantial differences”, so either the work has been done twice (first by the WLRC, then by Roth and Lemann), or the purpose of this sentence is still not clear.

L41 “correlating” -> correlated

L107 “one climatic station” what is the brand and the model of the 3 climate stations? See also previous P11056 L19 comment.

L127-130 The authors did not include in the submitted revised version what they claim in the reply letter. Which methods were used in the study to estimate both runoff and flow routing?

L152: “All HRUs were defined using a zero percentage threshold area, which means that all land use, soil, and slope classes were used in the process” As suggested by the reply letter, the authors could explain that the size of their catchments is small and that they have detailed land use maps at the plot scale, so that using a threshold for land use would decrease the available information and increase the uncertainty.

L155: My comment was misunderstood. Did the author parameterize their weather generator with WLRC data? If not, what data did they use?

L162 “Personnel at the research station are instructed to take grab samples only during rainfall events, when the river is turning brown” The authors still do not answer my question: how did the personnel grab sample? Which tools do they use? Do they filter the water sample? Which material do they use to filter the water sample? What is the size of the filter mesh? What happens next: drying? Weighting? And again, what is the sampling frequency? I.e. during a high flow event with brown water, what is the sampling frequency? Every 10 minutes? Every 10 hours?

L165 “eragrostis teff” -> *Eragrostis tef*

L241 “3.1.1 Seasonal comparison of rainfall data” why embedding section 3.1.1 within 3.1?

L261 There is no Table 7 in the manuscript.

L268 in the manuscript, p-factor is written “p-factor”, “P-factor” and “*p-factor*”. It should be homogenized. Please also check r-factor.

L326 “satisfacroy” -> satisfactory; see also “unsatisfactory” in L327

Figure 1: (1) Hurni is called twice in the caption. (2) Showing the CFSR stations is an indication of the respective size of the 3 catchments, but as suggested in my previous comment, the authors could also include the shapes (= catchment contours) of their 3 study cases within the shape of the Blue Nile catchment, so that the reader can have an idea of the relative size of the catchments. (3) I also highlight that it is very difficult to distinguish 10 green shades and the 2 blue shades. I know this does not affect the understanding of the paper, but it could improve the lecture of the map (which is nice!)

Figure 4 caption: tons per month (t/month)

Table 1: “ha” is missing of the swat-delineation size of both Anjeni and Maybar

Table 4 is not called within the text. Why using bold by the way? Do you think the guidelines of Moriasi et al also apply to rainfall pattern comparison?

Table 5 Some statistics related to the use of CFSR data (PBIAS in discharge validation) exceed the “satisfactory” threshold but are not highlighted in bold.

P11063 L24: The authors removed the whole paragraph. I believe it could have been relevant just to move it to the method section. It is part of the method for SWAT-CUP implementation and it gives weight to their work to know they made 500 runs for each iteration.