Reply to referee comment L. Boithias

We would like to thank L. Boithias for her time and effort spent reviewing out manuscript. We are very grateful for the clear, structured, and relevant remarks.

On the following pages we respond to all comments, questions and remarks. The first column contains the question or the comment from the referee, the second column is our response and clarification to said question and the third column is changes we made to our manuscript.

Referee comment	Author comment	Changes in manuscript
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).	There must be a mistake from our side. We did not intent to skip any referee comments; however it is possible that some changes were not carried into the final version of the manuscript. We have checked every current suggested referee comment and made the changes accordingly in this manuscript.	
P11057 L19: "other regions" could be introduced before "Ethiopia" to make the references clear.	P11057 L19 → This suggestion is not entirely clear to us. Changed the sentence according to our understanding. Moreover we added several references in the introduction, which include several other countries.	P11057 L19 → **[…] performance and parameterization in the US, China, Switzerland, Kenya, Ethiopia and other countries […]**
P11058 L3-4 and L5-6: Which method did the authors use?	P11058 L3-4 and L5-6→ Used methods have been added to manuscript. See also L127-130 below.	P11058 L3-4 and L5-6 → **Runoff was estimated using the SCS-CN method and flow routing was estimated using the variable storage coefficient method**
P11059 L16 and L19: "Daily river flow and sediment concentration: : :" What is the sampling material? What is the sampling frequency? This information may be useful to later broaden the discussion on the modelling quality. Can sediment concentration be "visible"? Aren't the authors talking about turbidity? Then what was the turbidity threshold to describe it	P11059 L16 and L19→ Details were added to manuscript. See also L162 below	P11059 L16 and L19 → **Grab samples are taken by hand with one litre bottles which are then filtered through ashless filter papers (retention capacity 7-12 µm). The filtered sediment samples are later transported to their respective research centres which oven-dry and weights them. Sampling frequency is every 10 minutes at rising water levels and every 30 minutes**

"visible"? Was it kind of experts' knowledge?		**after peak water level.**
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 [SCRP]) 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.	Because this sentence seems to be hard to understand and because we do not think the manuscript will lose information by removing it, we deleted the sentence.	Sentence deleted
L41 "correlating" -> correlated	Corrected the misspelling	Worqlul et al. (2014) **correlated** conventionally recorded […]
L107 "one climatic station" what is the brand and the model of the 3 climate stations? See also previous P11056 L19 comment.	Although we feel it is not entirely common to give the brand of the rainfall and climate station used and literature shows equal numbers of papers giving brands as not we are happy to add the brand of our climatic stations here.	The conventional or measured data contains daily rainfall and maximum and minimum temperature from one climatic station for each watershed **(Lambrecht Rain Gauge Hellman type with chart recorder, Piche tube evaporimeter and thermometers)**.
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?	We thought we included the used methods in chapter **2.5 SWAT model setup**. An additional sentence was added for that purpose in section 2.5.	**Runoff was estimated using the SCS-CN method and flow routing was estimated using the variable storage coefficient method.**
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	An explanatory phrase was added to the paragraph.	**Using a threshold with this kind of combination of small catchments in combination with highly detailed land use maps would have decreased the available level of information and increased**

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.		**uncertainty for modeling**
L155: My comment was misunderstood. Did the author parameterize their weather generator with WLRC data? If not, what data did they use?	The WLRC data was used as main weather input factor using the Hargreaves method for evaporation and the SWAT weather generator was used to fill in the gaps in the WLRC data. So yes, we parameterized the weather generator with WLRC data.	**The SWAT weather generator was used to fill the gaps in the WLRC data set for rainfall and temperature.**
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?	We think that adding this level of detail might be over the top somehow. Nonetheless we are happy to add some details.	**Grab samples are taken by hand with one litre bottles which are then filtered through ashless filter papers (retention capacity 7-12 μm). The filtered sediment samples are later transported to their respective research centres which oven-dry and weights them. Sampling frequency is every 10 minutes at rising water levels and every 30 minutes after peak water level.**
L165 "eragrostis teff" -> Eragrostis tef	Adapted	**Teff (*Eragrostis tef*) a widely cultivated and highly nutritional crop […]**
L241 "3.1.1 Seasonal comparison of rainfall data" why embedding section 3.1.1 within 3.1?	Because the seasonal rainfall comparison is a sub-section of the rainfall comparison we embedded it in 3.1. We first start with a general appreciation of the rainfall comparison before comparing the main seasonal differences. We would like to keep it that way.	The sectioning was not changed
L261 There is no Table 7 in the manuscript.	This is a typograph. The authors of course refer to table 4.	For Anjeni, Andit Tid, and Maybar the CFSR model performed unsatisfactorily (**see table 4**) with NSEs […]

Comment	Response	Revision
L268 in the manuscript, p-factor is written "p-factor", "P-factor" and "p-factor". It should be homogenized. Please also check r-factor.	Thank you for mentioning this. We did change all occurrences of p-factor and r-factor into italic and lowercase p-factor and r-factor.	**L268:** For the goodness of fit two indices called **"p-factor"** and **"r-factor"** are used. The **p-factor** is the fraction of measured data inside the 95PPU band […] **L178:** The *p-factor* is the fraction of measured data […] **L272:** With **a *p–factor*** of 0.71 and a**n *r-factor*** of […] **L315-316:** […] including SWAT–Cup *p–factor* and *r–factor* are summarised in […] **L332:** With a *p–factor* of 0.40 and an *r–factor* of […] **L179:** The *r-factor* is the ratio of the average width of the 95PPU band
L326 "satisfacroy" -> satisfactory; see also "unsatisfactoy" in L327	Both typographs have been corrected.	*Satisfacroy* → **satisfactory** *Unsatisfactory* → **unsatisfactory**
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!)	(1) Double citation of Hurni has been removed (2) The authors tried different versions of this map including the one suggested by the referee. Simply the differences in size between the research catchments and the Blue Nile Basin are so large that the research catchment's contours cannot be seen on the overview map. Therefore the authors prefer to show the overview map including the CFSR data even without including the shapes of the research catchments. (3) Thank you for that comment. While we agree with the referee's remark on shades and colours we would like to keep the map exactly like because our organization has	(1) Figure 1. Map overview of Blue Nile (Abbay) Basin with the WLRC research stations, agro-ecological zones **according to Hurni (1998)** and emplacements of CFSR stations. (2) Map is kept the same (3) Colors are kept the same

	been using this shading and colouring for many years now and this has a wide recognizable value, which we would like to keep.	
Figure 4 caption: tons per month (t/month)	Caption adapted	Modelled SWAT soil loss compared to measured soil loss (blue) for WLRC (red) and CFSR (green) input data and the 95 Percent Prediction Uncertainty (light blue). Each sub-figure contains the calibration and the validation period. Results are given in tons **per month (t/month)**.
Table 1: "ha" is missing of the swat-delineation size of both Anjeni and Maybar	"ha" has been added to both Anjeni and Maybar	Andit Tid Anjeni Maybar 466.78 ha 105.23 **ha** 101.98 **ha**
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?	This is a typograph in the LaTeX file. Bold highlights have been removed.	See table 4 for details
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.	All statistics have been checked and bold highlights have been added where necessary.	See table 5 for details
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.	Moved the sentence to the end of the method section.	Each model was calibrated with one to five iterations using 500 simulations each.