Reply to referee comment T. Baker

We would like to thank T. Baker 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.

Question/Comment	Response	Changes in manuscript
Comment 1: Ultimately this paper doesn't suggest a solution other than we should all be better off to collect high quality, high resolution data. Unfortunately, such luxuries of adequate high quality data are rare.	We do not entirely agree with this comment. This paper should suggest that the CFSR data cannot be used indiscriminately and in any case. We would like to point out that it is not always the best solution to use globally modelled climate data in very small catchments and that if used nonetheless results should be checked carefully.	None
Comment 2: Consider more regionally appropriate terminology for rainy seasons: short rains are *belg* season and long rains are *kremt* season.	Thank you for this comment. We have referred to *belg* and *kremt* in Section 2.3. Thereafter we refrained from using the terms for two reasons. First of all, there are not common for a large audience and second they do not occur in all three research watersheds, which makes it easier to refer to as *short and long rainy season.*	None
Comment 3: It is important to discuss that the belg in particular is quite erratic in Ethiopia with respect to its timing and volume of rainfall as well. You get at this a bit in Section 3.1, but it would be advisable to bring this up in your study area description with pertinent references as to why this important	Thank you for bring this up. We explicitly added Figure 2 to section 3.1 to compare CFSR and WLRC rainfall distribution. The figure, unfortunately, was moved to the back of the paper in the online version. The figure shows clearly that rainfall distribution at our research watersheds is quite clearly distributed for the belg season and much less for kremt season. It also shows that the CFSR dataset is much more erratic for the belg season. The boxplots for the WLRC data was compiled out of the data indicated in Table 3, which was also moved to the back of the online paper. We agree that this	None

	is an unfortunate display of figures and tables and that this needs cleaning up.	
Comment 4: Regarding HRUs: water is not routed between HRUs. Water is routed using the methods you describe between sub-basins. At the HRU level, the water balance is calculated but this is all then summed at the sub-basin level and not routed, which you imply in the way your discussion reads going from HRUs to routing without clarifying	We entirely agree with this comment, thank you very much for highlighting this. The text has been changed accordingly.	Lines (24-26, p. 2117): **ArcSWAT** divides the catchment into hydrological response units (HRUs) based on unique combinations of soil type, land use, and slope classes that allow for a high level of spatial detail simulation. **Runoff is predicted separately for each HRU and routed at subbasin level to obtain the total runoff for the watershed (Neitsch et al., 2011).**
Comment 5: Is there a citation for the 2m DEM used?	This DEM was produced by the SCRP and the CDE in 1981 based on orthophotos, a stereometric photo analysis and an extensive field survey. The reference was added to the manuscript.	**Added reference**: SCRP: Digital Elevation Model (DEM) of Anjeni, Gojam. Soil Erosion Conservation Database, Addis Abeba, 2000. SCRP: Digital Elevation Model (DEM) of Maybar. Soil Erosion Conservation Database, Addis Abeba, 2000. SCRP: Digital Elevation Model (DEM) of Andit Tid. Soil Erosion Conservation Database, Addis Abeba, 2000.
Comment 6: You discuss land use map, but nothing regarding land management practices. Did you modify SWAT for this in the database? You need information on management practices to adequately represent plant growth otherwise this will impact both discharge and sediment. And, what of livestock? In many of these areas overgrazing is an issue and certainty influences sediment and erosion. Also, gullying is a problem. How did you approach this challenge and the large amounts of sediment delivered in this way?	Thank you for pointing this out. The land use data were represented with field-scale surveys of crops and planting times of every field in the catchments. Furthermore, the traditional tilling tool “Maresha” was added to the SWAT database according to Temesgen et al. (2008) and Dile and Srinivasan (2014). Planting times were adapted according to surveys carried out in the watershed by WLRC and ourselves. We have added a small paragraph in “2.4 SWAT model setup” However, as the entire section on sediment loss was removed this is only of minor relevance.	**Section 2.4:** The planting and harvesting times were averaged over the entire period and planted at similar dates for the entire simulation. To simulate crop growth we used the heat unit function in ArcSWAT. Teff, for example, was planted beginning of July and harvested beginning of December with several tillage operations preceding planting. Tillage operations were adapted to the usage of the traditional Ethiopian plough called “Maresha” according to Temesgen et al. (2008), and Dile & Srinivasan (2014).
Comment 7: What about curve numbers? How did you handle the monsoonal climate and issues with soil water content? This has been	As stated in Section 2.2 curve numbers were defined in the SWAT database according two SCRP soil reports (Kejela 1995, Weigel 1986) and from one doctoral thesis (Zeleke 2000).	None

shown to be a challenge that must be addressed when modeling in such areas	There was no further calibration of curve numbers because the main goal of this paper was to compare the raw climate input without further calibration.	
Comment 8: "The sub-basin sizes were fixed at 2000 ha." Is this a typographical error? Or, is there an error in how the study site sizes were reported. Perhaps stick to similar units throughout (either km2 or ha, but not mixed).	This is a typographic error. Thank you very much for indicating it. We have changed the text accordingly.	The sub-basin sizes were fixed at **2 ha**.
Comment 9: To generate weather, you used the SWAT weather generator but also indicate that the local weather had numerous gaps. How did you address this in developing the underlying statistics for the weather generator? For example, how many missing days were there for the different variables? For variables that you may not have data, how did you develop statistics?	Thank you for pointing out this lack of clarity in the text flow. We introduced Table 3 in this section, but it was moved to the end of the text in the online version. The table shows which years were available for the WLRC data set. Statistics were compiled using the sub-daily rainfall data and daily temperature data. Weather data was only compiled for the missing years of the WLRC data and not for the entire time .	None
Comment 10: Regarding sediment: While I can completely understand that you may only have sediment data after rainfall events, it is a bit of a jump to then state there is no sediment during the dry season or outside rainfall events. First of all, flows will continue to move sediment after and event and secondly, there will still be some sediment movement in a river during such times. Low, perhaps but not entirely absent. And, if you only collect during rainfall events, how can you know this one way or another? Also, wouldn't there be some data collection up to some point after a rainfall event?	We are grateful for this remark. It is true that sediment collection would have had to be better explained. As we, based on suggestion of another reviewer, decided to remove all references to sediment loss modelling due to it limited significance for present manuscript we will not add anything to the manuscript. Nonetheless, we would like to explain that in the present research context some abstractions had to be made concerning sediment load. The SCRP and WLRC have been collecting sediment loss samples in 1 litre jars, filtering, drying and weighting it. The setup in Ethiopia does not allow for high-level sediment detection. Therefore, we adhered to the concept of the	Removed all references to sediment loss in manuscript.

	SCRP and assumed no sediment during the dry season. We do not contest the referees objections concerning sediment load, but we point to present research reality.	
Comment 11: Again, it is important to indicate the completeness of local data and any issues regarding accuracy (if known).	Data completeness was described in Table 3, which was unfortunately moved to the end of the text in the online version. In the final version this will be fixed.	None
Comment 12 Maybe in section 3.1, you can show basic summary statistics with information on observed data regarding missing days and times of year when such data are missing? You say that MAE and MSE were also computed. Where are they? This needs to stand out more among your results to paint a clearer picture. Again, in themselves they do not tell us much, but you are trying to paint a bigger picture here of challenges faced.	Basic summary statistics were computed initially but removed. Our approach was to compute performance statistics with seasonal values. CFSR and WLRC rainfall data were compared for belg, kremt and dry season. We think that it is more relevant to the subject to compare CFSR and WLRC rainfall data for these three specific seasons as they are very relevant to discharge and soil erosion modeling. MAE and MSE were removed entirely from the manuscript.	Reference to MSE and MAE removed: **Section 2.5: Model evaluation** Lines 10 and 11 (p. 2121): ~~For further time series evaluation, mean absolute error (MAE) and mean square error (MSE) were computed.~~
Comment 13: Looking at uncalibrated results in Tables 5 and 6, it is unclear to me how you can jump to the conclusion that the model performs poorly. In regard to sediment, similarly, for both Andit Tid and Anjeni the CFSR and WLRC produce adequate uncalibrated results, but neither performs well at Maybar.	The conclusion is drawn not only from Tables 5 and 6 but from the tables in combination with Figures 4, 5 and 6. The figures clearly show where and how strongly CFSR data diverges from the WLRC data – this is clearly supported by Tables 5 and 6. The combination of both leads to said conclusion. As the sediment loss discussion has been removed entirely, this is merely irrelevant.	Removed sediment loss discussion and data comparison.
Comment 14: Overall, I don't find the discussion on sediment to add much to this paper. Sediment delivery in the Upper Blue Nile is a serious challenge and perhaps your work can add to a separate and more complete discussion on this particular challenge in	Thank you for this indication. We, based on suggestion of another reviewer, have decided to remove all references to sediment loss from this manuscript, because it does not add significant information.	Removed sediment loss discussion and data comparison.

modelling the region		
Comment 15: I think it is hasty to make the following statements: *"Our results clearly show that no adequate discharge and/or sediment loss modelling was possible with the CFSR data"* *"Thus, contrary to Dile and Srinivasan (2014), this study suggests that CFSR data may not be applicable for small-scale modelling in data-scarce regions…"* My concerns with these specific statements are that 1) you did achieve adequate (uncalibrated) results using CFSR in some instances and 2) you make a jump to the conclusion that this therefore implies the data are inadequate in data-scarce regions generally. You have not given us other examples of using these data in other data-scarce regions (either your own research or citations to other works). Also, due to the unique challenges posed by the high variability of East African climate, why would this then apply to all other data-scarce regions?	We do not entirely agree with your statements. If you look at the entire picture comparing rainfall distribution data (figure 2) and subsequently modelled discharge (figures 4-6) then the following deductions apply: (1) CFSR rainfall data is either very much overestimating monthly rainfall or the seasonal distribution is wrong. Consequently modelled discharge is up to three times higher (Anjeni) or has a wrong seasonal distribution, which is balanced over the year, but dangerously wrong nonetheless. (2) The only thing we implied is that in data-scarce regions, the CFSR data is predestined to be used but at the same time this is exactly where comparison is not possible, as data is lacking. Now, if you compare the Anjeni data, you would simply calibrate your model with 3000 mm of rainfall a year instead of 1600 mm. Our implication only states that for example Dile & Srinivasan (2014) achieved a good result for their watershed, which was larger than ours. We assumed that the CFSR data, being a global model, might work better for larger scale watersheds.	None
Comment 16: You state that "there is no substitute for high quality conventional weather data." I doubt many would argue with argue on this point; however, in remote data-scarce regions of the developing world we rarely have such data available.	We agree with your statement, but nonetheless we would like to highlight possible issues when using CFSR data. It is not always the best solution to use data just because it is available. We think it would be wise to use CFSR data at least with caution, especially in data scarce regions where no subsequent calibration is possible because of lack of data.	None
Comment 17: Even in this work you indicate that the data	The data sets were described in Table 3, which was unfortunately taken out of the text and put	None

sets you were using have gaps, suggesting issues and challenges but then there are no details about the data sets.	at the end of the manuscript.	
Comment 18: And, just because this happens (right results for the wrong reasons), it doesn't necessarily follow that the data or models should be abandoned but rather an exploration of where there may be systematic errors in the data that can be improved upon or a discussion on when and where data are more or less applicable.	Thank you very much for this pertinent indication. Our manuscript was intended as a comparison of measured and modelled rainfall data sets. We did not intend to explore systematic errors of the CFSR data set, because we only have very limited data on three watersheds. Our comparison only suggests that the data are not applicable to our watersheds, and we suggest that other small-scale watersheds might have similar problems.	None
Comment 19: In this study, spatial scale is quite small and so therefore errors will be more pronounced, especially given that rainfall and its distribution are always a concern. And at such small scales, a greater consideration should be given to driving processes and ensuring those are adequately represented in the model.	Thank you very much for this comment. This is exactly our point of view. This is also how we intended our sentence, which you commented on under "Comment 15".	None
Comment 20: As noted above, consider how you refer to seasons (e.g., belg and kremt).	See "Comment 2".	None
Comment 21: A datum is but data are . . . please be mindful of noun-verb agreement	Thank you for this important remark. We have changed all references to "data" according to Question 3 in Reply to referee 1.	See "*Question 3*" in "*Reply to referee 1*"
Comment 22: "and/or" This convention has no place in English writing from a truth value perspective. "Or" is sufficient because in a statement where it is used, if one or both parts may be true, then the statement is true. "And" statements, on the other hand, are more restrictive, and are true only when	Thank you for this remark. As all reference to sediment loss was removed from this manuscript, we have changed the text accordingly.	Our results clearly show that no adequate discharge **[…]** modelling was possible with the CFSR data.

both parts of a statement are true. However, this scenario is also covered under the use of the term ”or”. As such, the “and/or” convention is extraneous and unnecessary.		