Reply to referee comment C. Baffaut

We would like to thank Claire Baffaut 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
Question 3 Overall, given the poor simulation results with sediment with either data set, I would suggest to remove these results for the paper and concentrate on the discharge. The story would be much stronger. The conclusions would be similar because with poor discharge, there is really no hope of obtaining meaningful sediment results.	Thank you for this highly appreciated hint. The idea behind the sediment loss comparison was to show that the use of CFSR rainfall data not only has an influence on discharge but that sediment loss modelling results tend to have an increasing variation from observed sediment loss when using CFSR data. Nonetheless this is a very valuable comment, which, according to our statistics and the procedure for sediment modelling is difficult to defend; we have adapted the title and the manuscript respectively and removed references to sediment loss modelling throughout the paper.	**Title:** Comparing CFSR and conventional weather data for discharge **[…]** modelling with SWAT in small catchments in the Ethiopian Highlands. **Abstract** (p. 2114): - Lines 1 and 2: Accurate rainfall data is the key input parameter for modelling river discharge **[…].** - Lines 9 and 10: To this end, the Soil and Water Assessment Tool (SWAT) was set up to simulate discharge **[…]** using CFSR and conventional weather data, in three small-scale watersheds ranging from 102 to 477 ha. - Lines 11 and 12: Uncalibrated simulation results were compared to observed river discharge **[…]** over a period of 25 years - Lines 15-17: Sentence removed entirely - Lines 17-19: Overall, the simulations with the conventional data resulted in far better results for discharge **[…]** than simulations with CFSR data. **1 Introduction** (p. 2116): - Lines 15 and 16: The CFSR and WLRC rainfall data are subsequently used to simulate river discharge **[…]** in three watersheds using SWAT - Lines 16-18: Uncalibrated CFSR modelled discharge **[…]** is then compared to uncalibrated WLRC modelled discharge **[…],** and the applicability of the CFSR data for hydrological predictions is statistically evaluated. **2 Methods** (p. :2116 - Lines 23 and 25: Second, the impact of spatial and temporal variability of rainfall on hydrology **[…]** was assessed by modelling discharge **[…]** with SWAT. - Lines 25-29: This second analysis provided an evaluation of how the

		change in rainfall input data affects discharge **[…]** modelling with SWAT. Third, the rainfall (in mm), discharge (in m3 s−1) **[…]** data were converted to mean monthly millimetres for all years and then compared visually and statistically. **2.1.1 Hydrologic Model**: - Lines 6-9 (p. 2118): removed entirely **2.3 Hydrometric Data** (p. 2119) - Lines 8-11 (p. 2119): removed entirely - Lines 3-5 (p. 2120): removed entirely **2.4 SWAT model setup**: - Lines 16-18 (p. 2120): Daily river flow **[…]** data were measured at the outlet of the three WLRC watersheds. - Lines 19-21: removed entirely **2.5 Model evaluation**: - Lines 16-18 (p. 2122): removed entirely - Lines 9 and 10 (p. 2123): Adequate visual agreement between observed and simulated data was compared on discharge **[…]** plots on a monthly and a monthly mean basis. **3.2.1 Mean monthly results with WLRC data**: - Lines 4-6 (p. 2126): The comparison of the mean monthly rainfall data and subsequently the mean monthly discharge **[…]** showed far better results than the comparison of daily and monthly data. - Lines 11 and 12: removed entirely - Lines 17-21: removed entirely - Lines 28-31: removed entirely **3.3.1 Mean monthyl results with CFSR modelled data:** - Lines 1 and 2 (p. 2128): removed entirely - Lines 6-10: removed entirely - Lines 16-22: removed entirely **4 Conclusions**: - Lines 1-3 (p. 2129): Finally, we modelled discharge **[…]** for the three stations with the SWAT model and compared uncalibrated results from CFSR rainfall and conventional rainfall. - Lines 10-13 (p. 2129): removed entirely - Lines 5-16 (p. 2130) removed entirely - Lines 17 and 18: Our results clearly show that no adequate discharge

		[…] modelling was possible with the CFSR data. - Lines 22-25: In addition, discharge **[…]** modelling showed that usage of CFSR weather data not only resulted in substantial deviation in total discharge, but also in the seasonal rainfall pattern.

Question 1: The reason for the poor performance with sediment simulation might be in inaccuracies with management data. Sediment simulation is very sensitive to tillage operations or to over-grazing. The authors mention that *"land use is dominated by smallholder rain-fed farming-systems with grain-oriented production, ox-plough farming and uncontrolled grazing practices."* How were these represented? For small watersheds of that size, would the quantity and timing of plough operations have an impact on simulation results?	We agree that this was not mentioned strongly enough. The land use data were represented with field-scale mapping 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. 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 interest.	**Section 2.4: SWAT model setup** 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).
Question 2: The authors then continue the analysis by aggregating more and looking at mean monthly results (Table 5 and 6, and figures 4 to 6). Analyses of mean monthly results with performance measures such as NSE, r2, and RSR are questionable because there are only 12 data points and these points have expected seasonal variation. At the minimum, there is no justification to apply the performance criteria proposed by Moriasi et al., which were defined for the comparison of measured and simulated time series. Mean monthly values can be calculated, plotted, and discussed as done in section 3.2. However, performance measures threshold values have no validity.	Thank you for this input. We agree with the second part of the comment and have removed references and comparisons to performance criteria by Moriasi. Nonetheless, we think that the comparison with NSE, r2, and RSR are suitable, without applying the performance criteria, because they present a valuable statistical comparison. The mean monthly values are henceforth discussed without performance criteria.	**Section 3.2.1: Mean monthly results with WLRC data** - Lines 7-9 (p. 2126): The mean monthly discharge data for Anjeni showed **very high agreement** (see Fig. 6) **[…].** - Lines 13-15 (p. 2126): For Andit Tid, the mean monthly discharge modelled with WLRC rainfall showed a **satisfactory agreement […]**. - Lines 21-24 (p. 2126): For Maybar, the mean monthly discharge modelled with WLRC rainfall showed a good agreement with the observed data. **[…]** Observed and modelled discharge… **Section 3.3.1: Mean monthly results with CFSR modelled data** - Lines 23-24 (p. 2127): For Anjeni, the discharge modelling with the CFSR rainfall input **showed an unsatisfactory agreement.** - Lines 2-5 (p. 2128): For Andit Tid, the mean monthly CFSR discharge modelling performance **showed an unsatisfactory agreement.** - For Maybar the **CFSR modelled discharge indicated very unsatisfactory agreement**. The hydrograph showed a strong overestimation […]. **Section 4: Conclusions** - Lines 20-29 (p. 2129): The monthly mean data comparison of CFSR data showed an unsatisfactory result for discharge in Anjeni and Maybar, while Andit Tid showed a good agreement only because underestimation and overestimation were in balance. Simulations with the CFSR data lead to a minor underestimation of the total yearly discharge for Andit Tid and Maybar and a very strong overestimation of discharge in Anjeni.

		- Lines 26-29/1-4 (p. 2129/2130): The measured WLRC climatic data provided very high agreement for modelled discharge results for Anjeni and Andit Tid and good agreement for Maybar. The simulations with the conventional data lead to an overestimation of discharge for all three stations: Anjeni, Andit Tid, and Maybar. However, the hydrographs show clearly that for all three catchments the problem of overestimation comes mainly from the three months after the main rainy season, where the SWAT modelled discharge takes much longer to reach baseflow level than observed data and not from the modelling of the main rainy seasons.
Question 4: It would be nice to present the results for each micro-watershed in the same order each time, whether in the text, in the tables, or in the figures. Similarly, the format of tables that have similar information should be similar. For example, it would help the reader to have the same order of performance measures in tables B1 and B2. Why are tables A1 through B2 not cited directly in the text? Are they supplementary material? I think they are quite critical to the understanding and interpretation of the results and should not be relegated to supplementary material.	Table B1 and B2 have been adapted – there was an error of sequence in the table. Thank you for pointing this out. Concerning citations in the text we believe we have referenced the respective figures adequately, e. g. paragraph 3.1.1 contains reference to A2, and paragraph 3.2 contains reference to B1. Table B2 was removed entirely as, according to Question 3, the entire reference to sediment loss was removed.	- Table B1 and B2 re-arranged. - Table B2 removed
Question 5: Finally, figures 4, 5, and 6 should be introduced in increasing order, e.g., introduce figure 4 before figure 5. These figures are also difficult to read because of the superposition of colors in the bar charts	The idea of this graph was to show the implications of using the CFSR data on discharge and sediment loss. We agree that the reading of the graph might be challenging. Therefore we have simplified the figures by removing sediment loss data (see Question 3) and by adapting the legends.	- Graphs adapted according to referee question. - Sediment loss data were removed. - Order of figures introduction was adapted
Comment 1: I don't see anything in the manuscript that supports the last statement of the abstract: "and might be better adapted to larger	This statement is the result of a comparison to results by Dile & Srinivasan (2014). In the Conclusion paragraph (p. 2130, line 27) we state that our results are contrary to the ones	None

watersheds than the ones used in this study". Please remove	of Dile & Srinivasan and thus we consider that the CFSR data might be better adapted to the larger watershed size of Dile & Srinivasan than to small-size watersheds like ours. Therefore we do not think that the sentence should be removed as it is a central conclusion of the entire research.	
Comment 2: How can tons of sediment be converted to millimetres? Doing so would require an assumption on the density of the sediment. Is it what was done? What is the assumption? Furthermore, I don't see the necessity of converting a mass to a depth. What does it bring? Why not using tons per hectare, a common unit used to evaluate sediment loss from an area?	The idea behind this conversion was to bring rainfall, discharge and sediment loss to a common denominator to be able to consider ranges and variations on the same scale in one single graph. For this purpose sediment data was assumed to have a density of 1 and resulting volume in m^3 were divided by the area in m^2 to obtain mm. For this manuscript this does not apply anymore as all references to sediment yield have been removed for lack of accuracy. See Question 3 for details.	Removed all references to sediment loss in the manuscript.
Comment 3: In section 2.1.1, the authors mention that SWAT divides the catchment into HRU. Technically, SWAT does not do this. The HRU delineation can be done with the ArcSWAT interface but it could also be done with different tools.	Thanks for pointing this out. We did indeed omit mentioning that we used ArcSWAT and not SWAT for watershed delineation and HRU division. We have adapted the text accordingly.	2.1.1 Hydrologic Model - Lines 24-26: **<u>ArcSWAT</u>** 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
Comment 4: Again in this section (2.1.1), the authors state that SWAT predicts individual HRU hydrology using the water balance equation. That is not technically true. Each component of the water cycle, i.e., runoff, ET, aquifer recharge, and subsurface flow, is calculated individually. The water balance equation can be used to validate these calculations.	Again, thank you very much for highlighting this. We have adapted the text accordingly.	- Lines 26 & 1 (p. 2117 and 2118): **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: In section 2.2, please detail how the surveys were conducted. Were those from interviews, observations? If references are available, give them. What is the size of the individual holdings? Describe the process of generating a generic land use map from the 2008 and 2010 land use maps	Land use is derived from yearly land use mappings which are generated by observations and interviews. The generic land use map was generated by combining the 2008 and 2010 landuse surveys, which were partially conducted by the authors, with findings from 2008, 2012, and 2014. The phrase in question was adapted	**Section 2.2: Spatial data:** Land use data were adapted from yearly surveys carried out by SCRP and WLRC **through land use mapping and interviews** and by own surveys in 2008 and 2012. To adapt to annually changing land use patterns, a generic map was adapted from the WLRC land use maps of 2008**, 2012, 2014** (Anjeni), and 2010**, 2012, 2014** (Andit Tid, Maybar)
Comment 6 In section 2.3, how are the one-litre samples collected? Are they grab samples? Flow proportioned samples? What is the protocol?	The litre samples are indeed grab-samples. The samples are collected by hand using 1 litre bottle, which have a wide opening at the top. Samples are filtered, dried and finally weighted. As sediment calibration was entirely removed from the manuscript (see Question 3) this is now irrelevant.	Removed all references to sediment loss from manuscript.
Comment 7: In section 2.4, is the sub-basin size really fixed to 2000 ha? What does that mean for micro watersheds that range from 100 to 500 ha? There must be a mistake somewhere.	Thank you very much for pointing this out. This is obviously a typographic mistake. The sub-basin size was fixed at 2 ha and not at 2000 ha. Corrections were made to the text.	**Section 2.4: SWAT model setup:** The sub-basin sizes were fixed at **2 ha**.
Comment 8: Section 2.4, line 21: “During the dry season and outside rainfall events the monitored rivers are sediment free”. Really no sediment at all? It would be very difficult to visually distinguish a low sediment concentration (up to 100 mg/l) from no sediment at all. The assumption of no sediment might be justified on the basis that the concentrations are low and the transport is insignificant compared to what rainfall events transport. But it is probably not sediment free.	We agree to this but please keep in mind that these are research stations with only very basic monitoring possibilities. The sediment monitoring was therefore reduced to times with high and visible sediment contents e.g. during rainfall events from initial water level increase back to baseflow level. Outside storm events, no sediment data can be measured (we simply cannot measure that small an amount of sediment with our instruments), which implies the assumption that rivers are “quasi sediment-free”. As sediment calibration was entirely removed from the manuscript (see Question 3) this is now irrelevant.	Removed all references to sediment loss from manuscript.

Comment 9: Section 2.5, line 20: NSE is not always the best objective function for reflecting overall fit. In particular, it is not very indicative of performance when measured data have low variance. When there is high variance, NSE is biased toward high values. I am not contesting its use here, only that it is presented as the best.	Thank you for highlighting this. As the text shows, this is a citation from Servat & Dezetter (1991). We have corrected the passage accordingly and changed the word "best" to "good"	**Section 2.5: Model evaluation:** The NSE is recommended because for one it provides **a good** objective function for reflecting the overall fit of a hydrograph (Servat et al., 1991) and second, because it is very commonly used, it provides extensive comparable information on reported values (Moriasi et al., 2007).
Comment 10: Section 3.2: A graph of simulated and measured daily or monthly discharge values would greatly help.	This is a very interesting comment. We explicitly decided beforehand not to publish a graph with daily or monthly discharge values, as we considered it as not suitable in terms of visualization and communication of our results in the paper itself. In our opinion daily or monthly discharge data over a 28 year period simply do not reveal quality information about calibration. In consideration of communicating relevant results we refrained from doing that explicitly. Nonetheless, we will happily provide these graphs for an electronic annex to provide the readers with the full information.	Added graphs for simulated and measured monthly discharge values for Anjeni, Andit Tid, and Maybar
Comment 11: Technical comments: The word "data" is plural. Correct usage is: "data are available", or "where they are available". Please correct throughout the paper.	Thank you for this comment. The entire text has been adapted accordingly.	**Abstract:** - Line 1: Accurate rainfall data **are** the key input parameter for […] - Lines 2 and 3: Remote areas of Ethiopia often lack adequate precipitation data and where **they are** available, there might be substantial temporal or spatial gaps. 1 Introduction: - Line 10: The CFSR data **are** based on a spectral model - Line 25: […] conclude that CFSR data **are** most deficient […]
Comment 12: Page 2115, line 4 "modelled rainfall data": there is a contradiction between modelled and data.	The idea of "modelled rainfall data" was to raise comprehension concerning CFSR rainfall data, which is not measured but originates from a "*global weather model*". To call this data "modelled rainfall data" in opposition to	**Introduction:** This makes it necessary to use other sources of **[…]** rainfall data for SWAT modelling.

	"*measured WLRC rainfall data*" seems adequate. Nonetheless we adapted the sentence accordingly and removed "modelled".	

References

Gupta, H.V., Kling, H., Yilmaz, K. K., Martinez, G. F. (2009): Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology. 377, pp. 80-91.

Gupta, H.V., Wagener, T., Liu, Y., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes 22, 3802–3813

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences 5, 89–97.

McCuen, R.H., Knightm, Z., Cutter, A.G., 2006. Evaluation of the Nash–Sutcliffe efficiency index. Journal of Hydrologic Engineering 11 (6), 597–602

Schaefli, B., Gupta, H.V., 2007. Do Nash values have value? Hydrological Processes 21, 2075–2080

Jain, S.H., Sudheer, K.P., 2008. Fitting of hydrologic models: a close look at the Nash– Sutcliffe index. Journal of Hydrologic Engineering. 13 (10), 981–986.