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	Authors answer	Changes in manuscript
However, the paper lacks of rigor in some sections (see my detailed comments hereafter) and I still have this question pending: what is the added value of hydrological modelling to show that CFSR data are not reliable enough to model small river basins? The Figure 2 is enough to show that both modelled and measured rainfall datasets are strongly different. The paper would gain much if this particular point could be more discussed.	The detailed comments have been answered on the following 8 pages and the sections have been adapted and modified accordingly Concerning the pending question by the reviewer we would like to emphasize the following: The SWAT website seems to suggest that the CFSR data is available for any place on the globe and that hydro-meteorological data can be downloaded and used without preoccupation. There is no warning about discrepancies or variations in CFSR data, which could lead to very wrong modelling results and subsequently wrong deductions. What we wanted to show was not only deviations in rainfall patterns (which are obvious), but also discrepancies in seasonal patterns and their implications for SWAT discharge and sediment loss modelling. We wanted to clearly show that despite calibration of SWAT rainfall data has a strong influence, which has a multiplying effect on discharge and sediment yield. - We have added a paragraph to the Introduction section concerning the importance of this subject - We have added a paragraph to the conclusion section concerning this subject	- see subsequent answers to detailed comments from referee - The particular point about the added value of hydrological modelling has been added to the Conclusion section: **The SWAT modelling showed that CFSR rainfall pattern and rainfall yearly total amount variations were so significant that SWAT model calibration could not adequately represent measured discharge and sediment yield.**
Overall, the method should be revised to merge similar topics together and avoid repetition (see e.g. section 2, section 2.1.1, section 2.2, section 2.3).	The 'Method' section has been adapted to make it more coherent. See comments for **P11057 L15, P11055 L18-etc.**	
P11057 L15: why embedding section 2.1.1 within 2.1?	'2.1.1 Hydrologic model' was changed to '2.2 Hydrologic model'	**2.2 Hydrologic model**
P11055 L18-etc: "Several studies evaluating the CFSR data: : :" this section is interesting since it gives examples of successful and unsuccessful uses of CFRS data. However, key information	Thanks for this hint., the 'Introduction' section.as been adapted accordingly	**However, the applicability of the CFSR data for small-scale catchments in the Ethiopian Highlands has not been adequately investigated yet. Aforementioned studies did not focus on small-scale watersheds but mainly on large basins, which tend to**

is missing for the authors/readers to compare the present study to the previous studies: what were the sizes of the modelled catchments? Additional literature assessing CFSR data: Bressiani et al. 2015; Alemayehu et al. 2015.		**balance errors from CFSR data.** **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) in Ethiopia, formerly the Soil Conservation Research Programme [SCRP]) has shown substantial differences in daily, monthly, and annual rainfall. So far, few studies have been conducted in the Ethiopian context on the impact of rainfall data on streamflow simulations. Fuka et al. (2013) used CFSR data in a 1200 km^2 watershed in Ethiopia with SWAT suggesting CFSR data performs as good as or even better than conventional precipitation. Worqlul et al. (2014) correlating conventionally recorded rainfall with CFSR data over the Lake Tana basin (15'000 km^2). They suggested that seasonal patterns could adequately be captured although the CFSR data did uniformly overestimate and underestimate measured rainfall. A recent study from Dile and Srinivasan (2014) evaluated the use of CFSR data for hydrological prediction using SWAT in the Lake Tana basin, Ethiopia. The study achieved satisfactory results in its simulations for both CFSR and conventional data. While the outcome was better with conventional weather data, the study concludes that CFSR could be a valuable option in data-scarce regions. Other studies using CFSR data not in the Ethiopian context (Alemayehu, 2015 and de Almeida Bressiani, 2015) and with large to very large catchments (13'750 to 73'000 km^2) concluded that CFSR data gave good to very good results and the SWAT model responded reasonably. One CFSR application in the Dongi and Puli river basins in China by Yang et al. (2014) with watershed sizes from 366 to 1098 km^2 concluded that CFSR data was significantly different and that the CFSR data spatial distribution might be the cause for the weak performance.**
Another key point I want to raise is the potential improvement of the discussion. For instance, in the conclusion the authors claim (**P11069 L12-13**) that "CFSR data may not be applicable for small-scale modelling". Based on my own experience of CFSR data I totally agree with it, but the authors should previously extend their discussion comparing their results to the results of the other studies using CFSR data, including the size of the case-study basins. It seems intuitive that for larger basins, CFSR data errors are balanced and the hydrological modelling achieves better quality, but the authors should explore it through the literature.	- Discussion has been expanded to include more general comparison, and specifically on the issues of size of study areas - Two sentences have been added to the Introduction section for better understanding.	
P11054 L11: minimal surface is not consistent with Table 1 and with P11057 L8	We have added clarifications to the text and to the table.	Added: - P11054 L11 (abstract): no changes - Table 1: Added "Size WLRC" and "Size SWAT-delineation" - P11057 L8: corrected typographic error
P11054 L24-25: "is one of the most important input parameters…" that is true, but needs to be supported by adequate references. In addition, this is probably true for all other hydrological models, it could be good at that point to broaden the introduction before introducing SWAT	Sentence was adapted for better understanding and several references were added to the Introduction section. Arnold et al. (1998, 2012), Worqlul (2014), Fuka (2014), Dile & Srinivasan (2014), de Almeida Bressiani et al. (2015), Alemayehu (2015)	Accurately represented, spatially distributed **hydro-meteorological and hydro-climatic data** are the most important input parameters for hydrological modelling with the Soil and Water Assessment Tool (SWAT).
P11054 L26:	References have been added.	[…]for hydrological modelling with the Soil and Water Assessment

references to SWAT papers would be welcome here for non-users e.g. Arnold et al., 1998, 2012 (see ref list below)	Arnold (1998, 2012), Douglas-Mankin et al. (2010)	Tool**, called SWAT hereafter (Arnold et al., 1998, 2012; Douglas-Mankin et al., 2010).**
P11055 L4: "for SWAT website" what do the authors mean? The sentence should be rephrased to avoid ambiguity	The sentence was changed and a reference was added containing the corresponding URL in the bibliography. Furthermore the website has been added to the "Methods and materials" section.	Changed the sentence: **The Climate Forecast System Reanalysis (CFSR, 2014) readily provides, for any coordinated on the globe, a climate data set adapted to SWAT**.
P11055 L11: ";" are not appropriate here	Punctuation changed	Replaced ";" with ","
P11055 L12-14: this sentence is hazardous: if the work has already been done, why doing it again?	We are not entirely sure we understand the referee's intention here, as we deem it necessary to state that first there was a rainfall comparison and only afterwards the SWAT modelling results were compared. Both are important and both have their right to exist. Therefore we adapted the sentence according to our best understanding.	Added […] **a first evaluation, carried out by our research group, of CFSR-modelled rainfall data** […].
P11056 L11: "35 years" -> 33 years? This is however not consistent with P11059 L21 and Table 1, and not all the 3 stations have the same "year of construction".	Added details for a better understanding of the sentence	[…] compared to CFSR data over a **maximum** period of **34** years from 1981 to 2014 **(for Maybar, 33 years for Andit Tid and 32 years for Anjeni)**
P11056 L12: "annual, interannual, and seasonal cycles" this is one of the added value of the paper and I recommend the authors to better highlight it, for instance by extending the introduction on this particular point.	We extended the Introduction section accordingly and tried to highlight the importance of seasonal cycles in analysing and validating the CFSR data set.	First the CFSR data were statistically compared to measured WLRC rainfall data for accurate representation of annual, interannual, and seasonal cycles. **This is important because temporal occurrence of rainfall has a great impact not only on discharge but moreover on sediment yield generation. Many crop types are sowed at the beginning of the rainy season(s), which implies extensive extensive ploughing beforehand, which leaves fields unprotected for the first few rainfall events. Hence, is clear that temporal occurrences of annual, interannual and seasonal cycles play a crucial role for the validation of a data set like the CFSR climatic data.**
P11056 L19: Method: somewhere the authors have to describe the material used for rainfall and temperature measurements, their frequency, the spatial resolution of the CFSR data as put into the model, the number of measurement stations respect to the number of CFSR stations, etc.	Changed title of section to "Methods and materials" Added description of frequency and spatial resolution	**2.2 Hydrometeorological data** **The hydrometeorological data consists of two sets. The conventional or measured data contains daily rainfall and maximum and minimum temperature from one climatic station for each watershed. These climatic stations have been installed in the early 1980s and span the period until 2014 with some larger gaps (see Table 1 for details) mainly from 2000 to 2010.** **The CFSR data (The Texas A&M University spatial sciences website, globalweather.tamu.edu) was obtained for the entire Blue Nile Basin (Bounding box: latitude 8.60°, -12.27°N and longitude 33.94°, -40.40°E) before choosing the four closest stations for each watershed. It includes daily rainfall, maximum and minimum**

		temperature as well as wind speed, relative humidity, and solar radiation for 12 locations (see Figure 1 for details).
P11056 L25 and L27: "calibrated": what is the time-step? What about validation process mentioned in Table 1?	Thank you for pointing this out. We modified the sentence to integrate the validation process.	The SWAT model was calibrated **and validated** for discharge once using WLRC climatic data **set and in another run** using the CFSR climatic data set.
P11057 L8: minimal surface is not consistent with Table 1 and with P11054 L11.	Changed minimal surface to reflect size from table 1.	Changed 102 to **112 ha**
P11057 L9: altitudes are not consistent with Table 1.	Changed altitudinal ranges to fit table 1	Changed 2400 to 3548 masl to **2406 to 3538 masl**
P11057 L12: "divided: : : into one: : :" is not much elegant, I guess the sentence can be rephrased.	Very true. Thank you very much for this observation. We changed the sentence for more "elegance" and introduced "belg" and "kremt" at the same time.	Changed the sentence to: **Anjeni has a unimodal rainfall pattern while Andit Tid and Maybar have a bimodal rainfall regime with a small rainy season from March to May (belg) and a main rainy season from June to September (kremt) followed by a long dry season from October to March.**
P11057 L16: I doubt ArcSWAT "was used to assess the impact of different rainfall patterns: : :" but I believe SWAT was instead used for it and the version of SWAT used for it should be mentioned. ArcSWAT is just the interface to build the SWAT model to be run. To make it clear, the authors should separate what ArcSWAT does and what SWAT does.	- Sentence was modified. - Added SWAT version - Clarified the structure	**SWAT (SWAT2012 rev. 620) was used** to assess the impact of different rainfall patterns on run-off and sediment loss dynamics **through the ArcSWAT interface** (Version 2012.10_1.14).
P11057 L19: "other regions" could be introduced before "Ethiopia" to make the references clear.	This suggestion is not entirely clear to us. Changed the sentence according to our understanding.	[…] and parameterization **all over the world and in** Ethiopia […]
P11057 L22: "specific" what do you mean? You could also say that SWAT is a semidistributed hydrological model.	This is a citation from the theoretical documentation of SWAT by Arnold et al. (2012). We would like to leave the sentence as it is.	No changes to the sentence.
P11058 L3-4 and P11058 L5-6: Which method did the authors use?	Changed the sentence to reflect only used methods and removed unnecessary enumerations	The surface runoff **was estimated using the Natural Resources Conservation Service Curve Number** (SCS-CN) method (USDA-SCS, 1972). […]
P11058 L28: ": : :heat unit function: : :" this is not "Spatial data", the authors could rename the section title or move the sentence in the section where they describe the model (section 2.1.1)	This is very much appreciated. We removed the sentence to a more appropriate section.	Moved the sentence to "SWAT model setup", where it is appropriate.
P11058 L28: what is "Teff"? More details would enlighten the reader	Added details for Teff.	Teff **(eragrostis teff), a widely cultivated and highly nutritional crop native to Ethiopia**, was planted beginning of July […]

who is not familiar with African inter-tropical agronomy.		
P11059 L6: "satellite images" which ones? How many? Were images captured during low flow or high flow?	There was one satellite image for each watershed, hence the plural. As the river beds in these catchments do not vary at all, the stream network compatibility check consisted merely of checking general errors in stream network.	Added: […] **(one stallite image for each watershed).**
P11059 L7: "The sub-basin sizes were fixed at 2 ha" what do the authors mean? Is it the minimal drainage area?	Removed the sentence entirely as there is no gain in information here.	Removed: The sub-basin sizes were fixed at 2 ha.
P11059 L8: "All HRUs were defined: : :" The authors should explain why they kept such accuracy. Did for instance the authors use a detailed land cover map?	We used the zero percentage threshold area because of the very detailed land use/land cover map.	No changes to sentence.
P11059 L13: Which data where used as input into the SWAT weather generator? Only measured? Only CFSR? Both depending if measured or CFSR rainfall data was used? Did the authors compare their temperature measurements to the CFSR temperature	Thanks again for this observation. We modified the three sentences for a better understanding.	The CFSR time series were complete from 1979 to 2014. The WLRC data had substantial gaps in the time series, mostly in the early 1990s and after 2000 (see Table 1 for details). The SWAT weather generator was used to fill the **gaps in the WLRC data set** for rainfall and temperature. Otherwise daily precipitation and minimum and maximum temperature data were used to run the model.
P11059 L16: "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.	Unclarity here is due to the way sediment is collected in the SCRP/WLRC research stations: personnel are instructed to take grab samples only when the river turns brown and to continue taking samples until the rivers water turn clear again. Outside rainfall events SCRP/WLRC assumes there is no sediment in the river. In this paper we are describing the procedure only. We added a sentence for clarification to the section.	The flow observations are available throughout the entire year while **measured** sediment concentrations **from grab samples** are only available during rainstorm events. G**rab samples have only been collected during rainfall events, when the river is turning brown.**
P11059 L19: Can sediment concentration be "visible"? Aren't the authors talking about turbidity? Then what was the turbidity threshold to describe it "visible"? Was it kind of experts' knowledge?		
P11060 L1: The section title is a kind of mix up that the authors should clarify according to what the section is dealing with. See also my previous comment: merge similar topics together.	Adapted	Change section title to: **Calibration setup, validation, and sensitivity analysis**
P11060 L2: The Abbaspour reference should point to SWAT-CUP, not to SUFI2.	Thank you for this observation	Moved reference to point at SWAT-Cup
P11060 L14: "Nash-Sutcliff" -> Nash-Sutcliffe; "standardized Root: : :" -> "the standardized Rood: : :"	Adapted	Changed sentence to: […] used the Nash-Sutcliff**e** Efficiency (NSE), **the** standardized Root Mean Square Error […]
P11060 L15: "All are very commonly: : :" the sentence is slightly	We do agree that the sentence might be clumsy, but what we wanted to point out is that these parameters are	Changed the sentece to: **These are well-known statistical parameters, which are often used**

clumsy! The authors may give a stronger justification for using those criteria, instead of just considering it's good to use them because everybody do so. The authors could also just remove the sentence.	commonly used for comparison of time series, especially for modelling results, which makes their application useful as our results can then be compared to other studies.	**for comparison of time-series especially in hydrological modelling (Dile & Srinivasan, 2014; Abbaspour, 2015; Moriasi et al., 2007; Starks and Moriasi, 2009; De Almeida Bressiani et al., 2015, Gebremicael et al., 2013, Alemayehu et al., 2015) and therefore help others to compare our modelling results to previous studies.**
P11060 L22: "and a better accuracy of observations: : :" this is not clear	Sentence modified	Changed sentence to: […] **with a perfect concordance of modelled to observed data at 1, a balanced accuracy at 0 and a lower accuracy of modelled data below zero.**
P11060 L22: "RMSE" the authors should detail RMSE here.	Details added to the RMSE	The RSR is a standardized **Root Mean Square Error (RMSE, standard deviation of the model prediction error),** which is calculated […]
P11060 L27: ": : :model simulation to a large: : :" -> ": : :model simulation, to a large: : :" a comma is needed here to make the sentence clear, or the sentence should be rephrased.	Sentence rephrased	RSR varies from the optimal value of 0, which indicates zero RMSE or residual variation, **to a large positive value, which indicates large RMSE or residual value and therefore worse model simulation performance**
P11061 L10: "Legates and McCabe" these authors don't need to be called twice.	Sentence modified	As suggested by **Legates and McCabe (1999)** […]
P11061 L17: "Belg" and "Kremt" timing and durations should be described in the method section.	The timings of belg and kremt have been added to section "Study area" and to table 4.	**Anjeni has a unimodal rainfall pattern with a main rainy season from June to September while Andit Tid and Maybar have a bimodal rainfall regime with a small rainy season from March/April to May (belg) and a main rainy season from June to September (kremt) followed by a long dry season from October to March.**
P11063 L23 and P11065 L23: Table 5: What about calibration and validation results? In Table 5 I guess only overall goodness-of-fit indices are given, what about the specific values for calibration and validation? Is hence the model truly validated? The authors should discuss it.	We are not sure what is meant by "specific value" in this context? Therefore we do not know how to answer this question. The values given in Table 5 are commonly agreed goodness-of-fit statistical parameters, which define how well a model fits observed values. Although we feel the "Results" and the "Discussion" section contain an adequate amount of information already, we added the validation data to the table and the Results section.	**Validation data added in Table 5** Sentences added to the results section: For each station we added validation evaluation in the form of: - **Validation of discharge for Maybar with WRLC data showed *good* results with RSR: 0.56, NSE: 0.74 and PBIAS 17.3 and *unsatisfactory* results for the CFSR dataset with RSR: 0.98, NSE: 0.04, and *very good* PBIAS: --1.9.** - **Validation of sediment yield for XY with WRLC data showed a marginally *satisfacroy* result with RSR: 0.68, NSE: 0.51 and *unsatisfactoy* PBIAS --64.3 indicating a general overestimation and *unsatisfactory* results for the CFSR dataset with RSR: 1.39, NSE: --0.94, and *satisfactory* PBIAS: --11.9 indicating underestimation.**

		Sentence added to the conclusion section: **The WLRC rainfall data set resulted in three calibrated and validated discharge models while the CFSR data resulted in none. For the sediment loss modelling the WLRC rainfall data resulted in two out of three calibrated and validated models while none could be adequately calibrated for the CFSR data set.**
P11063 L24: "Each model: : :" this is method.	Thank you for this observation. The paragraph was modified accordingly.	Removed: Each model was calibrated with one to five iterations using 500 simulations each. The data was split into calibration and validation periods, which contained similar amplitudes (see Fig. 3 for further details) over their respective periods. Parameters initially contained original ranges, which were gradually adapted according to modeling results.
P11064 L8: "were maximized, but still inside SWAT absolute values" what do the authors mean?	Sentence was removed as there is no additional information gained from it. What we initially meant is that parameter ranges were maximized or minimized to one of the two ends of the initial parameter range but kept inside the physically defined absolute values suited for the parameter.	Removed: Parameter ranges settings were maximised, but still inside SWAT absolute values (Abbaspour, 2007).
P11064 L16: ": : :rainfall data proved impossible" what do the authors mean?	Sentence was modified to improve understanding	**Satisfactory calibration could not be reached with CFSR data and neither baseflow, nor peaks could be adequately represented.**
P11065 L8: ":" are not appropriate	Modified sentence and removed the ":"	Nonetheless, *satisfactory* results were achieved for discharge with RSR**,** NSE**,** and PBIAS […]
P11065 L10-13: "The hydrograph: : : increasing." This sentence is not clear and needs to be rephrased.	Sentence was modified for improved clarity	**Figure 3 shows regular discharge peaks from February to March, in accordance to rainfall pattern deviation as seen on Figure 2, when no increase of discharge was measured at the research station. At the same time, the kremt season is regularly underestimated when using the CFSR rainfall input, while the measured discharge is increasing during the same period.**
P11065 L19-22: "Sediment loss modelling: : :untouched" this is method. However, one can criticize this method: since sediment loss is driven by the hydrology, then calibration process could also be more reliable when calibrating both discharge and sediments at the same time. The authors could explain their method choice.	Yes and no. According to Abbaspour (2015) and Arnold et al. (2012) this is the appropriate method to use for calibration of sediment loads. Therefore we added two references indicating this approach for calibration. Maybe we did not state that clearly enough but the method consists of calibration of both at the same time. The hydrology is calibrated first and then calibrated hydrologic **parameter ranges** are left untouched. This only means that hydrologic parameter ranges are not further modified outside a calibrated parameter range.	**Soil loss modelling was calibrated using the same set of 9 parameters for each catchment including the calibrated discharge parameter ranges (see Table 2 for description (Abbaspour, 2015 and Arnold et al. 2012)**.

	We added references and we added the word “ranges” to further clarify the sentence.	
P11067 L5: Conclusion: the 7 first paragraphs are an extended summary of the results that is not strictly appropriate for a conclusion. Key outputs from this study are coming in L28 and should be highlighted.	The seven first paragraphs of the conclusion have been deleted, as they are a repetition, thank you for that observation. We did, however, add a paragraph referring to the first referee comment at the top of this file regarding the added value of the hydrological modelling to show discrepancies of CFSR to conventional rainfall data.	**The SWAT modelling showed that CFSR rainfall pattern and rainfall yearly total amount variations were so significant that SWAT model calibration could not adequately represent measured discharge and sediment yield.**
Table 1: What does “Year of construction” refer to? The year of construction of the gauging station? Did the measurements started just after construction? What is the meaning of “field scale” for land use and soil maps? About “daily sediment loss”: what was measured, the concentration or the load? In guess the concentration in the river (at the gauging station) is slightly different from the load lost from the hillslope. Sources could be given as a table footnote.	- Thank you for these observations. Some of the problems are tackled in the text, but we are aware that the table has to be improved. This is why we added some table notes and additional information. - Details for “Year of construction” - Sources for watershed sizes - Details for “field scale” - Watershed sizes calculated by the ArcSWAT delineation tool Source for the soil map	Table 1. **Description of study sites, data sources and time series and gaps. The subdivision of data relates to calibration and validation periods.**
Table 2: The title should explain the meaning of “a” and “v” in SWAT-CUP. The table should also show the initial value ranges to remind section 2.4. A “_” is missing to a_CN2.mgt and “hu” needs to be corrected to “hru”.	Thank you for these very useful observations. These were very unfortunate mistakes and they have been corrected. - a__ and v__ meanings have been added to the table as a table note. - “Initial ranges” column has been added to the table - The second “_” has been added to a__CN2.mgt - “hu” has been changed to “hru”	Table 2. **Swat parameters used for discharge and sediment loss calibration with initial ranges and fitted final parameter ranges.** No further changes to table 2
Table 2: What is the meaning of the very small changes (e.g. -0.0038, 0.0023 for ESCO, -0.084 for SURLAG…) ?	We are not quite sure about this question – this is why we separated it from the question above to be able to respond in a clear and concise manner: - ESCO has absolute values only from 0 to 1, which means, changes will most probably be very small. - The very small changes are also the result of changes suggested by SWAT-Cup for a parameter that is not highly sensitive. In order to minimize the relative width of the 95PPU we accepted SWAT-Cup parameter range reduction for less sensitive parameters	
Table 3: To my opinion, this table is useless. If one wants the detail of Moriasi’s paper, then he can read his paper. But what	We do not entirely agree with that comment. Even though the referee’s comment is pertinent, we feel that performance criteria are better suited in a table than in the	Bold highlights have been added to **table 5** where statistical performance ratings meet the “satisfactory” criteria by Moriasi. Table 3 is kept in the manuscript as is.

to my opinion would be really useful, is to highlight (using bold, italic or whatever) the values that meet the satisfactory criteria values in Table 5.	text itself or as a reference. If readers need to find Moriasi's paper first, and then the table inside that paper it feels more straighforward to keep the table in the manuscript. Therefore we will keep that table.	
Table 4: The title should be more detailed. What are the 2 rainfall datasets compared? What is their duration, which periods are compared? Which region of the world are we talking about? What are Kremt and Belg? In general, the title of a Table or a Figure should give enough information to the reader that he does not necessarily need to go through the manuscript to understand the table/ figure.	Thank you very much for this observations. We have added necessary information to the table itself and to the title of the table. - Changed title of table 4 - Added duration in brackets in table - Added definition of seasons in table - Added highlighting of satisfactory performance ratings	Table 4. **Seasonal comparison of rainfall time series of daily rainfall amounts. Satisfactory performance ratings are highlighted in bold. Details for duration and gaps can be found in table 1.**
Table 5: See my comment to Table 3 and my comment to P11063 L23 and P11065 L23. Bold/Italic highlights should be explained in the title. The title should also remember that calibration and validation periods are given in Table 1.	Thank you again for this helpful comment. We agree that the table is much more concise and clear like that. We did refrain from including validation results because they show the exact same tendency while cluttering the table.	**Table 5:** - Bold highlights added where statistical performance ratings meet the "satisfactory" criteria - Title was adapted to show only "Calibration" results
Figure 1: It's difficult to get an idea of the relative scale of the 3 small sub-catchments of interest. Reporting the shapes in the main figure could enlighten the reader. If the sub-catchments are too small then another representation should be considered. Berha, Kolla, Dega, Wurch are not described in the manuscript. The title should say the map is a land use map (I guess ?) and give the year corresponding to the land use shown in the Figure.	We agree with the comment of referee LB. Therefore we removed the sub-catchment representation and adapted the map and caption accordingly. 1. Changed map content and legend content 2. Adapted title 3. Added source for representation 4. Added title "Agroecological zones" on map 5. Added locations of CFSR rainfall stations used for comparison	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.**
Figure 2: WLRC stations are not starting in 1979. Why not putting Dry Season, Kremt and Belg in those figures and referring to it throughout the manuscript?	- Changed the title of figure 2 and removed reference to time series length - Added a sentence in figure caption to refer to Table 1 for details on time series length - We did not add Kremt and Belg in the image because we want to preserve the information as it is. We think that adding more information would overstrain the statement and create confusion.	Title: **Precipitation distribution** Figure 2. **Monthly CFSR and WLRC rainfall distribution of all station as boxplots with monthly rainfall distribution. CFSR data from 1979 to 2014 and WLRC data from 1981/1982/1984 to 2014. See Table 1 for details.**
Figure 3: Again, the title should be more concise. The figure shows both observed and modelled discharge, discharge is simulated from both WLRC and CFSR rainfall datasets, etc.	- Changed the caption of figure 3 to better reflect figure content and added some more information.	Figure 3. **Modelled SWAT discharge compared to measured discharge (blue) for WLRC (violet) and CFSR (pink) input data and the 95 Percent Prediction Uncertainty (light blue). Each sub-figure contains the calibration and the validation period. Results are given in m^3/s.**

Figure 4: See comment to Figure 3.	- Changed the caption of figure 4 to better reflect figure content. - Added some more information on content.	Figure 4. **Modelled SWAT sediment loss compared to measured sediment loss (blue) for WLRC (red) and CFSR (gree) 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 (t).**