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).**

Reply to referee comment C. Baffaut

We would like to thank C. 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.

Referee comment	Authors answer	Changes in manuscript
The paper clearly shows significant discrepancies between the CFSR and WLRC precipitation data (section 3.1 and table 4). I don't understand why the authors go further and present the results of the modelling using CFSR data as inputs. What are the chances to have useful results? Why is there a need to go through the analysis of model results with CFSR data? Is there evidence in the literature or in the policy world that these considerations are not well taken into account?	This is indeed a valuable question and we would like to respond in the same way we answered the that question in the other review: 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.	- see answers to detailed comments from referee LB - The particular point about the added value of hydrological modelling has been added to the "Introduction" section and the "Conclusion" section: Introduction: **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 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. Conclusion: **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 2: Unless a reader is familiar with the SWAT-CUP specific notation, the parameter names and values will not be understood.	Thank you very much for that comment. Table 2 has been adapted accordingly	- Added description for parameters - Added initial values of parameters

Manuscript prepared for Hydrol. Earth Syst. Sci.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 30 December 2015

Comparing CFSR and conventional weather data for discharge and ~~sediment~~ soil loss modelling with SWAT in small catchments in the Ethiopian Highlands

Vincent Roth[1,2,3] and Tatenda Lemann[1,2,3]

[1]Centre for Development and Environment (CDE), University of Bern, Switzerland
[2]Sustainable Land Management Research Group, University of Bern, Switzerland
[3]Water and Land Resource Centre (WLRC), Addis Abeba, Ethiopia

Correspondence to: Vincent Roth (vincent.roth@cde.unibe.ch)

Abstract. Accurate rainfall data are the key input parameter for modelling river discharge and ~~sediment~~ soil loss. Remote areas of Ethiopia often lack adequate precipitation data and where it is available, there might be substantial temporal or spatial gaps. To counter this challenge, the Climate Forecast System Reanalysis (CFSR) of the National Centers for Environmental Prediction (NCEP) readily provides weather data for any geographic location on earth between 1979 and 2014. This study assesses the applicability of CFSR weather data to three watersheds in the Blue Nile Basin in Ethiopia. To this end, the Soil and Water Assessment Tool (SWAT) was set up to simulate discharge and ~~sediment~~ soil loss, using CFSR and conventional weather data, in three small-scale watersheds ranging from ~~102~~ 112 to 477 ha. Calibrated simulation results were compared to observed river discharge and observed ~~sediment~~ soil loss over a period of 32 years. The conventional weather data resulted in very good discharge outputs for all three watersheds, while the CFSR weather data resulted in unsatisfactory discharge outputs for all of the three gauging stations. ~~Sediment~~ soil loss simulation with conventional weather inputs yielded satisfactory outputs for two of three watersheds, while the CFSR weather input resulted in three unsatisfactory results. Overall, the simulations with the conventional data resulted in far better results for discharge and ~~sediment~~ soil loss than simulations with CFSR data. The simulations with CFSR data were unable to adequately represent the specific regional climate for the three watersheds, performing even worse in climatic areas with two rainy seasons. Hence, CFSR data should not be used lightly in remote areas with no conventional weather data where no prior analysis is possible.

1 Introduction

Accurately represented, spatially distributed ~~rainfall is one of~~ hydro-meteorological and hydro-climatic data are the most important input parameters for hydrological modelling with the Soil and Water Assessment Tool~~(SWAT)~~, called SWAT hereafter (Arnold et al., 1998, 2012; Douglas-Mankin et al., 2010) . Although a great deal of effort is being invested into rainfall and climatic data collection, many areas of Ethiopia have no adequate precipitation data, and where such data are available, the monitoring network contains substantial temporal and spatial gaps. This makes it necessary to use other sources of modeled rainfall data for SWAT modelling. The ~~Global Weather Data for SWAT website~~ Climate Forecast System Reanalysis (CFSR) readily provides, for any ~~coordinates~~ coordinated on the globe, a ~~Climate Forecast System Reanalysis (CFSR) data set for download~~climate data set adapted to SWAT. This data set is the result of the close cooperation between two United States organizations, the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR), which have completed a global climate data reanalysis over ~~35~~ 36 years from 1979 through 2014. The CFSR data is based on a spectral model which includes the parametrisation of all major physical processes as described in detail in Kalnay et al. (1996)~~; Kistler et al. (2001) ;~~ , Kistler et al. (2001) , and Saha et al. (2010).

However, ~~a first comparison~~ the applicability of the CFSR data for small-scale catchments in the Ethiopian Highlands has not been adequately investigated yet. Aforementioned studies did focus on large basins with numerous CFSR stations, which tend to balance errors in rainfall patterns. 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. So far, few studies have been conducted in the Ethiopian context on the impact of rainfall data on streamflow simulations. ~~The impact of spatial variability of precipitation on model run-off showed that standard uniform rainfall assumptions can lead to large uncertainties in run-off estimation (Faurès et al., 2000) . Several studies evaluating the CFSR data set have suggested that climatic models tended to overestimate interannual variability but underestimate spatial and seasonal variability (Diro et al., 2009)~~ Fuka et al. (2014) used CFSR data in a 1200 km^2 watershed in Ethiopia with SWAT suggesting CFSR data performs as good 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 ~~(Dile and Srinivasan, 2014)~~ 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 (De Almeida Bressiani et al., 2015; Alemayehu et al. 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 to the data set. One CFSR application in China (Yang et al., 2014) with meso-scale watersheds (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.

The impact of spatial variability of precipitation on model run-off showed that standard uniform rainfall assumptions can lead to large uncertainties in run-off estimation (Faurès et al., 2000) . Several studies evaluating the CFSR data set have suggested that climatic models tended to overestimate interannual variability but underestimate spatial and seasonal variability (Diro et al., 2009) . In another study Cavazos and Hewitson (2005) performed statistical downscaling of daily CFSR data with Artificial Neural Networks, and their predictions showed low performance in near-equatorial and tropical locations, which led them to conclude that the CFSR data is most deficient in locations where convective processes dominate. Another study found the CFSR data set performed well on a continental scale but that it failed to adequately reproduce some regional features (Poccard et al., 2000). A study in China performed streamflow simulations by SWAT using different precipitation sources in a large arid basin using rain gauge data combined with Tropical Rainfall Measuring Mission (TRMM) data (Yu et al., 2011). The study established that streamflow modelling performed better using a combination of TRMM and rain gauge, as opposed to rain gauges only. Different interpolation schemes with the use of univariate and covariate methods showed that Kriging and Inverse Distance Weighting performed similarly well when used with the SWAT model (Wagner et al., 2012).

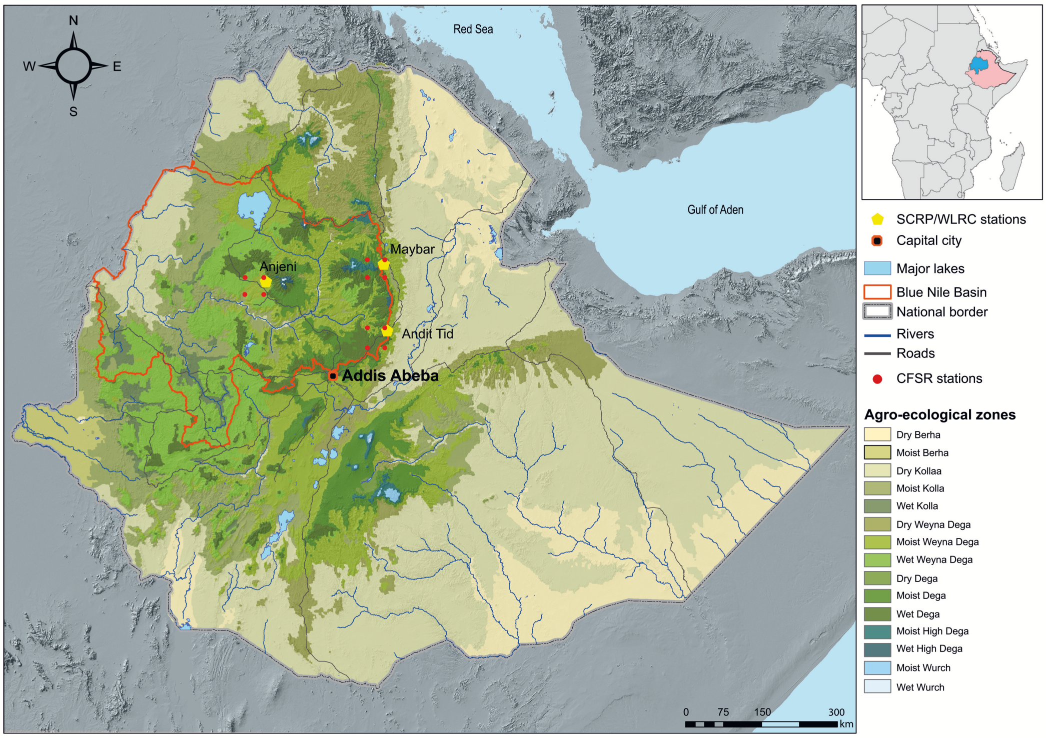

Figure 1. Map overview of Blue Nile (Abbay) Basin with the WLRC research stations, agro-ecological zones according to Hurni Hurni (1998) and emplacements of CFSR stations.

In this paper, WLRC and SCRP rainfall data (hereafter called WLRC data) are compared to CFSR data over a ~~period of 35~~ maximum period of 34 years from 1981 to ~~2014.~~ 2014 (Maybar, 33 years for Andit Tid and 32 years for Anjeni). The main objective of this paper is to compare the two data sets for annual, interannual, and seasonal cycles and subsequently to compare the effects on discharge and ~~sediment~~ soil loss modeling when using these data sets in three locations in the Ethiopian highlands (see figure 1). Calibrated CFSR modeled discharge and ~~sediment~~ soil loss is then compared to calibrated WLRC modelled discharge and ~~sediment~~ soil loss, and the applicability of the CFSR data in small-scale catchments for hydrological predictions is statistically evaluated and compared.

2 Methods and materials

The effects of spatial and temporal variability in the CFSR rainfall data set for the study area were examined in several steps. 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. Second, the impact of spatial and temporal variability of rainfall on hydrology and soil loss was assessed by modeling discharge and ~~sediment~~ soil loss with the SWAT model. The SWAT model was calibrated for discharge once using WLRC ~~rainfall~~ climatic data and once using the CFSR ~~rainfall~~ climatic data set. Afterwards ~~sediment~~ soil loss was calibrated for each catchment. In a last step discharge and ~~sediment~~ soil loss on a monthly basis were statistically and visually compared using performance ratings established by Moriasi et al. (2007).

2.1 Study area

The study areas of the three micro-scale catchments are located in the eastern and central part of the Blue Nile Basin. The Anjeni (AJ) and the Andit Tid (AT) are sub-basins of the Blue Nile Basin, which drains towards the west into the main Nile at Khartoum. The Maybar (MA) catchment drains into the Awash river to the East of the Ethiopian highlands. The catchment sizes range from ~~104 ha to 447~~ 112 ha to 477 ha and their altitudinal ranges extend from ~~2400 to 3548~~ 2406 to 3538 masl (see table 1 on page 6 for details). The catchments have a sub-humid to humid climate with an annual temperature ranging from 12° C to 16° C and a mean annual rainfall ranging from 1211 mm to 1690 mm. ~~The rainy seasons are divided into two seasons for~~ Anjeni has a unimodal rainfall pattern with a main rainy season from June to September while Andit Tid and Maybar ~~and into one for Anjeni~~have a bimodal rainfall regime with a small rainy season from April to May (belg) and a main rainy season from June to September (kremt) followed by a long dry season from October to March. Land use is dominated by smallholder rain-fed farming-systems with grain-oriented production, ox-plough farming, and uncontrolled grazing practises.

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, 4 for each watershed (see Figure 1 for details).

2.2.1 ~~Hydrologic model~~

~~ArcSWAT (Version 2012.10.14~~

2.3 Hydrologic model

SWAT (SWAT2012 rev. 620) was used to assess the impact of different rainfall patterns on run-off and ~~sediment loss dynamics (Arnold et al., 2012)~~ soil loss dynamics through the ArcSWAT interface (Version 2012.10_1.14). Here, we present the SWAT model only briefly, as it has been widely used in the past, with extensive review of its performance and parameterization in Ethiopia and other regions (Gessesse et al., 2014; Mbonimpa, 2012; Betrie et al., 2011; Tibebe and Bewket, 2011; Lin et al., 2010; Stehr et al., 2008; Schuol and Abbaspour, 2007). SWAT is a physically-based river basin or watershed modelling tool. The SWAT model requires specific information about weather, soil properties, topography, vegetation, and land management practices occurring in the watershed (Arnold et al., 2012). 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). The surface run-off is estimated in the model using one of two options (1) the Green and Ampt method (Green and Ampt, 1911) or (2) the Natural Resources Conservation Service Curve Number (SCS-CN) method (USDA-SCS, 1972). The flow routing is estimated using the variable storage coefficient method (Williams, 1969), or the Muskingum method (Chow, 1959). ~~Sediment~~ soil loss for each HRU is calculated through the Modified Universal Soil Loss Equation (MUSLE). Sediment routing in channels is estimated using

Table 1. ~~Study~~ Description of study sites, ~~model input~~ data ~~,~~ sources and ~~available~~ time series and gaps. The subdivision of data relates to calibration and validation periods.

	Andit Tid	Anjeni	Maybar
Year of construction[a]	1982	1983	1981
Location	9.815° N	10.678° N	10.996° N
	37.711° E	37.530° E	39.657° E
Size WLRC	477.3 ha[b]	113.4 ha[c]	112.8 ha[b]
Size SWAT-delineation	466.78 ha	105.23	101.98
Altitudinal range	3040–3538 masl	2406–2506 masl	2530–2857 masl
	Data resolution		
DEM	2m		
Land use map[d]	field scale		
Soil map[e]	5x5m		
Climatic data	Daily precipitation		
	Daily min. and max. temperature		
Hydrology data	Daily discharge		
Soil loss data	Daily soil loss		
Sources	SCRP/WLRC/CDE/own		
	Data availability		
	Andit Tid	Anjeni	Maybar
Precipitation data	1982–2004	1984–2004	1981–2001
	2006	2010–2014	2004–2006
	2010–2014		2010–2014
Temperature	1982–1993	1984–1993	1981–1993
	1997–2002	1998–2004	1995–1998
	2010–2013	2010–2013	2010–2013
Discharge	1982–1993	1984–1993	1981–1993
	1995–1997	1995–2000	1997–2006
		2011–2014	2010–2014
Sediment	1982–1993	1984–1993	1981–1991
	1995–1997	1995–1998	1995–2006
	2011–2014	2011–2014	2011–2014
	Subdivision of data		
Calibration	1984 – 1993	1986 – 1998	1983 – 2006
Validation	1994 – 1997	2010 – 2014	2008 – 2014

[a] Year of construction is the year the station was built and monitoring started.

[b] Source: (Bosshardt, 1999)

[c] Source: (Bosshardt, 1997)

[d] Every field in the watershed was attributed a land use type on the map

[e] Source: (Belay, 2014)

stream power (Williams, 1980) and deposition in channels is calculated through fall velocity (Arnold et al., 2012; Gassman et al., 2007).

2.4 Spatial data

The spatial data used in ArcSWAT for the present study included the digital elevation model (DEM), land use data, and soil data (see table 1 for details). The DEM for the three WLRC watersheds was developed by the Centre for Development and Environment (CDE) of the University of Bern, Switzerland, for the former SCRP (SCRP and CDE, 2000a, b, c) and has a resolution of 2 m. The spatial distribution of soils for Anjeni was adapted from a soil survey carried out by the SCRP (Kejela, 1995) and a PhD dissertation by Gete Zeleke (2000). The physical and chemical parametrisation of the soil was adapted from the soil database in Zeleke's thesis and from Kejela's report. The soil characteristics for Maybar were adapted from the SCRP's Soil Conservation Research Report 7 (Weigel, 1986) and for Andit Tid from the SCRP's Research Report 3 (Bono and Seiler, 1984). 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). ~~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).~~

2.5 SWAT model setup

The watersheds were delineated using the Arc–SWAT delineation tool and its stream network compatibility was checked against the stream network from satellite images ~~. The sub-basin sizes were fixed at 2 ha.~~ (one satellite image for each watershed). SWAT compiled 1038 HRUs for Anjeni, 1139 HRUs for Maybar, and 728 HRUs for Andit Tid respectively. 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. ~~Daily precipitation and minimum and maximum temperature data at three WLRC stations were used to run the model with conventional weather inputs. All three WLRC stations~~ 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 on page 6~~ 1 for details). The SWAT weather generator was used to fill the gaps ~~for rainfall, temperature, solar radiation, and relative humidity~~in the WLRC data set for rainfall and temperature. Otherwise daily precipitation and minimum and maximum temperature data were used to run the model. Potential evapotranspiration (PET) was estimated using the Hargreaves method (Hargreaves et al., 1985). Daily river flow and sediment concentration data were measured at the outlet of the three WLRC watersheds. The flow observations are available throughout the entire year while ~~sediment concentrations~~ calculated sediment concentrations from grab samples are only available during rainstorm events ~~, when sediment concentrations are visible in the river~~ and are extrapolated over the whole time period. Personnel at the research station are instructed to take grab samples only during rainfall events, when the river is turning brown. 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 (eragrostis teff), a widely cultivated and highly nutritional crop native to Ethiopia, 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) with a tillage depth of 20 cm and a mixing efficiency of 0.3. ~~During the dry season and outside rainfall events the monitored rivers are assumed sediment free.~~

The model was run for 32 years from 1983 to 2014 with daily data inputs but monthly outputs. Calibration and validation periods were chosen equally balanced regarding high-flow and low-flow years in all three catchments. The model was first calibrated and validated for discharge and then calibrated and validated for ~~sediment~~ soil loss (see Table 1 on page 6 for details).

2.6 Calibration setup, ~~parameterization~~ validation, and ~~uncertainty~~ sensitivity analysis

The SUFI-2 algorithm ~~(Abbaspour et al., 2004, 2007)~~ in SWAT-Cup (Abbaspour et al., 2004, 2007) was used for the calibration and validation procedure and for sensitivity, and uncertainty analysis. SWAT-Cup calculates the 95% prediction uncertainty band (95PPU) in a iterative process. 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, and varies from 0 to 1 where 1 indicates perfect model simulation. The r-factor is the ratio of the average width of the 95PPU band and the standard deviation of the measured variable. There are different approaches regarding balance of p-factor and r-factor. The p-factor should preferably be above 0.7 for discharge and the r-factor value should be below 1.5 (Abbaspour, 2015), but when measured data are of lower quality other values apply. Once acceptable p-factor and r-factor are reached statistical parameters for time series analysis are compared.

For this study we used the ~~Nash-Sutcliff~~ Nash-Sutcliffe Efficiency (NSE), the standardized Root Mean Square Error (RSR), and the Percent Bias (PBIAS). ~~All are very commonly used statistical parameters~~These are well-known statistical parameters, which are often used for comparison of time-series especially in hydrological modelling (Starks and Moriasi, 2009; Gebremicael et al., 2013; Dile and Srinivasan, 2014; Ab therefore help others to compare our modeling results to previous studies. This study refers to the model evaluation techniques described by Moriasi et al. (2007), who established guidelines for the proposed statistical parameters (see table 3 below for details). The NSE is a normalised statistic that indicates how well a plot of observed versus simulated data fits the 1:1 line and determines the relative magnitude of the residual variance compared to the measured data variance (Nash and Sutcliffe, 1970). NSE ranges from $-\infty$ (negative infinity) to 1, with a perfect concordance of modelled to observed data at 1, a balanced accuracy at 0 and a better accuracy of observations below zero. The RSR is a standardized ~~RMSE,~~ Root Mean Square Error (RMSE, standard deviation of the model prediction error), which is calculated from the ratio of the RMSE and the standard deviation of measured data. RSR incorporates the benefits of error index statistics and includes a scaling factor. RSR varies from the optimal value of 0, which indicates zero RMSE or residual variation, ~~which~~

Table 2. SWAT parameters used for discharge and soil loss calibration with initial ranges and fitted ~~value~~ final parameter ranges.

Variable	Parameter name	Definition	Initial range	Fitted parameter ranges		
				Andit Tid	Anjeni	Maybar
Discharge	~~a~~[*]a__CN2.mgt	Curve number	–25 to 15	16.7 to 18.7	–7 to –4	15 to 25
	[**]v__GWQMN.gw	Threshold depth of water in shallow aquifer required for return flow to occur	0 to 5000	4761 to 4990	0 to 1611	2500 to 5000
	a__ESCO.hru	Soil evaporation compensation factor	0 to 1	-0.0038 to 0.046	0.0023 to 0.067	0 to 0.35
	v__GW_REVAP.gw	Groundwater "revap" coefficient	0.02 to 2	0.18 to 0.19	0.17 to 0.21	0.15 to 0.2
	a__CH_K2.rte	Effective hydraulic conductivity in channel	–0.01 to 500	6 to 13	–11 to 58	–0.01 to 15
	a__CH_N2.rte	Manning’s "n" value for the main channel	-0.01 to 0.3	0.0012 to 0.067	–0.15 to 0.062	0.025 to 0.065
	a__SURLAG.bsn	Surface runoff lag time	0.05 to 24	-0.084 to 3.98	0 to 6.63	0.05 to 12
	a__RCHRG_DP.gw	Deep aquifer percolation fraction	0 to 1	0.36 to 0.66	–0.51 to 0.23	0 to 1
	v__EPCO.hru	Plant uptake compensation factor	0 to 1	0.78 to 1.55	0.22 to 0.745	0 to 1
	v__SOL_AWC(1).sol	Available water capacity of the soil layer	0 to 1	0.13 to 0.22	0.19 to 0.47	0 to 1
Sediment	a__SLSUBBSN.hru	Average slope length	–10 to 45	8.85 to 42.34	–6.24 to -4.60	–5 to 5
	a__HRU_SLP.hru	Average slope steepness	–0.1 to 0.4	–0.16 to –0.04	–0.12 to –0.09	–0.5 to 0.72
	a__USLE_K(1).sol	USLE equation soil erodibility (K) factor	–0.11 to 0.24	0.079 to 0.14	0.44 to 0.49	0.04 to 0.31
	a__USLE_C.plant.dat	Min value of USLE C factor applicable to the land cover/plant	0.04 to 0.24	0.0009 to 0.004	0.48 to 0.5	0.34 to 0.626
	a__USLE_P.mgt	USLE equation support practice	0.42 to 0.79	–0.41 to 0.19	0.16 to 0.26	0.09 to 0.92
	v__SPCON.bsn	Linear parameter the maximum amount of sediment that can be reentrained	0.0001 to 0.01	0.005 to 0.007	0.0067 to 0.010	–0.01 to 0.01
	v__SPEXP.bsn	Exponent parameter for calculating sediment reentrained	1 to 1.5	1.27 to 1.5	1.32 to 1.37	1.23 to 1.35
	v__CH_COV1.rte	Channel cover factor	–0.05 to 0.6	0.2 to 0.39	0.057 to 0.099	–0.05 to 0.02
	v__PRF_BSN.bsn	Peak rate adjustment factor for sediment routing in the main channel	0 to 2	0.9 to 1.1	1.2 to 1.6	0.89 to 1.2

[*] a__ means a given value is added to the existing parameter value

[**] v__ means the existing parameter value is to be replaced by a given value

~~indicates perfect model simulation~~ to a large positive value, which indicates a large residual value and therefore worse model simulation performance (Moriasi et al., 2007).

The PBIAS measures the average tendency of the simulated values to be larger or smaller than their observed counterparts. The optimal value of PBIAS is zero. PBIAS is the deviation of data being evaluated, expressed as a percentage. A positive PBIAS value indicates the model is under–predicting measured values, whereas negative values indicate over–predicting.

For this article the recommendations for reported values were strictly applied for discharge calibration and lowered for ~~sediment loss~~ soil loss calibration.

Table 3. General performance ratings recommended by Moriasi et al. (2007)

Performance Rating	RSR	NSE	PBIAS	
			Streamflow	Sediment
Very good	$0.00 \leq RSR \leq 0.50$	$0.75 < NSE \leq 1.00$	$PBIAS < \pm 10$	$PBIAS \leq \pm 15$
Good	$0.50 < RSR \leq 0.60$	$0.65 < NSE \leq 0.75$	$\pm 10 \leq PBIAS < \pm 15$	$\pm 15 \leq PBIAS < \pm 30$
Satisfactory	$0.60 < RSR \leq 0.70$	$0.50 < NSE \leq 0.65$	$\pm 15 \leq PBIAS < \pm 25$	$\pm 30 \leq PBIAS < \pm 55$
Unsatisfactory	$RSR > 0.70$	$NSE \leq 0.50$	$PBIAS \geq \pm 25$	$PBIAS \geq \pm 55$

The model performance was also evaluated using the hydrograph visual technique, which allows a visual model evaluation overview. As suggested by Legates and McCabe ~~(Legates and McCabe, 1999)~~ (1999) this should typically be one of the first steps in model evaluation. Adequate visual agreement between observed and simulated data was compared on discharge and ~~sediment~~ soil loss plots on a monthly basis.

3 Results and discussion

3.1 General comparison of CFSR and WLRC rainfall data

The raw CFSR and WLRC rainfall input data showed significantly different patterns and rainfall amounts. For Andit Tid, situated on the eastern escarpment of the Blue Nile Basin, the belg and kremt rainfall seasons were temporally adequately represented; i.e., the timely occurrences of the rainy seasons were correctly represented through the CFSR data. However, total CFSR rainfall amounts were far from ~~adequately represented~~measured values: while the belg rainfall season in the CFSR data showed some overestimation, the total rainfall and length of the kremt rainy season were strongly underestimated. WLRC data distinctly show a main rainy season from July to September and a light rainy season from March to May, while the CFSR data only show mildly increased rainfall in March, April, July, and August but no distinct rainy season (see figure 2 on page 11 for comparison).

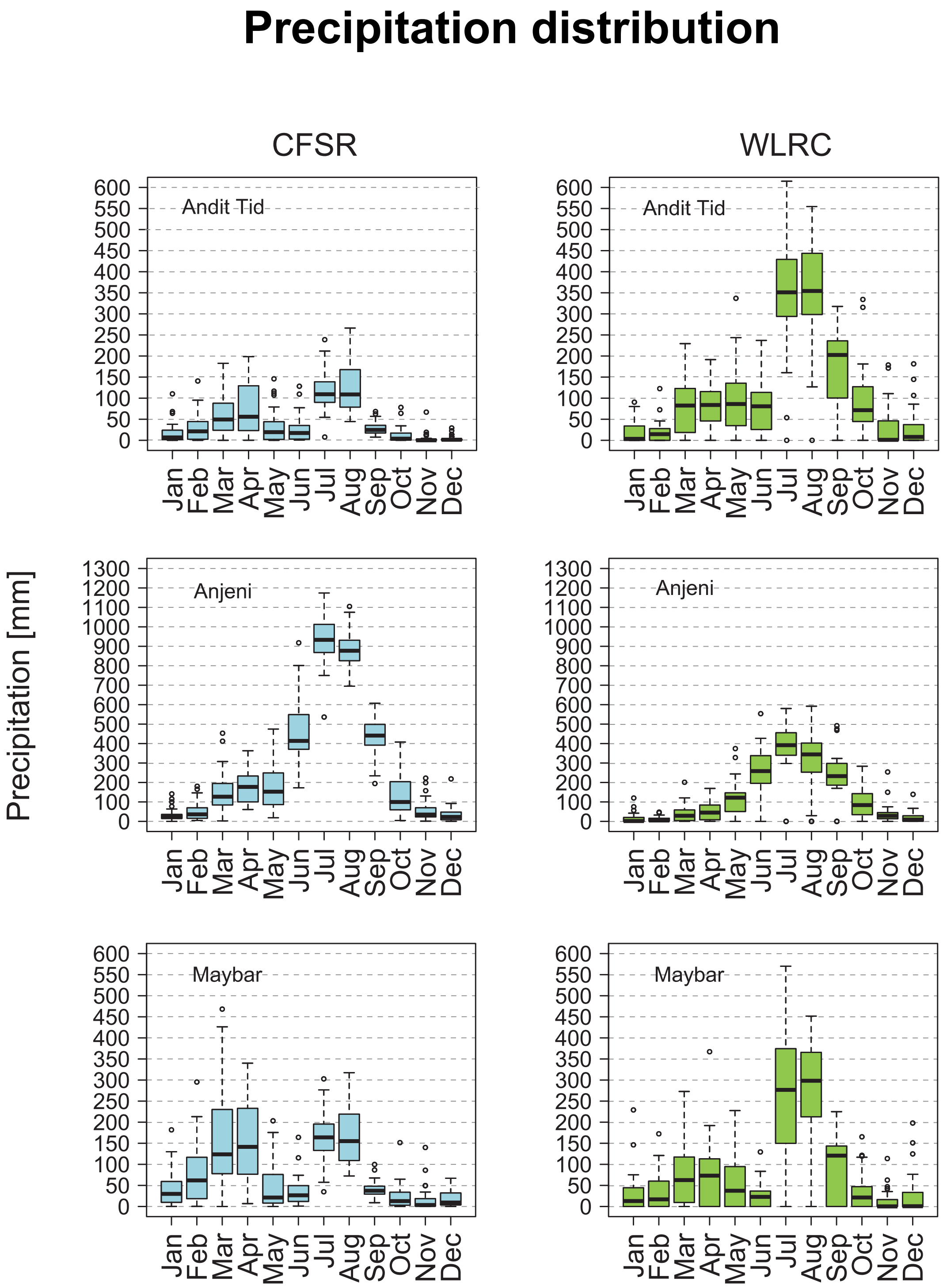

Figure 2. Monthly CFSR and WLRC rainfall distribution of all ~~stations (1979–2010), Andit Tid, Anjeni, Maybar~~station as boxplots with monthly rainfall distribution. CFSR data from 1979 to 2014 and WLRC data from 1981/1982/1984 to 2014. See Table 1 for details.

The CFSR data for Anjeni highly overestimated rainfall in the region. While WLRC data showed a clear trend towards only one main rainy season from May/June to September with average monthly rainfall ranging from 100 mm (May) to 380 mm (July), the CFSR data showed a pronounced main rainy season with monthly averages ranging from 400 mm to 1000 mm from June to September and a distinct small rainy season from March to May with monthly averages three times as high as the WLRC rainfall data. The total annual CFSR rainfall was three times the WLRC annual rainfall.

WLRC Maybar data showed a clear seasonality, with two rainy seasons, one in March and April, and one from July to August. The belg rainy season showed only mild increase of average rainfall to around 75 mm/month and the kremt rainy season showed a distinct increase of rainfall to an average of 270 mm/month. From the CFSR rainfall data, no clear distinction could be made between the belg and the kremt rainy season – both showed a rainfall increase to around 150 mm/month and the total annual rainfall was strongly underestimated.

In general, all CFSR rainfall patterns showed a similar composition: data variability was more uniformly distributed and the distinct seasonality of the WLRC data was not well represented. CFSR data underestimated the bimodal rainfall climates and strongly overestimated the unimodal rainfall climate. The WLRC data has a highly variable rainfall range in the bimodal rainfall locations, which is not reflected by the CFSR data. In general, the CFSR rainfall data does not represent the high variability of rainfall measured by WLRC data.

3.1.1 Seasonal comparison of rainfall data

The seasonal components of the CFSR rainfall were assessed for the three stations by breaking the monthly data into seasons (dry season from October to March, small rainy season (belg) from April to May, and large rainy season (kremt) from June to September) and by comparing only these. The comparison of measured rainfall to modelled rainfall for the dry season from October to March was unsatisfactory ($\mathrm{NSE} < 0.50$) with negative NSEs for three stations (AT: –1.92, AJ: –12.19, MA: –0.77). The PBIAS indicated model underestimation for Anjeni and Maybar (AJ: 134.2, MA: 30.7) and an overestimation of the rainfall for Andit Tid (AT: –55.2). The RSR showed large positive values (AT: 1.68, AJ: 3.55, MA: 1.3) indicating a low model simulation performance and again an unsatisfactory rating (see table).

For the belg rainy season from April to May the model performed badly. Surprisingly, the model performed worst in Anjeni, where no small rainy season occurs. The CFSR model performance for Anjeni was unsatisfactory, with an NSE of –5.42, a PBIAS of 106.1, and an RSR of 2.48. The CFSR model overestimated the monthly rainfall in all but 5 out of 22 years. Andit Tid and Maybar were slightly more adequate but still unsatisfactory. NSE was –0.79 and –0.24 respectively, indicating unsatisfactory performance. PBIAS was –39.4 and 24.3, respectively. RSR was 1.31 and 0.85, which again indicates an unsatisfactory result.

Table 4. Seasonal comparison of rainfall ~~data~~time series of daily rainfall amounts. Satisfactory performance ratings are highlighted in bold. Details for duration and gaps can be found in table 1

	Andit Tid (1982-2014)	Anjeni (1984-2014)	Maybar (1981-2014)
	Dry season Oct-Nov-Dec-Jan-Feb-Mar		
RSR	1.68	3.55	1.3
NSE	–1.92	–12.9	–0.77
PBIAS	55.2	134.2	~~30.7~~ **30.7**
	Belg Apr-May		
RSR	1.31	2.48	0.85
NSE	–0.79	–5.42	–0.24
PBIAS	–39.4	106.1	~~24.3~~ **24.3**
	Kremt Jun-Jul-Aug-Sep		
RSR	3.23	7.0	2.03
NSE	–9.79	–50.09	–3.28
PBIAS	–69.2	128	–47.1

The kremt rainy season from June to September is the season with the heaviest rainfall throughout the year. On average some 77% of the yearly rain falls within this time period. This is also the time period where the heaviest soil erosion occurs induced by rainfall. For Anjeni, Andit Tid, and Maybar the CFSR model performed unsatisfactorily (see Table 7 and Figure 3 in appendix A) with NSEs below 0.50 (AT: –9.79, AJ: –50.09, MA: –3.28), RSRs above 0.70 (AT: 3.23, AJ: 7.0, MA: 2.03), and PBIAS values ranging from –69.2 (AT) and –47.1 (MA) to +128 (AJ).

The kremt rainy season was underestimated by the CFSR model for the bimodal rainfall pattern in Andit Tid and Maybar, while the unimodal rainfall pattern was heavily overestimated by the CFSR model.

3.2 Discharge modeling with WLRC and CFSR data

The performance ratings for each of the three catchments including SWAT–Cup *p–factor* and *r–factor* are summarised in Table 5 on page 20. The table is divided into discharge comparison and ~~sediment~~ soil loss comparison. ~~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 figure 3 for further details) over their respective periods. Parameters initially contained original ranges , which were gradually adapted according to modeling results. The final ranges~~ Final parameter ranges are presented in Table 2 on page 9.

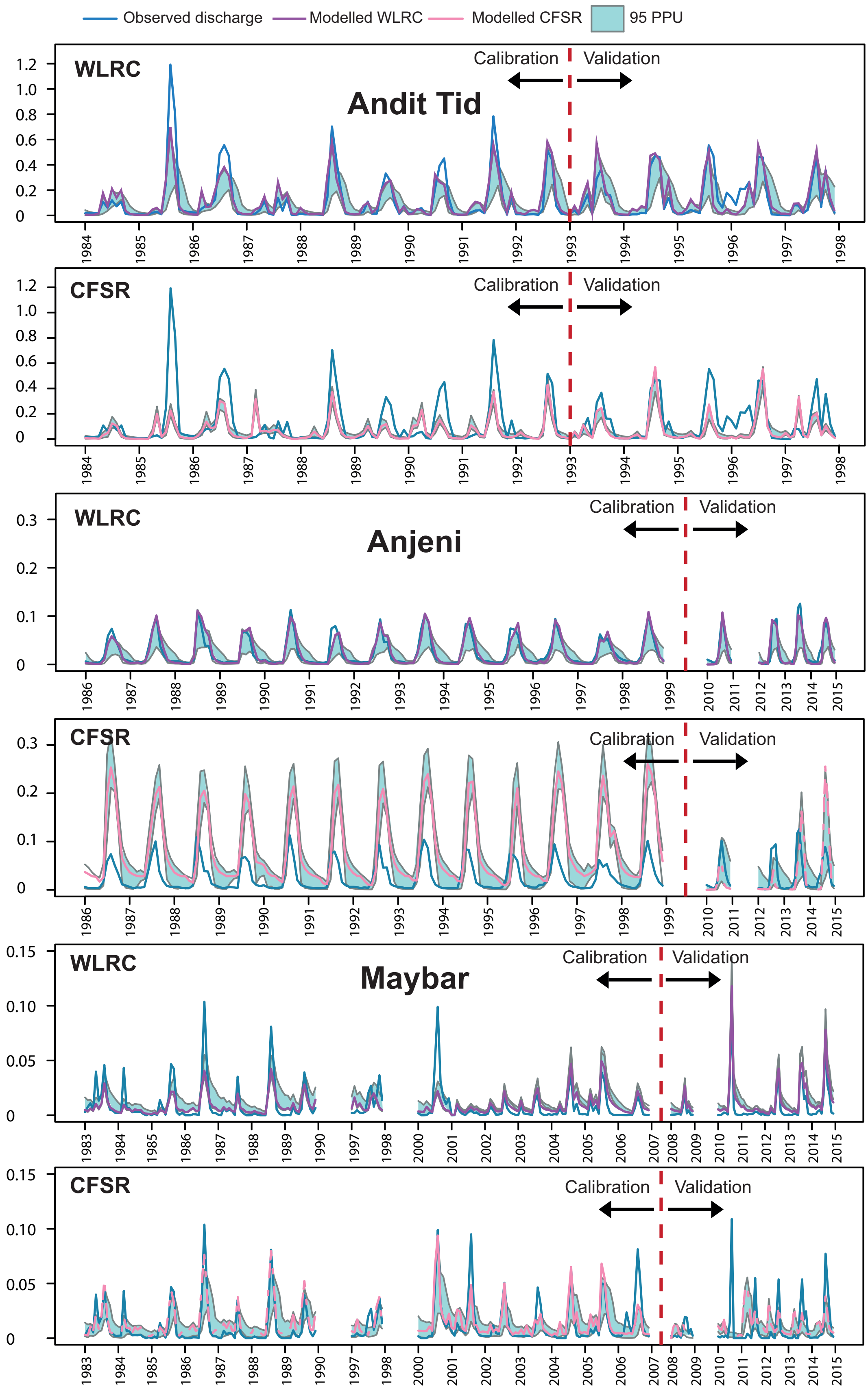

Figure 3. ~~Calibration and validation of~~ Modelled SWAT discharge ~~with~~ compared to measured discharge (blue) for WLRC (violet) and CFSR (pink) input data and the 95 Percent Prediction Uncertainty (light blue). ~~Data~~ Each sub-figure contains the calibration and the validation period. Results are given in m^3/s.

3.2.1 Andit Tid

Calibration of Andit Tid with WLRC rainfall data yielded *very good* results. With an *p–ractor* of 0.71 and a *p–factor* of 0.53 (see Chapter 2.6 on page 8 for performance rating) the statistical parameters RSR, NSE and PBIAS yielded *"very good"* results (0.46, 0.79, 3.1 respectively). ~~Validation for Andit Tid yielded in satisfactory results with~~ The CFSR rainfall data, which underestimated the WLRC rainfall pattern, yielded *unsatisfactory* results with RSR, NSE, and PBIAS of 0.80, 0.36, and 31.4. ~~Parameter ranges settings were maximised, but still inside SWAT absolute values ((Abbaspour et al., 2007) .~~ The hydrograph on page 15 shows that the underestimation of rainfall amounts for Andit Tid did result in a constant underestimation of peak flows and of base flows throughout the whole time period.

Validation of discharge for Andit Tid with WRLC data showed *very good* results with RSR: 0.46, NSE: 0.79 and PBIAS 9.6 and marginally *unsatisfactory* results for the CFSR dataset (RSR: 0.74, NSE: 0.45, PBIAS: 37.9).

3.2.2 Anjeni

Anjeni showed *very good* result for calibration with WLRC rainfall data. RSR, NSE and PBIAS were well inside the optimal performance ratings (0.39, 0.85, and 3.7 respectively), see table 3 on page 10 and figure 3 on page 15 for comparison.

~~Calibration~~ Satisfactory calibration could not be reached with CFSR data ~~, where the CFSR rainfall data did strongly overestimate the measured rainfall data proved impossible. With parameter ranges set to maxima,~~ and neither baseflow, nor peaks could be adequately represented. With a *p–factor* of 0.49 and an *p–ractor* of 1.91 the statistical parameters were *unsatisfactory* (RSR: 2.70, NSE:-6.27, and PBIAS: -226.0). The hydrograph (Figure 3 on page 15) shows that the strong overestimation of CFSR rainfall data during belg lead to a modelled discharge with extreme peaks during kremt, which do not correspond to the discharge regime of measured WLRC data.

Validation of discharge for Anjeni with WRLC data showed *very good* results with RSR: 0.41, NSE: 0.83 and PBIAS –6.7 and *unsatisfactory* results for the CFSR dataset with RSR: 1.24, NSE: –0.53, and *very good* PBIAS: 8.1.

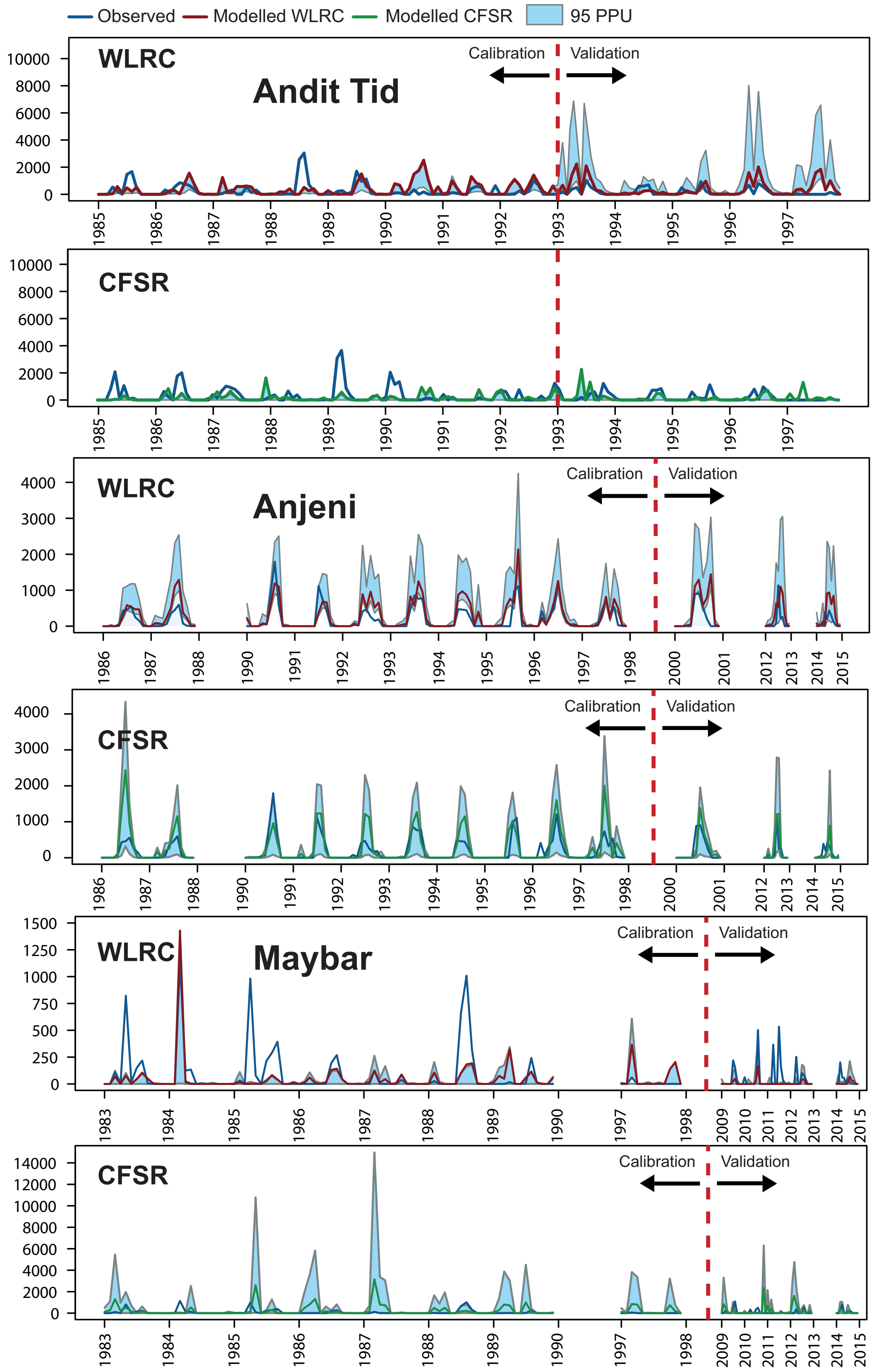

Figure 4. ~~Calibration and validation of sediment~~ Modelled SWAT soil loss ~~with~~ compared to measured soil loss (blue) for WLRC (red) and CFSR (gree) input data and the 95 Percent Prediction Uncertainty (light blue). ~~Data~~ Each sub-figure contains the calibration and the validation period. Results are given in tons (t).

3.2.3 Maybar

Calibration of Maybar with WLRC rainfall data proved to be less straight forward than Anjeni and Andit Tid. The rugged topography of Maybar combined with a inadequate cross-section proved challenging to model. Nonetheless, *satisfactory* result were achieved for discharge with RSR, NSE, and PBIAS of 0.63, 0.60, and –23.4 respectively.

The CFSR rainfall data yielded an *unsatisfactory* discharge simulation result with RSR~~: , NSE: ~~, NSE, and PBIAS~~:~~. As the CFSR modelled rainfall shows two similar rainy seasons where WLRC rainfall data has distinct belg and kremt rainy season, SWAT modelled discharge showed similar trends. ~~The hydrograph with CFSR data on page 15 shows~~ Figure 3 shows regular discharge peaks from February to ~~April for every year, when there are none measured while showing only small CFSR peaks for the main rainy season from June to September, when measured discharge is significantly increasing. Again, the SWAT modelled discharge reflected the~~ March, in accordance to rainfall pattern deviation as seen on Figure 2, when no increase of discharge was measured at the research station. The SWAT model reflected input rainfall pattern adequately, which lead to discharge peaks during belg, when there are none in the measured data. At the same time it lead to reduced discharge peaks during kremt, when the measured WLRC data are clearly pronounced.

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.

3.3 ~~Sediment~~ Soil loss modelling with WLRC and CFSR data

~~Sediment~~ Soil loss modelling was calibrated using the same set of 9 parameters for each catchment (see Table 2 for description). Calibration of soil loss was conducted using the parameter ranges for discharge calibration, and adapting the sediment parameters while leaving discharge parameters untouched. Performance ratings for each of the three catchments including SWAT–Cup *p–factor* and *p–ractor* are summarised in table 5 on page 20 and visually represented on Figure 4 on page 17. Performance rating levels were considerably lowered for ~~sediment~~ soil loss modeling. Threshold for the *p–factor* was set at 0.40 with an *r–factor* below 1.80 and standard performance ratings for RSR, NSE and PBIAS.

3.3.1 Andit Tid

The good results from WLRC discharge modeling facilitated ~~sediment~~ soil loss calibration and resulted in *satisfactory* performance ratings for RSR, NSE (0.69, 0.65), and an *unsatisfactory* PBIAS, which was slightly below threshold with –56.3. Graphic representation showed good visual results (see figure 4 on page 17) in general, but also showed constant overestimation of the modelled data except for three years 1988, 1989, and 1994.

~~Sediment loss modelling with CFSR data reflected the results from discharge modeling~~Validation of sediment yield for Andit Tid 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.

3.3.2 Anjeni

~~Sediment~~ Soil loss modeling with WLRC rainfall data and calibrated discharge yielded *satisfactory* results.With a *P–factor* of 0.40 and an *r–factor* of 0.65, and statistical parameters RSR: 0.67, NSE: 0.55, and PBIAS: -19.9 the model was just *satisfactory*. The graphic showed adequate results with a constant overestimation of the model except for two years in the early nineties. Modelling with CFSR data, resulted in strongly unsatisfactory results (RSR: 1.01, NSE: –0.02, and PBIAS: -33.9), which can easily be explained with the strong model overestimation of rainfall and subsequently discharge. Parameters could not be adapted further to achieve better results as they were already set to the edge of the possible ranges.

Validation of sediment yield for Anjeni with WRLC data showed *satisfactory* results with RSR: 0.67, NSE: 0.64 and PBIAS –14.1 indicating a general overestimation and *unsatisfactory* results for the CFSR dataset with RSR: 1.02, NSE: –0.03, and *satisfactory* PBIAS: –1.9 indicating underestimation.

3.3.3 Maybar

~~Sediment~~ Soil loss calibration with WLRC rainfall data and calibrated discharge resulted in *unsatisfactory* statistical results (RSR: 1.24, NSE: –0.54, PBIAS: –34.1). *P–factor* and *r–factor* were 0.42 and 0.60, respectively.

Calibration in Maybar with CFSR rainfall data yielded *unsatisfactory* results (RSR: 1.02, NSE: –0.03, PBIAS: 54.4). As described in the discharge calibration section (Section 3.2.3), CFSR rainfall data in Maybar tended towards overestimation of belg and underestimation of kremt, which resulted in overestimation of monthly discharge during belg and underestimation during kremt. This trend was redrawn with sediment calibration resulting in small but distinct peaks during belg and smaller peaks than measured during kremt. There was no satisfactory calibration possible with CFSR rainfall data.

Validation of sediment yield for Maybar with WRLC data showed *satisfactory* results for both data sets with a very strong overestimation from the CFSR data set and an equally strong overestimation from the WLRC data set.

Table 5. Calibration and validation results ~~,~~ of monthly CFSR and WLRC modelled discharge and soil loss. Values that meet at least the "satisfactory" criteria are highlighted in bold

	Andit Tid		Anjeni		Maybar	
	CFSR	WLRC	CFSR	WLRC	CFSR	WLRC
Discharge - Calibration						
p-factor	0.49	0.71	0.49	0.92	0.41 0.61	
r-factor	0.20	0.53	1.91	0.46 0.54	0.96	
RSR	0.83	**0.46**	2.70	**0.37**	1.16	**0.53**
NSE	0.31	**0.79**	–6.27	**0.86**	–0.35	**0.72**
PBIAS	46.1	**3.1**	–226.0	**2.0**	29.6	**1.5**
Discharge - Validation						
p-factor	0.30	0.66	0.69	0.69	0.38	0.61
r-factor	0.29	0.54	1.41	0.57	0.52	1.11
RSR	0.74	**0.46**	1.24	**0.41**	0.98	**0.56**
NSE	0.45	**0.79**	–0.53	**0.83**	0.04	**0.74**
PBIAS	37.9	**9.6**	8.1	**–6.7**	–1.9	**–17.3**
Soil loss - Calibration						
p-factor	0.33	0.45	0.32	0.40	0.44	0.28
r-factor	0.19	0.59	1.30	0.65	4.47	0.28
RSR	1.02	**0.67**	1.01	**0.67**	2.55	0.84
NSE	-0.03	**0.64**	-0.02	**0.55**	–5.51	0.29
PBIAS	54.4	**–14.1**	**–33.9**	**–19.9**	180.5	**39.2**
Soil loss - Validation						
p-factor	0.30	0.39	0.38	0.38	0.23	0.15
r-factor	0.51	1.60	1.61	1.10	2.67	0.06
RSR	1.39	**0.68**	1.08	**0.62**	2.24	0.98
NSE	–0.94	**0.51**	-0.17	**0.62**	-4.04	–0.03
PBIAS	11.9	–64.3	**–30.5**	**–31.3**	–94.7	92.8

4 Conclusions

In this paper we studied the applicability of CFSR weather data to three small-scale watersheds in the Ethiopian highlands with the goal of assessing the usability for future modelling in data-scarce regions. First, we compared CFSR and WLRC rainfall data at three stations in the Ethiopian

Highlands and therefore rainfall data was compared on a monthly basis with boxplots. Second, we modelled discharge with the SWAT model; once with WLRC data and once with CFSR rainfall data. Third, we modelled ~~sediment~~ soil loss for the three stations with the SWAT model and compared calibrated results ~~from CFSR rainfall and conventional rainfall~~ to measured data.

~~The rainfall data comparison for CFSR and WLRC data showed strong discrepancies in seasonal and monthly rainfall amounts for all three catchments. For Andit Tid, both, belg and kremt rainy season were levelled downwards resulting in *unsatisfactory* results for each season with strongest deviations for kremt (see Tabel 4 on page 13 for details). Anjeni rainfall data from the CFSR model overestimated the measured WLRC rainfall very strongly. This resulted in strong deviations with performance ratings well below *satisfactory* thresholds. Maybar rainfall data from CFSR showed the highest deviation for the representation of seasonality. Neither belg, nor kremt or the dry season were adequately modelled. Deviation ranged from slight (dry season) to overestimation of belg season and a strong underestimation of kremt season. All in all the CFSR model could not adequately render rainfall patterns for Maybar.~~

~~Discharge simulation comparisons with WLRC data produced very good results: the three catchments could be modelled with *very good* performance ratings for RSR and NSE except for the PBIAS, which was only *satisfactory* in the case of Maybar and *very good* for Andit Tid and Anjeni.~~

~~Discharge simulations with CFSR~~ The WLRC rainfall data set resulted in three calibrated and validated discharge models while the CFSR data resulted in ~~*unsatisfactory* performance ratings for the three catchments. Discharge modelling results yielded results in line with rainfall data comparison: Anjeni discharge was highly overestimated, Andit Tid discharge was underestimated and Maybar discharge had overestimation of belg discharge and underestimation of kremt discharge.~~

~~Sediment loss modeling with~~ none. For the soil loss modeling the WLRC rainfall data ~~and calibrated discharge~~ resulted in two ~~*satisfactory* (Andit Tid and Anjeni) and one *unsatisfactory* (Maybar) calibrations. For Andit Tid the model could render sediment loss adequately except for some peaks in the mid and late nineties. For Anjeni the model performed even better with a slight overestimation over the whole period. For Maybar calibration failed.~~ out of three calibrated and validated models while none could be adequately calibrated or validated for the CFSR data set. 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.

~~Sediment loss performance ratings from simulations with CFSR rainfall data and calibrated discharge yielded in *unsatisfactory* results for the three catchments. The same deviation patterns observed in discharge calibration ensued in sediment loss calibration. Catchments with high discharge model overestimation resulted in high sediment loss overestimation and catchments with displaced seasonal discharge patterns resulted in displaced sediment loss patterns.~~

~~Andit Tid sediment loss modelling with CFSR data resulted in *unsatisfactory* results. The underestimation of the discharge modelling did not allow for satisfactory sediment loss calibration. The hydrograph~~

~~(see Figure 4 on page 17) shows that the general underestimation of rainfall data lead to underestimation of discharge, which lead to reduced sediment loss modelling. Sediment loss modeling with CFSR data in Anjeni resulted in a constant overestimation of sediment loss and performance ratings were *unsatisfactory*. For Maybar the misplaced seasonal rainfall lead to higher discharge for belg and lower discharge for kremt, which resulted in a shift of sediment loss peaks from kremt to belg. Performance ratings were *unsatisfactory*.~~

Our results clearly show that adequate discharge and ~~sediment~~ soil loss modelling was not possible, in present case, with the CFSR data~~in present case~~. This suggests that SWAT simulations in small-scale watersheds in the Ethiopian highlands do not perform well with CFSR data in every case, and that sometimes there is no substitute for high-quality conventional weather data. Such weather data – with high spatial and temporal climatic data resolution – were available for the three small-scale catchments used in the study but are not in many other cases. In these other cases one should carefully check CFSR data against similar climatic stations with conventionally measured data. In addition, discharge and ~~sediment~~ soil loss modelling showed that usage of CFSR weather data not only resulted in substantial deviation in both total discharge and total ~~sediment~~ soil loss, but also in the seasonal rainfall pattern. The seasonal weather pattern is one of the major drivers of ~~sediment~~ soil loss and is especially pronounced in the Blue Nile Basin, with one long rainy season occurring as fields are ploughed and sowed. Thus, contrary to ~~Dile and Srinivasan (2014)~~previous studies for the Ethiopian Highlands, this study suggests that CFSR data may not be applicable in any case for small-scale modelling in data-scarce regions: the authors even suggest that outcomes of SWAT modelling with CFSR data ~~alone~~ for small-scale catchments may yield erroneous results which cannot be verified and may lead to wrong conclusions. Nonetheless, the advantage of CFSR data is its completeness over time, which would allow for comprehensive watershed modelling in regions with no conventional weather data or with longer gaps in conventionally recorded rainfall records.

References

Abbaspour, K. C.: SWAT-CUP 2012: SWAT calibration and uncertainty programs-A user manual, Tech. rep., Swiss Federal Institute of Aquatic Science and Technology, Eawag, 2015.

Abbaspour, K. C., Johnson, C. A., and van Genuchten, M. T.: Estimating Uncertain Flow and Transport Parameters Using a Sequential Uncertainty Fitting Procedure, Vadose Zone Journal, 3, 1340–1352, doi:10.2113/3.4.1340, 2004.

Abbaspour, K. C., Vejdani, M., Haghighat, S., and Yang, J.: SWAT-CUP Calibration and Uncertainty Programs for SWAT, in: The fourth International SWAT conference, pp. 1596–1602, Delft, Netherlands, 2007.

Alemayehu, T., Griensven, A., and Bauwens, W.: Evaluating CFSR and WATCH Data as Input to SWAT for the Estimation of the Potential Evapotranspiration in a Data-Scarce Eastern-African Catchment, Journal of Hydrologic Engineering, p. 5015028, doi:10.1061/(ASCE)HE.1943-5584.0001305, 2015.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hydrologic modeling and assessment Part 1: model development, Journal of the American Water Resources Association (JAWRA), 34, 73–89, 1998.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., Van Griensven, A., Van Liew, M. W., Kannan, N., and Jha, M. K.: SWAT: Model use, calibration and validation, Transactions of the ASABE, 55, 1491–1508, 2012.

Belay, G.: Gerda Watershed Soil Report, Tech. Rep. February, Water and Land Resource Centre (WLRC), Addis Abeba, Addis Abeba, 2014.

Betrie, G. D., Mohamed, Y. A., Van Griensven, A., Srinivasan, R., and Mynett, A.: Sediment management modelling in the Blue Nile Basin using SWAT model, Hydrology and Earth System Sciences, 15, 807–818, doi:10.5194/hess-15-807-2011, 2011.

Bono, R. and Seiler, W.: The Soils of the Andit-Tid Research Unit (Ethiopia) Classification, Morphology and Ecology, with soil map 1:10'000, University of Bern, Switzerland, Bern, 1984.

Bosshardt, U.: Catchment Discharge and Suspended Sediment Transport as Indicators of Physical Soil and Water Conservation in the Minchet Catchment, Anjeni Research Unit, Soil Conservation Research Programme, Bern, Switzerland, 1997.

Bosshardt, U.: Hydro-sedimentological Characterisation of Kori Sheleko and Hulet Wenz Catchments, Maybar and Andit Tid Research Units, Soil Conservation Research Programme, Bern, Switzerland, 1999.

Cavazos, T. and Hewitson, B. C.: Performance of NCEP – NCAR reanalysis variables in statistical downscaling of daily precipitation, Climate Research, 28, 95–107, 2005.

CFSR: Global Weather Data for SWAT, http://globalweather.tamu.edu/.

Chow, V. T.: Open-channel hydraulics, McGraw-Hill, Boston, MA [etc.], doi:ISBN 07-010776-9, 1959.

De Almeida Bressiani, D., Srinivasan, R., Jones, C. A., and Mendiondo, E. M.: Effects of different spatial and temporal weather data resolutions on the stream flow modeling of a semi-arid basin, Northeast Brazil, International Journal of Agricultural and Biological Engineering, 8, 1–16, doi:10.3965/j.ijabe.20150803.970, 2015.

Dile, Y. T. and Srinivasan, R.: Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: an application in the Blue Nile River Basin, JAWRA Journal of the American Water Resources Association, 50, 1226–1241, doi:10.1111/jawr.12182, 2014.

Diro, G. T., Grimes, D. I. F., Black, E., O'Neill, A., and Pardo-Iguzquiza, E.: Evaluation of reanalysis rainfall estimates over Ethiopia, International Journal of Climatology, 29, 67–78, doi:10.1002/joc.1699, 2009.

Douglas-Mankin, K., Srinivasan, R., and Arnold, J. G.: Soil and Water Assessment Tool (SWAT) Model: Current development and applications, American Society of Agricultural and Biological Engineers, 53, 1423–1431, 2010.

Faurès, J.-M., Goodrich, D. C., Woolhiser, D. A., and Sorooshian, S.: Impact of small-scale spatial rainfall variability on runoff modeling, Journal of Hydrology, 173, 309–326, 2000.

Fuka, D. R., Walter, M. T., MacAlister, C., Degaetano, A. T., Steenhuis, T. S., and Easton, Z. M.: Using the Climate Forecast System Reanalysis as weather input data for watershed models, Hydrological Processes, 28, 5613–5623, doi:10.1002/hyp.10073, 2014.

Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G.: The Soil and Water Assessment Tool: Historical development, applications, and future research directions, American Society of Agricultural and Biological Engineers, 50, 1211–1250, 2007.

Gebremicael, T. G., Mohamed, Y. A., Betrie, G. D., van der Zaag, P., and Teferi, E.: Trend analysis of runoff and sediment fluxes in the Upper Blue Nile basin: A combined analysis of statistical tests, physically-based models and landuse maps, Journal of Hydrology, 482, 57–68, doi:10.1016/j.jhydrol.2012.12.023, 2013.

Gessesse, B., Bewket, W., and Bräuning, A.: Model-based characterization and monitoring of runoff and soil erosion in response to land use/land cover changes in the Modjo watershed, Ethiopia, Land Degradation & Development, pp. 1–14, doi:10.1002/ldr.2276, 2014.

Green, W. A. and Ampt, G. A.: Studies on Soil Phyics, The Journal of Agricultural Science, 4, 1–24, doi:10.1017/S0021859600001441, 1911.

Hurni, H.: Agroecologial belts of Ethiopia: Explanatory notes on three maps at a scale of 1:1,000,000, Soil Conservation Research Programme (SCRP); Centre for Development and Environment (CDE), Bern, Switzerland, 1998.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J. E., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynalds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437–471, 1996.

Kejela, K.: The soils of the Anjeni Area - Gojam Research Unit, Ethiopia, University of Bern, Switzerland, Bern, Switzerland, research r edn., 1995.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., Dool, H. V. D., Jenne, R., and Fiorino, M.: The NCEP – NCAR 50-Year Reanalysis : Monthly Means CD-ROM and Documentation, Bulletin of the American Meteorological Society, 82, 247–268, 2001.

Legates, D. R. and McCabe, G. J.: Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation, Water Resources Research, 35, 233–241, doi:10.1029/1998WR900018, 1999.

Lin, S., Jing, C., Chaplot, V., Yu, X., Zhang, Z., Moore, N., and Wu, J.: Effect of DEM resolution on SWAT outputs of runoff, sediment and nutrients, Hydrology and Earth System Sciences Discussions, 7, 4411–4435, doi:10.5194/hessd-7-4411-2010, 2010.

Mbonimpa, E. G.: SWAT Model Application to Assess the Impact of Intensive Corn-farming on Runoff, Sediments and Phosphorous loss from an Agricultural Watershed in Wisconsin, Journal of Water Resource and Protection, 04, 423–431, doi:10.4236/jwarp.2012.47049, 2012.

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, American Society of Agricultural and Biological Engineers, 50, 885–900, 2007.

Nash, J. E. and Sutcliffe, J. V.: River Flow Forecasting through Conceptual Models: Part 1. - A Discussion of Principles, Journal of Hydrology, 10, 282–290, 1970.

Poccard, I., Janicot, S., and Camberlin, P.: Comparison of rainfall structures between NCEP/NCAR reanalyses and observed data over tropical Africa, Climate Dynamics, 16, 897–915, doi:10.1007/s003820000087, 2000.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, Bulletin of the American Meteorological Society, 91, 1015–1057, doi:10.1175/2010BAMS3001.1, 2010.

Schuol, J. and Abbaspour, K. C.: Using monthly weather statistics to generate daily data in a SWAT model application to West Africa, Ecological Modelling, 201, 301–311, doi:10.1016/j.ecolmodel.2006.09.028, 2007.

SCRP and CDE: Digital Elevation Model (DEM) of Anjeni, Gojam, 2000a.

SCRP and CDE: Digital Elevation Model (DEM) of Andit Tid, Shewa, 2000b.

SCRP and CDE: Digital Elevation Model (DEM) of Maybar watershed, Wello, 2000c.

Starks, P. J. and Moriasi, D. N.: Spatial resolution effect of precipitation data on SWAT calibration and performance: implications for CEAP, American Society of Agricultural and Biological Engineers, 52, 1171–1180, 2009.

Stehr, A., Debels, P., and Romero, F.: Hydrological modelling with SWAT under conditions of limited data availability : evaluation of results from a Chilean case study, Hydrological Sciences Journal, 53, 588–601, 2008.

Temesgen, M., Rockström, J., Savenije, H. H. G., Hoogmoed, W. B., and Alemu, D.: Determinants of tillage frequency among smallholder farmers in two semi-arid areas in Ethiopia, Physics and Chemistry of the Earth, 33, 183–191, doi:10.1016/j.pce.2007.04.012, 2008.

Tibebe, D. and Bewket, W.: Surface runoff and soil erosion estimation using the SWAT model in the Keleta Watershed, Ethiopia, Land Degradation & Development, 22, 551–564, doi:10.1002/ldr.1034, 2011.

USDA-SCS: Design Hydrographs, in: National Engineering Handbook, chap. 1-22, p. 824, USDA Soil Conservation Service, Washington D.C., Washington D.C., http://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/national/water/?cid=stelprdb1043063, 1972.

Wagner, P. D., Fiener, P., Wilken, F., Kumar, S., and Schneider, K.: Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions, Journal of Hydrology, 464-465, 388–400, doi:10.1016/j.jhydrol.2012.07.026, 2012.

Weigel, G.: The Soils of the Maybar Area, Wello Research Unit, Ethiopia, University of Bern, Switzerland, Bern, research r edn., 1986.

Williams, J. R.: Flood routing with variable travel time or variable storage coefficients, Transactions of the ASABE, 12, 100–103, 1969.

Williams, J. R.: SPNM, a model for predicting sediment, phosphorus, and nitrogen yields from agricultural basins, Water Resources Bulletin, 16, 843–848, 1980.

Worqlul, A. W., Maathuis, B., Adem, A. A., Demissie, S. S., Langan, S., and Steenhuis, T. S.: Comparison of TRMM, MPEG and CFSR rainfall estimation with the ground observed data for the Lake Tana Basin, Ethiopia, Hydrology and Earth System Sciences Discussions, 11, 8013–8038, doi:10.5194/hessd-11-8013-2014, http://www.hydrol-earth-syst-sci-discuss.net/11/8013/2014/, 2014.

Yang, Y., Wang, G., Wang, L., Yu, J., and Xu, Z.: Evaluation of Gridded Precipitation Data for Driving SWAT Model in Area Upstream of Three Gorges Reservoir, PLoS ONE, 9, e112 725, doi:10.1371/journal.pone.0112725, 2014.

Yu, M., Chen, X., Li, L., Bao, A., and Paix, M. J. D. L.: Streamflow Simulation by SWAT Using Different Precipitation Sources in Large Arid Basins with Scarce Raingauges, Water Resources Management, 25, 2669–2681, doi:10.1007/s11269-011-9832-z, 2011.

Zeleke, G.: Landscape Dynamics and Soil Erosion Process Modeling in the North-Western Ethiopian Highlands., Phd thesis, University of Bern, Swizterland, 2000.