Response to comments of Reviewer 2

We are delighted with the response reviewer 2. We appreciate the many detailed and helpful comments. It allowed us to improve the manuscript significantly. We implemented all suggested changes. Below, we first cite the comment from the reviewer and follow that with our response in "blue". The "red" color are words that we inserted in the text of the original manuscript. For ease of reading we do not show the deleted text.

General comments

Comment: The study indicates the importance of local calibration of empirical models for estimating sediment load and concentration. However, the research question is not clearly indicated and the discussion is very shallow. Overall with major modification the article can be accepted.

Response: The main research question was whether the existing sediment rating curves developed by the Ministry of Water Resources and Electricity (Mookie) of Ethiopia could be used to describe sediment concentrations which depend on saturation of watersheds (i.e. ratio of cumulative effective rainfall with the maximum threshold effective precipitation). Sediment concentration in the highlands of Ethiopia decreases with the progress of the rainy phase of monsoon and with increment of discharge. We agree with the reviewer that the research question on last paragraph of page 1423 was not well formulated and reworded it as follows:

> "Since the traditional method of determining rating curves for sediment loads assume that the sediment concentrations are a unique function of the discharge, this method cannot be used in environmental applications for predicting sediment concentrations when the sediment concentration decreases throughout the season for a given amount of discharge. The objective of this paper was, therefore, to develop a realistic method to determine the decreasing sediment concentration with the progression of the monsoon using the limited data common in most of the tropics. The study is carried out in the Blue Nile basin, in the Ethiopian highlands, where four major rivers and their watersheds were selected to test how well the relation performs for a range of scales."

Comment: Introduction: It is not properly address what is lacking from the previous scientific studies. It looks like the study was conducted because you have sediment-discharge data.

Response: We agree and we added on page 21, line 10 the text in red to the paragraph indicating the previous scientific studies

> "In the Blue Nile Basin in the Ethiopian highlands, where the construction of the Grand Ethiopian Renaissance Dam (GERD) is underway near the border of Sudan and other planned hydroelectric dams upstream of it, determining sediment loads is becoming more urgent. At the same time, concern for the environment has been increasing and it has been noted that the fish production in Lake Tana is decreasing due to increasing sediment concentrations (Vijverberg et al., 2012). Thus, the ability to predict accurately the sediment concentration in and load to the lakes and man-made reservoir has become important. These are not available in the Ethiopian highlands."

In addition, we added the following to the end of the paragraph that follows the above paragraph.

> "However, it is cumbersome to obtain the required data for these models especially in developing countries. Therefore, previously when concurrent concentration and discharge measurement are taken at irregular intervals; rating curves are often the preferred choice for predicting sediment loads in the past (e.g., Walling, 1990) and more recently (e.g., Horowitz, 2010); Kokpinar et al., (2015); Choi and Lee, 2015; Kheirfam and Vafakhah, 2015). The abundance of papers on load rating curves in the literature should be not surprising since purpose of the measurements was to determine the amount of sediment that potentially could be deposited in rivers and reservoirs. In the published refereed literature, a limited number of articles developed sediment rating curves. These few studies were carried out in Sweden (Fenn et al., 1985); Ontario Canada (Irvine and Drake, 1987), British Columbia in Canada (Sichingabula, 1998), South Australia (Sun et al, 2001) and for the Himalayan glacier in India (Arora et al., 2014).Thus, compared to the sediment load rating curves that are available throughout the world for many rivers, there are very few sediment concentration rating curves and none for a tropical monsoon climate.

Comment: Some of the citations are out dated.

Response: We were aware that we used "aged" citations, since we were referring the original authors for the rating curves or the models as they were the founders of the methods or the models. A can be seen in the response to the previous comment, we have incorporated recent literature.

Comment: Methods: What kind of instruments was used by MoWIE for sediment sampling, what was its accuracy?

Response: We obtained the following information from MoWIE for the Lake Tana watersheds: Discharge was measured by measuring the stage, cross-section of the channel and flow velocity at the gaging station. At the same time, grab sample of the river water from the gaging station was taken for sediment concentration analysis. This was done by collecting the water samples using standard plastic bottles and transporting it to the laboratory for sediment concentration analysis. The sediment concentration was determined by using gravimetric analysis.

Comment: Indicate the coordinate of the study area in Figure 2. Use different symbols for the gauging stations at the catchment and at the small plots.

Response: Thanks for the comment. We corrected the figures based on the comment as indicated below.

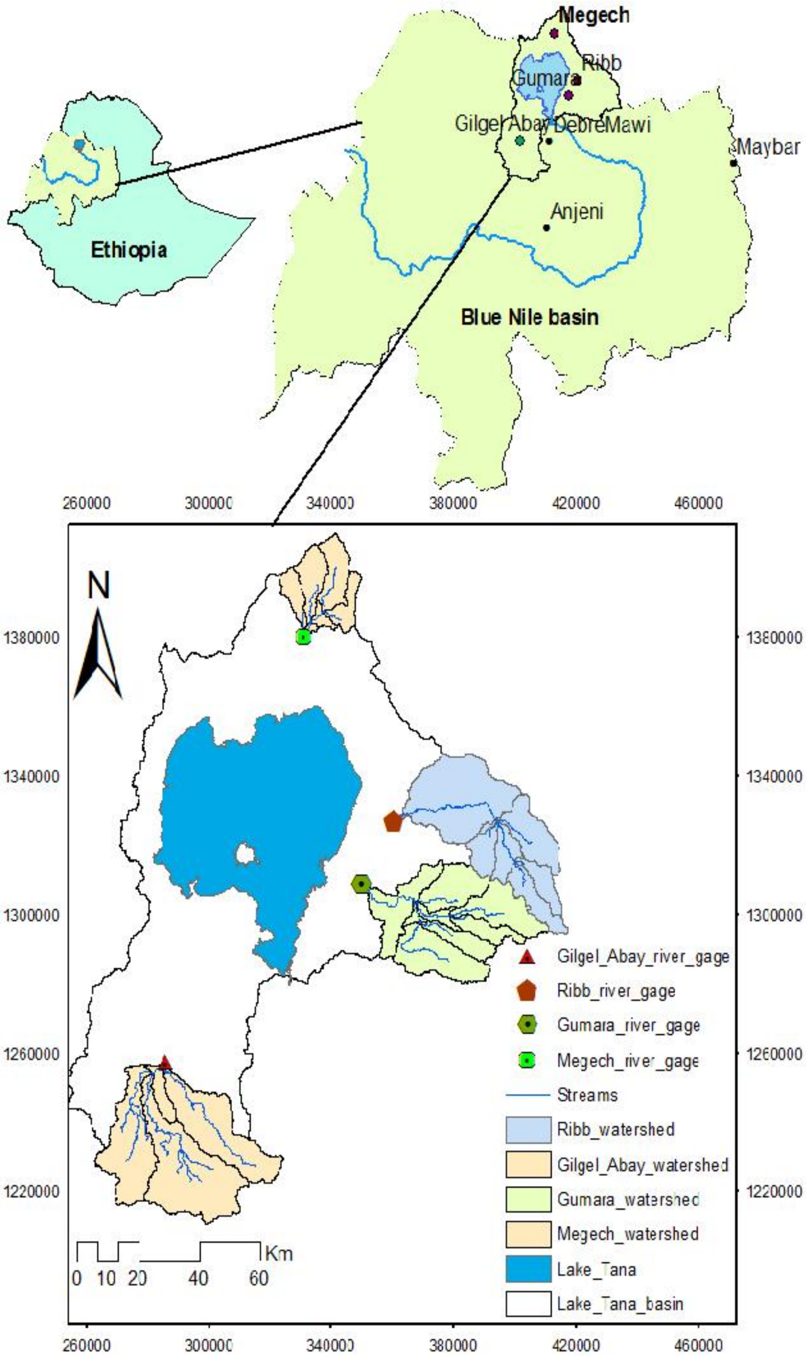
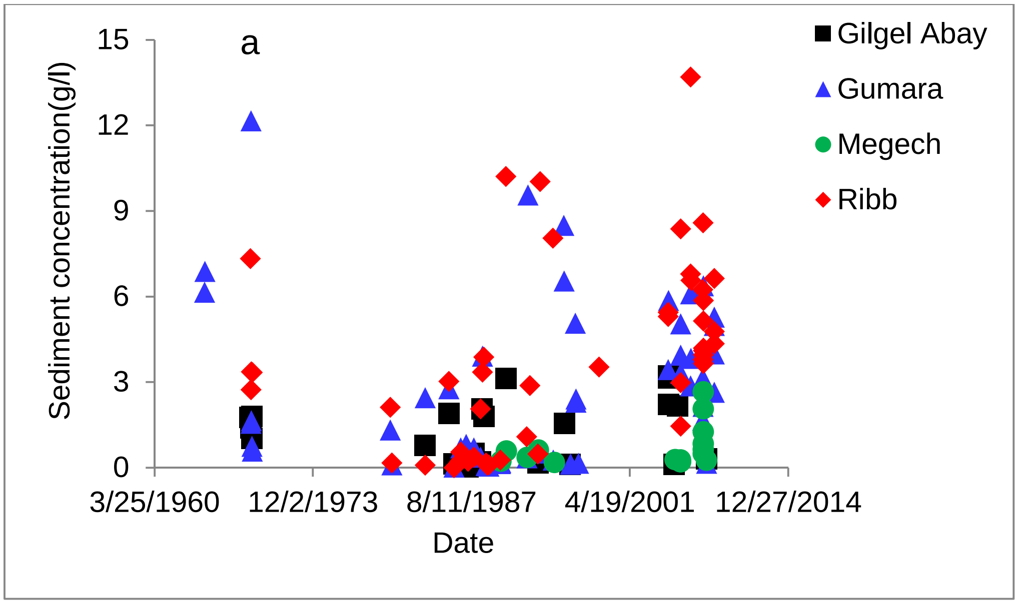
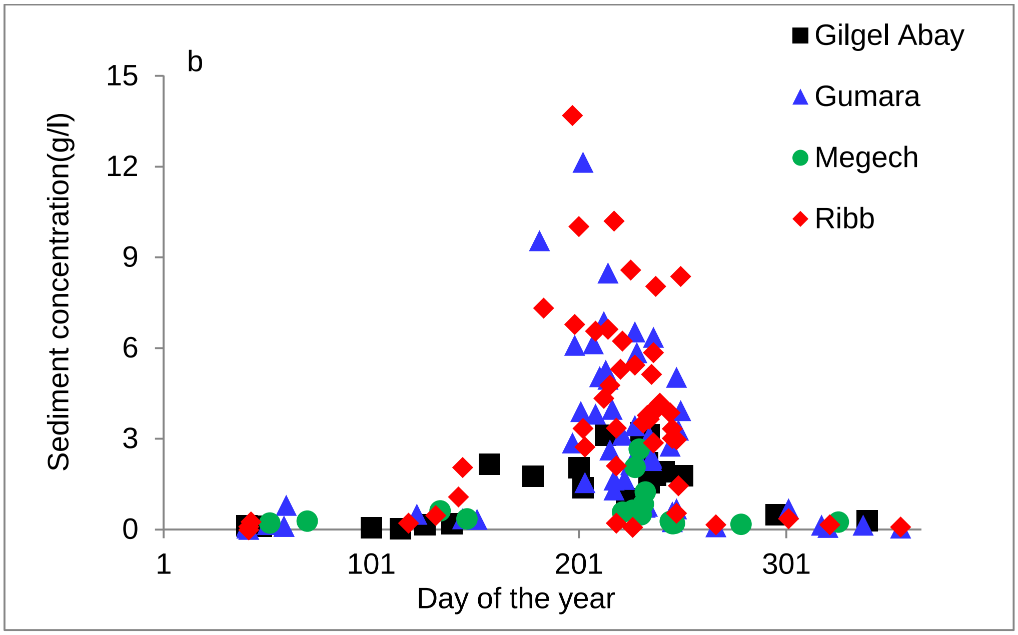


Figure 2. Location maps of the Lake Tana watersheds (Gilgel Abay, Gumara, Ribb and Megech) and the 100 ha watersheds (Debre Mawi, Anjeni and Maybar) in or close to the Blue Nile Basin.

Comment: Results: Use different symbols for the graphs in case the printing is in Black and White. Use proper scale for the Y and X axis at the 1:1 graphs

Response: Thanks for the comment. We corrected as proposed and included in the manuscript. The graphs in which the symbols have been changed are presented below.





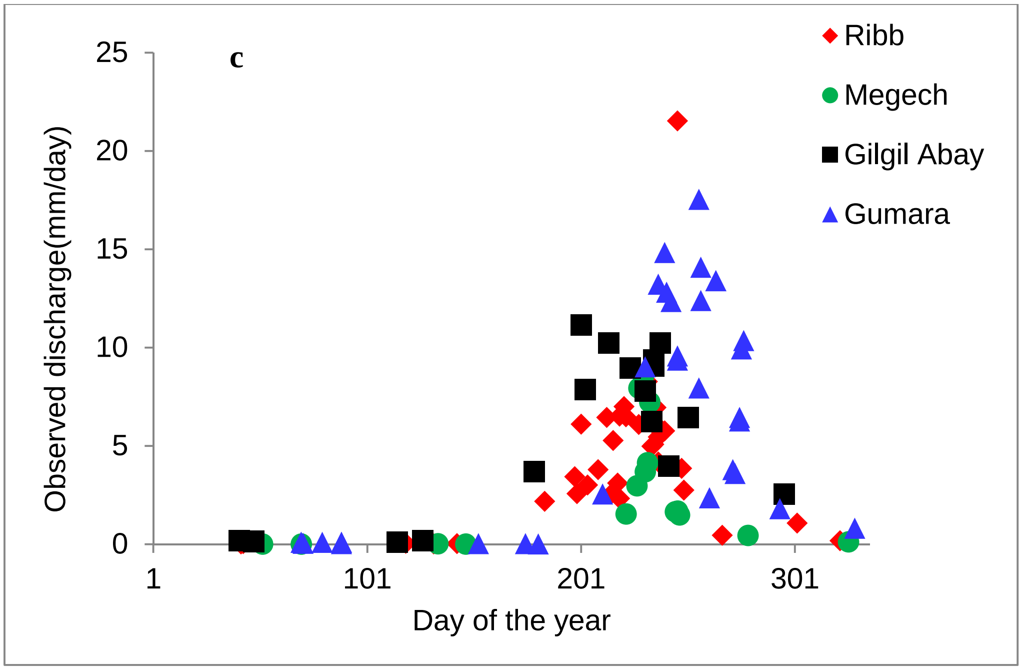
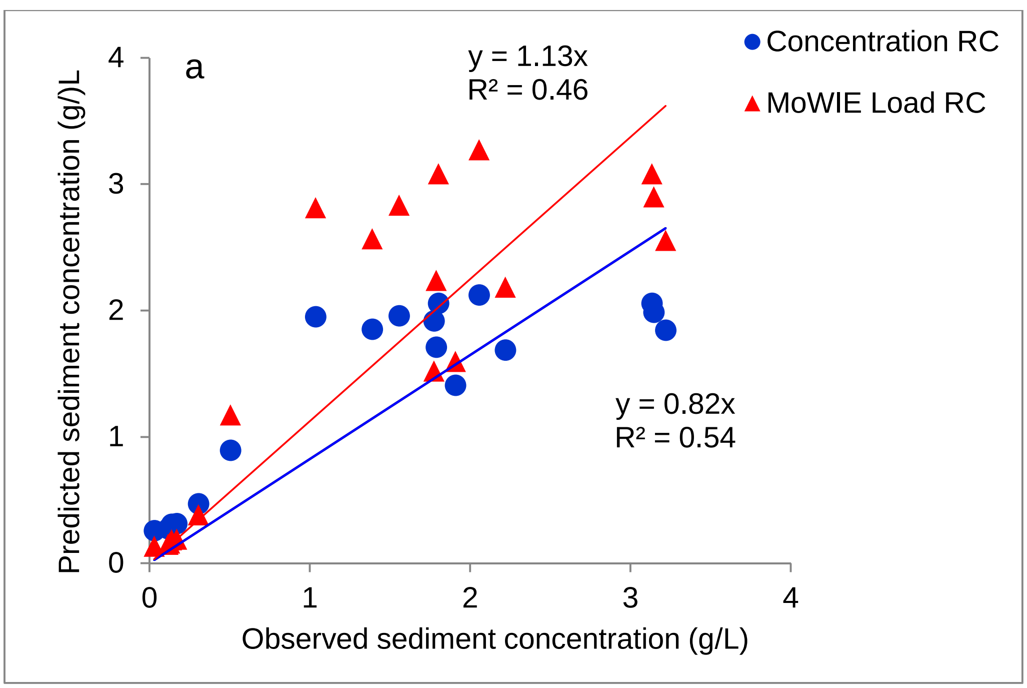
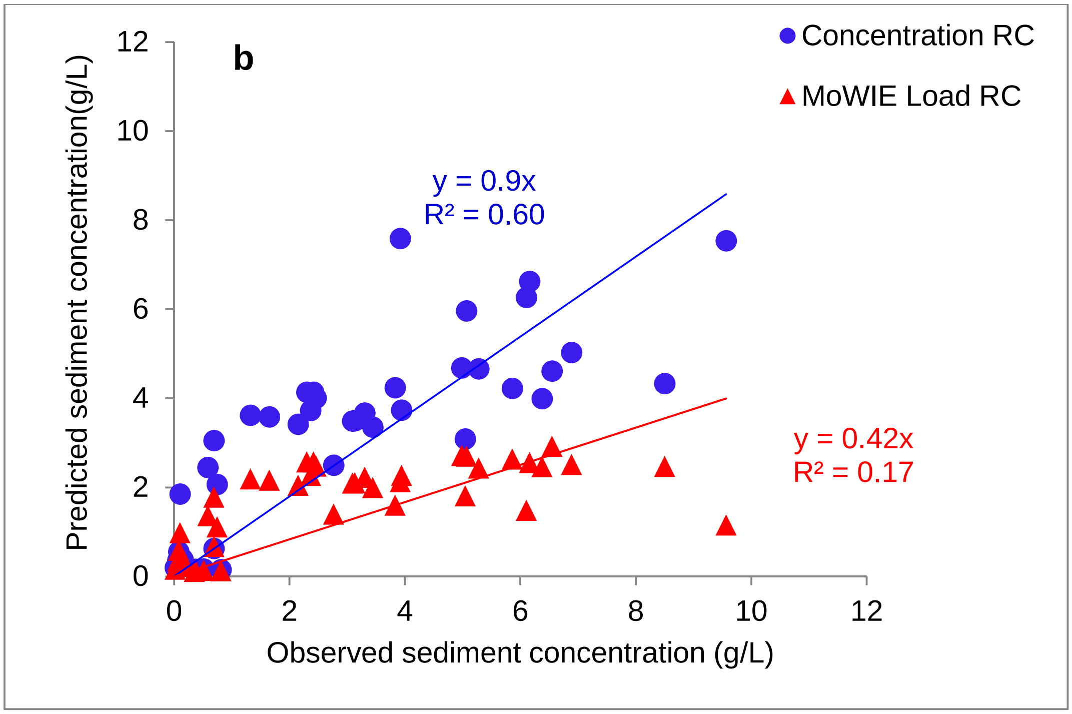
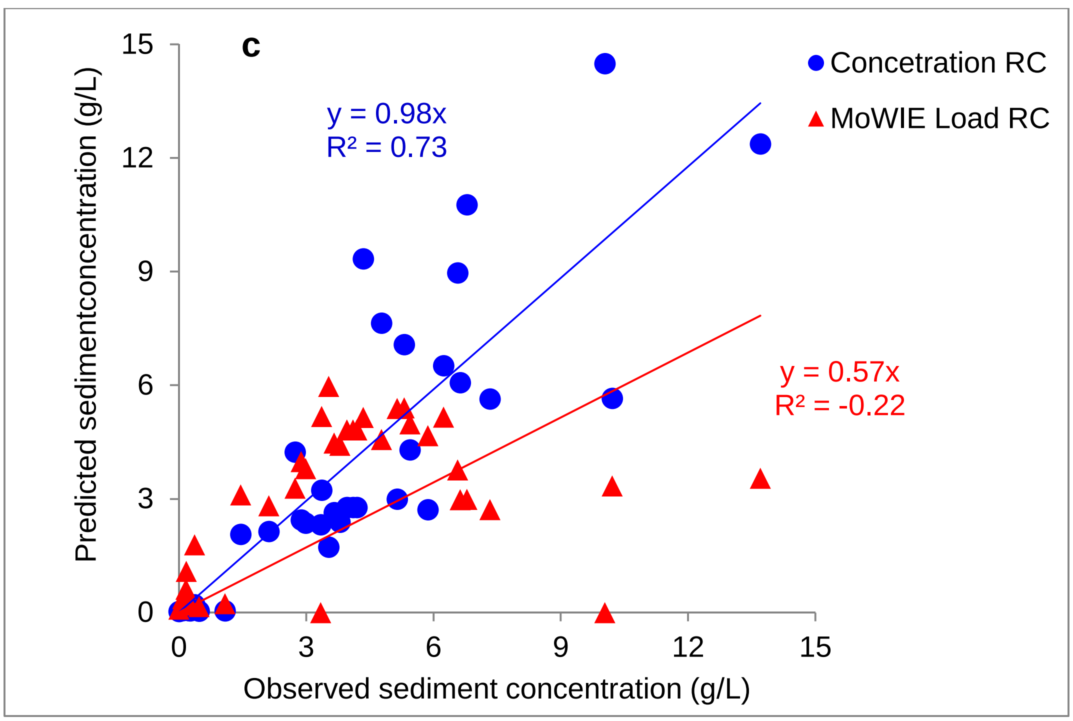


Figure 3. Observed sediment concentration and discharge for the four Lake Tana watersheds: Gilgel Abay, Gumara, Megech and Ribb. **a**. sediment concentration vs date of sampling **b**. sediment concentration as a function of day of sampling independent of the year, and **c**. observed discharge plotted vs sampling day.




b
Concentration RC
MoWIE Load RC
y = 0.9x
R² = 0.60
y = 0.42x
R² = 0.17
Predicted sediment concentration(g/L)
Observed sediment concentration (g/L)
0
2
4
6
8
10
12


c
Concetration RC
MoWIE Load RC
y = 0.98x
R² = 0.73
y = 0.57x
R² = -0.22
Predicted sedimentconcentration (g/L)
Observed sediment concentration (g/L)
0
3
6
9
12
15

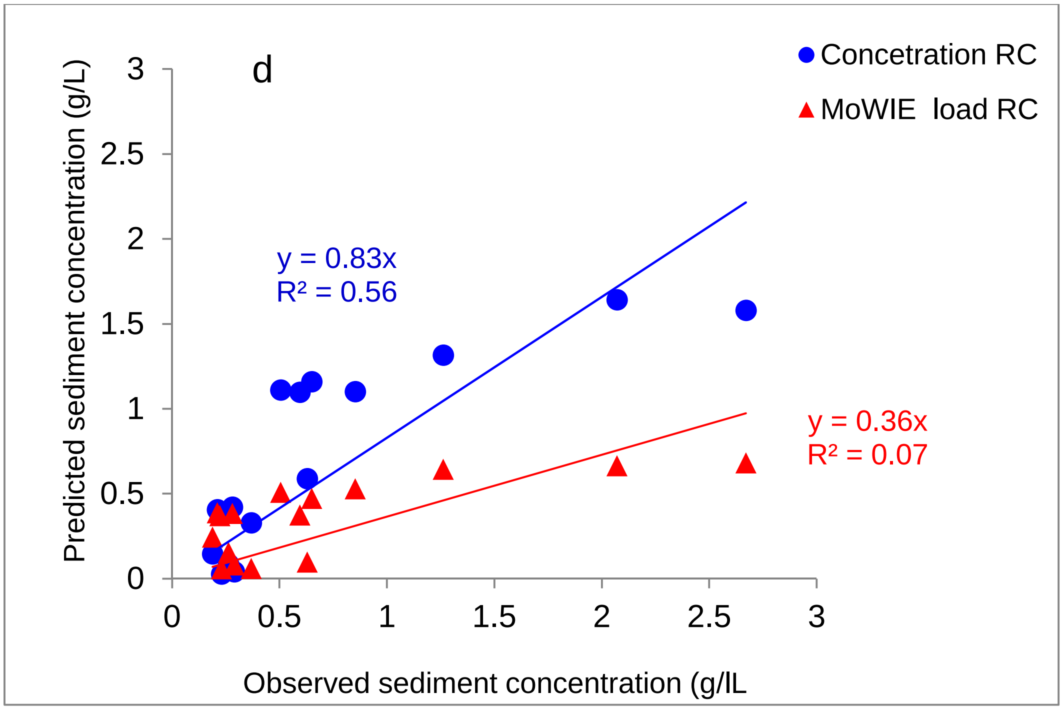
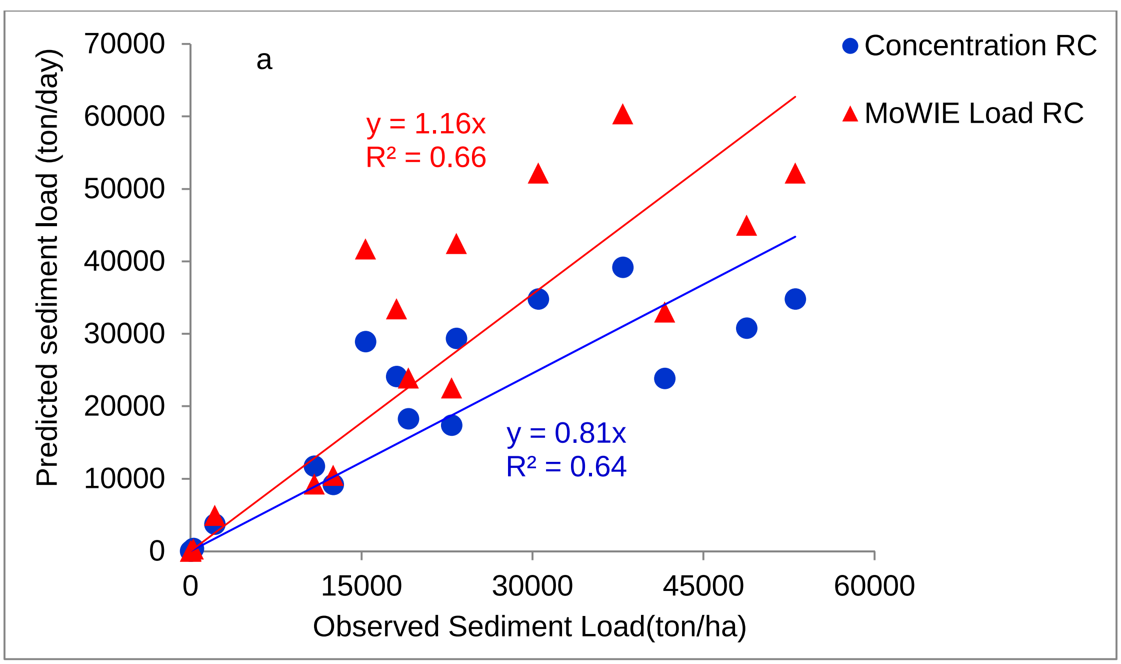
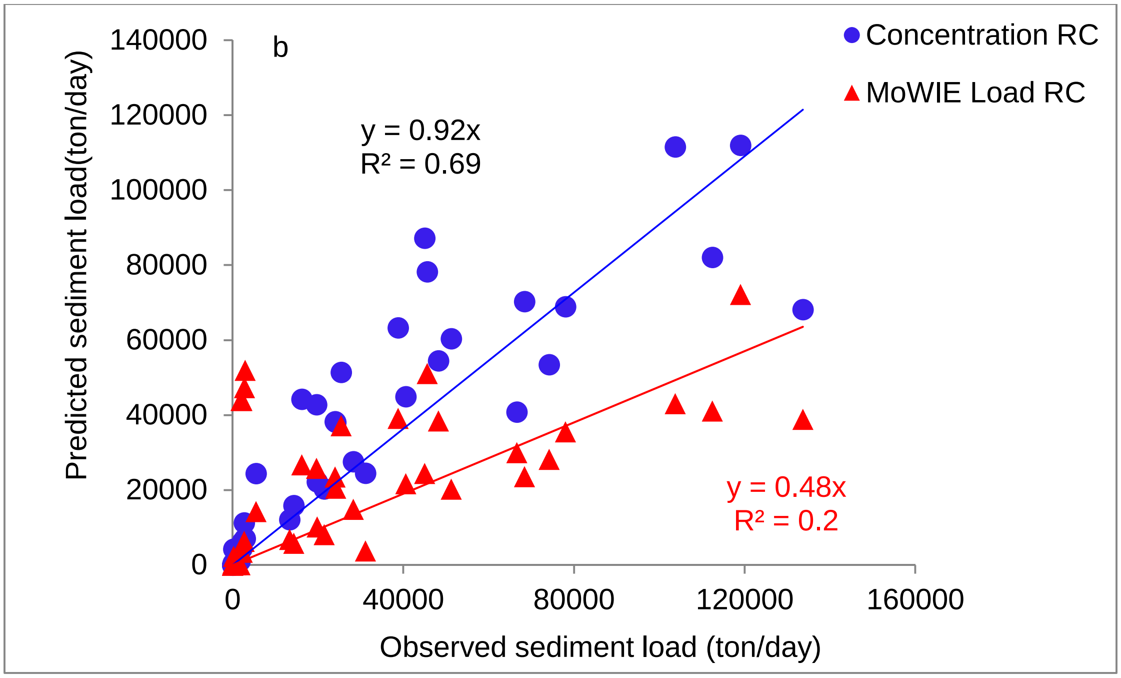
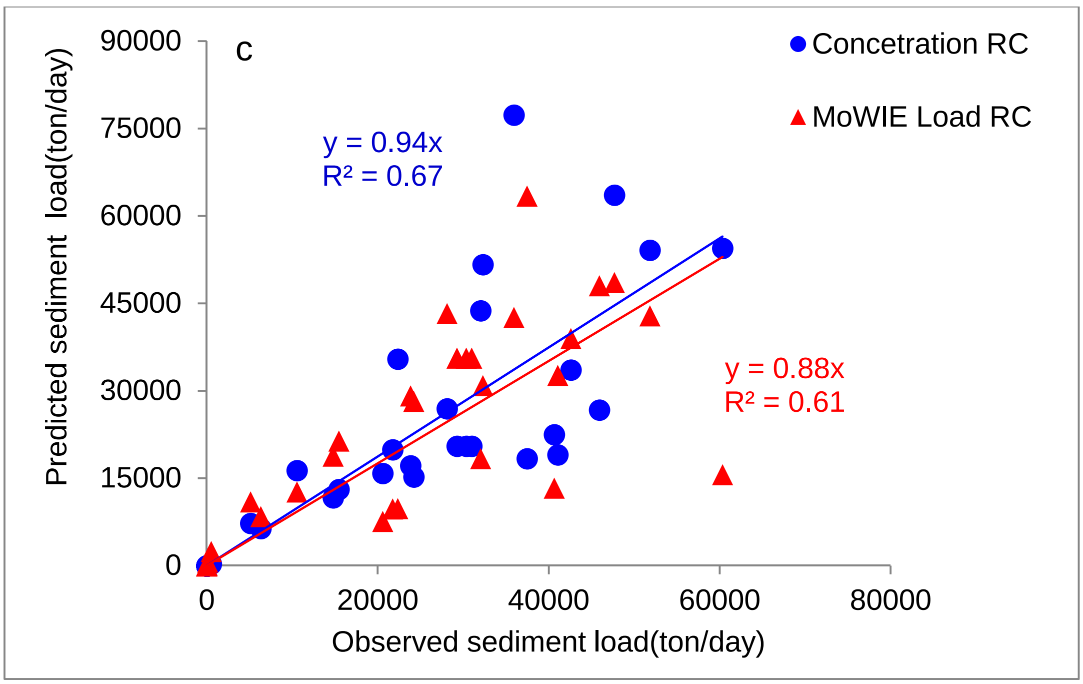


Figure 4. Predicted versus observed sediment concentration using concentration rating curve and MoWIE load rating curve for the Lake Tana watersheds (a) Gilgel Abay, (b) Gumara, (c) Ribb, (d) Megech





b
Concentration RC
MoWIE Load RC
y = 0.92x
R² = 0.69
y = 0.48x
R² = 0.2
Predicted sediment load(ton/day)
Observed sediment load (ton/day)
140000
120000
100000
80000
60000
40000
20000
0
0
40000
80000
120000
160000



c
Concetration RC
MoWIE Load RC
y = 0.94x
R² = 0.67
y = 0.88x
R² = 0.61
Predicted sediment load(ton/day)
Observed sediment load(ton/day)
90000
75000
60000
45000
30000
15000
0
0
20000
40000
60000
80000

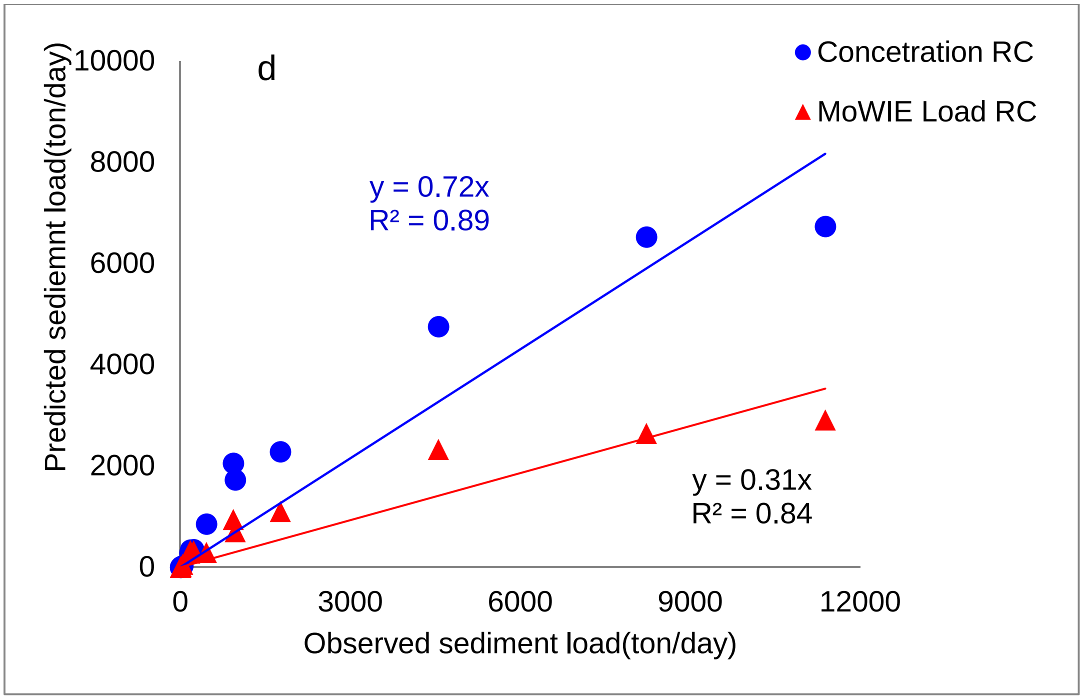


Figure 6. Predicted versus observed sediment load using concentration rating curve and MoWIE load rating curve for the Lake Tana watersheds (a) Gilgel Abay, (b) Gumara, (c) Ribb, (d) Megech

Comment: Figure 2: I couldn't see the logic behind the fitting of the Load/concentration rating curves from catchment based (> 500 km2) to micro-watershed (0.1-11 km2). These catchments are expected to have different morphologic and fluid transport mechanisms. So can we say that the correlation happens by chance?

Response: Our result of good correlation for the new rating curves at various scales did not happen by chance. It indicates that there is some similarity in response by the two groups of watershed sizes (100's km^2 and 100 ha). The similarity can be explained as follows: Such similarities exist because the watersheds behave similarly when they dry out and wet up. When they dry up, all upland watersheds will be plowed and vegetation cover is poor. At this time, sediment is available that can be transported within the watersheds. After all rills have established (after P_T), sediment in the watershed is limited as the soil is saturated and becoming more cohesive and vegetation cover improves.

Specific comments:

Comment: Page-1: Line 5-9: Need to be combined properly.

Response: Apologies of not being more clear. The statement was re-written starting on page 1421 line 2 of the original manuscript.

> "In order to determine sediment loads in the absence of these measurements, models have been used. Knowing the total sediment loads from rivers is essential for evaluating the siltation of reservoirs (Ali et al., 2014) and assessment of soil erosion and nutrient loss (Walling, 1977). As a result Knowledge of sediment concentration is vital in most environmental applications among others as it hampers fish reproduction and reduces the esthetic value of surface waters (Vijverberg et al., 2012)"

Comment: Page-1: Line 10-13: Should be rewritten

Response: We rewrote the statement as follows and added on page 1421 line 10 in the original manuscript.

> "Construction of the Grand Ethiopian Renaissance Dam is underway in the Blue Nile basin and other hydroelectric dams upstream are in the planning stage. Therefore, controlling the high sediment losses is of high importance."

Comment: Page 1: Line 19-21: The USLE and its derivatives are RUSLE and MUSLE.

Response: Thanks for the comment we corrected as proposed and included in the original manuscript on page1421 line 21

> "The USLE and it derivatives are RUSLE (Renard et al., 1991) and MUSLE (Williams and Berndt, 1977)"

Comment: Page 1: Line 21 – 2(Page 2): It is not clear what these area? Indicate the cons and pros of these models as it is or based on a general category y (empirical, process based, hybrids...).

Response: The statement was re-written as follows and inserted in the original manuscript on page 1422, line 5

"The challenges in using these models can be seen in two ways i) The models have been developed in temperate climate and arid areas where the runoff mechanism is governed by Infiltration excess unlike the highland areas where saturation excess runoff is dominating (Steenhuis et al., 2009; Bayabil et al., 2010; Tilahun et al., 2013) and ii) almost all of the models need intensive data with many parameters that might be available centrally in developed countries but not in developing countries such as Ethiopia .

Comment: Page-2: Line 5-6: What kinds of data are important for these models that are difficult to obtain in developing countries?

Response: Thank you for the comment. We included the following text in the manuscript on page 1422 line 2 to specifically to explain the data scarcity.

"The reason is that these models were originally developed for areas where soils and land use data have been gathered since the early nineteen hundreds and can be obtained from a central location. However in Ethiopia these types of data are not (yet) available. For example, the land use and land cover map do not list the leaf area index required as input to SWAT. Similarly, the soil data in Ethiopia is coarse and is missing basic input data such as soil texture, hydraulic conductivity and other soil physical parameters. Therefore rating curves are often the preferred choice for predicting sediments loads in developing countries using periodically measured sediment concentration and discharge data "

Comment: page-2 Line 6-8: There is no proper transition of this section from the previous idea.

Response: Thanks for pointing out the missing link. We corrected it by adding the following paragraph between on page 1422 after line 5:

"There is direct link between models and rating curves in sediment studies in the Ethiopian highlands. Because of the limited data, rating curves are used to validate models (e.g. Easton et al., 2010 and Setegn et al., 2009b). Here we developed concentration rating curves so that sediment concentrations can be generated for calibration and validation from observed flow. In addition, it can be used as a model to predict sediment concentration and load."

Comment: Page 2: What is the source of Eq.2?

Response: The source of Eq. 2 is from Eq.1 by dividing the estimated load to the corresponding discharge. As stated in in the text

> "The concentration, *C*, can be found by dividing the load (Eq. 1) with the discharge *Q*"

Comment: Page 9: It is not clear why you exclude the time from 1964-1967.

Response: Thanks for the comment. It was a misprint and we corrected it as follows:

> "The sediment concentrations in the Lake Tana watershed have been increasing since the initial measurement were made in 1964. We selected the following periods for analysis 1964-2008 for Gilgel Abay, Gumara and Rib. For Megech the data was only available for 1990-2007 and the analysis was made with this data."

Comment: Page 10 (1427 line 11): Line 1: To what extent is the data "Good"?

Response: No rigorous tests were performed. Based on past experience these years had good data mainly because the country was politically stable at the time.

Comment: Page 10: (1472 line '14) for how long you collected the rainfall data?

Response: For larger watersheds, the existing rainfall data from 1994-2008 was obtained from National Meteorological Agency Bahir Dar Branch. For micro watersheds (100 ha watersheds) the rainfall data was collected at the same time when discharge and sediment data was measured. This was explained in the available data section on page 1427, line 14

Comment: Page 10: Line 15-16: For the calculation of the effective precipitation on a daily basis, initial abstraction is more important than the ET.

Response. We agree that initial abstraction plays a role on days that rains especially when rainfall is small. We decided not include it because we felt that the effect was likely small during intermediate and large storms especially since the rain variability is great for monsoon rainfall. We did not include the rainfall variability either due to lack of available data.

Comment: Page 10: Line 6: Is there a reason for using Thiessen polygon in the study area?

Response: We used Thiessen polygon method for estimating the areal average rainfall from the watershed. We have chosen this method because it was simple and doesn't require additional information like topography and distance than the other methods. Based on the comment, we rewrote the sentences as follows. This is included in the original manuscript on page 1427, line16.

> "The areal rainfall was calculated using the Thiessen-polygon method for the available rainfall stations in the Lake Tana watersheds as these watersheds have two and more than two rainfall stations. The method was chosen because of it was simple and does not require additional or more information. Details of station weights based on the method are given in the supplementary materials (Supplement, Table A1)"

Comment: Page 10: Line 10-11: For Page 12: Line 4-6: The range of goodness of fit for the NSE should be indicated.

Response: Thank you for the comment. The "goodness of fit" of NS value in the manuscript was included on page 1429, line 6 as follows:

> "According to Moriasi et al. (2007), the goodness of fit for model performance was considered as very good for NS>0.75; good for NS values between 0.65 and 0.65; satisfactory when NS values were between 0.65 and 0.5 and poor wen less than 0.5"

Comment: Page.... Line 13-15: What was the reason for dividing the period into three groups?

Response: It is known that there are two distinct seasons in the highland: dry and monsoonal rainy period. During the rainy period, there is a threshold at which the watersheds saturates and behave differently from the beginning of the rainy period. Since there is a different behavior in the dry, beginning and end of rainy periods, we used these periods to develop the model. (We were referring the data collection periods in developing the rating curve and we found this information from MoWIE).

Comment: Page 13: Line 1-5: The hyetograph of each watershed should be plotted on the upper axis and the Pe and PT should be indicated clearly.

Response: Thanks for the comment. The hyetograph and the PT are different for each of the watersheds and would be therefore be difficult to include in the figure.

Comment: Page 13, Line 15-16: What do you mean by under predicting? Which one is reliable? The measured/observed or predicted?

Response: We were comparing the sediment load and sediment concentration predicted by concentration rating curve and by MoWIE load rating curve with the measured sediment load and concentrations. In this specific case, we were referring that, the MoWIE load rating curves under predicts the sediment concentration compared to concentration rating curve that we developed. We changed the text as follows to make it clearer.

> "For the Lake Tana watersheds, the sediment concentrations predicted by the MoWIE load rating curve were less than the observed values (Fig 4). The concentration rating curve predicted the concentrations relatively well with Nash Sutcliff values of ranging from 0.55 to 0.65 and R^2 values between 0.57 to 0.69 of with slopes close to one"

Comment: Page 13, Line 16-18: Why you include R2 in the "Observed-Predicted" graphs?
Response: It was to indicate the strength of relationships between the observed and predicted sediment concentrations and load by using concentration and MoWIE load rating curves.

Comment: Page 13, Line 18-19: For Gilgel Abay the MoWIE curves perform better. This should be indicated clearly in this section.
Response: Comment accepted. The paragraph reads now and included in the manuscript. Explanation was included on page 1431 line 6 as follows:

> "For the Lake Tana watersheds, the sediment concentrations predicted by the MoWIE load rating curve were less than the observed values (Fig 4). The concentration rating curve predicted the concentrations relatively well with Nash Sutcliff values ranging from 0.55 to 0.65 and R^2 values between 0.57 to 0.69 of with slopes close to one. The only exception was the sediment load predictions for the Gilgel Abay (Fig. 6a) that was slightly better predicted by the MoWIE load curve than the concentration rating curve. This was partly due to the fact that there were relatively few measurements made during the beginning of the rain phase when sediment concentration where high.

Comment: Page 14: Line 9-10: How do you measure "Good"?
Response: "Good" in this case is to indicate the rating curves performance based on Moriasi et. al., (2007). The definition of good is included in the revised watershed as indicated in a previous comment

Comment: Page 15: Line 13-15: Do you think that having too much parameter means "Better" estimating capacity

Response: Since we fixed the power of b to 0.4, there is only one more parameter to fit. So indeed it might make the fit “better”. Since the intent of the paper was obtaining a better fit, it can be justified using one more parameter.

Additional References

Kokpinar, M. A., Altan-Sakarya, A. B., Kumcu, S. Y. , and Gogus, M. (2015).Assessment of sediment yield estimations for large watershed areas: a case study for the Seyhan, Demirkopru and Hirfanli reservoirs in Turkey, Hydrological Sciences Journal-Journal Des Sciences Hydrologiques,60,2189-2203:DOI: 10.1080/02626667.2014.959954.

Choi, S.U., and Lee, J.: Assessment of total sediment load in rivers using lateral distribution method, Journal of Hydro-Environment Research, 9,381-387, 2015: DOI: 10.1016/j.jher.2014.06.002.

Kheirfam, H., and Vafakhah M.: Assessment of some homogeneous methods for the regional analysis of suspended sediment yield in the south and southeast of the Caspian Sea, Journal Of Earth System Science, 124, 1247-1263,2015:DOI: 10.1007/s12040-015-0604-7.

Fenn, C. R., Gurnell, A. M. and Beecroft, I. R.: An Evaluation of the Use of Suspended Sediment Rating Curves for the Prediction of Suspended Sediment Concentration in a Proglacial Stream, Geografiska Annaler, Series A Physical Geography
67, 71-82, 1985.

Irvine, K. N. and Drake, J. J. Process-Oriented Estimation of Suspended Sediment Concentration, JAWRA Journal of the American Water Bulletin, 23, 1017- 1025, 1987.

Sichingabula, H. M.: Factors controlling variations in suspended sediment concentration for single-valued sediment rating curves, Fraser River, British Columbia, Canada By: Hydrological Processes, 12, 1869-1894, 1998.

Sun, H., Cornish, P. S., and Daniell, T. M.: Turbidity-based erosion estimation in a catchment in South Australia By: Journal of Hydrology Volume, 253, 227-238, 2001: DOI: 10.1016/S0022-1694(01)00475-9.

Arora, M., Kumar, R., Kumar, N., and Malhotra, J.: Assessment of suspended sediment concentration and load from a large Himalayan glacier, Hydrology Research, 45,292-306, 2014.

Renard, K.G., Foster, G. R., Wessies, D. K., and Yoder, D. C.: Prediction of soil erosion by water: A guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE). Report ARS 703, Agricultural Research Service, US Department of Agriculture, 1991.

Williams, J. R. and Berndt, H. D.: Sediment yield prediction based on watershed hydrology. *Trans. Am. Soc. Agric. Engrs,* 20, 1100–1104, 1977.

Moriasi, D., Arnold, J., Van, L.M., Bingner, R., Harmel, R. and Veith T.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. Asabe, 50, 885-900, 2007.