Reply to comments of Anonymous Referee #1

The authors would like to thank the Anonymous Referee #1 for the comments; they are very helpful to improve our manuscript. Our responses to the comments are provided in bold (below each comment).

General comments

The paper develops a rainfall-runoff model of medium complexity, distinguishing between groundwater, direct runoff and interflow; and splitting the catchments into three using topography. The parameter estimation uses a combination of calibration and estimation of parameters based on soil properties. The work is a brave attempt to develop and test a model for an area that suffers from limited flow, precipitation and hydrological properties data. However the paper does not really provide significant advances in understanding hydrological responses or innovation in modeling. The quality of model outputs is declared good, but this is arguable and a detailed analysis of model errors has not been reported.

We do agree that the type of model used does not fully allow to gain additional insight in hydrological responses, and the framework used is not innovative in modeling as well, as many conceptual models have been developed and used. However, we believe that the innovation of this paper should not be sought in these aspects but rather in the fact that through a simplified model (i.e. a conceptual model), we are able to assess hydrological processes in the Lake Tana basin, which would be complex to do if a more sophisticated model were used (one would immediately encounter problems due to data scarcity). The insights that are gained learn a lot about the hydrology of this system, which up till now was less understood. We hope that the reviewer may agree that this objective of the paper is sufficiently novel to be published.
Concerning the quality of the model outputs and the errors associated with them, they can be evaluated based on model performance indicators like the Nash-Sutcliffe Efficiency (NSE), Root Mean Squared Error (RMSE), the coefficient of determination (R^2), Bias, etc. The visual comparison (Plots) can also give an overall judgment. The authors used NSE, RMSE and R^2 in addition to the plot comparison of the observed and simulated outputs. The statistical results for NSE and R^2 were greater than 0. 7, which shows the good performance of the model.
Model uncertainty arises from a variety of sources, such as model parameterization, process representation, equifinality, and calibration accuracy. We agree the importance of a detailed analysis of model errors. In this regard, we carried out more than 2500 iterations using the Particle Swarm Optimization (PSO) technique to reach to the optimal model parameters and minimize calibration errors. We made sensitivity analysis of the model parameters to identify the important parameters and rank parameters that have significant impact on the model outputs (Fig.7 in the paper). Each and every step of the model development has been discussed with the possible limitations.

We agree with the reviewer to further investigate model errors and model clarity based on the relevant comments suggested by the reviewer that will be shown on the response to the more detailed comments of the reviewer and revise the manuscript.

While recognizing the data issues, the authors claim that periods of data are relatively high quality; however this has not been shown, for example the reader cannot judge the degree of rainfall and flow data errors. The model is rather complex given the data restrictions, shown by the sensitivity analysis. The conclusions about hydrological processes cannot be justified given the data issues, the use of text book values of parameters of unknown applicability here, and the apparent limited performance of the model.

The authors' claim of relatively high quality discharge data for the calibration of the models emancipates from the data acquisition methodologies used. Unlike the previous water level measurement of twice a day using staff gauges, in this case, the water level measurements were made using Mini-Divers, automatic water level recorders (every 10min), and manual readings from a staff gauge (three times a day, at 7 a.m., 1 p.m. and 6 p.m.). Moreover, rating curves were produced using recent survey of river x-sections. We do not dare to say that the data is absolutely prefect. However, there is a significant improvement to what was available before. We agree that we did not show the rating curve plot in the manuscript together with the range of levels to which the rating curves were applied. In the revised version, we will elaborate more on the accuracy of the rating curves that were established. Further in this rebuttal, we will elaborate in more detail on this issue.

With respect to the rainfall data, the authors used the available rainfall data from the rain gauges in and around the study catchments and discussed the accuracy of the data. In the revised version, we will include the location of the rain gauges. This figure is also included further in this rebuttal where a more detailed discussion on this issue is given.

The reviewer commented on the limitations of the use of text book values of parameters of unknown applicability. We agree with the comment and understand the limitation. Unfortunately, the study area faces high scarcity of data. In such instances, it is normal to consider data from relevant sources with caution. Accordingly, porosity and field capacity of the soils were derived from the study area soil texture data based on literature recommendation. Similar procedures are followed for the saturated hydraulic conductivity for the deep percolation by identifying the likely aquifer formation of the study area. We tried to optimize the literature recommended values to get better performance of the model (again, we refer to answers to more detailed comments of the reviewer).

There are various gaps in the description of the method, as I explain in my comments below. Overall, I am not confident that this model or the conclusions made about processes are justified, and all the evidence points to the model being over-complicated. The authors may have been better using a stricter application of the methods of Fenicia et al. 2008 to gradually build up the complexity of the model to the justified level, with more explicit attention to errors in inputs and outputs. Below are a few more detailed comments that may help in a revised version; however in my opinion the aims and approach need re-thought.

We note the need for the clarification of some of the descriptions of the method that were not clear to the reviewer and to the other readers. These will be addressed in reply to the detailed comments of the reviewer point by point. We appreciate the suggestion of the reviewer to use a stricter application of the methods of Fenicia et al. (2008) in the building up of the model and we will consider it as a bench mark model to compare the performance of our topography driven model in the revised version of the paper.

More detailed comments (Specific comments)

5293, 3. Model is modified from what? Not clear what is being modified. 5293, 10. Differently from what?

As explained on page 5293 in the paper, our model is developed based on the works of Jothityangkoon et al. (2001), Krasnostein and Oldham (2004) and Fenicia et al. (2008). However, we made modifications on some of their model concepts and equations. The major modification (variations) made in this paper can be seen from three cases.

i. The catchment bucket representation concept

The works of Jothityangkoon et al. (2001), Krasnostein and Oldham (2004) and Fenicia et al. (2008) considered the catchment bucket to consist of the soil reservoir and the groundwater reservoir. In our model, in addition to the soil and groundwater reservoirs, we included the other component that considers the impermeable part of the catchment. So, the catchment is divided into the soil and groundwater reservoirs part and the impermeable part. As we know that our study catchments have impermeable surfaces (with little or no soil cover), we needed to consider this separately in the rainfall-runoff process representation of the conceptual model.

ii. Soil surface Catchment characterization

Catchment characterization was made based on topography. Hence, the catchment was divided as steeply, hilly and level and the input data to the model were determined accordingly. This is because the model is not a fully distributed model and hence topography is considered as a major landscape characteristics to determine the other catchment features required for the model.

iii. Percolation to the groundwater table and hydraulic conductivity for the interflow

In the soil and groundwater reservoirs, we modified the equations of deep percolation and hydraulic conductivity for the interflow component of the soil reservoir.
For example, in the case of Fenicia et al. (2008), percolation to groundwater reservoir is modelled as:

$$P_s = P_{\max}(\frac{S_u}{S_{fe}})$$

where $P_{\max}$ **is maximum percolation,** S_u **soil storage and** S_{fe} **is maximum soil storage. For details, refer to Fenicia et al. (2008). In our paper, this is conceptualized differently as given in the paper. Moreover, we made a distinction between the upper and deep soil hydraulic conductivities such that the hydraulic conductivity for the interflow component of the soil reservoir is dealt separately in our modelling approach. Details of the equations are shown in the paper.**

Eq 8 and 9. Equations applicable at hill-slope scale? Needs some further justification.

Equations 8 and 9 are universal equations. Equation 8 is a universal equation for velocity.

$$velocity = \frac{Displacement}{Time}$$

This equation is applicable anywhere as long as the displacement and time are determined accurately. In our case, the displacement is assumed to be the average slope length of the catchment (distance subsurface flow travels) and the time is the subsurface flow response time (the time the subsurface flow takes to reach to its exit).
Equation 9 is Darcy's equation that describes the flow of a fluid through a porous medium. It is applicable for a porous medium as long as the flow is laminar (which generally is the case in the case of a natural groundwater flow). Similar application of Darcy's Law to the groundwater aquifer within a planar hillslope has been indicated in Jothityangkoon et al. (2001).

How can all these parameters be justified? Why are there only seven – they need estimated for each of the three slope classifications?

The model parameters are justified from calibration, validation, sensitivity analysis and performance studies of the model. From the model development, we identified seven parameters and these were calibrated using the Particle Swarm Optimization (PSO) technique. From the model sensitivity analysis, we showed that three of the seven model parameters are poorly sensitive and there is little confidence in the model's correspondence with these parameters and they can be reduced without appreciable impact on the model (This is shown on pages 5306 and 5307 in the paper under Section 6.3). The seven parameters for the three slope classifications are reached as follows.

Parameters for the recharge (α_1 and α_2)

In the three slope classification, α_1 is to consider for the recharge from the steeply slope into the medium slope surface and α_2 is for the recharge from the medium slope surface into the flat slope surface. There is no parameter for the steeply slope surface since there is no surface that recharges it. So, there are two parameters for the three slope classifications.

Parameter for the impermeable surface of the catchment (λ)
In this case, the catchment is divided into two surfaces (impermeable surface with no or little soil cover and the soil surface). The parameter λ is introduced to represent the fraction of impermeable surface within the total catchment and this part of the catchment is not classified as steeply, medium and flat slopes since the classification of this part of the catchment into such classes is not important. So we have one parameter.

The parameters β, γ, k_1 and $K_{s,u}$

These parameters β and γ are introduced to account variability of permeability and deep percolation of soil with soil water storage. k_1 relates discharge and storage for the ground water and $K_{s,u}$ is the saturated hydraulic conductivity in the upper soil layer. We assumed that these parameters are less influenced by topography and each model parameter is assumed to be same for each slope classification of the catchment. Moreover, it looks quit inconsistent to separate the groundwater system in the catchment and we preferred all the three slope based classified catchments to share the same groundwater reservoir.
In this perspective, we will have in total seven parameters for the three slope classifications. We agree with the reviewer that we did not provide this explanation in the paper. In the revised version of the paper, we will add these clarifications.

5300, 1-4. Local relevance of the text book values? Really the textbook should provide ranges, which are fed into calibration (further increasing the calibration problem).

We estimated porosity and field capacity of the soils and the saturated hydraulic conductivity for the deep percolation from literature recommendations. We agree with the comment and understand the limitation. However, owing to the high scarcity of data in the study area, it still remains necessary to consider data from relevant sources with caution. Hence, the soil texture class data of the catchments were used to estimate porosity and field capacity of the soils. From studies by Cosby et al. (1984), we note that soil texture is closely related to the variability in soil moisture parameters (porosity and field capacity of the soils). Similar procedures are followed for the saturated hydraulic conductivity for the deep percolation in that its value is estimated by identifying the likely aquifer formation of the study area. In fact, the literatures provide a range of values. In such instances, we considered average values and we tried to optimize the values by iterating to get the best model performance results.

5300, 12. We need to see location map of these gauges – as precipitation is the key input –and know something about their accuracy. Was the PE spatially variable?
What assumptions have been made about stream flow routing and stream-groundwater interactions?

The location map of the rain gauges is provided below and will be included in the revised version of the paper.
Generally rainfall data are obtained on daily basis. The data for most of the stations are consistent and continuous, particularly for first class stations like Dangila, Adet and Debretabor. However, we encountered gaps in some stations like Sekela Station for some periods in the year. In such instances, only the rainfall data from the other stations is considered. As discussed in the paper, most of the rainfall stations in Gilgel Abay catchment are installed at the water divides and there is no station in the middle of the catchment. In this regard, the Gumara catchment is with higher density of rainfall stations. PE is also spatially variable.

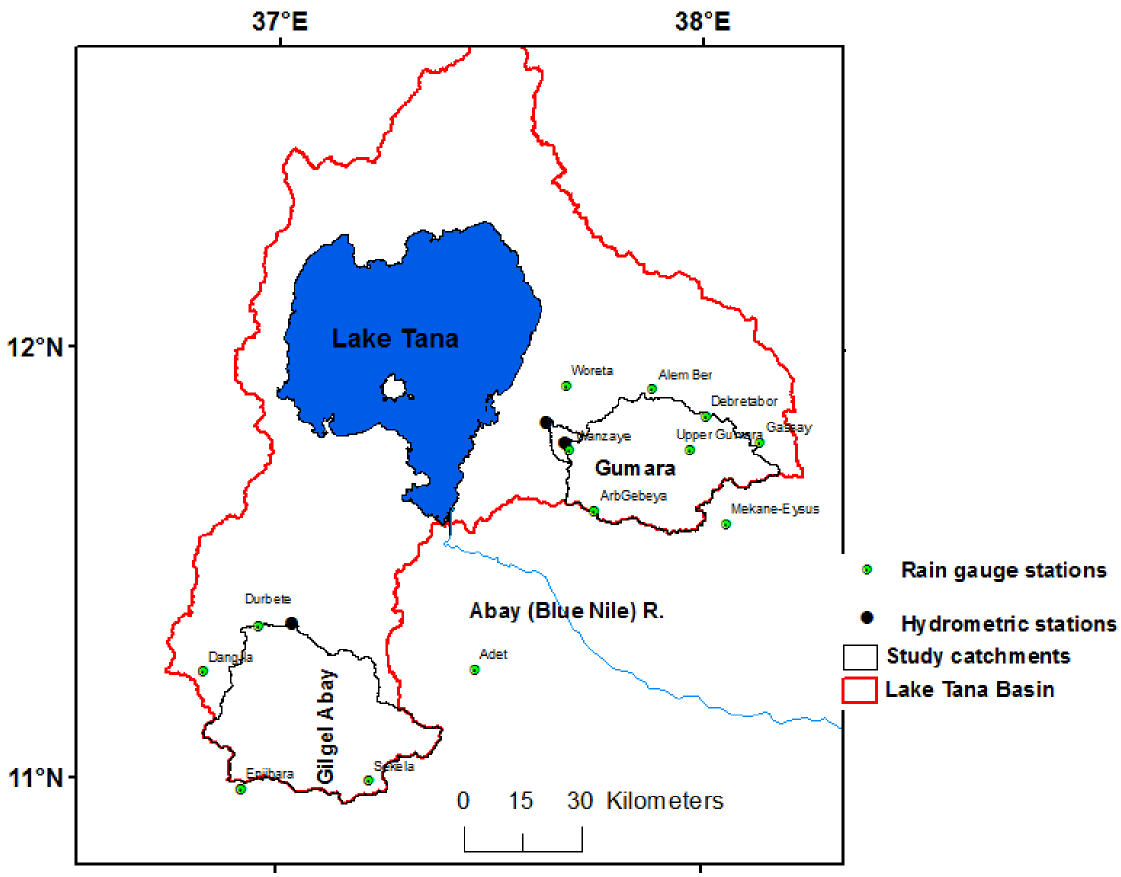


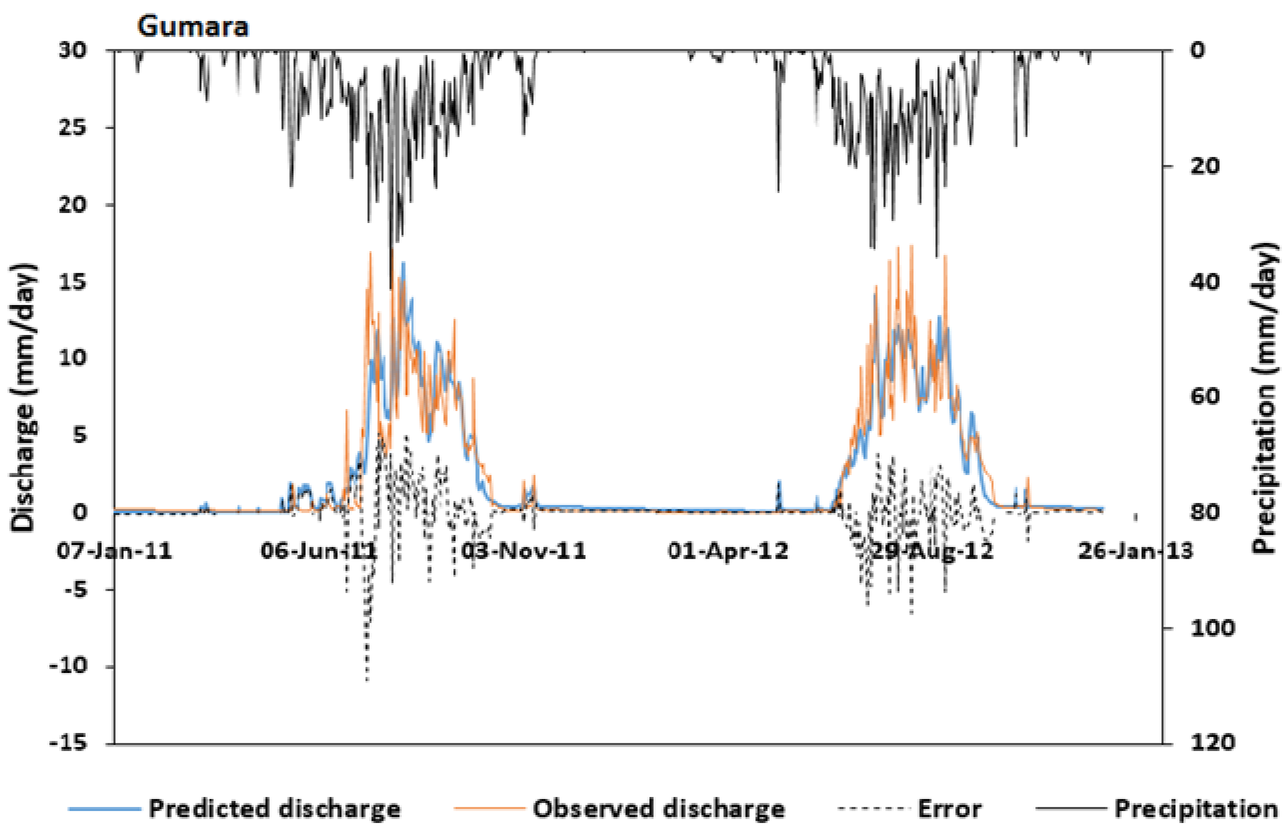
Fig. 1C. Location map of rainfall stations for the study catchments

In this paper of hydrological modeling, stream-groundwater interactions are assumed to be minimal and the groundwater is assumed to recharge the streams from deep percolation of rainfall on the catchments that produces baseflow of the

rivers/streams. The storage effect of the streams when considered on the basis of average daily flows of the streams is assumed to be negligible and hence streamflow routing was not considered for such smaller streams.

Figure 6. Does not look like great performance to me. Needs some more insightful plots to elucidate magnitude and nature of errors.

The quality of the model outputs and the errors associated with them are usually evaluated based on model performance indicators like the Nash-Sutcliffe Efficiency (NSE), Root Mean Squared Error (RMSE), the coefficient of determination (R^2), Bias, etc. The visual comparison (Plots) can also give an overall judgment. The authors used NSE, RMSE and R^2 in addition to the plot comparison of the observed and simulated outputs. The statistical results for NSE and R^2 were greater than 0.7, which shows the good performance of the model. The plots of the simulated and observed discharges of Figure 6 in the paper can show this, but it is true that there are some deviations of the simulated discharge from the observed ones at some points in the time series. There can be various reasons for this, as explained in the paper. One instance can be the rainfall data. As can be seen from Fig.1C, there are no rain gauges in the middle of the Gilgel Abay catchment and given the high spatial variability of the rainfall in the whole Blue Nile basin, this can create its own uncertainty on the model performance.
Fig.2C. below shows plot of the errors (difference of predicted discharge to observed discharge) to give more insight to the errors for Figure 6 in the paper.



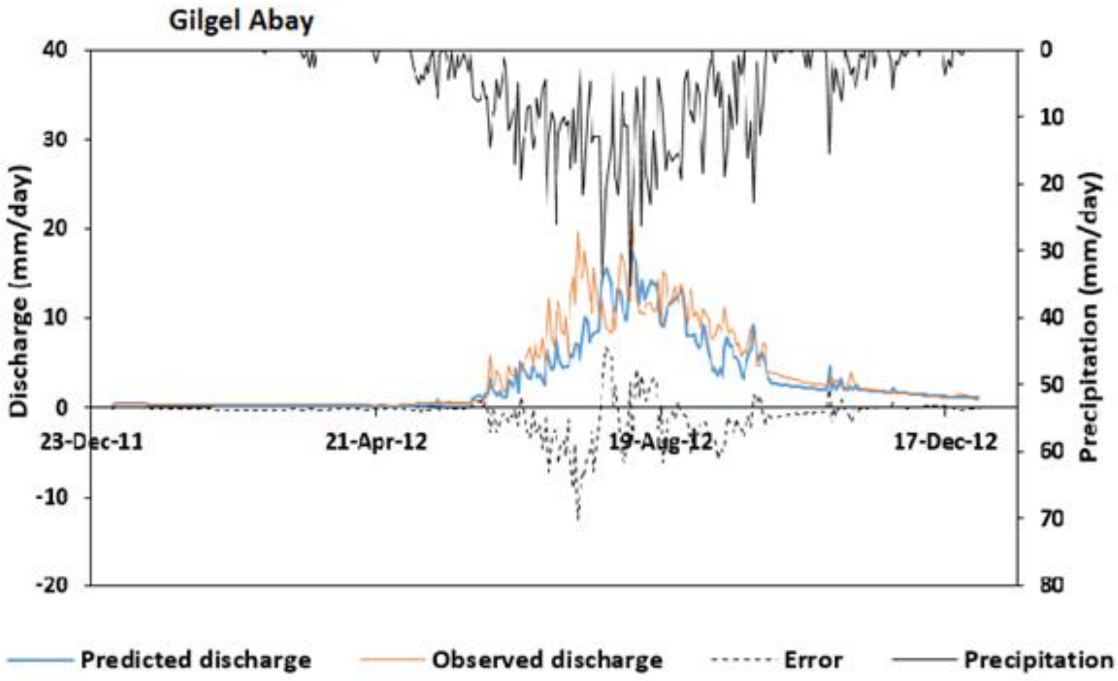


Fig.2C. Comparison of predicted and observed discharge together with the error (difference of the two) and precipitation of the Gumara and the Gilgel Abay catchments for the calibration period

Eq 20, 21. Authors claim that the gauged flow data are high accuracy – it would be useful for the reader to see the rating curves, together with the range of levels to which the rating curves were applied.
Stochastic optimization implies the stochastic nature of the input errors were considered? How are rainfall errors considered? Stochastic optimization gives stochastic outputs, which is misprepresented, or at least under-utilized, by reporting optimal parameter values.

The rating curves together with the range of levels to which the rating curves were applied have been provided below (Fig.3C and Fig.4C). We will also incorporate the figures in the revised paper.
In the model calibration, we did not use stochastic optimization that depends on one or more of the input data subject to randomness. The input data (for example rainfall) are observed data (soil data have been estimated from relevant sources when observed data are absent). For the model calibration, we used the particle Swarm Optimization (PSO) technique. PSO optimizes a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity. Each particle's movement is influenced by its local best known position but, is also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles.

This is expected to move the swarm toward the best (global) solutions. We used 30 particles in the PSO.

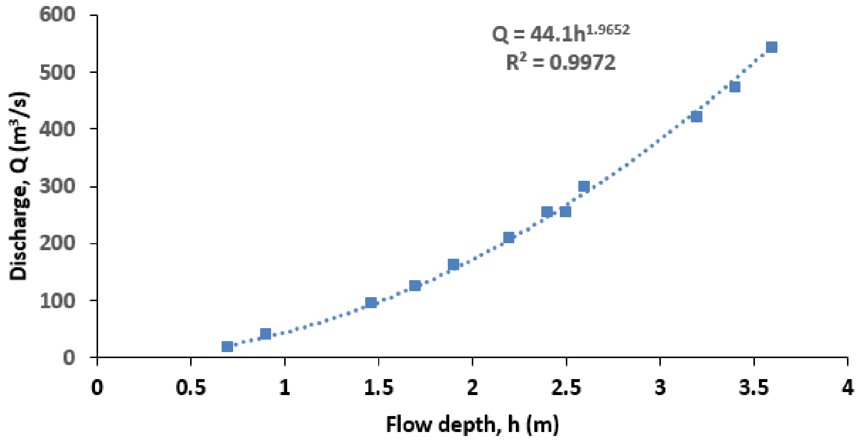


Fig. 3C. Stage-Discharge relationship (Rating curve) for Gumara at Wanzaye Station

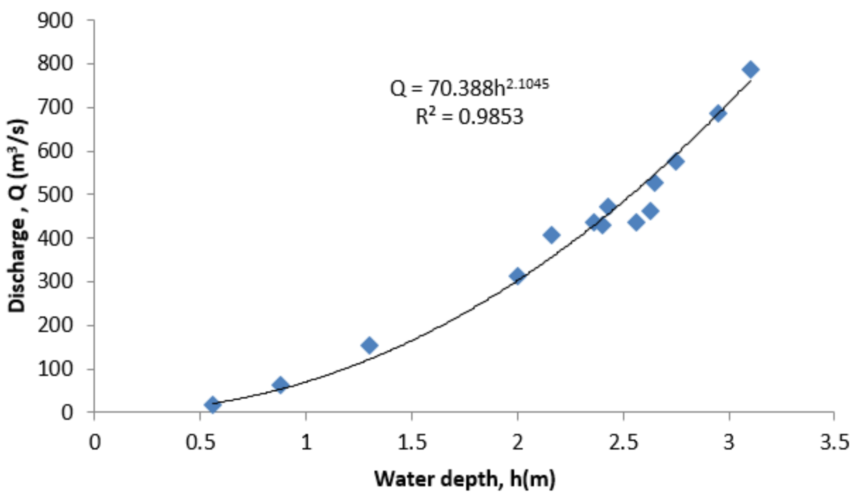


Fig. 4C. Stage-Discharge relationship (Rating curve) for Gilgel Abay at Picolo Station

5302, 5. Why only 7 parameters? Each catchment was split into different runoff production units to represent variation in catchment properties using topography, so why not 21 parameters?
Can the splitting into three areas be shown on a map, e.g. using color coding?

This comment is similar to a comment given by the reviewer above. The explanation is given below.
The seven parameters for the three slope classifications are reached as follows.

Parameters for the recharge $(\alpha_1$ **and** $\alpha_2)$

In the three slope classification, α_1 is to consider for the recharge from the steeply slope into the medium slope surface and α_2 is for the recharge from the medium slope surface into the flat slope surface. There is no parameter for the steeply slope surface since there is no surface that recharges it. So, there are two parameters for the three slope classifications.

Parameter for the impermeable surface of the catchment (λ)
In this case, the catchment is divided into two surfaces (impermeable surface with no or little soil cover and the soil surface). The parameter λ is introduced to represent the fraction of impermeable surface within the total catchment and this part of the catchment is not classified as steeply, medium and flat slopes since the classification of this part of the catchment into such classes is not important. So we have one parameter.

The parameters β, γ, k_1 and $K_{s,u}$
These parameters β and γ are introduced to account variability of permeability and deep percolation of soil with soil water storage. k_1 relates discharge and storage for the groundwater and $K_{s,u}$ is the saturated hydraulic conductivity in the upper soil layer. We assumed that these parameters are less influenced by topography and each model parameter is assumed to be same for each slope classification of the catchment. Moreover, it looks quit inconsistent to separate the groundwater system in the catchment and we preferred all the three slope based classified catchments to share the same groundwater reservoir.
In this perspective, we will have in total seven parameters for the three slope classifications. We agree with the reviewer that we did not provide this explanation in the paper. In the revised version, these clarifications will be included.

Fig.5C. shows the splitting of the Gilgel Abay and Gumara catchments (the study catchments) into three slope categories (steeply, medium and flat slope surfaces).

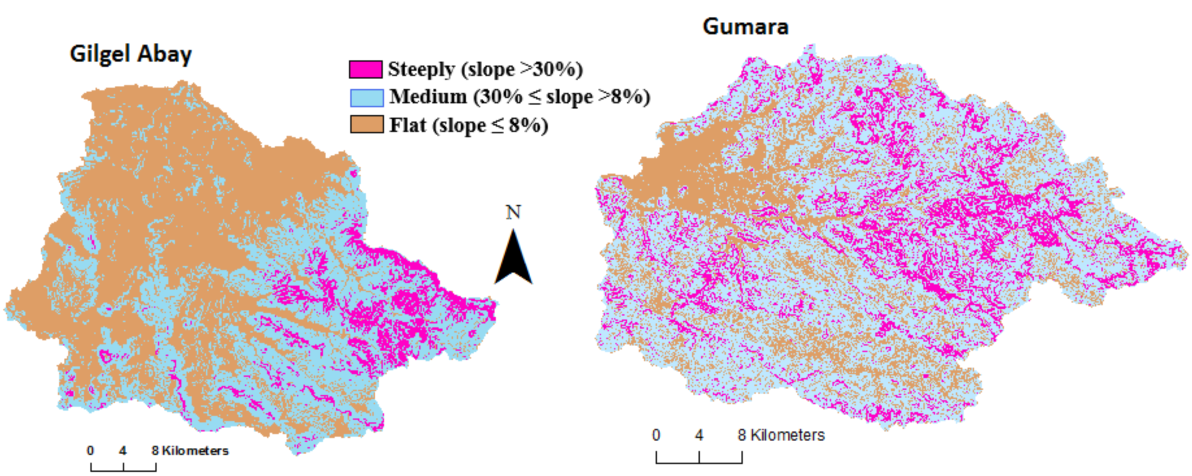
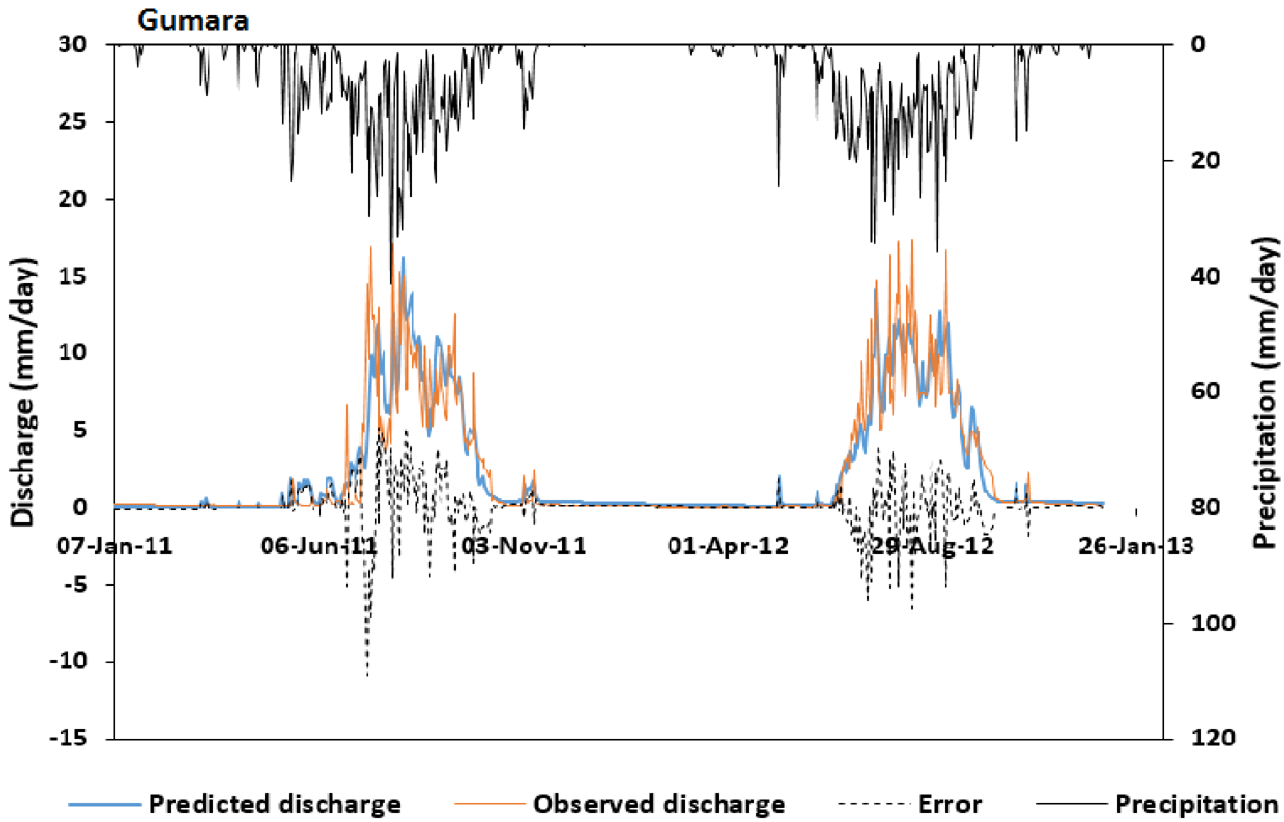


Fig.5C. The three slope categories for the Gilgel Abay and Gumara catchments

5304, 5. Figures 5 and 6 do not show this very well. Some more insightful plots about the errors are needed. In Figure 5, it seems there are some rather serious errors. E.g. the wetting up period deserves some discussion, In Fig 6, I cannot really see the nature or magnitude of the errors; however there are clearly some systematic errors that need critical discussion. The flow regime / climate in the validation period seems quite similar to the calibration period, so comparable performance is expected.
Validation should ideally test the model to breaking point.

This comment is similar to one of the comments above. Figures 5 and 6 in the paper are plots of predicted and observed discharge and precipitation of the Gumara and the Gilgel Abay catchments for the calibration and validation periods. We still believe that the plot simulates well the general behavior of the observed streamflow hydrographs. Fig.6C and Fig.7C below show plot of the errors (difference of predicted discharge to observed discharge) to give more insight to the errors. Generally, the errors look to have no trend. However, we notice that the model errors tend to increase during wetting up periods in most instances. Initially, the soils are relatively dry and most of the rainfall during the beginning of the rainy season is not effective to produce runoff in the model as the soil reservoir has to be filled first to generate the faster component of the runoff. In the model, mostly average conditions prevail owning to average input data (rainfall, soil, catchment characteristic, etc.). Besides model uncertainties, the rainfall data quality can also affect the model performance, mainly in the case of the Gilgel Abay catchment as discussed in the paper.
The flow data used for validation is from 2000-2005 (6 years data) and for calibration is the 2011 and 2012 years data for Gumara and the 2012 data for Gilgel Abay. Each year data is different, depending on the climate of the year and catchment conditions. However, the trend is similar each year such that there is high discharge in the rainy season (June to September) and a decreasing trend of discharge after September in line with the dry season. The 6 years discharge data is considered sufficient to run validation tests.



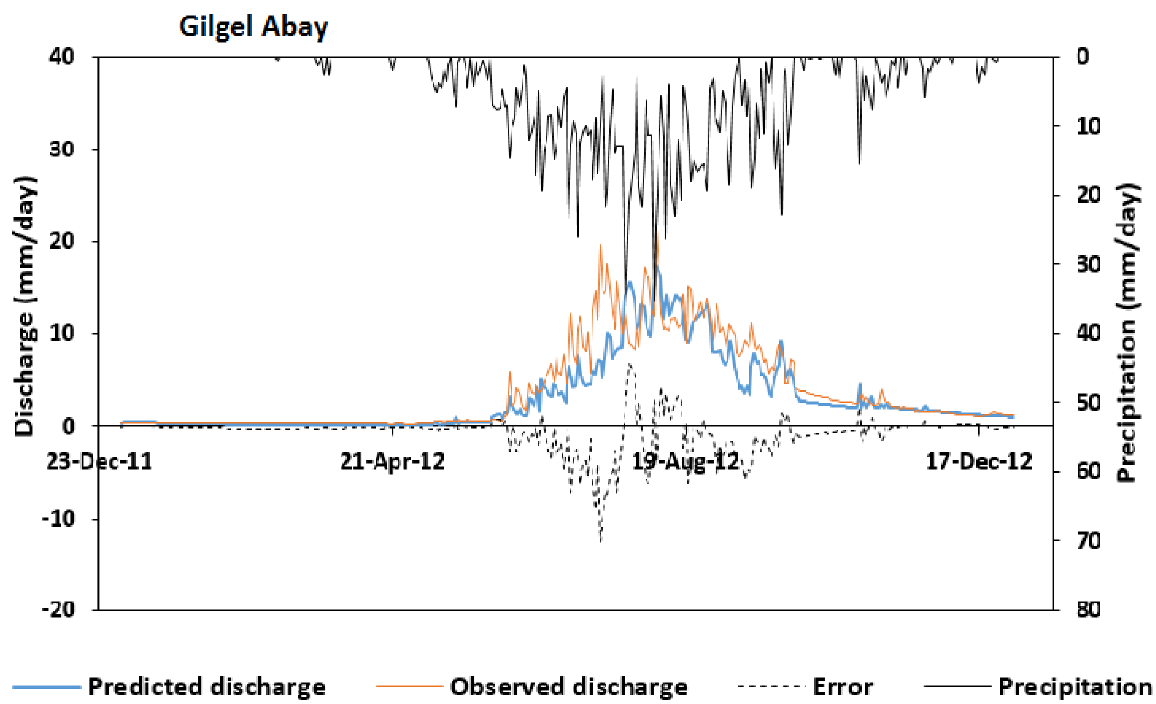


Fig. 6C. Comparison of predicted and observed discharge together with the error (difference of the two) and precipitation of the Gumara and the Gilgel Abay catchments for the calibration period

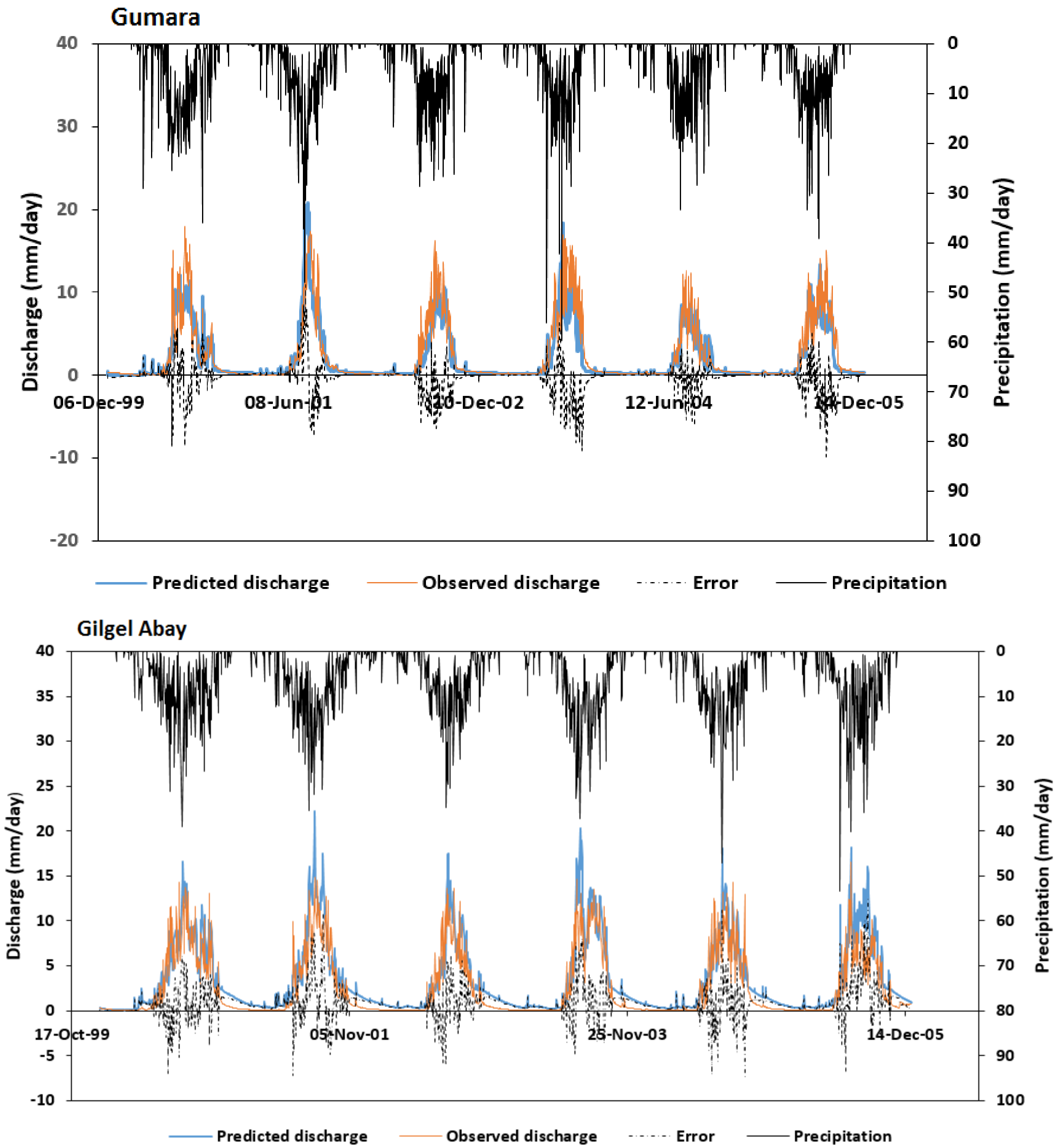


Fig.7C. Predicted and observed discharges together with the error (difference of the two) and precipitation of the Gumara and the Gilgel Abay catchments for the validation period.

5304, 14. I didn't follow what this meant. Which data are averaged over the year?

The modelled discharges appear to be less variable over time than the observed discharges. Therefore, the sentence on page 5304, line 14 in the paper is to explain

this. We used average daily rainfall data, average soil data (e.g. porosity, field capacity, and soil depth), average catchment characteristics data (e.g. slope, slope length) to mention some for the model inputs. Hence, this averaged condition may be one source of error such that the model may not exactly mimic extremes like peak discharges. We include these clarifications in the revised paper.

5305-5306. I don't see how these observations are meaningful given the errors in the model. There seem to be large errors in the flow peaks, so the model cannot be used as a basis for concluding upon importance of direct runoff.

Generally, the model performance and the model errors have to be explained based on commonly employed model performance indicators. In modeling practice, the usual practice is that if the model performance indicator results are above a recommended value (for example > 0.5 for NSE and R^2), then the model is taken as acceptable and model results are considered meaningful. Our modeling approach is not different from this. The authors used NSE, RMSE and R^2 in addition to the plot comparison of the observed and simulated outputs. The statistical results for NSE and R^2 were greater than 0.7, which shows the good performance of the model. The plots of the simulated and observed discharges of Figures 5 and 6 (in the paper) or Fig. 6C and 7C (above) can show this, but it is true that there are some deviations of the simulated discharge from the observed ones at some points in the time series. This is the limitation of the model. But still we understand that the model results are very important clues to understand the runoff processes in this data scarce region of the Upper Blue Nile basin and for the general water resources planning in the area. On Pages 5305-5306 in the paper, we discussed the runoff processes in the study areas based on the model results, but we do not dare to say that the results are absolutely perfect.

Figure 7. Local sensitivity analysis – value of this is unclear give high uncertainty in parameter values. Global analysis would be more useful.
5307 - Sensitivity analysis results support the view that the model is too complex; or at least components of it are too complex

The reviewer's idea is not very clear here. In one case, he (she) states as the sensitivity analysis is unclear, but in another case gives comments based on the sensitivity analysis. The optimal model parameters are obtained using the particle Swarm Optimization (PSO) algorithm, which performs global analysis. In Figure 7 (in the paper), we investigated the sensitivity of each model parameter when the parameter value is different from the optimal one, keeping the other model parameter values constant (equal to the global optimal value). Based on the comments of reviewer 1, we also made global sensitivity analysis and results are depicted in Fig.8C.

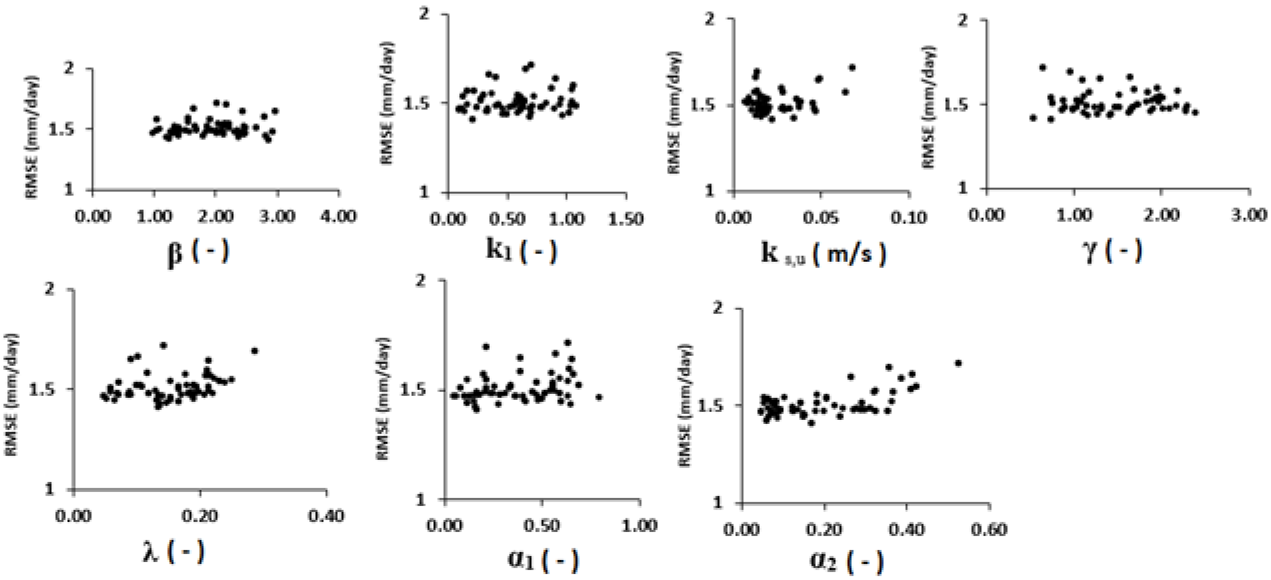


Fig.8C. Model parameter sensitivity analysis for Gumara catchment. Parameters are explained in table 2 in the Paper.

5307, 23. This is not an encouraging performance. Probably a two or three-parameter model could achieve this.

As shown in the paper, the results of NSE and R^2, for the direct parameter transferability test to other catchment were 0.58 and 0.6 respectively. The authors' suggestion of encouraging performance is based on these results. As it can be seen, the results are not bad. But the authors still stressed the need for further tests on similar catchments, as shown in the paper. We understand that various types of models with different number of parameters can be considered. Probably a two or three-parameter model could also give acceptable performance results, but such types of models are black box types and may not help for understanding the runoff processes in a particular catchment.

5309, 10. This conclusions is not justified from the results. The effect of the topographically-based division of the catchment has not been explored at all?

In the paper, the effect of the topographically-based division of the catchment is reflected mainly with respect to the input data to the model. Since the model was not a fully distributed model, it was necessary to use average catchment data. For this, we used topography as a proxy for the variability of most of the catchment characteristics like soil data (soil depth, porosity and field capacity) and undertake catchment classification. The explanations on page 5309, lines 9 and 10 in the paper are to emphasize this role of topography in the model. Moreover, we also showed the effects of topography on runoff and we obtained that hillslopes (medium and steep slope areas) generated almost no direct runoff as saturated excess flow. We will further elaborate the effect of the topographically-based division of the catchment in the revised version of the paper to make this clearer.

References

Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A statistical exploration of the relationship of soil moisture characteristics to the physical properties of soils, Water Resour. Res., 20, 82–690, 1984.

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: Understanding catchment behavior through stepwise model concept improvement, Water Resour. Res., 44, W01402, doi: 10.1029/2006WR005563, 2008.

Jothityangkoon, C., Sivapalan, M., and Farmer, D. L.: Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development, J. Hydrol., 254, 174–198, 2001.

Krasnostein, A. L. and Oldham, C. E.: Predicting wetland water storage, Water Resour. Res., 40, W10203, doi: 10.1029/2003WR002899, 2004.