Rebuttal Letter to manuscript hess-2016-290:

"Water budget modeling of the Upper Blue Nile basin using the JGrass-NewAge model system and satellite data"

Wuletawu Abera; Giuseppe Formetta; Luca Brocca and Riccardo Rigon

Dear Editor Professor Dominic Mazvimavi, and dear reviewers,

We would like to thank you for your comments and suggestions, which gave us the opportunity to improve the paper. In the revised manuscript (MS), we hope to solve all the issues raised. In this document we answer to all the reviewers questions . Comments are shown in bold font, followed by our answer/comment in normal font. The major corrections/changes in the manuscript are displayed between " ".

Editor's comment:

The Reviewers have submitted very detailed and important comments about this manuscript. The authors are encouraged to submit a revised paper that ADEQUATELY addresses the comments of the reviewers.

The revised paper will be referred to the referees to establish if all the comments have been adequately addressed.

Dear Editor,

We thank you for the comment given to our MS which obviously further improves the quality of our paper. In the revised manuscript (MS), we tried our best to address adequately the issues raised by the two reviewers.

Reviewers' comment:

Anonymous Referee #1:

This manuscript proposes a method to improve water budget modelling by using the available, but sparse, hydrometerological data and satellite products. The current manuscript provides a good try to predict hydrological process in data scarce regions or ungauged basins. Although there are publications related to such topic in ungauged basins, the intent of the manuscript is worthy and significant, and is of

interest to readers of HESS. Seeing the potential of this study, I am in general supportive of publication if the following comments are addressed in the resubmission.

Dear reviewer #1, we thank you for the general appreciation of our work, the comments and suggestions you give that helps to further improve our MS. In the following, your comments are answered one by one:

Major concerns:
1. I would encourage the authors to rewrite the methodology section. Give a clear message to the reader what you did and how you did. For example, the manuscript entitled as 'JGrass-NewAge model system'. However, I could not find detail or key information about the method. What's the theory of the method based on? What's the advantage of the method? The headings in method section are the same as those in section 5.

Regards to the JGrass-NewAge system, it is built on the object modeling system v3 (OMS3) informatics, which aims to deploy modern modeling solutions, with the philosophy of promoting reproducible research. The best way to have general information about it is the paper Formetta et al., 2014. JGrass-NewAGE is a collection of various modeling solutions for all hydrological compartments or fluxes. The detail of each component are presented and validated in various papers: rainfall-runoff modeling (Formetta et al. 2011), shortwave solar radiation modeling (Formetta et al. 2013), longwave solar radiation modeling (Formetta et al. 2016), and digital watershed modeling (Formetta et al. 2014b; Abera et al. 2014). We believe the level of details about JGrass-NewAge in page 4 and 5 are enough, but we revised the section for clarity. Here is the new paragraph about JGrass-NewAGE:

"UBN water budget is estimated using the JGrass-NewAGE hydrological system. It is a set of modelling components, reported in table 1, that can be connected at runtime to create various modelling solutions. Each component is presented in details and tested against measured data in the corresponding papers cited in the table 1. Similar study using JGrass-NewAge system, but using mostly in-situ observations, has been conducted in Posina river basin (northeast Italy), and the model performance is assessed positively (Abera et al., submitted). Brief descriptions on the components used in this study are provided in the following sections. In this study, the shortwave solar radiation budget component (section 3.3), the evapotranspiration component (Priestley and Taylor, section 3.3), the Adige rainfall-runoff model (section 3.4), and all the components illustrated in figure 2 are used to estimate the various hydrological flows."

Table 1. JGrass-NewAge system components and respective references. The components in bold are the ones used in this study.

Role	Component Name	Description
Basin partitioning	**GIS spatial toolbox and Horton Machine**	A GIS spatial toolbox that uses DEM to extract basin, hillslopes, and channel links for NewAge-JGrass set-up (Formetta et al., 2014a; Abera et al., 2014)
Data interpolation	Kriging, Inverse Distance Weighting, and JAMI	Interpolates meteorological data from meteorological stations to points of interest according to a variety of kriging algorithms (Goovaerts, 2000; Haberlandt, 2007; Goovaerts, 1999; Schiemann et al., 2011), Inverse Distance Weighting (Goovaerts, 1997)
Energy balance	**Shortwave radiation**, Longwave radiation	Calculate shortwave and longwave radiation, respectively, from topographic and atmospheric data (Formetta et al., 2013, 2016).
Evapotranspiration	Penman-Monteith, **Priestly-Taylor**, Fao-Evapotranspiration	Estimates evapotranspiration using Penman-Monteith (Monteith et al., 1965), Priestly-Taylor (Priestley and Taylor, 1972), and Fao-Evapotranspiration (Allen et al., 1998) options
Runoff	ADIGE (**Hymod**)	Estimates runoff based on Hymod (Moore, 1985) algorithm (Formetta et al., 2011)
Snow melting	Snow melt	Modelling snow melting using three temperature and radiation based snow algorithms (Formetta et al., 2014b)
Optimization	**Particle Swarm Optimization**, DREAM, LUCA	Calibrate model parameters according to Particle Swarm Optimization (Kennedy and Eberhart, 1995), DREAM (Vrugt et al., 2009), LUCA (Hay et al., 2006) algorithms respectively.

Regarding to the titles of the subsections in Methodology and in Results, the titles are the same because they refer to same water budget term (precipitation, evaporation, discharge, and storage, sequentially), and is given in both sections because we think the correspondence helps the understanding. We apologise, instead for the mistake we did in subsections' hierarchy, which is now corrected (please, see answer #12 to revier #2).

1.a. Some parts in the results analysis and discussion section are more suitable to be in the methodology section. For instance, it would be better to introduce the indices (i.e., KGE, PBIAS, r) in section 4.

It is true that goodness-of-fitness (GOF) indices can be in methodology section. However, since those indices are common in literature, maintaining their details in the main text is, in our opinion, distractive. That is the reason we decided to move description of the indices in the appendix section. However, we added a phrase that refers to the appendix also in the methodology section. This sentences in section 3.4 i.e. "The objective function used to estimate the optimal value of the parameter is the Kling-Gupta efficiency (KGE, Kling et al. (2012)). The KGE is preferred to the commonly used Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe (1970)) because the NSE has been criticized for its overestimation of model skill for highly seasonal variables by underestimating flow variability (Schaefli and Gupta, 2007; Gupta et al., 2009). For evaluation of the model performances, in addition to the KGE, two other goodness-of-fit (GOF) methods (percentage bias (PBIAS) and correlation coefficient) used in this study are described in Appendix A."

For validation statistics, the following sentence is added in section 3.6:

"and three goodness-of-fit (KGE, PBIAS, r) are used as comparative indices (for detailed information please see Appendix A)"

In addition, what's the spatial resolution of the HRU? When performing simulation, what are the time step and the spatial resolution of output?

The mean spatial resolution of the HRU is about 430 km^2 and we use daily time steps. This size is a trade-off between the resolution of the satellite data and the need to group some of them to have some statistical significance. The simulation results are therefore one for each HRU and at each time step of one day. The HRU estimates should be considered as a spatial average. Discharges however, are simulated at the nodes of the river networks. In the introduction section, the following phrases are added to better describe both the spatial and temporal resolution of the simulation:

"It obtains, at relatively small spatial scales and at daily time step, all the water budget components."

In addition, we have mentioned the number of subbasins used, and the mean ± standard deviation, as follows:
"In this study, the UBN basin is divided into 402 subbasins (HRUs of mean area of 430± 339 km2) and channel links, as shown in figure 1b."

"The index k = 1,2,3… is the control volume where the water budget is solved."

"The water budget components are estimated for each HRU and, subsequently, a routing scheme is applied to move the discharges to the basin outlet through the channel network." (section 3.1)

1.b. There are different hydrometerological data and satellite products, but it is difficult to readers to obtain their information (e.g., what kind of satellite products). I would suggest the authors providing a table to show all the data and their spatiotemporal resolutions. How did you deal with the different resolutions (especially spatial resolution) of input parameters?

A table was added as requested by the reviewer. The approach we followed on the description of the satellite products is to use a single 'best' satellite product, based, in the case of precipitation, on Abera et al., 2016. For the other water budget terms we were mostly constrained by products availability. Any product is described in the methodological section along with the description of the methods used to estimate the component. In summary, we used SM2R-CCI for precipitation, GLEAM for ET (but we have provided appropriate comparison with MODIS in supplementary material), in-situ hydrometer data for discharge (no other choice possible), and GRACE for storage change (no other choice possible). In the revised MS, the following table describing all the satellite products used in the paper and its spatial and temporal resolutions has been added at the end of the methodology section:

Table 1. Short summary of the list of remote sensing products used in this study.

Satellite products	Spatial resolution	Temporal resolution	Reference	used as
SM2R-CCI	0.25	daily	Brocca et al. (2014, 2013); Abera et al. (2016)	input for Precipitation
GLEAM	0.25 degree	daily	Miralles et al. (2011a); McCabe et al. (2016)	verification for evapotranspiration
MODIS ET (MOD16)	1-km	8-days	Mu et al. (2007, 2011)	verification for evapotranspiration
GRACE TWS	1 degree	30-days	Landerer and Swenson (2012)	Verification for storage change
CM-SAF	0.25 degree	daily	Schulz et al. (2009)	input for evapotranspiration component

The methods for processing and estimating the data at each HRU level are described in methodology section for each component (section 3). The reference spatial resolution for model inputs and validation is the area of each HRU. So, for each HRU, we estimate the weighted average of the quantity weighted by how much of the pixel area overlaps with the HRU polygon. For precipitation, this comments was already mentioned at page 6 line 11, while, in the revised MS, we have added the following sentence regards to ET:

"For comparison with NewAge ET, we estimated area weighted average GLEAM ET for each HRU polygon."

2. Discussion should be enhanced. What's the disadvantage of the method when applying in data-scarce regions with large area? For example, results of figure 5 indicated that the simulated runoffs were underestimated. What's the reason? Was it caused by uncertainties/errors in precipitation products? I could not find any quantitative information about errors of SM2R-CCI. Meteorological stations should observe precipitation, radiation, and etc. Why didn't you use them for validation and discussion?

Unfortunately, the meteorological stations seem not to provide any further information besides precipitation. It is true that the model underestimation is most likely due to the

underestimation of SM2R-CCI, as described on the page 11 line 29. Abera et al., 2016 by comparing with in-situ observations, shows that SM2R-CCI slightly underestimates the total cumulative rainfall in the study area. i.e. “Generally, the model predicts both the high flows and low flows well, with slight underestimation of peak flows (figure 5 a), which is likely due to the underestimation of SM2R-CCI precipitation data for high rainfall intensities (Abera et al., 2016).” Additional source of error can also be caused by model inconsistency due to averaging out input data over large areas.
This sentence is added in the revised MS: “Additional source of error can also be caused by model inconsistency due to averaging out input data over large areas”

3. The authors claimed that the JGrass-NewAGE system are described in a series of papers and not re-discussed in this manuscript. What’s the difference between this study and the previous papers? What’s the main contribution of this work?

The previous papers contain description of the single components that were validating separately on other catchments of small size where there was relatively abundant ground meteorological information. Those papers cover the informatics of the system, DEM treatment and river network schematization, and finally radiation, runoff, and snow modeling.
In this paper those components are linked in a unique modelling solution and work all together cooperatively to solve the water budget closure.
In addition, another important contribution of this paper is the application of the obtained modeling solution in a large basin using various data (satellite and in-situ), which is what NewAge was originally developed for.. In poorly gauged area, modeling in our opinion, working in this way is the only way to obtain spatially distributed water resource information that can be used reliably for management purpose.

Specific comments:
1. 1-21. ‘up to 2000 mm per year’. It would be much clearer by adding precipitation.

The point here is to emphasize that some parts of the Nile basin (i.e. parts in Upper Blue Nile and in the equators) receive 2000 mm per year, while others have insufficient precipitation. We rephrased:

“Most of the countries within the basin, such as Egypt, Sudan, Kenya, and Tanzania, receive insufficient fresh water (Pimentel et al., 2004). Exceptions to this are the small areas at the equators and the Upper Blue Nile basin in the Ethiopian highlands, which receives up to 2000 mm per year (Johnston and McCartney, 2010).”

2. 3-1. It should have space between 'given' and '('. The authors should proof read the manuscript to avoid such mistakes.

Space has been added; we removed such errors in the revised manuscript.

3. 3-6. 'the river enters a deep a canyon' contains grammatical errors.
Thank you for this, we corrected it. Now it is: "After about 150 km, the river enters to a deep canyon, and slowly changes direction to the south."

4. 3-18. The elevation values show certain difference compared to those in page 2 line 3.
Thanks you for spotting this. The one in page 2 line 3 was takes from literature value, and the page 3 line 18 was taken from SRTM digital elevation data. Since different values (small differences) were reported in various literatures, we used our SRTM value in both cases.

5. 3-30. It may mislead to conclude 'the seasonal variability of the basin is very high' because the authors claimed that the temperature has small seasonal variability.
We explicitly mentioned that the seasonal variability of precipitation (and evapotranspiration) is high at line 3-27, and that the variability of temperature is small at 3-28. Since it does not provide new information, we decided to remove this sentence.

6. 4-1. Figure 1. I suggest adding units for axes (also other figures) as well as enlarging the schematic map (at least the text). What does the colour represent in figure 1b?
We re-draw the figure to improve its clarity. The colors in figure 1b represent the mean elevation of HRU in the basin, which is now illustrated by a legend.

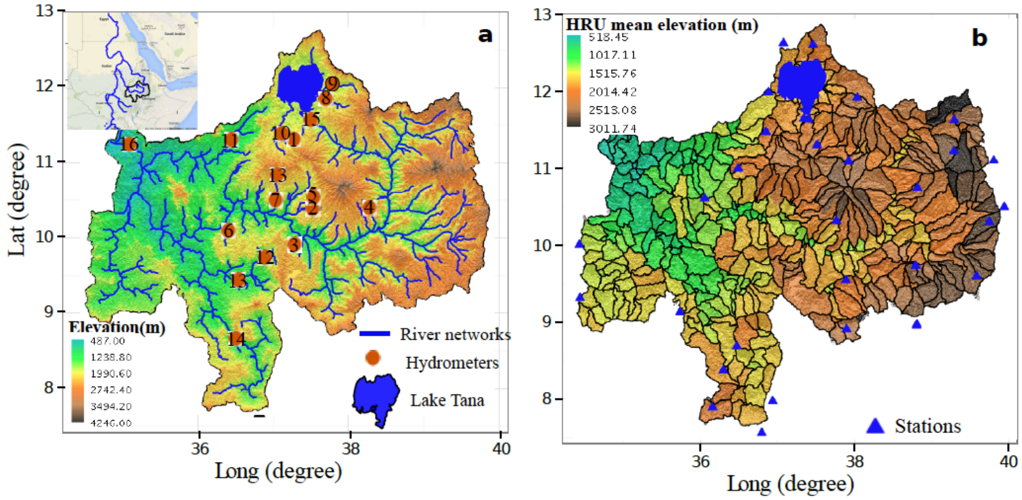


Figure 1. The Upper Blue Nile basin digital elevation map, along with the gauge stations (a); and subbasin partitions and meteorological stations used for simulation (b). Numbers inside the circles (figure a) designates the river gauging stations. The name of the basin referring to the numbers are provided in table 3.

7. 4-15. It seems that the citation appeared in the first time, and 2014b should change to 2014a. The authors should proof read the manuscript to avoid such mistakes.

The citations are in alphabetic order.

8. 5-4. What does GIS mean? Please consider defining the abbreviation.

Thank you for this. GIS refers to geographic information system. We have defined GIS in the revised MS.

9. 5-9. How did you divide the basin into 402 subbasins? According to what kind of rules? I'm not sure whether figure 1b is your results or not.

The partition of the basin into 402 subbasins is ontained by means of standard watershed partition techniques, and the specific procedures for JGrass-NewAge which are described in detail in Formetta et al., 2014 and Abera et al 2014. In the manuscript, it is also briefly presented at page 5 line 3 to 5 line. We revised the section as follows for clarity:

"The SRTM 90 m X 90 m elevation data is used to generate the basin Geographic Information System (GIS) representation. The basin topographic representation in GIS, as detailed in (Formetta et al., 2014a; Abera et al., 2014; Formetta et al., 2011), is based on the Pfafstetter enumeration (Formetta et al., 2014a; Abera et al., 2014). The basin is subdivided in Hydrologic Response Units (HRUs), where the model inputs (i.e. meteorological forcing data), and hydrological processes and outputs (i.e.

evapotranspiration, discharge, net radiation) are averaged. A routing scheme is applied to move the discharges from HRUs to the basin outlet through the channel network."

Yes, figure 1b is the subbasin partition results as mentioned in the caption.

10. 5-13. Figure 2 is difficult to read. The texts were small and difficult to guess their meaning. I suggest the authors redraw it.

We have increase the text font and thickness of the lines of the figure. In addition, we revised the text for clarity by removing some technical terms (such as .CSV, G.C, F.C), as follows:

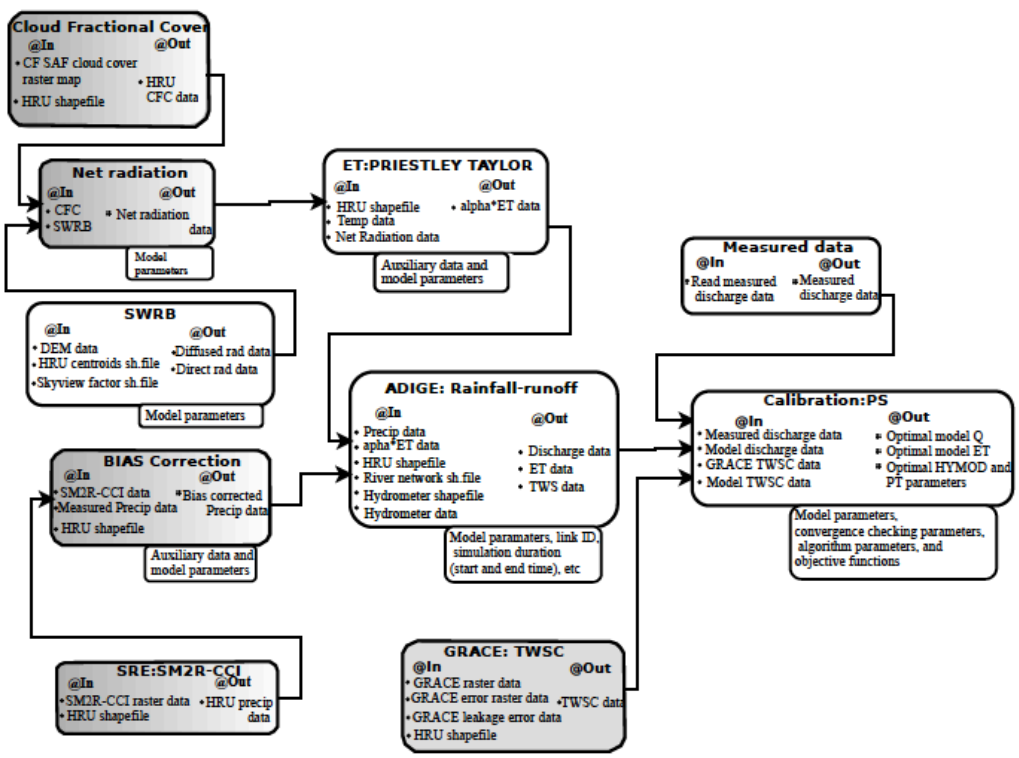


Figure 2. Workflow with a list of NewAge components (in white), and remote sensing data processing parts (shaded in grey, not yet included in JGrass-NewAGE and currently performed with R tools) used to derive the water budget of the UBN. It does not include the components used for the validation and verification processes.

11. 6-23. Works cited in a manuscript should be accepted for publication or published already. There are many publications describing psychometric constant.

We have replaced with appropriate citation (i.e. Brutsaert, 2005).

12. 6-27. What's the relation between S(t) and T_B in equation 3? Can you explain more?

There is no relation between S(t) and T_B, at least for what related to equation (3). S(t) is the water (storage) present in a HRU. Instead, T_B, the Budyko time, affects the alpha in equation (3), because the value of alpha is obtained for balancing the water budget (i.e equation (1)) in such a way that after T_B years the storage equals the initial one, i.e. S(TB) = S(0). This implies the use of an optimisation procedure, and such alpha is obtained together with the other parameters of the overall modelling solution (including runoff production, evapotranspiration, etc.) within the calibration procedure. Detail note on this is available to our under reviewer paper i.e. Abera et al. submitted (Advanced in Water Resources). To explicitly put some notes on relationship between s(t) and T_B, and description of the concept, we have added the following sentence and cited the paper under review as follows:

"In this procedure, given that S(t) is not measured, the assumption that there is null water storage difference after a long time, named Budyko's time, T_B , (Budyko, 1978), is required. So, here, what is searched is a time duration (T_B) such that the water storage assumes again the initial value (Abera et al., submitted). Once T_B is fixed, automatic calibration can be set to produce the set of parameters, including α_{PT} and Smax, for which, besides discharge is well reproduced, is also $S(T_B) = S(0)$. In this study, $T_B = 6$ years.."

13. 7-26. Semicolon should be replaced with 'and'.

Semicolon is replaced with 'and'.

14. 8-4. What does KGE mean? Please consider defining the abbreviation.

Thank you; in the revised MS we have introduced the KGE in the methodological section, as follows:

"The objective function used to estimate the optimal value of the parameter is the Kling-Gupta Efficiency (KGE, Kling et al., 2012)."

15. 8-8. What does 'described in A' mean? Does 'A' represent 'Appendix'?

Thank you, we have added Appendix before 'A'.

16. 9-18. It is curious to use J representing precipitation. In addition, precipitation, evapotranspiration, and discharge are components of water budget. Why did you use different section headings (i.e., 5.1, 5.1.1, 5.1.2, . . .)?

We adopted J for precipitation to be consistent with other papers of our research group (for instance, Rigon et al. 2016). Yes, there is error in the heading sections, and we revised to use the same level of heading for all the components.

17. 9-21. I would suggest the authors adding 'the Oromia region (or other mentioned places)' into Fig.1.
Thank you for this. However, we argue that the important idea here is to show the spatial pattern within the natural basin. We already verified that adding regional boundaries (information) makes figure 1 very crowded. It seemed better to us to delete the Oromia name from the text, as it is the only one mentioned.

18. 10-1. Figure 3a indicates precipitation is highest in southern region. However, figure 3b showed a different pattern (i.e., east shared highest precipitation), especially in JJA.
The two figures are different. Figure 3a shows the long-term mean precipitation as perceived by reviewer 1. Figure 3b, however, shows the level of percentage share of precipitation falls by seasons. In the east part of the basin, the highest percentage share (of its lower annual precipitation) falls in summer (JJA) in comparison to the other parts.

19. 11-4. How and why did you select only some subbasins? Did you consider r and PBIAS (figure 4, e.g., high r and low PBIAS, and low r but high PBIAS)?
We didn't consider r or PBIAS to select the subbasins. We select the three sub basins systematically to cover the basin spatial distribution: one from eastern, center, and western part of the basin. The following sentences has been added to clarify this:

"Figure 4 a shows the comparisons of the ET time series from 1994-2002 (aggregated at daily, weekly, and monthly, from top to bottom) between NewAge and GLEAM. The Figure specifically refers to three selected subbasins representing different ranges of elevations and spatial locations."

20. 11-10. 'while the it tends to' contains grammatical errors.
We removed 'the' from this sentence.

21. 11-23. 'within the basin at the internal channels (2)'. What does '(2)' mean?
It is changed to "(Table 2)" in the revised manuscript.

22. 11-27. I do not think r2=0.92 is lower than r=0.93 or r=0.94. I suggest the authors to unify the index.

It is very difficult to find similar index across all the papers. But, having PBIAS and r are relatively common, we decided to use r and PBIAS for comparison, in addition to KGE which is our primary index of model evaluation. Thank you for the comment, and here we convert the r2 index values report in literature in to r for unifying the indexes. We are also prudent to do comparison with other studies. So in this section, we just indicate the comparative performances:

"At the outlet, even during the validation period, the model is able to capture the dynamics of the basin response very well (KGE=0.92, PBIAS = 2.4, r = 0.93). The results show that the performances of the NewAge simulation are similar to the performances reported by Mengistu and Sorteberg (2012), with slightly lower PBIAS value (PBIAS=8.2, r =0.95)".

23. 13-1. Are all the parameters unitless? Why are two [−]? Furthermore, I could not find table 1 in the context.

The three parameters (with [-]) are unitless and for others it is length and time, which is given by [L] and [T] respectively in the table. Thanks for indicating the confusion between the two α[-]. In the revised manuscript the first and second α[-] has been changed to α_{hymod}[-] and α_{PT} [-] respectively. The following sentence has been added in the MS to refer to the table:

"The optimized parameters of the Adige model, obtained using automatic calibration procedure of NewAge, are given at table 3."

24. 13-2. Can you number the hydrometer stations and then add these IDs into figures 1b and 5?

Thank you we have labeled ID both in the figure 1, table 3 and figure 5 (please see the answer to specific comment 6).

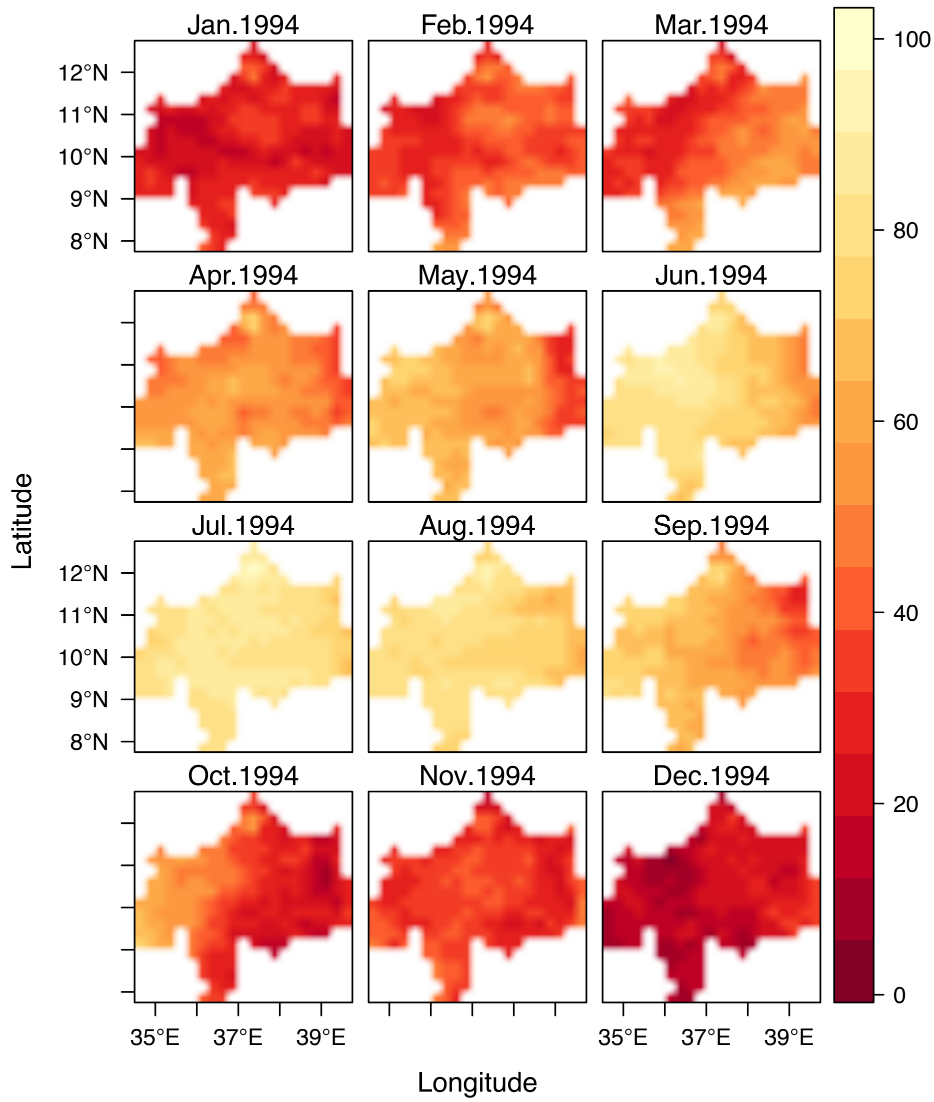
25. 14-8. Are Wase-Tana and FlexB commonly used models? Please consider defining the abbreviation.

It is true the two models are not common. We cited the papers where the models are described.

"Similarly, without calibration for the Gilgel Abay river, the NewAge simulation performance is better than the results of Wase-Tana (Wosenie et al., 2014, PBIAS=34)) and FlexB (Fenicia et al., 2008, PBIAS=77.6) or comparable to SWAT (PBIAS=5)."

26. 18-5. Can you provide some radiation, cloud, and wind observations? This may be better to draw the conclusion.

We don’t have observations of radiation, cloud and wind. We used JGrass-NewAge shortwave component to estimate the radiation data, together with the information of cloud fractional cover (CFC) from EUMETSAT Climate Monitoring Satellite Application Facility (CM SAF) project (Schulz et al., 2009). Wind data is not used at all in this study. It is true that including the radiation estimates and cloud data provides more insight to understand the conclusion given at this particular line. Providing spatial maps of these data in the manuscript, however, reduce its readability. Here are some samples (monthly mean for the year 1994) of the cloud cover map for the basin:



But also available at blog: http://ecohydrogeomorpho-metry.blogspot.it/2016/04/cloud-coveron-surface-net-radiation.html

27. 19-9. What does S mean?

We changed this into ds/dt.

28. 19-11. The number of decimal places was set to 3 for precipitation. Is it necessary? I suggest the authors unify the number of decimal places.
Of course it is not important. We removed all the decimal number throughout the paper.

29. 21-12. ‘figure’ should be ‘figures’.
We changed it to ‘figures’.

30. 26-6. ‘et al.’. The authors should list all the authors of a citation and unify the citation style. The authors should proof read the manuscript to avoid such mistakes.
We corrected this and other citation errors.

31. Texts of most of the figures are unclear. I would suggest the authors redraw the figures.
In the new manuscript, we improved the figures for clarity.

Anonymous Referee #2:

GENERAL AND IMPORTANT COMMENTS ABOUT THE MANUSCRIPT

The Manuscript (MS) is an attempt to integrate various sources of satellite remote sensing data towards macro-scale hydrologic modelling in a region in Africa. Such a concept is novel considering the eminent data limitations pertaining to lack or limited observed in-situ hydro-meteorological data important for model calibration and validation purposes. In this study, the authors seem to be interested in historical cases of the water budget, and hence may elect to put this is the title, or justify why they are not interest in forecasting. From the present standpoint, however, the paper can be considered for publication in the near future, but only after addressing some serious technical issues that degrade the novel concept proposed and applied by the authors. In this respect, and to improve and make the MS much better, I wish to recommend major revisions before further consideration. The following are some of the major comments that need readress:

We thank reviewer #2 for the appreciation of our work. When performing our studies we analyzed historical data, as any other hydrological study. We are, obviously, interested in forecasting the hydrological cycle components, but this necessarily relies on the availability of the meteorological forcings. It is possible to forecast (in the sense of meteorology) discharges (for instance) if we have rainfall (and other meteorological) data. This assumes that we have access to real time data in the basin, which we do not have. More relaxed forecast, or better, projection, could be made after acquiring appropriate climate projections. But for this, to have a model system which is validated for a given basin is the first step. This is actually one of the goals of the present paper. However, we used as much as possible the suggestions given by the reviewer to improve our new manuscript.

Major concerns

(a). Language Limitation: the MS is poorly written and generally very difficult to read right from the abstract to the conclusions. This may be due to language limitation/culture of the authors, but considering that the MS will have a bigger readership; it would be nice to English edit the MS so that the actual intentions-technical and linguistic-can come out clear. The way the results, especially the statistics and maps, are presented makes one question the objective of the work. In some cases, it is difficult to understand it the authors intend a comparative assessment at various spatial scales of the regions in the basin? There is also the random use of difficult expressions appearing from nowhere without prior definition, i.e. in defining the table in page 15, he used Figure 5, Table 2 which is difficult to understand.

We used all the suggestions of the two reviewers, and revise the manuscript accordingly. In page 15 there are not Tables. There are Tables in page 13, and we assume the reviewer refers to them. In the revised MS, we modified the introduction to emphasize the objective and novelty of the study, and the figures are revised for clarity.

(b). the author claim that his research is motivated by data limitation. However, he seems to have some stations with streamflow data as by the hydromet stations in the study area map or otherwise, the hydrographs used in the validation exercise. This begs the question: So where is the boundary of this data limitation he is claiming?

Data limitation does not mean total absence of data. Certainly we have some precipitations and discharge data. However these data are in 35 locations for precipitation data in an area of 175 thousand square kilometers. Meaning, just a station every 5000 square kilometers or areas of around seventy by seventy square kilometers of side (on average). Convective processes generating precipitation can be as small as 10 kilometers square, so the optimal gauge network distribution should be as small as that, to capture all the relevant phenomena. Considering this fact, almost any region in the world is data-scarce, but some regions such as the Upper Blue Nile basin are even more data-scarce regions than others. For discharge analysis, the numbers of hydrometer stations are very few (16 hydrometers) with a data set having lots of missing data and gaps. So for the objective outlined in the study, the estimation of spatially and temporally hydrological information of the basin, UBN surely can be characterized as data limited basin.

Could it be possible to use the available data to parameterize the model and later regionalize the model? Or is it possible to develop criteria to extrapolate the results after calibration and validation of the satellite estimates with the limited but available observed data-sets?

Yes, this is actually what it was done. We use all the data available in a period to calibrate the model and we modeled all the data (hydrological information) by means of NewAGE in the inner points. Actually, if with regionalisation the reviewer means statistical techniques, we did not use any of them. If the reviewer asks for the transferability of our approach, we can confirm that it can be extrapolated to any basin with similar or larger size.

The authors may also need to justify why 402 sub watershed were delineated considering the limited river gauging stations shown in the study area map.

Even if hydrometeorological data are available in fewer stations, satellites allow us to have rainfall forcing at a much finer scale. Partition of the basins in 402 parts is functional to use all the rainfall spatial information we have, in a trade-off with a reasonable computational demand. It also serves to accounts for the morphological structure of the river network, which, obviously counts very much in forming the hydrologic response. On the latter topic, the last author co-authored some papers that can support this fact.

If he wants to retains them, then he should define use a criteria to choose at least 10-15 sub-catchments and provide their morphometry together with the simulated values of the water balance components in the results section, for consistency and clarity. A table (and not maps) in this respect would quickly help things out here.

If we did not clearly communicate the objective of the paper, obviously, it is our fault. However, the objective of the paper is to estimate spatio-temporally distributed water budget of the UBN basin. Hence, the methodology followed and the results presented are for the whole basin, not for only some specific sub-catchments. When in-situ data is available, that specific sub-catchment is used to verify the performance of the model estimations. In other words, to assess the discharge predictive capacity of the model, those subbasins with observed discharge data are selected (about 16 subbasins), and GOF indexes are presented a table (Table 3). But for the rest of the analysis, we wanted to do water budget closure for each subbasin in the whole basin.

c. Considering data uncertainties, would it be wise to believe the higher model reliability and hence results?

We considered ground measures as true. The data provided by the model solution we used show that there is consistency between discharge gauges and rainfall estimates, and the model works satisfactorily also for the validation periods. Model and data are consistent (once the model is calibrated). Abera et al (2016) tried to answer the question of the reliability of the satellite rainfall data comparing with in-situ data. We agree with the reviewer suggestion, and added the following sentences in the conclusion section:

“Despite the good results obtained, it is important to note that this study is limited by the lack of in-situ ET observation and low resolution GRACE data for confirmation of storage. To these regards, the results of this study would benefit from basin specific assessments of ET and ds/dt RS products based on ground measurements, as done in Abera et al. (2016) for precipitation."

The authors need a good and elaborate justification of how the errors cancelled out during the simulation.

Errors do not cancel. When possible, any of the modelling components used was validated separately. We have checked the functioning of each of them in many other cases, as testify by our own literature (as already detailed for the reviewer #1), even if in those cases data were less scarce. In this specific case, precipitation from satellites is verified and corrected using the available few in-situ observations, storage (at least at the whole basin scale) is verified using GRACE data, discharge is verified at about 16 hydrometer stations. So we know that each component, besides implementing sound science, works fine with the appropriate data. That is what we can trust. When we calibrate hydrological model just on discharge data, parameters' values become a collector of uncertainties (a garbage collector, as some colleague calls it), but we assume that this is well understood and does not require a further disclaimer.

Furthermore, the author seems to be using some part of the available data for calibration, and the same half plus the rest within the time frame for validation.

We don't. We used some part of the available data to calibrate the model at the main outlet, and used the other part for validation. In addition, the other data sets available in the interior hydrometer stations are used for validation the model capability to estimate discharge at each links of the river network of the basin. This is clarified in section 3.6, as follows:

"At the basin outlet (Ethiopia-Sudan Border), the ADIGE rainfall-runoff component (i.e. HYMOD model) is calibrated to fit the observed discharge during the six years of calibration period (1994-1999) at daily time steps."

"Discharge simulation is validated for separate time-series data at the outlet at Ethiopia-Sudan Border, where the model is calibrated. In addition, the simulation of NewAge at the internal links is validated where in situ data are available. The evaluations at the internal links provide an assessment of model estimation capacity at ungauged locations."

In my opinion, the conventional way would be to divide the data-sets into two, one for calibration and the other for calibration.
Correct! That is what we did.

Could this be the reason for the good efficiency realised? The authors need to justify this methodology very strongly.
As we said, we did not use the same data for both validation and calibration. Hence, we believe that the reason for good model performance is due to the explicit characterization of inputs component and the goodness of the modeling solutions adopted.

(1) TITLE

1 - The title is okay and acceptable, but may sound better if the authors consider the conventional way of staring a sentence with a verb i.e. Modeling/Estimation/Assessing of the Water Balance etc. This is however trivial at this moment.
We agree with the reviewer. We changed the title to: "Modelling the water budget of the Upper Blue Nile basin using the JGrass-NewAge model system and satellite data"

(2) ABSTRACT

2 - In my opinion, the first sentence can be made simple and realistic i.e. . . .by saying the region is one of the data scarce regions is the developing regions (but not in the world as this raise a lot of questions and may temp one to ask for proof of review in the introduction. Are there basins in the UNRB that have data? Is the justification of one of the data scarce regions in the world thus still valid?

Yes, we have changed it: "The Upper Blue Nile basin is one of the most data-scarce regions in developing regions."

In my opinion, the water budget components of study can be explicitly mentioned in the sentence without the brackets, and the tools used well captured and summarized. This makes the section clear and easy to read. Considering that modeling procedure employed, and the possible uncertainties involved, the results need to be rounded off i.e. by saying that precipitation values between 1000-1600mm were estimated depending on seasonality etc. Generally, the abstract can be well written and summarized in good English language, and only important content.
We revised the sentence as follows:
"In this study we develop a methodology that can improve the state-of-art by using the available, but sparse, hydrometerological data and satellite products to obtain the estimates of all the components of the hydrological cycle (precipitation, evapotranspiration, discharge, and storage)."

We presented the uncertainty by mean plus/minus (i.e. for precipitation we used 1360 ± 230), and we prefer our to represent the uncertainties and long-term annual mean value.

(3) INTRODUCTION

3 -This section can be language edited and the phrases backed with the latest references. The references also need to be ordered either from the latest to the oldest or vice versa as required by the journal.

The following sentence is taken from the journal authors' guidelines, and states that citation can be ordered based on relevance, and that is what we followed:

“In terms of in-text citations, the order can be based on relevance, as well as chronological or alphabetical listing, depending on the author's preference.”

4 - In my opinion, the text in lines 4-10 can be summarised and well captured within the text without using bullets or points.
In the revised manuscript, we tried to synchronize them in shorter sentences.

5 - Lines 27-28: the sentence beginning with [The use of RS precipitation products...] can be well written, more content added and justified. Here the authors can introduce and justify the use of other products such as GLEAM, MODIS data products etc for simulation. The author seems to neglect this section/paragraph and YET it forms the basis of their novel idea of using RS for data scarce regions. In my opinion, ‘at least two paragraphs’ on this section should be added to improve and justify his methodology where he has introduced a lot of RS products from nowhere. For instance, how have these RS tools and methods been applied in other regions of data scarcity? What were the results achieved? Can the methods be replicated in the current study basin? Has the JGrass NewAge (JGNA) model been applied elsewhere and what were results and strengths etc? This section should a major part of the MS and if not well captured then it can be concluded that the MS contributes very little value to hydrological science.

We wanted to avoid the description of various remote sensing (RS) products, and instead suggest that the readers should look for this information in the appropriate papers about the use of RS for hydrology that we cited better in the revised manuscript. However, a review of the overwhelming number of applications of various RS products for hydrology is not the subject of the paper. The justification of the particular remote sensing data for a particular component is explained it the respective section. For instance, the justification as to why we used SM2R-CCI for precipitation is given in detail at section 3.2; the GLEAM for evapotranspiration is given at section 3.3 etc. But, we accept that the general comment on the use of RS for water budget modelling and its prospect can be commented at this section. Hence, in the revised MS, the following paragraph is added:

“To overcome data scarcity, large scale hydrological modelling can be supported by remote sensing (RS) products, which fill the data gaps in water balance dynamics estimation (Sheffield et al., 2012). For instance, a considerable number of researches has been carried out in the last two decades in developing satellite rainfall estimations procedures (Hong et al., 2006; Bellerby, 2007; Huffman et al., 2007; Kummerow et al., 1998; Joyce et al., 2004; Sorooshian et al., 2000; Brocca et al., 2014).
RS is also a viable option to fill the gaps for basin scale evapotranspiration estimation. Global satellite evapotranspiration products have been available by applying energy balance and empirical models to satellite derived surface radiation, meteorology and vegetation characteristics, and they are recognised to have a certain degree of reliability (e.g. Fisher et al., 2008; Mu et al., 2007; Sheffield et al., 2010).

Basin scale storage estimation is the most difficult task. Fortunately, the Gravity Recovery and Climate Experiment (GRACE) (Landerer and Swenson, 2012) came to fill this gap (e.g. Han et al., 2009; Muskett and Romanovsky, 2009; Rodell et al., 2007; Syed et al., 2008; Rodell et al., 2004). Guntner (2008), Ramillien et al. (2008) and Jiang et al. (2014) reviewed the use of GRACE data and positively recommended it for large scale water budget modeling. At the moment, satellite based retrievals of discharge are not available as operational or research products, but, potentially it can be retrieved from satellite altimetry and multispectral sensors (e.g. Tarpanelli et al., 2015; Van Dijk et al., 2016). Moreover, the Surface Water Ocean Topography (SWOT, Durand et al. (2010)) mission, which is expected to be launched in 2020, will provide river elevation (with an accuracy of 10 cm), slope (with an accuracy of 1 cm/1 km) and width that can be used in estimating river discharge (Paiva et al., 2015; Pavelsky et al., 2014).

Notwithstanding the availability of these RS products at various (spatial and temporal) resolutions and accuracy, their use is clearly a new paradigm in water budget closure estimations (Sheffield et al., 2009; Andrew et al., 2014; Sahoo et al., 2011; Gao et al., 2010; Wang et al., 2014)."

In the same mood, we do not want to add much information about JGrass-NewAGE that can be better accessed in previous papers by the same authors. The details provided in section 3.1 seem long enough to describe the model system. Regarding to previous applications of JGrass-NewAge, the following sentence has been added in the revised MS, at section 3.1:

"Similar study using JGrass-NewAge system, but using mostly in-situ observations, has been conducted in Posina river basin (northeast Italy), and the model performance is assessed positively (Abera et al., submitted)."

(4) THE STUDY AREA

6 - There are loose statements here and there that can be tightened and generalized. For instance, in line 5, one would ask: where is Bahir Dar where the river originates? Such loose statements assume and make the MS only fit for regional publication. In my opinion, one elaborate map of topography (DEM), river network and stream gauges can be sufficient here. I am also sure with good GIS skill, and added topological data, the rainfall stations can still be added without making the map look untidy and congested. Or else, he may also elect to take a map of the catchment delineations and the rainfall stations in the methodology, and use that chance to highlight the subcatchments…

Thank you, we improved figure 1. As suggested by the reviewer, we dedicated one map describing the DEM, river network, and stream gauges, with stream gauge stations labeled by ID number. Since the sub basins are the scale at which the water budget is estimated, we maintain this map along the former.

7 - (better more than 10) where he wants to focus his results using a table as

mentioned above already.

We do not think that adding more catchments' details is useful for the readability of the paper. However, DEM, important shape files to be used in GIS, and the list of catchments details is provided as supplementary material.

(5) METHODOLOGY

8 - On page 4 lines 12-15, the authors may want to choose one or two more applicable references of the co-author.

The lists of papers cited are describing different modeling solutions, each for one component of the JGrass-NewAge system. Since all components are used, it is important that we cited all of them. However, the sentence has been revised (see major comment #1 of reviewer #1).

9 - In page 5, Figure 2 needs simplifications and better explanations. The color coding shades used will not appear if the paper is printed in black and white.

Thank you, we improved the text and shadings.

10 - Some parts in section 3.2.1 ideally belong to the introduction. Let the authors focus on the data-sets used and why they were used.

Actually what has been written in the first and second paragraph was the explanation why and how we used SM2R-CCI precipitation data. Please see the answer for comment 5.

11 - The reference Abera et al., submitted is completely out placed and may not be necessary at this stage of the journal.

Since it contains similar efforts, with more details on the foundations of water budget closure studies using hydrological model, but using in-situ observations, it is helpful to cite this paper. In addition, the paper is revised and resubmitted.

12 - There are many good ways of structuring this section in hydrology. Let the authors develop a simple and flowing structure from section 3.1. For example, section 3.1 can be titled 'Data and Methods'. Section 3.1.1 can be on 'Water Balance Modeling'. Section 3.1.2 can be on 'The Modeling System'. Section 3.1.3 can be on 'Data and Modeling Procedure' etc. The authors are free to choose what structure they want to adopt. As it is at the moment, there is too much information everywhere, a majority of which is not well captured and explained.

We realized that sub-sectioning of section 3 and 4 went wrong. New subsections are:
3 Methodology
3.1 JGrass-NewAGE System setup
3.2 Precipitation
3.3 Evapotranspiration

3.4 Discharge
3.5 Water storage
3.6 Calibration and validation

4. Results and discussion
4.1 Precipitation
4.2 Evapotranspiration
4.3 Discharge
4.4 Water storage
4.5 Water budget closure

5. Conclusions

We think that in this way there is a clear relation between the topics of the two sections (section 3 and 4).

13 - Some content in section 3.2.3 on page 7 are not necessary and can be avoided generally.

Section 3.2.3 contains totally twelve lines. It is very difficult for us to understand what we can avoid to say. We give information about the algorithm we use for reproducing discharges, and the validation method. We believe that this information is necessary.

14 - Section 4 on calibration and validation can be renamed as section 3.2 and well elaborated as explained before. In this section, the authors need to JUSTIFY WHY the same data period used for calibration is also available for Validation? This may infer a technical limitation that can affect the model results purported by the authors.

Regarding about section renaming, please see specific comment 12. We did not use the same data for calibration and validation, as described in major comment C.

6. RESULTS AND DISCUSSION

15 - Generally, the results are not balanced and well presented. The spatial maps dominate all the results. Well structured tables may provide more information considering the many catchments of study.

Depends on the objective of the paper, the deliverability of the results need to be based on the maps. We think that one figures convey more than thousands words if well understood. Evidently we were not able to convey clearly their meaning. We have worked to improve figure captions and comments.

16 - The first paragraph in the results section may not be necessary, or better be summarized.

Thank you, we summarized it as follows:

"The results of the study are organized as follows: firstly, we present the results for 1) precipitation, 2) evapotranspiration, 3) discharge and 4) total water storage; secondly, the JGrass-NewAGE system is used to resolve the water budget closure at each subbasin, and the contribution of each term water budget term is further is analyzed."

17 - The authors should find a way of presenting the maps in a nice, simple and clear manner. As they are at the moment, the polygons dominate the results. An elaborated table with selected catchment justified in the methodology can be good enough. Only one or two maps can be used here for visualization and overall balance of presentation of the results.

Given our objective, the presentation of our results without maps is impossible. We limited one, if not two, figure (plot) for each component. Data are averaged over a subbasin and there is not internal spatial variability in the output. So it is clear that "polygons" stand out.

18 - In line 23-24 of page 9, is the discrepancy small as mentioned? Could it be that the SM2R-CCI was not properly corrected? Please explain into details.

The difference between annual long-term rainfall value of 1900 mm and 2049 mm, given by different data sources, can be considered small. Besides, if one considers the uncertainty pertinent to each data sources and estimation method, s/he should conclude that the difference is acceptable.

19 - The legend for Fig 3 needs to be well placed and elaborated.

We revised the legend and the caption were improved.

20- In section 5.1.1 of page 11, there is need for technical justification by the authors as this is a very strong section of hydrology. (i) If GLEAM has had validation in other areas, with a good match with observations, then I it would be ok to use it for plausibility checks. However, as it stands, the New Age simulation of ET highly over- or under-simulate the ET fluxes. Should the results thus be fully trusted with these graphs?

The detail information about the GLEAM is provided in the methodological section (page 11 line 17 to 27). GLEAM had several checks: "The performance of GLEAM is assessed positively in different studies (McCabe et al., 2016; Miralles et al.,2011b).

The literature checks of the product was not for a given area and were not based on hydrological modeling accurate as our. Hence we would not say that NewAGE over or under estimates the budgets. This would assume that GLEAM is the truth. As mentioned in the methodological section, both of them are estimates, which differ but are somewhat coherent. NewAge results also depend on various other inputs. However we: assessed rainfall inputs (in another paper), check the consistency of the water budget components (such that mass is conserved) , check the consistency of data and model outcomes. Therefore we are sure that our results are quite robust in comparison with previous ones, including GLEAM's.

21 - The author can elect to present one or two of the Graphs/Figures but well elaborated and discussed into details. As it is, figure 4(b) is of limited value and would rather be discussed in the text or annexed.

The whole paragraph (i.e second paragraph of section 4.2) is all about figure 4b, and we believe that it constitutes a sufficient comment. However, we revised the text as follows:

"The agreement/disagreement between the two estimations varies from subbasin to subbasin (figure 4). The spatial distribution correlation and PBIAS between the NewAge and GLEAM ET is presented in figure 4 b. Spatially, the correlation between JGrass-NewAGE and GLEAM is higher in the eastern and central parts of the basin, while it tends to decrease systematically towards the west (i.e. to the lowlands, see figure 4 b). The correlation between the two ET estimations increases when passing from daily to monthly time steps. The PBIAS between the two estimates ranges from -10% to 10%, with large numbers of subbasin being from -3% to 3%. Spatially, the comparison shows that GLEAM overestimates ET in the western parts of the basin (border to the Sudan) and underestimates ET in the northern parts of the basin (figure 4b). The overall basin correlation is 0.34 ± 0.07 (daily time step), 0.51 ± 0.08 (weekly time step), and 0.57 ± 0.10 (monthly time steps). Generally, except at daily time step, the two estimates have acceptable agreements (very low bias, and acceptable correlation). However, in comparison with the correlation (0.48 ± 0.15) and PBIAS ($14.5 \pm 18.9\%$) obtained between NewAge ET and MODIS ET Product (MODET16), as shown in the supplementary material, the correlation and PBIAS between NewAge ET and GLEAM ET is much better."

22- Considering the model/data uncertainties, a KGE of 93% may be theoretically high if not good enough. There is hence a need for a strong justification of how the errors cancelled out during calibration and validation.

The modeling components were tested separately from the whole, when possible. So rainfall estimation was estimated with rainfall measurements (we dedicated a paper to this). Storage was estimated against GRACE data, and so on. We do not believe that model/data uncertainties cancel each other. A better hypothesis is that the calibration procedure is able to mask systematic measurement errors.

23 - Fig 5 is not well represented. This can be avoided or the authors can choose the sub- catchments to illustrate 'a prior in the methodology section' as mentioned already. The challenge here is that with the many sub catchments, the author does not seem to know how to cluster them in a consistent manner throughout the paper.

We agree that we need to explain better what is shown in Figure 5. It seems that we did not clearly show what we wanted. We modeled daily discharge at all river links of the basin for 16 years. The results were presented in two ways: (1) Time series simulations at few links of the river network where we have observed discharge to compare with.
These comparisons are shown in the river network map to visualize the locations of these links within the basin (i.e. figure 5). The names of these locations are given in the caption, and information about them is also given in Table 2. (2) Figure 6, now moved to the supplementary material, presents a snapshot of discharge estimates for any river links of the basin. We tried to improve the Figure caption to help better the reader understanding

24 - The results on page 14 can be summarised and well written. On table 2, is the final outlet of Upper Blue Nile located at El Diem with an area of 174 000km2?
No idea!
We revised the section. Yes, it is the outlet of the basin. We have added a column to the table that connects the table with the spatial location in figure 1.

25 - Fig 6 on page 15 needs to be elaborated and well explained or else moved to the annex.
We moved figure 6 to the supplementary materials, as it does not provide any comparison or statistics with observations. However, it shows how we can obtain the discharge at each links.

26 - On page 16, it would be good to justify how the discharge in the entire basin was modelled. I.e. did you add/route all the upstream discharges and accumulated downwards? This as a technical consideration for the paper.

Thank you for this. In the methodological section (section 3.4) and the following sentence has been added to explain how the discharge routing is modeled:

"The NewAge Hymod component is applied to any HRU, in which the basin is subdivided and the total watershed discharge is the sum of the contribution of each HRU routed to the outlet."

27 - All the results needs to be discussed from a hydrological standpoint. This section is important for the authors to justify the publication, and provide key element of study that improves the knowledge in hydrology in such areas generally.

Thank you for the suggestions you gave all through the paper. We used all of them to improve the paper.

7. CONCLUSIONS

28 - The paper needs to be summarised in the context of the study. Considering the uncertainties, the results need to be reported with this recognition i.e. ET values between 650-750mm were estimated for various sections of the basin etc

There is need for more conclusions about the challenges of the study and the methods generally. This will form a basis for recommending future studies in areas with similar data limitation.
As it is, the section is completely lacking and does not provide future research directions in hydrology.

We revised the conclusion section being more specific on our results and uncertainties, and remarking the challenges we met in our studies (see the marked-up MS). However, we do not take responsibility to indicate future research directions. In our opinion we already show something that is a little beyond the state of art of the discipline. These improvements include the use of various satellite sources for verifying and/or assessing all the water budget terms, and the production of the same water budget at various time scale, verifying mass conservation through the cycle. Besides, we produced the software to obtain it, we made it available, and everybody can replicate our results.

8. REFERENCE
29- The references are not formatted to the Journal requirements as required by HESS. Check and realign all of them.
References formatting have corrected accordingly.

References:

Abera, W., Antonello, A., Franceschi, S., Formetta, G., and Rigon, R.: The uDig Spatial Toolbox for hydro-geomorphic analysis, British Society for Geomorphology, London, UK, in: clarke nield (eds.) geomorphological techniques (online edition) edn., 2014.

Abera, W., Brocca, L., and Rigon, R.: Comparative evaluation of different satellite rainfall estimation products and bias correction in the Upper Blue Nile (UBN) basin, Atmospheric Research, 178-179, 471-483, doi:1 2016.

Formetta, G., Mantilla, R., Franceschi, S., Antonello, A., and Rigon, R.: The JGrass-NewAge system for forecasting and managing the hydrological budgets at the basin scale: models of flow generation and propagation/routing, Geoscientific Model Development, 4, 943–10 955, 2011.
Formetta, G., Rigon, R., Chávez, J., and David, O.: Modeling shortwave solar radiation using the JGrass-NewAge system, Geoscientific Model Development, 6, 915–928, 2013.

Formetta, G., Antonello, A., Franceschi, S., David, O., and R., R.: The basin delineation and the built of a digital watershed model within the JGrass-NewAGE system, Bolet?n Geol?gico y Minero: Special Issue "Advanced GIS terrain analysis for geophysical applications, 2014a.

Formetta, G., Antonello, A., Franceschi, S., David, O., and Rigon, R.: Hydrological modelling with components: A GIS-based open-source framework, Environmental Modelling Software, 55, 190–200, 2014b.

Formetta, G., Kampf, S. K., David, O., and Rigon, R.: Snow water equivalent modeling components in NewAge-JGrass, Geoscientific Model Development, 7, 725–736, 2014c.

Mengistu, D. and Sorteberg, A.: Sensitivity of SWAT simulated streamflow to climatic changes within the Eastern Nile River basin, Hydrology and Earth System Sciences,16, 391–407, 2012.

Rigon, R., Bancheri, M., Formetta, G., de Lavenne, A. (2016). The geomorphological unit hydrograph from a historicalâ˘AR˘ critical perspective. Earth Surface Processes and Landforms, 41(1), 27-37.

Modelling the water budget of the Upper Blue Nile basin using the JGrass-NewAge model system and satellite data

Wulelawu Abera[1], Giuseppe Formetta[2], Luca Brocca[3], and Riccardo Rigon[1]

[1]Department of Civil, Environmental and Mechanical Engineering, University of Trento, Italy
[2]Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, UK
[3]Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, Italy

Correspondence to: Wuletawu Abera (wuletawu979@gmail.com); Riccardo Rigon(riccardo.rigon@unitn.it)

Abstract. The Upper Blue Nile basin is one of the most data-scarce regions in ~~the world~~developing countries, hence, the hydrological information required for informed decision making in water resources management is limited. The hydrological complexity of the basin, tied with the lack of hydrometerological data, means that most hydrological studies in the region are either restricted to small subbasins where there are relatively better hydrometeorological data available, or at the whole basin scale but at very coarse time scales and spatial resolutions. In this study we develop a methodology that can improve the state-of-art by using the available, but sparse, hydrometerological data and satellite products to obtain the estimates of all the components of the hydrological cycle (precipitation, evapotranspiration, discharge, and storage). To this scope, we use the JGrass-NewAGE system and various remote sensing products. The satellite products SM2R-CCI is used for obtaining the rainfall inputs; SAF EUMETSAT for cloud cover fraction for proper net radiation estimation; GLEAM for comparison with estimated ET; and GRACE gravimetry data for comparison of the total water storage amounts available. Results are obtained at daily time-steps for the period 1994-2009 (16 years), and they can be used as a reference for any water resource development activities in the region. The overall long term mean budget analysis shows that precipitation of the basin is 1360 ±230 mm per year. Evapotranspiration covers 56% of the yearly budget, runoff is 33%. Storage varies from minus 10% to plus 17% of the budget.

Key Words: Water budget, Upper Blue Nile, JGrass-NewAGE system, Satellite data, evapotranspiration

1 Introduction

Freshwater is a scarce resource in many regions of the world: the problem continues to be aggravated by growing populations and significant increases in demand for agricultural and industrial purposes. The Nile River basin is one such region, with relatively arid climate because of high temperatures and solar radiation which foster rapid evapotranspiration. Most of the countries within the basin, such as Egypt, Sudan, Kenya, and Tanzania, receive insufficient fresh water (Pimentel et al., 2004). Exceptions to this are the small areas at the equators and the Upper Blue Nile basin in the Ethiopian highlands, which receives up to 2000 mm per year (Johnston and McCartney, 2010). Particularly, the Upper Blue Nile (hereafter UBN) basin is the main sources of water in the region. Also, it is probably one of the most hydro-climatologically and socio-politically complex basins in the world. The water resources management in the basin face many pressures and challenges: (1) as the principal contributor (i.e 85%) to the main Nile basin, UBN supports the lives of hundreds of millions of people living downstream, and it is referred to as the "Water Tower" of northeast Africa; (2) locally, the basin is inhabited by 20 million people whose main livelihood is subsistence agriculture (Population Census Commission 2008); (3) topographically, the basin is very complex: it starts from mountains as high as 4,160 m a.s.l. and drains to lowlands of about 500 m a.s.l.; (4) the UBN is a part of trans-boundary river, hence its development and management require diplomatic discussions with many national governments; (5) many international and non-governmental organizations, each with different policies, legal regimes, and contrasting interests, are involved in the freshwater governance of the basin; (6) the Ethiopian government has started many water resource development projects, such as irrigation schemes and dams, among which the Grand Ethiopia Renaissance Dam (GERD), which, upon completion, will be one of the largest in Africa.

Tackling all these facts and challenges and developing better water use strategies is only possible by gathering quantitative information (Hall et al., 2014). Understanding the hydrological processes of UBN, therefore, is the basis for both the transboundary negotiations about sharing the water resources f the basin and for assessing the sustainability of ~~subsistence~~ farming systems in the region. In fact, because of the lack of hydrometeorological data and a proper modelling framework, ~~however,~~ the recent modelling efforts conducted within the basin have evident limitations in addressing these problems. ~~As a consequence, spatio-temporal hydrological information in the basin is very scarce.~~ Studies in the region are limited to small basins, particularly within the Lake Tana basin where there are relatively better hydrometeorological data (Rientjes et al., 2011; Uhlenbrook et al., 2010; Tekleab et al., 2011; Wale et al., 2009; Kebede et al., 2006; Bewket and Sterk, 2005; Steenhuis et al., 2009; Conway, 1997; Mishra et al., 2004; Mishra and Hata, 2006; Teferi et al., 2010),or at the whole basin scale, but in which case information on spatial variability is usually ignored (Kim et al., 2008; Kim and Kaluarachchi, 2009; Gebremicael et al., 2013; Tekleab et al., 2011). Other studies are limited to a specific hydrological process e.g. rainfall variability (Block and Rajagopalan, 2007; Abtew et al., 2009), time series and statistical analysis of in situ discharge/rainfall data (Teferi et al., 2010; Taye and Willems, 2011) or perform modelling at very low temporal resolutions (e.g. monthly) (Kim and Kaluarachchi, 2008; Tekleab et al., 2011). ~~Consequently,~~ Spatially distributed information on all the components of the water budget does not exist and ~~In a region where all the hydrological fluxes (precipitation, evapotranspiration and discharge) are important elements of the water budget, "traditional"~~ basin modelling approaches that are tailored to a single component do not provide an effective picture of the dynamics of the water resources within the basin.

To overcome data scarcity, large scale hydrological modelling can be supported by remote sensing (RS) products, which fill the data gaps in water balance dynamics estimation (Sheffield et al., 2012). For instance, a considerable number of researches has been carried out in the last two decades in developing satellite rainfall estimations procedures (Hong et al., 2006; Bellerby, 2007; Huffman et al., 2007; Kummerow et al., 1998; Joyce et al., 2004; Sorooshian et al., 2000; Brocca et al., 2014).

RS is also a viable option to fill the gaps for basin scale evapotranspiration estimation. Global satellite evapotranspiration products have been available by applying energy balance and empirical models to satellite derived surface radiation, meteorology and vegetation characteristics, and they are recognised to have a certain degree of reliability (e.g. Fisher et al., 2008; Mu et al., 2007; Sheffield et al., 2010).

Basin scale storage estimations is the most difficult task. Fortunately, the Gravity Recovery and Climate Experiment (GRACE) (Landerer and Swenson, 2012) came to fill this gap (e.g. Han et al., 2009; Muskett and Romanovsky, 2009; Rodell et al., 2007; Syed et al., 2008; Rodell et al., 2004). Guntner (2008), Ramillien et al. (2008) and Jiang et al. (2014) reviewed the use of GRACE data and positively recommended it for large scale water budget modeling. At the moment, satellite based retrievals of discharge are not available as operational or research products, but, potentially it can be retrieved from satellite altimetry and multispectral sensors (e.g. Tarpanelli et al., 2015; Van Dijk et al., 2016). Moreover, the Surface Water Ocean Topography (SWOT, Durand et al. (2010)) mission, which is expected to be launched in 2020, will provide river elevation (with an accuracy of 10 cm), slope (with an accuracy of 1 cm/1 km) and width that can be used in estimating river discharge (Paiva et al., 2015; Pavelsky et al., 2014).

Notwithstanding the availability of these RS products at various (spatial and temporal) resolutions and accuracy, their use is clearly a new paradigm in water budget closure estimations (Sheffield et al., 2009; Andrew et al., 2014; Sahoo et al., 2011; Gao et al., 2010; Wang et al., 2014).

This study is an effort contributing to answering the quantitative issues related to the aforementioned problems and aims to resolve the water budget of the UBN basin using a new hydrological modelling framework (see section 3.1) and remote sensing data improving the estimates of previous studies. It obtains, at relatively small spatial scales and at daily time step, all the water budget components. It is also a methodological paper, in that it delineates various methodologies to overcome the data scarcity ~~, and inherits from~~ (Abera et al., submitted).

The paper is organized as follows: firstly, descriptions of the study area and model setup are given (section 2), then the methodologies for each water budget component and the model set-up are detailed in section 3. The results and discussions of each component and the water budget are presented in section 4. Finally, the conclusions of the study are given (section 5).

2 The Study Basin

The Upper Blue Nile (UBN) river originates at Lake Tana at Bahir Dar, flowing southeast through a series of cataracts. After about 150 km, the river enters to a deep canyon, and changes direction to the south. After flowing for another 120 km flow, the river again changes its direction to the west and northwest, towards the El Diem (Ethiopia-Sudan border). Many tributaries draining from many parts of the Ethiopian highlands join the main river along its course. The total distance of the river within Ethiopia is about 1000 km.

The UBN basin represents up to 60% of the Ethiopian highlands contribution to the Nile river flows, which is itself 85% of the total (Abu-Zeid and Biswas, 1996; Conway, 2000). The area of the river basin enclosed by a section at the Ethiopia-Sudan border is about 175,315 km^2 (figure 1), covering about 17% of the total area of the country. The large scale hydrological behaviour of the basin is described in a series of studies (Conway, 1997, 2000, 2005; Conway and Hulme, 1993). Specifically, its hydrological behaviour is characterized by high spatio-temporal variability. Since the UBN basin has the lion's share of the total Nile flow, it is the economic mainstay of downstream countries (i.e. Sudan and Egypt). Moreover, the Ethiopian highlands are highly populated and have high water demands of their own for irrigation and domestic uses.

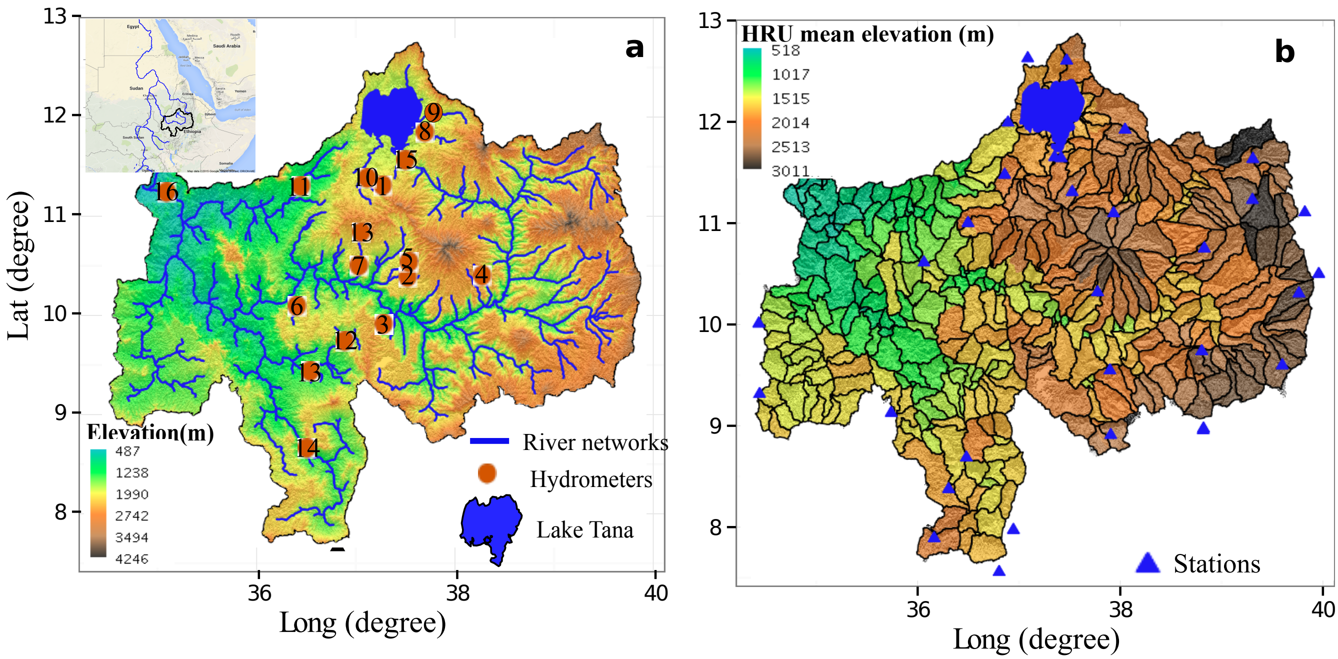


Figure 1. The Upper Blue Nile basin digital elevation map, along with the gauge stations (a); and subbasin partitions and meteorological stations used for simulation (b). Numbers inside the circles (figure a) designates the river gauging stations. The name of the basin referring to the numbers are provided in table 4.

The topographic distribution of the basin is shown in figure 1. The topography of UBN is very complex, with elevation ranging from 500 m in the lowlands at the Sudan border to 4160 m in the upper parts of the basin. Due to the topographic variations, the climate of the basin varies from cool (in the highlands) to hot (in the lowlands), with large variations in a limited elevation range. The hot season is from March to May, the wet season, with lower temperatures, is from June to September, while the dry season runs from October to February. There are three controlling mechanisms of rainfall in the UBN basin and in Ethiopia as whole (Seleshi and Zanke, 2004): the Intertropical Convergence Zone (ITCZ) that drives the monsoon rainfall during the wet season (Jun.-Sept.); the Saharan anticyclone that generates the dry and cool northeasterly winds in the dry season (Oct.-Feb.); and the Arabian highlands that produce thermal lows in the hot season (Mar.-May). Specifically studies have found that the interannual and seasonal variability of precipitation in the UBN basin is governed by the Southern Oscillation Index (SOI), the equatorial eastern Pacific sea level pressure, and the sea-surface temperature (SST) over the tropical eastern Pacific Ocean (Camberlin, 1997; Seleshi and Zanke, 2004). The mean annual rainfall and potential evapotranspiration of the UBN basin are estimated to be in the ranges of 1,200-1,600 mm and 1,000-1,800 mm, respectively (Conway, 1997, 2000), with high spatio-temporal variability. The annual temperature mean is 18.5^o, with small seasonal variability.

3 Methodology

Water budget simulation is essential to the estimation of both water storage and water fluxes (rate of flow) for given, appropriate, control volumes and time periods. It is given by:

$$\frac{\partial S_k(t)}{\partial t} = J_k(t) + \sum_{i}^{m(k)} Q_{ki}(t) - ET_k(t) - Q_k(t) \quad (1)$$

where $J(t)$ is rainfall, and $ET(t)$ is actual evapotranspiration, $Q(t)$ is discharge, $Q_{ki}(t)$ is the discharge from the contributing streams. The index $k = 1, 2, 3...$ is the control volume where the water budget is solved. In our case, the control volume is a portion of the basin (a subbasin) derived from topographic partitioning as described in section 3.1.

~~Different RS and in situ data are used to enforce or validate the modelling solutions implemented and to close the water budgets. The datasets used for the water budget are described in their respective sections.~~

3.1 JGrass-NewAGE system set-up

UBN water budget is estimated using the JGrass-NewAGE hydrological system. It is a set of modelling components, reported in Table 1, that can be connected at runtime to create various modelling solutions. Each component is presented in details and tested against measured data in the corresponding papers cited in the Table 1. Similar study using JGrass-NewAge system, but using mostly in-situ observations, has been conducted in Posina river basin (northeast Italy), and the model performance is assessed positively (Abera et al., submitted). Brief descriptions on the components used in this study are provided in the following sections. In this study, the shortwave solar radiation budget component (section 3.3), the evapotranspiration component (Priestley and Taylor, section 3.3), the Adige rainfall-runoff model (section 3.4), and all the components illustrated in figure 2 are used to estimate the various hydrological flows.

A necessary step for spatial hydrological modelling is the partitioning of the topographic information into an appropriate spatial scale. The SRTM 90 m X 90 m elevation data is used to generate the basin GIS representation.The basin topographic representation in Geographic Information System (GIS), as detailed in (Formetta et al., 2014a; Abera et al., 2014; Formetta et al., 2011), is based on the Pfafstetter enumeration (Formetta et al., 2014a; Abera et al., 2014). The basin is subdivided in Hydrologic Response Units (HRUs), where the model inputs (i.e. meteorological forcing data), and hydrological processes and outputs (i.e. evapotranspiration, discharge, shortwave solar radiation) are averaged ~~, using a threshold upstream drainage area concept~~ (Formetta et al., 2014a). ~~The water budget components are estimated for each HRU and, subsequently,~~ A routing scheme is applied to move the discharges from HRUs to the basin outlet through the channel network.

In this study, the UBN basin is divided into 402 subbasins (HRUs of mean area of 430 $\pm$ 339 km^2) and channel links, as shown in figure 1b. This spatial partitioning may not be the finest scale possible, however, considering the size of the basin, and the resolution of satellite products, it can be considered an acceptable scale to capture the spatial variability of the water budget.

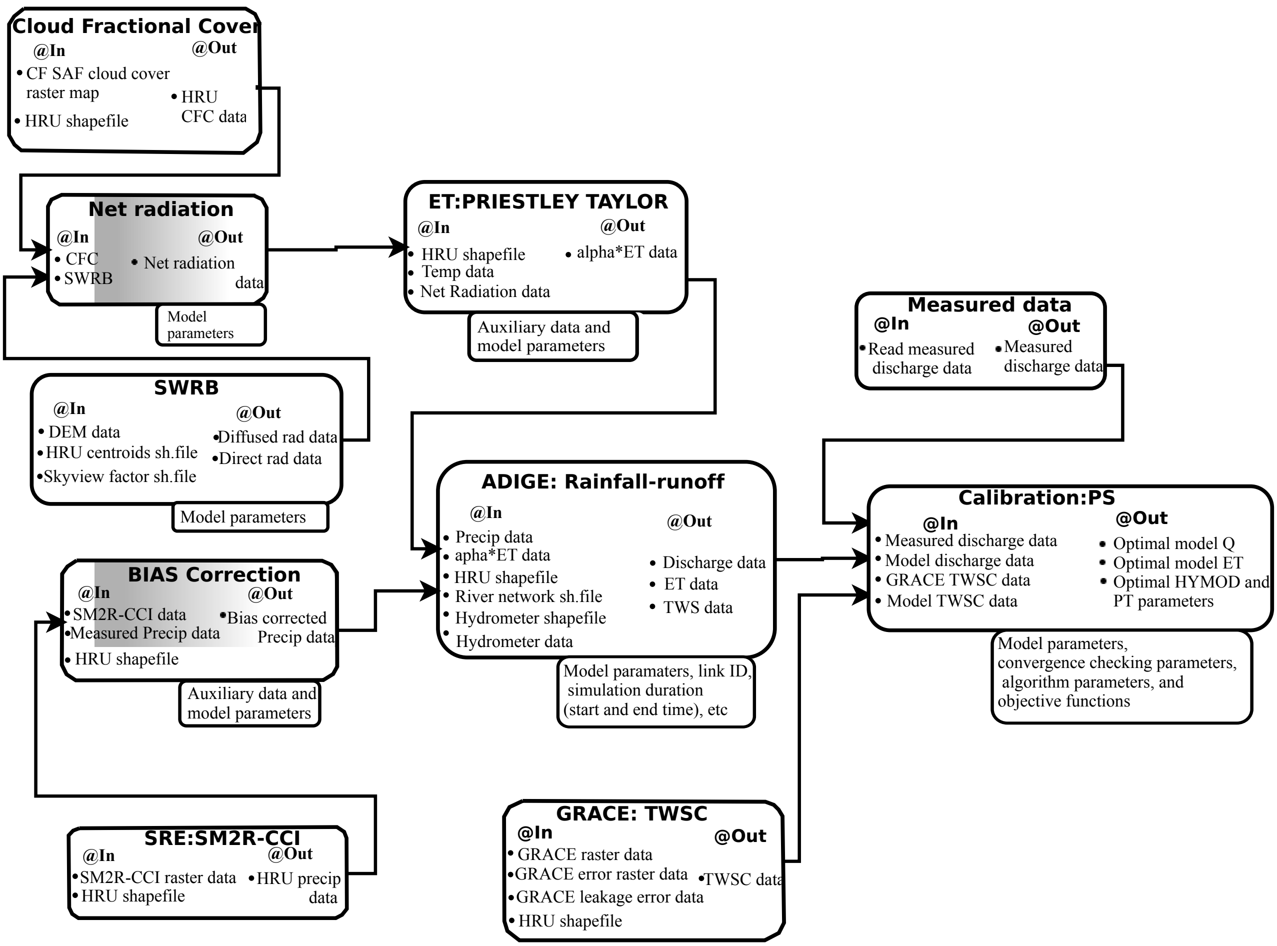


Figure 2. Workflow with a list of NewAge components (in white), and remote sensing data processing parts (shaded in grey, not yet included in JGrass-NewAGE and currently performed with R tools) used to derive the water budget of the UBN. It does not include the components used for the validation and verification processes.

Table 1. JGrass-NewAge system components and respective references. The components in bold are the ones used in this study.

Role	Component Name	Description
Basin partitioning	**GIS spatial toolbox and Horton Machine**	A GIS spatial toolbox that uses DEM to extract basin, hillslopes, and channel links for NewAge-JGrass set-up (Formetta et al., 2014a; Abera et al., 2014)
Data interpolation	Kriging, Inverse Distance Weighting, and JAMI	Interpolates meteorological data from meteorological stations to points of interest according to a variety of kriging algorithms (Goovaerts, 2000; Haberlandt, 2007; Goovaerts, 1999; Schiemann et al., 2011), Inverse Distance Weighting (Goovaerts, 1997)
Energy balance	**Shortwave radiation**, Longwave radiation	Calculate shortwave and longwave radiation, respectively, from topographic and atmospheric data (Formetta et al., 2013, 2016).
Evapotranspiration	Penman-Monteith, **Priestly-Taylor**, Fao-Evapotranspiration	Estimates evapotranspiration using Penman-Monteith (Monteith et al., 1965), Priestly-Taylor (Priestley and Taylor, 1972), and Fao-Evapotranspiration (Allen et al., 1998) options
Runoff	ADIGE (**Hymod**)	Estimates runoff based on Hymod (Moore, 1985) algorithm (Formetta et al., 2011)
Snow melting	Snow melt	Modelling snow melting using three temperature and radiation based snow algorithms (Formetta et al., 2014b)
Optimization	**Particle Swarm Optimization**, DREAM, **LUCA**	Calibrate model parameters according to Particle Swarm Optimization (Kennedy and Eberhart, 1995), DREAM (Vrugt et al., 2009), LUCA (Hay et al., 2006) algorithms respectively.

3.2 Precipitation $J(t)$

The spatio-temporal precipitation input term of Eq. 1 (J(t)), is quantified with RS-based approaches. Currently, there are several satellite rainfall estimates (SREs) available for free at varying degrees of accuracy and reliability. recently Abera et al. (2016) compared five of them with high spatial and temporal resolutions ~~in a series of hydrological applications~~ over the same basin. It was shown that SM2R-CCI (Brocca et al., 2013, 2014) is one of the best products, particularly in capturing the total rainfall volume. A comparative analysis of the effects of different SREs on basin water budget components is an interesting area of research that can be extended from this study, however, here only SM2R-CCI is used ~~as water budget modelling~~ for obtaining the precipitation input. The systematic error (bias) of SM2R-CCI is removed according to the ecdf matching techniques by Michelangeli et al. (2009) and Abera et al. (2016) by using ground measures. ~~Once SM2R-CCI is corrected for bias errors, it is used to examine the spatio-temporal precipitation variability of the basin and drive the JGrass-NewAGE modelling system to solve the discharge.~~ The subbasin mean precipitation is estimated by averaging all the pixels RS corrected data within each subbasin. In accordance with the basin partition described in section 3.1, the 1994-2009 daily precipitation set is generated for 402 subbasins.

3.3 Evapotranspiration ET

Evapotranspiration estimation is crucial for agricultural and water resources management as it is an important flux within a basin. The lack of in-situ data relating to ET impedes modelling efforts and makes it probably the most difficult task in water budget assessment. Here, ET is estimated according to the NewAge specific component. ~~evapotranspiration component is estimated using the radiation budget as input, as it is the main radiant energy available at the surface to drive the surface biophysical processes and evaporation~~. ~~This approach~~ It provides estimates at any temporal and spatial resolution required by using the Priestley and Taylor (PT) Formula (Priestley and Taylor, 1972), which is one of the more common models used ~~simplified models used to estimate ET~~. PT ~~is mainly based on net radiation, Rn, simplifying~~ simplifies all the unknowns into the α_{PT} coefficient, as shown in Eq. 2.

$$ET = \alpha_{PT} \frac{\Delta}{\Delta + \gamma} (Rn) \tag{2}$$

Where Δ is the slope of the Clausius-Clapeyron relations and γ is the psychometric constant (Brutsaert, 2005). In this study, however, the actual evapotranspiration, ET is constrained not only by the atmospheric demands as in (Eq. 2), but it uses storage information which can be obtained from the ADIGE rainfall-runoff component of JGrass-NewAGE. Hence, the ET equation is modified as (Abera et al., submitted):

$$ET(t) = \alpha_{PT} \frac{S(t)}{S_{max}} \frac{\Delta}{\Delta + \gamma} (Rn) \tag{3}$$

where $S(t)$ is the groundwater storage, and S_{max} the maximum storage capacity for each HRU. The important unknown coefficient α_{PT} (Pejam et al., 2006; Assouline et al., 2016) and the S_{max} ~~(maximum water storage capacity for each HRU)~~ are calibrated within the rainfall-runoff model component, as explained below.

In this procedure, given that $S(t)$ is not measured, the assumption that there is null water storage difference after a long time, named Budyko's time, T_B, (Budyko, 1978), is required. So, here, what is searched is a time duration (T_B) such that the water storage assumes again the initial value (Abera et al., submitted). Once T_B is fixed, automatic calibration can be set to produce the set of parameters, including α_{PT} and S_{max}, for which, besides discharge is well reproduced, is also $S(T_B) = S(0)$. In this study, $T_B = 6$ years.

In equation 3, Rn is the main input modulating the atmospheric demand component of ET. To this scope, the NewAge shortwave radiation budget component, SWRB (Formetta et al., 2013), is used to return a value for each subbasin in clear sky conditions. Irradiance in clear sky conditions, however, is unsuitable for all sky condition since surface shortwave radiation is strongly affected by cloud cover and cloud type (Arking, 1991; Kjærsgaard et al., 2009). Therefore, the clear sky SWRB estimated using NewAge-SWRB is cut ~~to Rn in all-sky conditions~~ by using the cloud fractional cover (CFC) satellite data set (Karlsson et al., 2013), processed and provided by EUMETSAT Climate Monitoring Satellite Application Facility (CM SAF) project (Schulz et al., 2009). In this case net radiation is generated only from the shortwave radiation and the cloud cover data, as in the following formulation (Kim and Hogue, 2008):

$$Rn = (1 - CFC) R_S \tag{4}$$

Where R_S is the net shortwave radiation ~~estimated using the NewAge-SWRB component for each subbasin,~~ and Rn is the net radiation. The daily CFC data originates from polar orbiting satellites, version CDRV001, using a daily temporal resolution and a 0.25^o spatial resolution from 1994 to 2009 (16 years). Satellite data are processed (Karlsson et al., 2013) to obtain the mean daily CFC for each subbasin. In comparison to CFC, the effects of surface albedo on Rn is minimal, particularly in highland areas with vegetation cover and no snow cover such as the UBN basin.

Once ET is estimated according to the methods described, it is useful to validate it with independently obtained ET estimates or data. In situ ET observations are not available for this basin, as is the case for most regions. Estimates of ET based on RS have been made by different algorithms (Norman et al., 1995; Mu et al., 2007; Jarmain, 2009; Fisher et al., 2008). In this study, the Global Land Evaporation Amsterdam Methodology (GLEAM) (Miralles et al., 2011a), a global, satellite-based, ET data set is used. The performance of GLEAM is assessed positively in different studies (McCabe et al., 2016; Miralles et al., 2011b). Differently from the NewAge approach, GLEAM also considers dynamic vegetation information to cut PT-based potential ET to actual ET (Miralles et al., 2011a). GLEAM is available at 0.25^o spatial resolution and daily temporal resolution. For comparison with NewAge ET, we estimated area weighted average GLEAM ET for each HRU polygon. The aim of the comparison is not for strict validation, but rather to assess the level of consistency between the two independent estimations. More details on GLEAM can be found in Miralles et al. (2011a, b) and McCabe et al. (2016). Comparison of the NewAge ET with MODIS standard ET product is also available in the supplementary material of the paper.

3.4 Discharge Q

For discharge estimation, the ADIGE rainfall-runoff component is used. It is based on the well-known HYMOD model (Moore, 1985) as runoff production component which also include the routing component, the artificial inflow-outflow management component. Detailed descriptions of HYMOD implementations in the NewAge model system are given at Formetta et al. (2011) and Abera et al. (submitted). The main inputs for the ADIGE model are $J(t)$ and $ET(t)$, as estimated in the previous sections. The NewAge Hymod component is applied to any HRU, in which the basin is subdivided and the total watershed discharge is the sum of the contribution of each HRU routed to the outlet. The ADIGE rainfall-runoff has five calibration parameters, and the calibration is performed using the particle swarm (PS) optimization. PS is a population-based stochastic optimization technique inspired by the social behaviour of flocking birds or fish schools (Kennedy and Eberhart, 1995). It is suited to obtaining a global optimal and less susceptible to getting trapped in local minima (Scheerlinck et al., 2009). The objective function used to estimate the optimal value of the parameter is the Kling-Gupta efficiency (KGE, Kling et al. (2012)). The KGE is preferred to the commonly-used Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe (1970)) because the NSE has been criticized for its overestimation of model skill for highly seasonal variables by underestimating flow variability (Schaefli and Gupta, 2007; Gupta et al., 2009). For evaluation of the model performances, in addition to the KGE, two other goodness-of-fit (GOF) methods (percentage bias (PBIAS) and correlation coefficient) used in this study are described in Appendix A.

3.5 Total water storage change ds/dt

The ds/dt in Eq. 1 is the water contained in the ground, soil, snow and ice, lakes and rivers, and biomass. It is the total water storage (TWS) change, calculated as the residuals of the water budget fluxes for each control volume. In this paper, the ds/dt estimation at daily time steps is based on the interplay of all the other components as presented in Eq. 1. There is no way to estimate areal TWS from in situ observations. The new Gravity Recovery and Climate Experiment (GRACE) data (Landerer and Swenson, 2012) has a potential to estimate this component, but at very low spatial and temporal resolutions. At large scale, however, it can still be used for constraining and validating data to the modelling solutions. Here, the performance of our modelling approach to close the water budget, i.e. estimating storage following the characterization of all the terms, is assessed using the GRACE estimation at the basin scale. Monthly data is obtained from NASA's Jet Propulsion Laboratory (JPL) ftp:// podaac-ftp.jpl.nasa.gov/allData/tellus/L3/land mass/RL05. The leakage errors and scaling factor (Landerer and Swenson, 2012) that are provided with the product are applied to improve the data before the comparison is made. The total error of GRACE estimation is a combination of GRACE measurement and leakage errors (Billah et al., 2015). Based on the data of these two error types, the mean monthly error of GRACE in estimating total water storage change (TWSC) in the basin is about 8.2 mm. Since the other fluxes, for instance Q and ET, are modelled as functions of basin water storage, the good estimation of water storage by a model has inference to its reasonable computation of other fluxes as well (Döll et al., 2014).

3.6 Calibration and validation approach

The satellite precipitation data set (SM2R-CCI) is error corrected based on in situ observations. At the basin outlet (Ethiopia-Sudan Border), the ADIGE rainfall-runoff component (i.e. HYMOD model) is calibrated to fit the observed discharge during the six years of calibration period (1994-1999) at daily time steps. Based on the approach described in the ET estimation section, α_{PT} is calibrated by by imposing that $S(t) = S(T_B)$ after $T_B = 6$ years. The value of six years is arbitrary but it was found to give good agreement with GRACE data (see below), so no other values were used. The simulation for each hydrological component is then verified using available in-situ or remote sensing data (Table 2), and three goodness-of-fit (KGE, PBIAS, r) are used as comparative indices (for detail information please see Appendex A), as follows:

- Discharge validation: Discharge simulation is validated for separate time-series data at the outlet at Ethiopian-Sudan border, where the model is calibrated. In addition, the simulation of NewAge at the internal links is validated where in situ data are available. The evaluations at the internal links provide an assessment of model estimation capacity at ungauged locations.

- ET validation: Once ET is estimated according to the procedures described above, GLEAM (Miralles et al., 2011a) is used as an independent data set to assess ET estimation. After GLEAM is aggregated for each subbasin, the GLEAM and the NewAge ET are compared and the goodness-of-fit (GOF) indexes are calculated, based on 16 years of data (1994-2009).

– ds/dt validation: The water storage change, ds/dt, estimated as residual of the water budget, is validated against the GRACE based data-set. The GRACE product was used to estimate the total water storage change for the the whole basin, since the error of GRACE increases if used at small scales. To harmonize and enable comparison between the model and the GRACE TWS data, it is necessary to do both time and spatial filtering. Following the GRACE TWSc temporal resolution, the model ds/dt is aggregated at monthly time steps.

Table 2. Short summary of the list of remote sensing products used in this study.

Satellite products	Spatial resolution	Temporal resolution	Reference	used as
SM2R-CCI	0.25	daily	Brocca et al. (2014, 2013); Abera et al. (2016)	input for Precipitation
GLEAM	0.25 degree	daily	Miralles et al. (2011a); McCabe et al. (2016)	verification for evapotranspiration
MODIS ET (MOD16)	1-km	8-days	Mu et al. (2007, 2011)	verification for evapotranspiration
GRACE TWS	1 degree	30-days	Landerer and Swenson (2012)	Verification for storage change
CM-SAF	0.25 degree	daily	Schulz et al. (2009)	input for evapotranspiration component

4 Results and Discussion

The results of the study are organized as follows: firstly, we present the results for ~~simulation performance and their comparison with independent data, along with brief spatio-temporal characteristics of the fluxes and storage.The simulated water budget components are described in the following order:~~ 1) precipitation, 2) evapotranspiration, 3) discharge and 4) total water storage; secondly, the JGrass-NewAGE system is used to resolve the water budget closure at each subbasin, and the contribution of each term water budget term is further is analyzed.

4.1 Precipitation J

The spatial distribution of mean, long-term, annual precipitation is presented in figure 3a. Generally, precipitation increases from the east (about 1000 mm/year) to the south and southwest (1800 mm/year). SM2R-CCI shows that the south and southwest parts of the basin receive higher precipitation than the east and northeast parts of the highlands. The rainiest subbasins are in the southern part of the basin.The precipitation data used correspond to a mean annual rainfall of about 1900 mm, while the mean annual precipitation reported for this region by Abtew et al. (2009) is about 2049 mm. The latter estimation, however, ~~small discrepancy could be that the estimation~~ is from point gauge data, while this study is based on areal data ~~from SM2R-CCI~~. ~~by available ground rain gauges.~~ Generally, precipitation increases from the east (about 1000 mm/year) to the south and southwest (1800 mm/year). This spatial pattern is consistent with the results of Mellander et al. (2013) and Abtew et al. (2009). To understand the spatial distribution of the seasonal cycle, the quarterly percentage of total annual precipitation, calculated from 1994 to 2009 in daily estimations, is presented in figure 3 b. During the summer season (June, July and August), while the subbasins in the north and northeast receive about 65% of the annual precipitation (figure 3 b), the subbasins in the south receive about 40% of total precipitation.

4.2 Evapotranspiration ET

Based on the approach detailed in our methodology, the ET is estimated for each subbasin at daily time steps. Figure 4 a shows the comparisons of the ET time series from 1994-2002 (aggregated at daily, weekly, and monthly, from top to bottom) between NewAge and GLEAM. The Figure specifically refers to three selected subbasins representing different ranges of elevations and spatial locations. NewAge estimates have higher temporal variability in comparison to GLEAM ~~, for all three time steps. It clearly shows that GLEAM underestimates ET during the peak periods and overestimates it during low periods~~. In the represented locations, GLEAM therefore accumulates a systematic growing difference in evapotranspired water volume, which could be not consistent with the estimated storage (see below).

The agreement/disagreement between the two ET estimations vary from subbasin to subbasin (figure 4). The spatial distribution correlation and PBIAS between the NewAge and GLEAM ET is presented in figure 4 b. Spatially, the correlation between JGrass-NewAGE and GLEAM is higher in the eastern and central parts of the basin, while it tends to decrease systematically towards the west (i.e. to the lowlands, see figure 4 b). The correlation between the two ET estimations increases when passing from daily to monthly time steps. The PBIAS between the two estimates ranges from -10% to 10%, with large numbers of subbasin being from -3% to 3%. Spatially, the comparison shows that GLEAM overestimates ET in the western parts of the basin (border to the Sudan) and underestimates ET in the northern parts of the basin (figure 4b). The overall basin correlation is 0.34 $\pm$ 0.07 (daily time step), 0.51 $\pm$ 0.08 (weekly time step), and 0.57 $\pm$ 0.10 (monthly time steps). ~~The PBIAS does not change over the three time steps (1.5 $\pm$ 9.1%).~~ Generally, except at daily time step, the two estimates have acceptable agreements (very low bias, and acceptable correlation). However, in comparison with the correlation (0.48 $\pm$ 0.15) and PBIAS (14.5 $\pm$ 18.9%) obtained between NewAge ET and MODIS ET Product (MOD16), as shown in the supplementary material, at 8-days time steps, the correlation and PBIAS between NewAge ET and GLEAM ET is much better.

4.3 Discharge Q

The optimized parameters of the Adige model, obtained using automatic calibration procedure of NewAge are given at Table 3. At the basin outlet, the automatic calibration of the NewAge components provided very good values of the GOF indices (KGE=0.93, PBIAS = 2.2, r = 0.94). The performances, at the outlet remain high also during the validation period, having

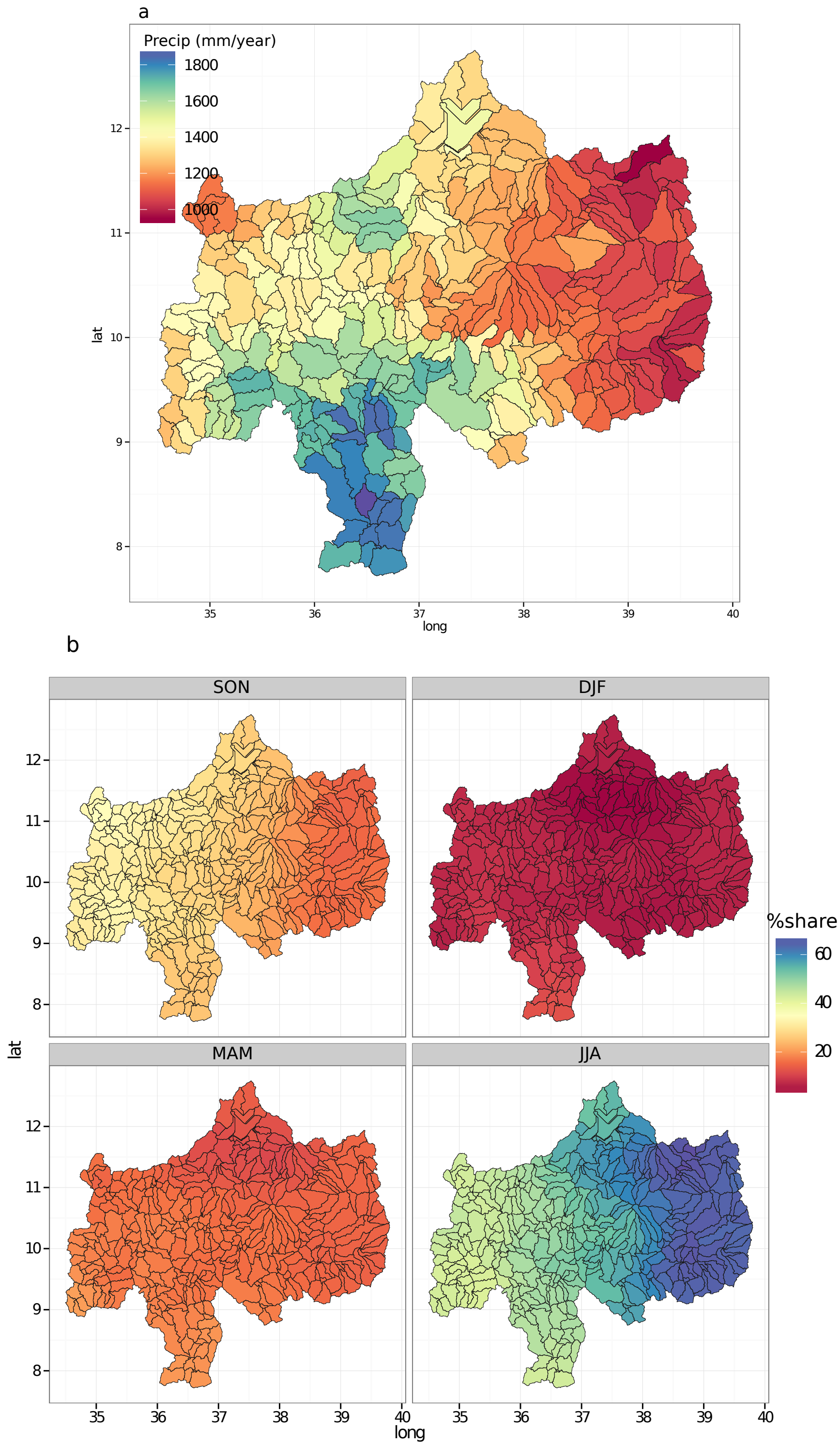


Figure 3. The spatial distribution of mean annual rainfall (a), and quarterly percentages share of the total rainfall (b) estimated from long term data (1994-2009): SON (September, October, and November), DJF (December, January, and February), MAM (March, April, and May), JJA (June, July and August). Note that high seasonality is observed in the eastern part of the basin.

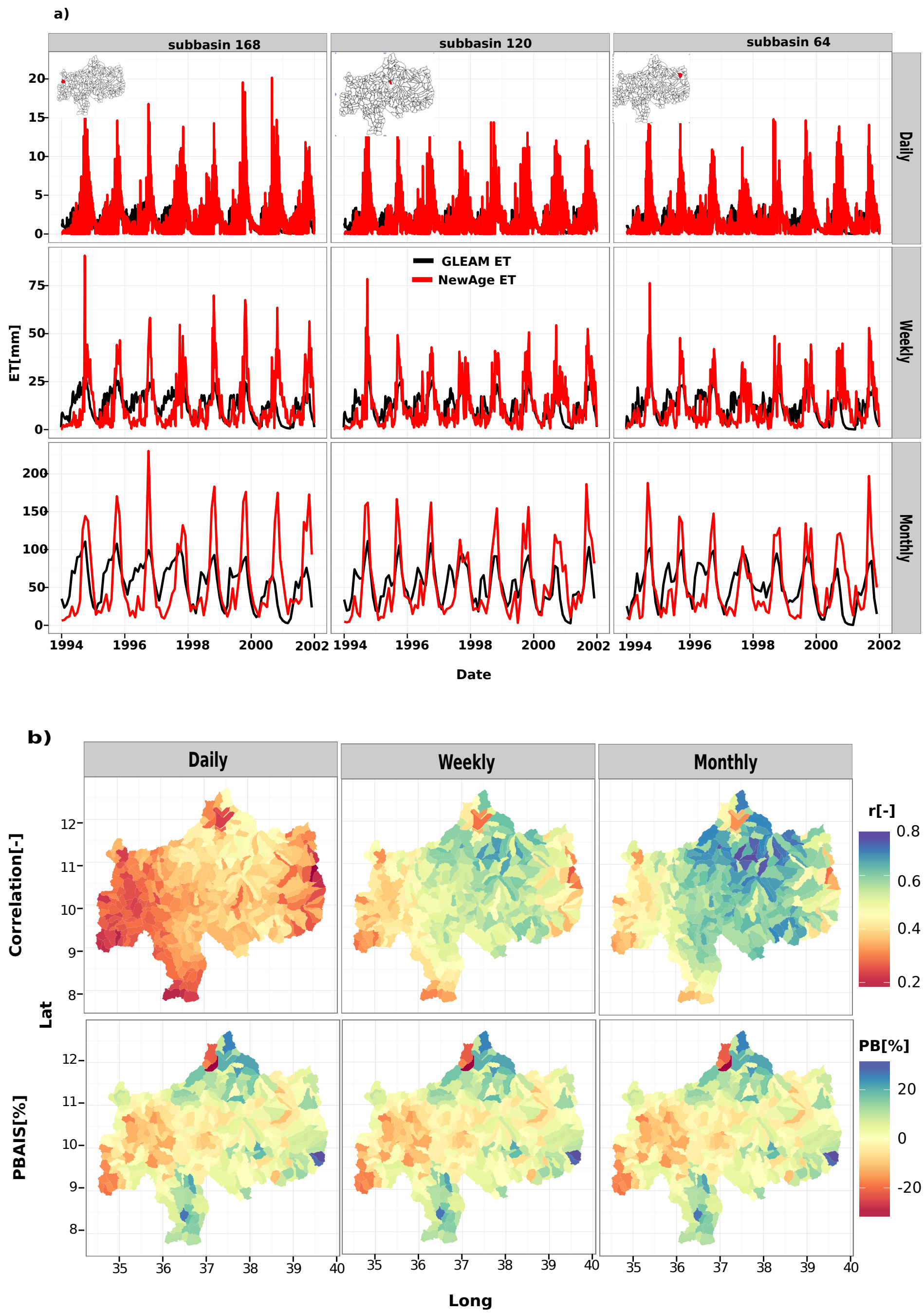


Figure 4. a:Time series ET estimation with NewAge and GLEAM for three subbasins: subbasin ID168 , subbasin ID120, and subbasin ID64 at daily, weekly and monthly time steps. The locations of the subbasins are indicated on the maps at the top of each column of plots. b: spatial distribution of correlation coefficient and PBIAS between NewAge and GLEAM estimations at daily, weekly and monthly time steps.

KGE=0.92, PBIAS = 2.4, and r = 0.93.. ~~At the same location, model performance is verified during the validation period and it is almost equal to the performance during the calibration period.~~

Model performances are also evaluated within the basin at the internal catchments outlets (Table 4) where stage measures are available. Figure 5 shows simulated hydrographs ~~channel links, when available,~~ along with the observed discharges for some locations. ~~At the outlet, even during the validation period, the model is able to capture the dynamics of the basin response very well (KGE=0.92, PBIAS = 2.4, r = 0.93).~~ The results show that the performances of the NewAge simulation are a little better than ~~similar to~~ the performances reported by Mengistu and Sorteberg (2012), with slightly lower PBIAS value (PBIAS=8.2, r=0.95). Generally, the model predicts both the high flows and low flows well, with slight underestimation of peak flows (figure 5 a). This is likely due to the underestimation of SM2R-CCI precipitation data for high rainfall intensities (Abera et al., 2016). Additional source of error can also be caused by model inconsistency due to averaging out input data over large areas.

Table 3. Optimized parameters obtained from daily ADIGE simulation during the calibration period (1994-1999). The last parameter is for the ET component.

Parameters	value
$C_{max}[L]$	694.18
$B_{exp}[-]$	0.64
$\alpha_{Hymod}[-]$	0.61
$Rs[T]$	0.086
$Rq[T]$	0.394
$\alpha_{PT}[-]$	2.9

Regarding the internal sites discharge forecasting, we remark some representative results. The hydrograph comparison between the NewAge simulated discharge and the observed one of the Gelgel Beles river, enclosed at the bridge near to Mandura with an area of 675 km^2, is shown in figure 5 b. The performance of the uncalibrated NewAge at Gelegel Beles has a correlation coefficient of 0.70, PBIAS is 11.40% and the KGE value is 0.68 (Table 4).

Simulation performances for the medium size basins, ~~based on model parameters calibrated at basin outlet,~~ such as the Ribb river, enclosed at Addis Zemen (area=1592 km^2, KGE = 0.81, PBIAS = 12% and r = 0.82, figure 5 c), and Gilgel Abay river, enclosed at Merawi (area = 1664 km^2, KGE=0.81, PBIAS=12%, r=0.93), are very good. For the Ribb river, the NewAge simulation performance can be compared with SWAT Model performances by Setegn et al. (2008) (r=0.74-0.76). Even though SWAT was calibrated for this specific subbasin, the results of our study are much better. Similarly, without calibration for the Gilgel Abay river, the NewAge simulation performance is ~~comparable~~ better than the results of Wase-Tana (Wosenie et al., 2014, PBIAS=34)) and Flex$_B$ (Fenicia et al., 2008, PBIAS=77.6) or comparable to SWAT (PBIAS=5).

To analyze the forecasting capacity of NewAge for the larger size basins, the performances at Angar river (area 4674 km^2), Lake Tana (area 15321 km^2), and Dedisa river basin (9981 km^2) are reported. The simulation analysis at the Angar river enclosed near Nekemt (KGE = 0.72, PBIAS = -14.10%, and r = 0.82), Lake Tana (KGE = 0.26, PBIAS = 5.10, and r =

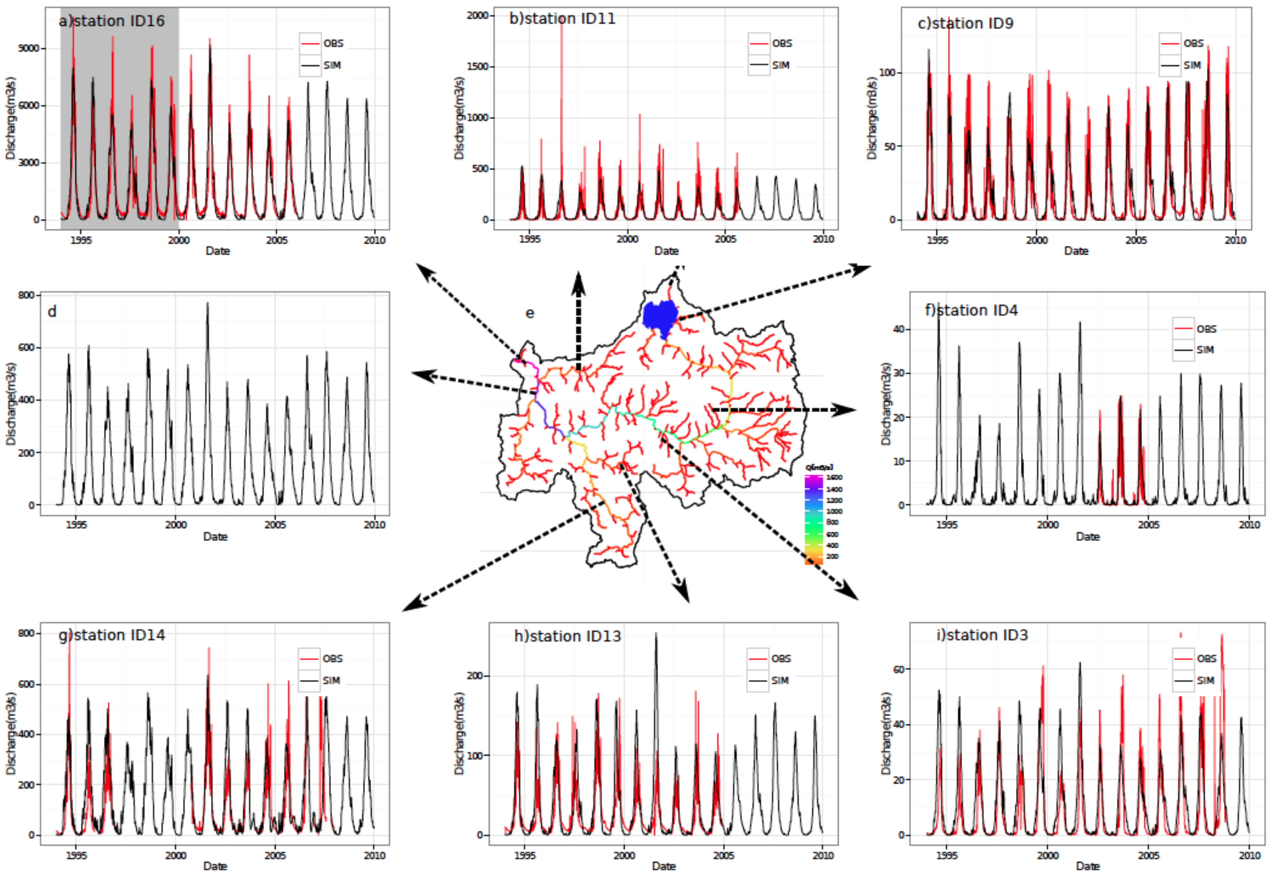


Figure 5. NewAge model forecasting validation at internal subbasins. The model calibrated and validated at El Diem (a) is used to estimate at each channel link and, where discharge measurements are available, they are verified: main Beles bridge (b), Ribb river enclosed at Addis Zemen (c), just simulation of the main Blue Nile before joining Beles river (d), Jedeb near Amanuel (f), Dedisa river basin enclosed near Arjo (g), Angar river basin enclosed near Nekemt (h), and Nesh near Shambu (i). Figure (e) shows the long term estimated daily discharge for all river links of the basin.

Table 4. The forecasting capacity of the NewAge Adige rainfall-runoff component at the internal sites, based on the optimized parameters calibrated at the outlet. The performance at the outlet (El Diem) is the model performance during validation period.

Hydrometer stations ID	River Name	Area (km^2)	KGE	PBIAS	r
1	Koga @ Merawi	244.00	0.67	-8.70	0.73
2	Jedeb @ Amanuel	305.00	0.38	40.80	0.53
3	Neshi @ Shambu	322.00	0.58	32.00	0.57
4	Suha @ Bichena	359.00	0.54	39.20	0.82
5	Temcha @ Dembecha	406.00	0.70	3.30	0.71
6	Gilgel Beles @ Mandura	675.00	0.68	11.40	0.70
7	Lower Fettam @ Galibed	757.00	0.67	-7.7	0.78
8	Gummera @ Bahir Dar	1394.00	0.19	-53.20	0.88
9	Ribb @ Addis Zemen	1592.00	0.81	12.00	0.86
10	Gelgel Abay @ Merawi	1664.00	0.81	12.00	0.93
11	Main Beles @ Bridge	3431.00	0.68	-1.70	0.74
12	Little Anger @ Gutin	3742.00	0.65	24.30	0.81
13	Great Anger @ Nekemt	4674.00	0.72	-14.10	0.82
14	Didessa @ Arjo	9981.00	0.55	19.60	0.81
15	Upper Blue Nile @ Bahir Dar	15321.00	0.26	5.10	0.60
16	Upper Blue Nile @ El Diem	174000.00	0.92	2.40	0.93

0.60), and Dedisa (KGE=0.55, PBIAS = 19.60, and r = 0.81) indicate that the performances are acceptable. The comparison of simulated and observed discharges, as well as the locations of the Angar (basin brief description (Easton et al., 2010)) and Dedisa rivers are shown in figure 5, in plots h and g respectively.

For most subbasins, because of the good model performances (i.e. KGE is higher than 0.5 and PBIAS is within 20%), the estimated discharges are deemed adequate for forecasting and estimating water resource at locations where gauges are unavailable. The model is also able to reproduce discharge across the range of scales. For instance, the model performances at the Ethiopia-Sudan border (175 315 km^2), Dedisa near Arjo (9981 km^2), main Beles (3431 km^2), and Temcha near Dembecha (406 km^2) are also acceptable, except for Lake Tana, where the discharge is regulated (figure 5 and table 4). Sample simulations at all the channel links of the study basin at daily time step are provided in the supplementary material.

4.4 Total water storage change

~~The~~ NewAge simulated ds/dt for 16 years for each subbasin ~~calculating it~~ is calculated as a residual of the flux terms. ~~We first compared t~~ The simulated ds/dt is represented and compared with the GRACE-based TWSC in Figure 6. ~~shows ds/dt time series for the whole basin, estimated by using NewAge and GRACE.~~ The storage change shows high seasonality over the basin, with positive change in summer and negative change in winter. The change varies from -100 to +120 mm/month. The model

ds/dt, aggregated at monthly time scale, is in accordance with the GRACE TWSC both in temporal pattern and amplitude. Over the whole basin a correlation coefficient of 0.84 is obtained. The good performances of the ds/dt component ~~also has an inference on the model capability to reproduce other components well, as it is the residual terms that balance the flux dynamics~~ is certainly caused also by the ability of NewAGE to well reproduce the other water fluxes. Due to the possible high leakage error introduced in GRACE TWSC at high spatial resolutions (Swenson and Wahr, 2006), statistical comparison at subbasin level is not performed. The spatial distribution of NewAge and GRACE ds/dt estimates can be found in the supplementary material.

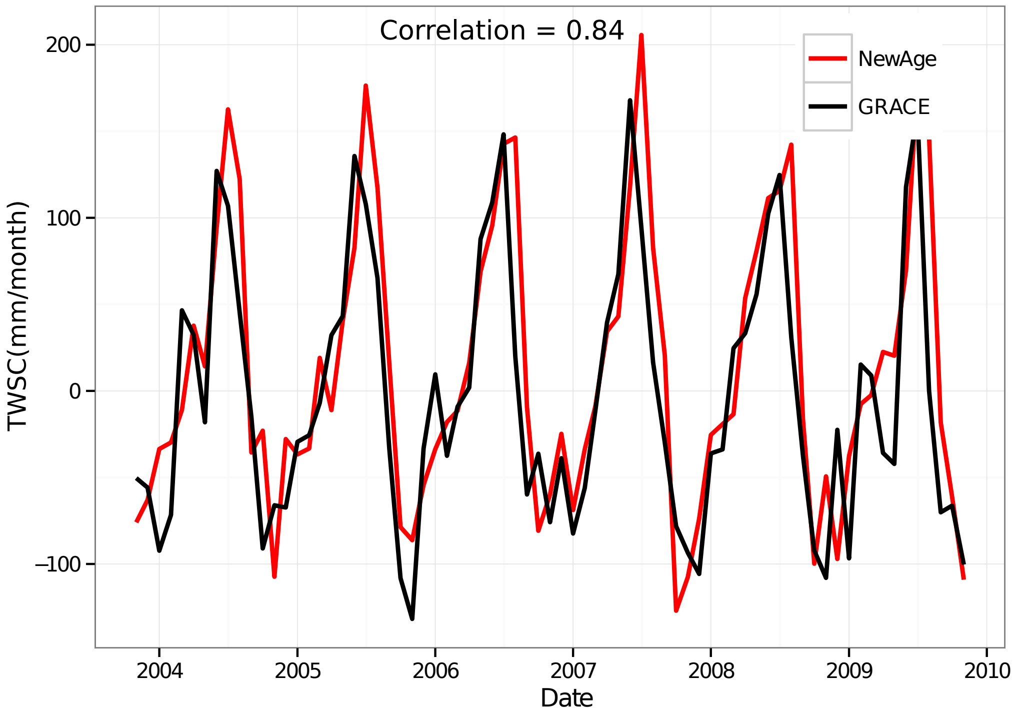


Figure 6. Comparison between basin-scale NewAge ds/dt and GRACE TWSC from 2004-2009 at monthly time steps.

4.5 Water budget closure

The water budget components (J, ET, Q, ds/dt) of 402 subbasin of the UBN are simulated for the period 1994-2009 at daily time steps. Figure 7 shows the long-term, monthly-mean, water budget closure derived from 1994-2009. The four months (January, April, July, and October) are selected to show the four seasons (Winter, Spring, Summer and Autumn). For all components, the mean seasonal variability is very high. Generally, the seasonal patterns of Q and ds/dt follow the J, showing the highest values in summer (i.e. July) and the lowest in winter (i.e. January). However, simulated ET shows distinct seasonal patterns with respect to the other components, the highest being during autumn (October), followed by winter (January). During the summer it is low, most likely due to high cloud cover.

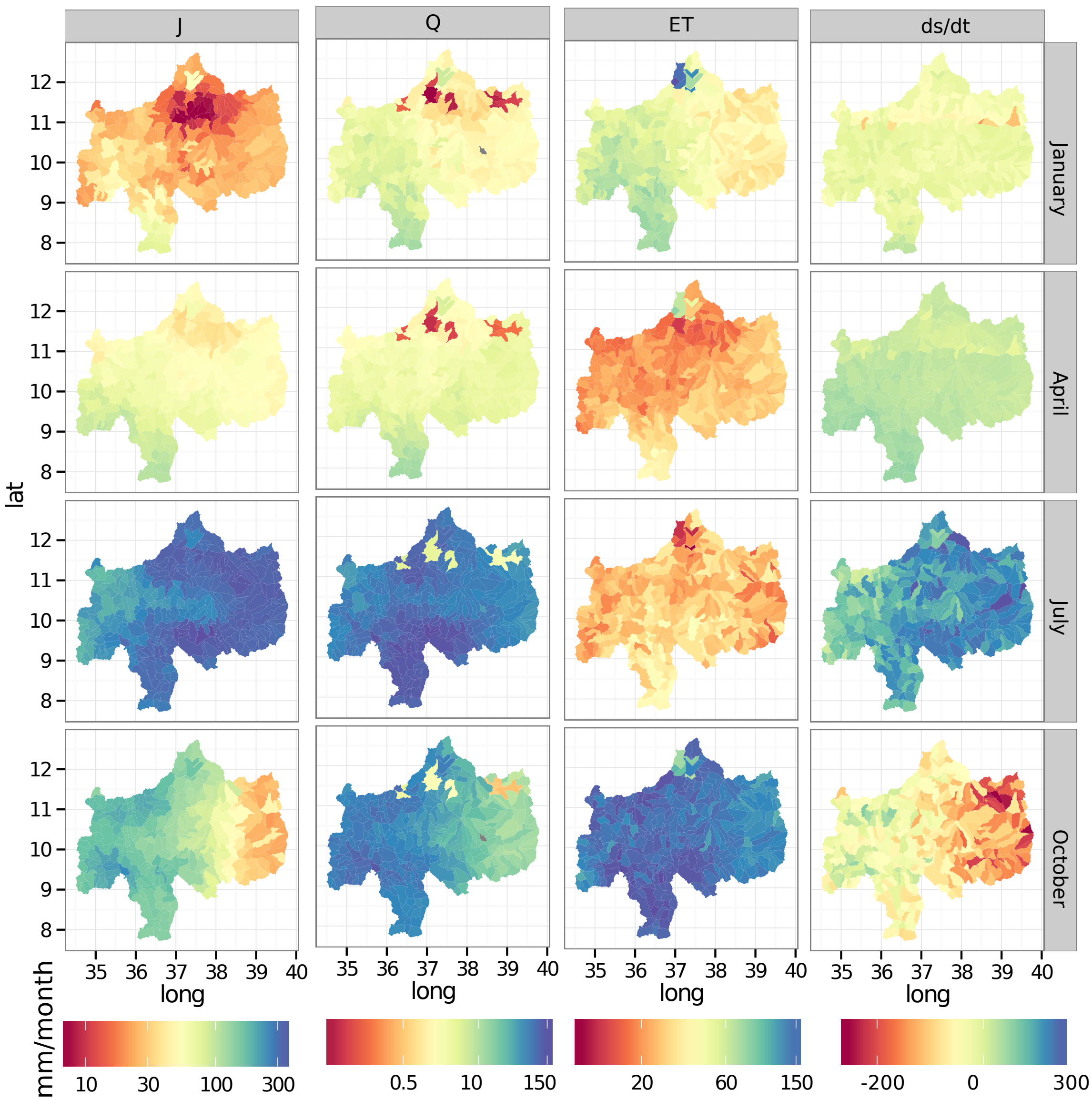


Figure 7. Spatial distribution of long term mean monthly water budget (January,April, July and October) in the UBN basin. For the sake of visibility, the legend is plotted separately and on logarithmic scale, except for the storage component.

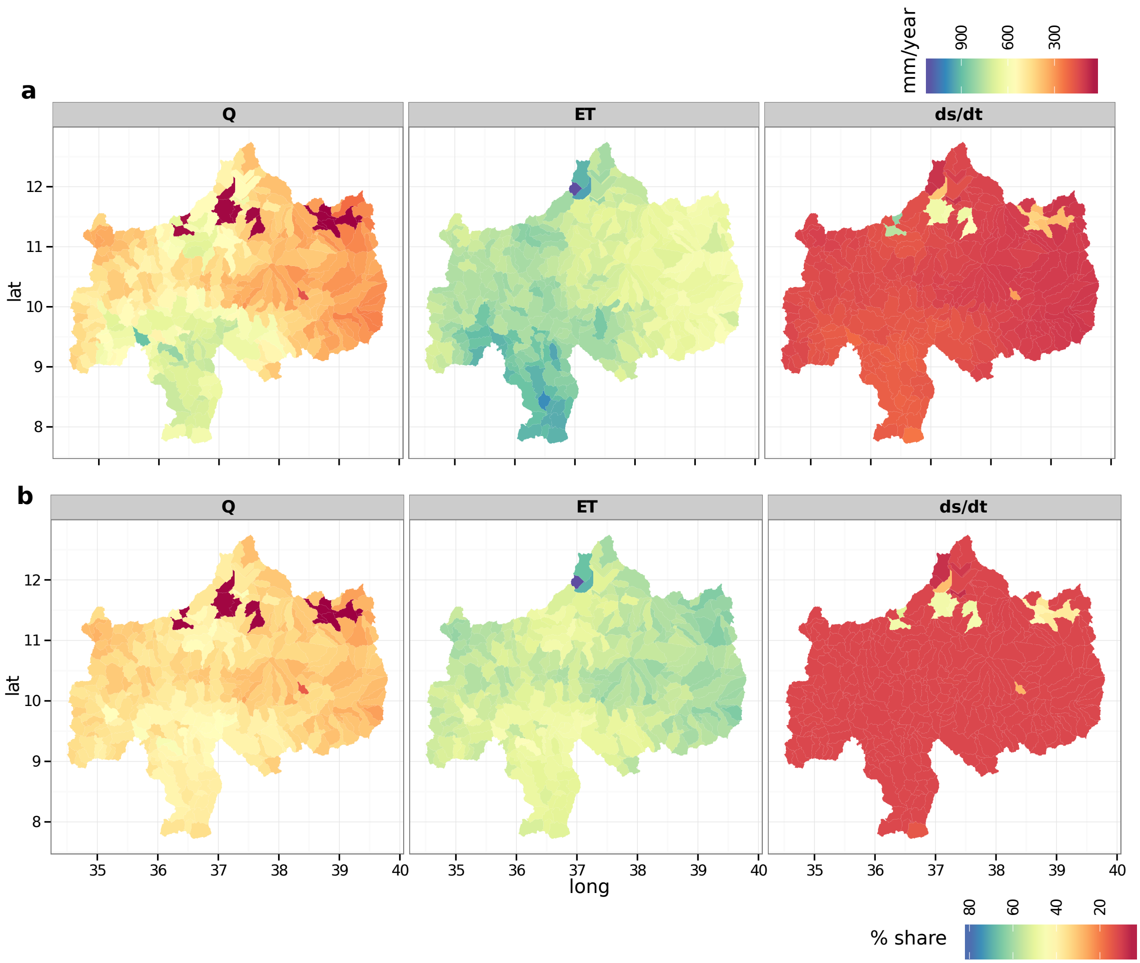


Figure 8. The spatial distributions of long term mean annual water budget closure: precipitation in mm (figure 3), the output terms (Q, ET, ds/dt) in mm (a), and the percentage share of the output term (Q, ET, ds/dt) of the total precipitation (b).

The variability between the subbasins is also appreciable. Generally, all components tends to increase from the east to the southwest part of the basin, except for the summer season (July). During summer, on the other hand, the eastern part of the basin receives its highest rainfall, stores more water, and generates high runoff as well. In general the dominant budget component varies with months. For instance, in January ET is the dominant while in June and July ds/dt is more dominant. After the summer season, Q and ET are the dominant fluxes. A regression analysis based on the results for all subbasins and all years shows that, at short time scales such as at daily or monthly, the variability in ET is not due to variablity in J (R^2=0.01). Conversely, at the yearly time scale, 78% of ET variance is explained by variability in J.

The spatial variability of the long term mean annual water budget closure is shown in figure 8. The spatial variability of J and Q is higher than ds/dt and ET. The higher Q and ET in the southern and southwestern part of the basin are due to higher J. Similarly Q is lower in the eastern and northeastern part of the basin. Focusing on the percentage share of the output term

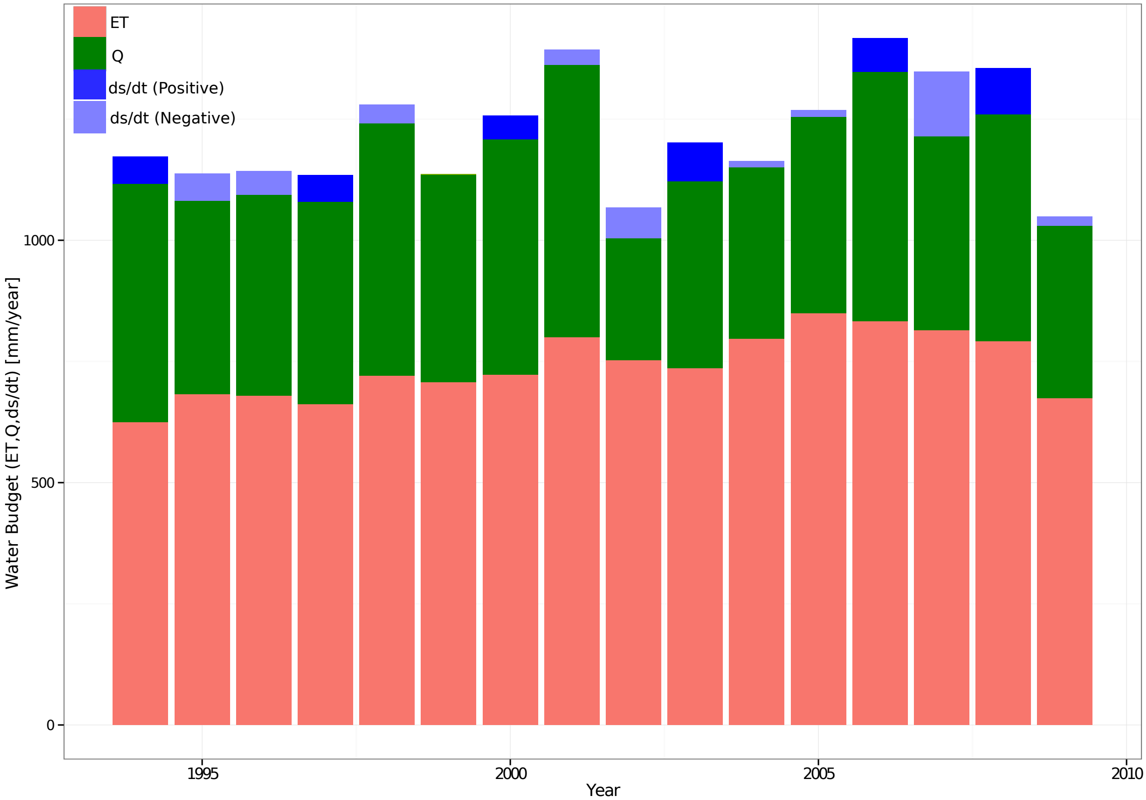


Figure 9. Water budget components of the basin and its annual variabilities from 1994 to 2009. The relative share of each of the three components (Q, ET and ds/dt) of the total available water J is represented by the length of the bars (N.B. the total length of the bar minus the negative storage is J). The positive and negative storage of the years are shown by dark blue and light blue respectively).

(Q, ET, ds/dt) of total J (figure 8 c), ET dominates the water budget, followed by Q. It is noteworthy that the eastern subbasins with low ET still have percentage share of ET due to low amount of J received.

The long-term basin-average water budget components shows: 1360 $\pm$ 230 mm of J, followed by 740 $\pm$ 87 mm of ET, 454 $\pm$ 160 mm of Q and -4 $\pm$ 63 mm of ds/dt. While the spatial variability of the water budget is high, the annual variability is rather limited. Higher annual variability is observed for J, followed by Q. 2001 and 2006 are wet years, characterized by high J and Q. Conversely, 2002 and 2009 are dry years with 1167 mm and 1215 mm per year of precipitation. Details on the two dry years (2002, 2009) of the region can be ~~consulted~~ read in Viste et al. (2013).

Figure 10 provides long term monthly mean estimates of water budget fluxes and storage. The basin scale mean budget is highly variable. The highest variability is mainly in J and S. During summer months, J, Q, and ds/dt shows high magnitude. ET is not highest in June, July and August, but, in October and December it is. The S accumulated in the summer season feeds the highest ET in autumn, and causes very high drops in S (figure 10).

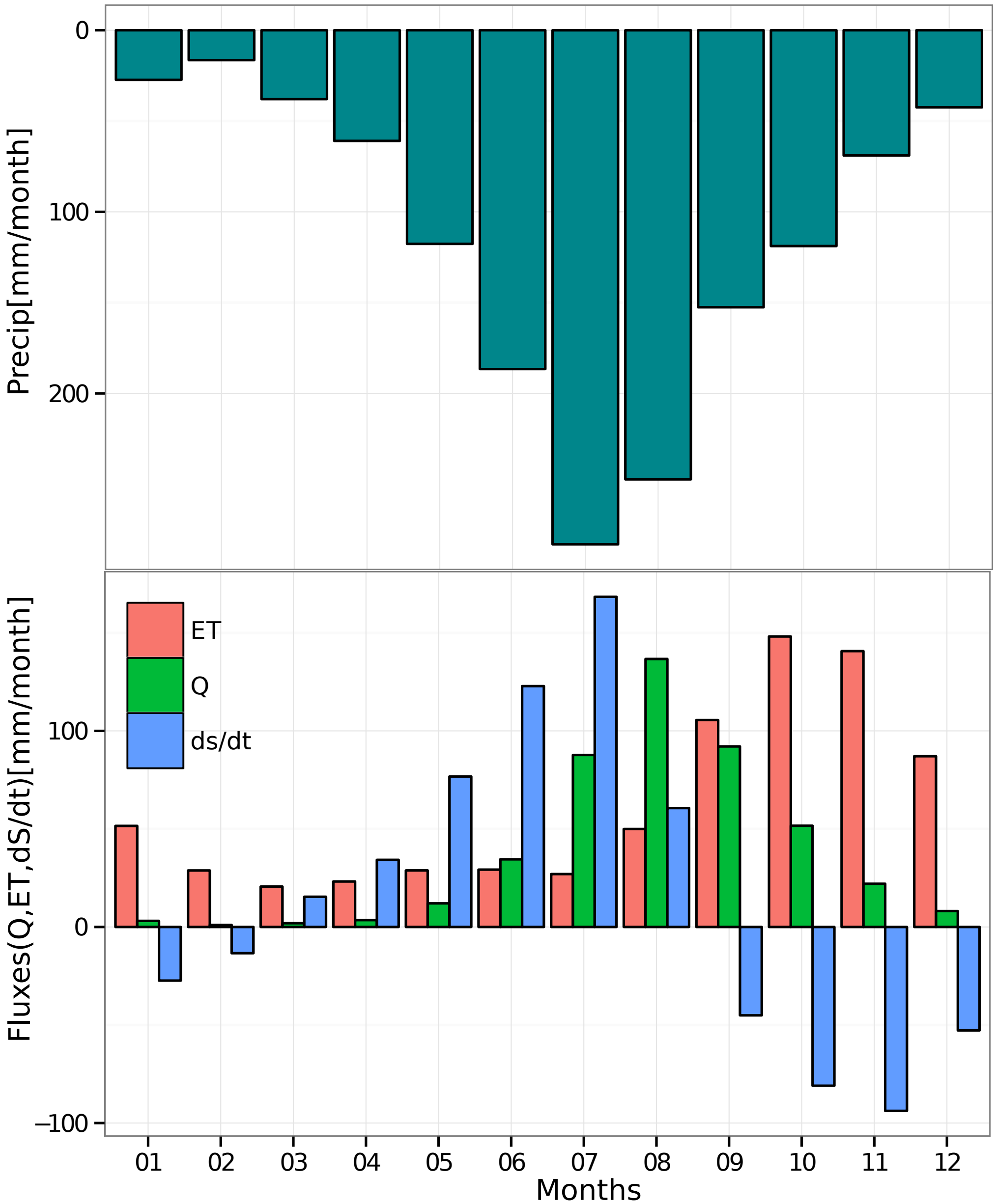


Figure 10. Monthly mean Water budget components at basin scale and long term, based on estimates from 1994 to 2009. The relative shares of the three components (Q, ET and dS/dt) of the total available water J are shown .

5 Conclusions

The goal of this study is to estimate the whole water budget and its spatial and temporal variability of the upper Blue Nile basin using the JGrass-NewAge hydrological ~~model~~ system and remote sensing data. The study covered 16 years from 1994-2009. Different remote sensing data (SM2R-CCI, SAF EUMETSAT CFC, GLEAM, GRACE) are used to force and verify the modeling results. The results can be summarized as follows. ~~Different remote sensing data (SM2R-CCI, SAF EUMETSAT CFC, GLEAM, GRACE) are effectively employed either to force the water balance modelling or to verify model results.~~

- The basin scale annual precipitation over the basin is 1360 $\pm$ 230 mm, and highly variable spatially. The southern and southwestern parts of the basin receive the highest precipitation, which tend to decrease towards the eastern parts of the basin (figure 3).

- The performances obtained ~~of the modelling solution~~ are promising (figures 5 and 6, and table 4) and often greatly improve previous results.

- Generally, the interannual variability of ET is high, and tends to be higher in autumn and lower in summer. The average basin scale ET is about 740 $\pm$ 87 mm, and is the larger flux in water budget in the basin.

- The comparison of simulated ET with the satellite product GLEAM shows that GLEAM has low temporal variability. The correlation between GLEAM ET and NewAge ET increases from daily time steps to monthly time steps, and spatially it is higher in the east and central parts of the basin. Comparison with MODIS products was also performed (reported in complimentary material). MODIS actually shows even more large departure from JGrass-NewAGE results. Both satellite products, however, seem to introduce a systematic bias which would not allow to close the budget.

- The NewAge ADIGE rainfall-runoff component is able to reproduce discharge very well at the outlet (KGE = 0.92). The long term annual runoff of the UBN basin is about 454 $\pm$ 160 mm. The verification results at the internal sites where measures are available reveal that the model can be used for forecasting at ungauged locations with some success.

- ~~Generally, the long term water budget simulation shows that the basin is in equilibrium around zero storage (-4 $\pm$ 63 mm) with minor departures even after the 16 years at which the condition imposed by the Budyko hypothesis, i.e. $T_B = 6$ years could have been forgiven.~~ The NewAge storage estimations and their space-time variability are effectively verified by the basin scale GRACE TWSC data which show high correlation and similar amplitude.

Despite the good results obtained, it is important to note that this study is limited by the lack of in-situ ET observation and low resolution GRACE data for confirmation of storage. To these regards, the results of this study would benefit from basin specific assessments of ET and ds/dt RS products based on ground measurements, as done in Abera et al. (2016) for precipitation.

Reproducibility

The forcing data used for NewAge simulation: SM2R-CCI is obtained from http://hydrology.irpi.cnr.it/people/l.brocca; the rain gauge precipitation and hydrometer discharge data were obtained from the National meteorological Agency and Ministry of Water and Energy of Ethiopia respectively, and it can be requested for research. The remote sensing data used for comparison: GLEAMS ET, MODIS ET and GRACE TWSC are freely available and can be downloaded at http://www.gleam.eu,http://www.ntsg.umt.edu/project/mod16, and ftp://podaac-ftp.jpl.nasa.gov/allData/tellus/L3/landmass/RL05 respectively. Modelling components used for the simulations are available and documented through the Geoframe blog http:geoframe.blogspot.com. Additional data (i.e. GIS database, topographic information, input data and additional results) and other notes regarding the paper can be found at Zenodo, DOI: Zenodo DOI:10.5281/zenodo.264004.

Acknowledgements. This research has been partially financed by the CLIMAWARE projects of University of Trento (/http://abouthydrology.blogspot.it/search/label/CLIMAWARE) and by European Union FP7 Collaborative Project GLOBAQUA (Managing the effects of multiple stressors on aquatic ecosystems under water scarcity, grant no. 603629-ENV-2013.6.2.1). We would like to acknowledge the National meteorological Agency and Ministry of Water and Energy of Ethiopia for providing us the gauge rainfall and discharge data. We also thank the two anonymous reviewer for their work that helped to enhance the initial manuscript with their comments.

Appendix A: Model performance criteria

The model evaluation statistics used in the paper are the goodness-of-fit (GOF) indices. The following indexes are used as objective function and comparison of estimations.

1. PBIAS: is the measure of average tendency of estimated values to be large or smaller that their measured values. The value near to zero indicates high estimation, whereas the positive value indicates the overestimation and negative values indicate model underestimation (Moriasi et al., 2007; Gupta et al., 1999).

$$PBIAS = \frac{\sum_{i=1}^{n}(P_i - O_i)}{\sum_{i=1}^{n} O_i} 100 \tag{A1}$$

The PBIAS value ranges from -20 to 20% is considered good, and values between ±20% and ±40% and those greater than ±40% are considered satisfactory and unsatisfactory respectively (Stehr et al., 2008).

2. Kling-Gupta efficiency (KGE) is developed by Gupta et al. (2009) to provide a diagnostically interesting decomposition of the Nash-Sutcliffe efficiency (and hence MSE), which facilitates the analysis of the relative importance of its different components (correlation, bias and variability) in the context of hydrological modelling. Kling et al. (2012) proposed a revised version of this index. It is given by

$$KGE = 1 - ED \tag{A2}$$

$$ED = \sqrt{(r-1)^2 + (vr-1)^2 + (\beta-1)^2} \tag{A3}$$

where ED is the Euclidian distance from the ideal point, β is the ratio between the mean simulated and mean observed flows, r is Pearson product-moment correlation coefficient, and v is the ratio between the observed (σ_o) and modelled (σ_s) standard deviations of the time series and takes account of the relative variability (Zambrano-Bigiarini, 2013). The KGE ranges from infinity to a perfect estimation of 1, but a performance above 0.75 and 0.5 is considered as good and intermediate respectively (Thiemig et al., 2013).

3. Pearson correlation coefficient (r): please refer Moriasi et al. (2007). The correlation coefficient is best as much as it is close to 1.

References

Abera, W., Antonello, A., Franceschi, S., Formetta, G., and Rigon, R.: The uDig Spatial Toolbox for hydro-geomorphic analysis, British Society for Geomorphology, London, UK, in: clarke & nield (eds.) geomorphological techniques (online edition) edn., 2014.

Abera, W., Brocca, L., and Rigon, R.: Comparative evaluation of different satellite rainfall estimation products and bias correction in the Upper Blue Nile (UBN) basin, Atmospheric Research, 178, 471–483, 2016.

Abera, W., Formetta, G., Borga, M., and Rigon, R.: Estimating the water budget components and their variability in a Pre-Alpine basin with JGrass-NewAGE, Advances in Water Resources, submitted.

Abtew, W., Melesse, A. M., and Dessalegne, T.: Spatial, inter and intra-annual variability of the Upper Blue Nile Basin rainfall, Hydrological processes, 23, 3075–3082, 2009.

Abu-Zeid, M. A. and Biswas, A. K.: River basin planning and management, Oxford University Press, 1996.

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, FAO, Rome, 300, 6541, 1998.

Andrew, M. E., Wulder, M. A., and Nelson, T. A.: Potential contributions of remote sensing to ecosystem service assessments, Progress in Physical Geography, 38, 328–353, 2014.

Arking, A.: The radiative effects of clouds and their impact on climate, Bulletin of the American Meteorological Society, 72, 795–813, 1991.

Assouline, S., Li, D., Tyler, S., Tanny, J., Cohen, S., Bou-Zeid, E., Parlange, M., and Katul, G. G.: On the variability of the Priestley-Taylor coefficient over water bodies, Water Resources Research, 2016.

Bellerby, T.: Satellite rainfall uncertainty estimation using an artificial neural network, Journal of Hydrometeorology, 8, 1397–1412, 2007.

Bewket, W. and Sterk, G.: Dynamics in land cover and its effect on stream flow in the Chemoga watershed, Blue Nile basin, Ethiopia, Hydrological Processes, 19, 445–458, 2005.

Billah, M. M., Goodall, J. L., Narayan, U., Reager, J., Lakshmi, V., and Famiglietti, J. S.: A methodology for evaluating evapotranspiration estimates at the watershed-scale using GRACE, Journal of Hydrology, 523, 574–586, 2015.

Block, P. and Rajagopalan, B.: Interannual variability and ensemble forecast of Upper Blue Nile Basin Kiremt season precipitation, Journal of Hydrometeorology, 8, 327–343, 2007.

Brocca, L., Moramarco, T., Melone, F., and Wagner, W.: A new method for rainfall estimation through soil moisture observations, Geophysical Research Letters, 40, 853–858, 2013.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., and Levizzani, V.: Soil as a natural rain gauge: estimating global rainfall from satellite soil moisture data, Journal of Geophysical Research: Atmospheres, 119, 5128–5141, 2014.

Brutsaert, W.: Hydrology: an introduction, Cambridge University Press, 2005.

Budyko, M.: 1., 1974: Climate and Life, International Geophysics Series, 18, 1978.

Camberlin, P.: Rainfall anomalies in the source region of the Nile and their connection with the Indian summer monsoon, Journal of Climate, 10, 1380–1392, 1997.

Conway, D.: A water balance model of the Upper Blue Nile in Ethiopia, Hydrological sciences journal, 42, 265–286, 1997.

Conway, D.: The climate and hydrology of the Upper Blue Nile River, Geographical Journal, pp. 49–62, 2000.

Conway, D.: From headwater tributaries to international river: observing and adapting to climate variability and change in the Nile basin, Global Environmental Change, 15, 99–114, 2005.

Conway, D. and Hulme, M.: Recent fluctuations in precipitation and runoff over the Nile sub-basins and their impact on main Nile discharge, Climatic change, 25, 127–151, 1993.

Döll, P., Fritsche, M., Eicker, A., and Schmied, H. M.: Seasonal water storage variations as impacted by water abstractions: comparing the output of a global hydrological model with GRACE and GPS observations, Surveys in Geophysics, 35, 1311–1331, 2014.

Durand, M., Fu, L.-L., Lettenmaier, D. P., Alsdorf, D. E., Rodriguez, E., and Esteban Fernandez, D.: The surface water and ocean topography mission: Observing terrestrial surface water and oceanic submesoscale eddies, Proceedings of the IEEE, 98, 766–779, 2010.

Easton, Z., Fuka, D., White, E., Collick, A., Biruk Ashagre, B., McCartney, M., Awulachew, S., Ahmed, A., and Steenhuis, T.: A multi basin SWAT model analysis of runoff and sedimentation in the Blue Nile, Ethiopia, Hydrology and earth system sciences, 14, 1827–1841, 2010.

Fenicia, F., McDonnell, J. J., and Savenije, H. H.: Learning from model improvement: On the contribution of complementary data to process understanding, Water Resources Research, 44, W06 419, 2008.

Fisher, J. B., Tu, K. P., and Baldocchi, D. D.: Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, Remote Sensing of Environment, 112, 901–919, 2008.

Formetta, G., Mantilla, R., Franceschi, S., Antonello, A., and Rigon, R.: The JGrass-NewAge system for forecasting and managing the hydrological budgets at the basin scale: models of flow generation and propagation/routing, Geoscientific Model Development, 4, 943–955, 2011.

Formetta, G., Rigon, R., Chávez, J., and David, O.: Modeling shortwave solar radiation using the JGrass-NewAge system, Geoscientific Model Development, 6, 915–928, 2013.

Formetta, G., Antonello, A., Franceschi, S., David, O., and Rigon, R.: Digital watershed representation within the NewAge-JGrass system, Boletín geológico y minero, 125, 369–379, 2014a.

Formetta, G., Kampf, S. K., David, O., and Rigon, R.: Snow water equivalent modeling components in NewAge-JGrass, Geoscientific Model Development, 7, 725–736, 2014b.

Formetta, G., Bancheri, M., David, O., and Rigon, R.: Performance of site-specific parameterizations of longwave radiation, Hydrology and Earth System Sciences, 20, 4641, 2016.

Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., and Lettenmaier, D. P.: Estimating the water budget of major US river basins via remote sensing, International Journal of Remote Sensing, 31, 3955–3978, 2010.

Gebremicael, T., Mohamed, Y., Betrie, G., 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, 2013.

Goovaerts, P.: Geostatistics for natural resources evaluation, Oxford university press, 1997.

Goovaerts, P.: Geostatistics in soil science: state-of-the-art and perspectives, Geoderma, 89, 1–45, 1999.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, Journal of hydrology, 228, 113–129, 2000.

Guntner, A.: Improvement of global hydrological models using GRACE data, Surveys in geophysics, 29, 375–397, 2008.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration, Journal of Hydrologic Engineering, 4, 135–143, 1999.

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

Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, Journal of Hydrology, 332, 144–157, 2007.

Hall, J., Grey, D., Garrick, D., Fung, F., Brown, C., Dadson, S., Sadoff, C., et al.: Coping with the curse of freshwater variability, Science, 346, 429–430, 2014.

Han, S.-C., Kim, H., Yeo, I.-Y., Yeh, P., Oki, T., Seo, K.-W., Alsdorf, D., and Luthcke, S. B.: Dynamics of surface water storage in the Amazon inferred from measurements of inter-satellite distance change, Geophysical Research Letters, 36, 2009.

Hay, L. E., Leavesley, G. H., Clark, M. P., Markstrom, S. L., Viger, R. J., and Umemoto, M.: Step wise, multiple objective calibration of a hydrologic model for a snowmelt dominated basin1, 2006.

Hong, Y., Hsu, K.-l., Moradkhani, H., and Sorooshian, S.: Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response, Water resources research, 42, 2006.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, Journal of Hydrometeorology, 8, 38–55, 2007.

Jarmain, C.: A Methodology for Near-real Time Spatial Estimation of Evaporation: Report to the Water Research Commission, Water Research Commission, 2009.

Jiang, D., Wang, J., Huang, Y., Zhou, K., Ding, X., and Fu, J.: The review of GRACE data applications in terrestrial hydrology monitoring, Advances in Meteorology, 2014, 2014.

Johnston, R. M. and McCartney, M.: Inventory of water storage types in the Blue Nile and Volta river basins, vol. 140, IWMI, 2010.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, Journal of Hydrometeorology, 5, 487–503, 2004.

Karlsson, K.-G., Riihelä, A., Müller, R., Meirink, J., Sedlar, J., Stengel, M., Lockhoff, M., Trentmann, J., Kaspar, F., Hollmann, R., et al.: CLARA-A1: a cloud, albedo, and radiation dataset from 28 yr of global AVHRR data, Atmospheric Chemistry and Physics, 13, 5351–5367, 2013.

Kebede, S., Travi, Y., Alemayehu, T., and Marc, V.: Water balance of Lake Tana and its sensitivity to fluctuations in rainfall, Blue Nile basin, Ethiopia, Journal of hydrology, 316, 233–247, 2006.

Kennedy, J. and Eberhart, R.: Particle swarm optimization, in: Proceedings of IEEE international conference on neural networks, vol. 4, pp. 1942–1948, Perth, Australia, 1995.

Kim, J. and Hogue, T. S.: Evaluation of a MODIS-based potential evapotranspiration product at the point scale, Journal of Hydrometeorology, 9, 444–460, 2008.

Kim, U. and Kaluarachchi, J. J.: Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia, Journal of Hydrology, 362, 39–56, 2008.

Kim, U. and Kaluarachchi, J. J.: Climate change impacts on water resources in the Upper Blue Nile River Basin, Ethiopia1, 2009.

Kim, U., Kaluarachchi, J. J., and Smakhtin, V. U.: Generation of monthly precipitation under climate change for the upper blue Nile river basin, Ethiopia1, 2008.

Kjærsgaard, J. H., Cuenca, R. H., Martínez-Cob, A., Gavilán, P., Plauborg, F., Mollerup, M., and Hansen, S.: Comparison of the performance of net radiation calculation models, Theoretical and applied climatology, 98, 57–66, 2009.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of Hydrology, 424, 264–277, 2012.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J.: The tropical rainfall measuring mission (TRMM) sensor package, Journal of atmospheric and oceanic technology, 15, 809–817, 1998.

Landerer, F. and Swenson, S.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resources Research, 48, 2012.

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D., Michel, D., and Wood, E. F.: The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data, Geoscientific Model Development, 9, 283–305, 2016.

Mellander, P.-E., Gebrehiwot, S. G., Gardenas, A. I., Bewket, W., and Bishop, K.: Summer rains and dry seasons in the Upper Blue Nile Basin: the predictability of half a century of past and future spatiotemporal patterns, PloS one, 8, 1932–6203, 2013.

Mengistu, D. and Sorteberg, A.: Sensitivity of SWAT simulated streamflow to climatic changes within the Eastern Nile River basin, Hydrology and Earth System Sciences, 16, 391–407, 2012.

Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, Geophysical Research Letters, 36, 2009.

Miralles, D., Holmes, T., De Jeu, R., Gash, J., Meesters, A., and Dolman, A.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453–469, 2011a.

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, Hydrology and Earth System Sciences, 15, 967–981, doi:10.5194/hess-15-967-2011, http://www.hydrol-earth-syst-sci.net/15/967/2011/, 2011b.

Mishra, A. and Hata, T.: A grid-based runoff generation and flow routing model for the upper Blue Nile basin, Hydrological sciences journal, 51, 191–206, 2006.

Mishra, A., Hata, T., and Abdelhadi, A.: Models for recession flows in the upper Blue Nile River, Hydrological processes, 18, 2773–2786, 2004.

Monteith, J. et al.: Evaporation and environment, in: Symp. Soc. Exp. Biol, vol. 19, p. 4, 1965.

Moore, R.: The probability-distributed principle and runoff production at point and basin scales, Hydrological Sciences Journal, 30, 273–297, 1985.

Moriasi, D., Arnold, J., Van Liew, 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.

Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sensing of Environment, 111, 519–536, 2007.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sensing of Environment, 115, 1781–1800, 2011.

Muskett, R. R. and Romanovsky, V. E.: Groundwater storage changes in arctic permafrost watersheds from GRACE and in situ measurements, Environmental Research Letters, 4, 045 009, 2009.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, Journal of hydrology, 10, 282–290, 1970.

Norman, J. M., Kustas, W. P., and Humes, K. S.: Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature, Agricultural and Forest Meteorology, 77, 263–293, 1995.

Paiva, R. C., Durand, M. T., and Hossain, F.: Spatiotemporal interpolation of discharge across a river network by using synthetic SWOT satellite data, Water Resources Research, 51, 430–449, 2015.

Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C., Allen, G. H., and Miller, Z. F.: Assessing the potential global extent of SWOT river discharge observations, Journal of Hydrology, 519, 1516–1525, 2014.

Pejam, M., Arain, M., and McCaughey, J.: Energy and water vapour exchanges over a mixedwood boreal forest in Ontario, Canada, Hydrological Processes, 20, 3709–3724, 2006.

Pimentel, D., Berger, B., Filiberto, D., Newton, M., Wolfe, B., Karabinakis, E., Clark, S., Poon, E., Abbett, E., and Nandagopal, S.: Water resources: agricultural and environmental issues, BioScience, 54, 909–918, 2004.

Priestley, C. and Taylor, R.: On the assessment of surface heat flux and evaporation using large-scale parameters, Monthly weather review, 100, 81–92, 1972.

Ramillien, G., Famiglietti, J. S., and Wahr, J.: Detection of continental hydrology and glaciology signals from GRACE: a review, Surveys in Geophysics, 29, 361–374, 2008.

Rientjes, T., Haile, A., Kebede, E., Mannaerts, C., Habib, E., and Steenhuis, T.: Changes in land cover, rainfall and stream flow in Upper Gilgel Abbay catchment, Blue Nile basin–Ethiopia, Hydrology and Earth System Sciences, 15, 1979–1989, 2011.

Rodell, M., Famiglietti, J., Chen, J., Seneviratne, S., Viterbo, P., Holl, S., and Wilson, C.: Basin scale estimates of evapotranspiration using GRACE and other observations, Geophysical Research Letters, 31, 2004.

Rodell, M., Chen, J., Kato, H., Famiglietti, J. S., Nigro, J., and Wilson, C. R.: Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE, Hydrogeology Journal, 15, 159–166, 2007.

Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F.: Reconciling the global terrestrial water budget using satellite remote sensing, Remote Sensing of Environment, 115, 1850–1865, 2011.

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

Scheerlinck, K., Pauwels, V., Vernieuwe, H., and De Baets, B.: Calibration of a water and energy balance model: Recursive parameter estimation versus particle swarm optimization, Water resources research, 45, 2009.

Schiemann, R., Erdin, R., Willi, M., Frei, C., Berenguer, M., and Sempere-Torres, D.: Geostatistical radar-raingauge combination with nonparametric correlograms: methodological considerations and application in Switzerland, Hydrology and Earth System Sciences, 15, 1515–1536, 2011.

Schulz, J., Albert, P., Behr, H.-D., Caprion, D., Deneke, H., Dewitte, S., Dürr, B., Fuchs, P., Gratzki, A., Hechler, P., et al.: Operational climate monitoring from space: the EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF)., Atmospheric Chemistry & Physics, 9, 2009.

Seleshi, Y. and Zanke, U.: Recent changes in rainfall and rainy days in Ethiopia, International journal of climatology, 24, 973–983, 2004.

Setegn, S. G., Srinivasan, R., and Dargahi, B.: Hydrological modelling in the Lake Tana Basin, Ethiopia using SWAT model, The Open Hydrology Journal, 2, 2008.

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., and McCabe, M. F.: Closing the terrestrial water budget from satellite remote sensing, Geophysical Research Letters, 36, 2009.

Sheffield, J., Wood, E. F., and Munoz-Arriola, F.: Long-term regional estimates of evapotranspiration for Mexico based on downscaled ISCCP data, Journal of Hydrometeorology, 11, 253–275, 2010.

Sheffield, J., Wood, E. F., and Roderick, M. L.: Little change in global drought over the past 60 years, Nature, 491, 435–438, 2012.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D.: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall, Bulletin of the American Meteorological Society, 81, 2035–2046, 2000.

Steenhuis, T. S., Collick, A. S., Easton, Z. M., Leggesse, E. S., Bayabil, H. K., White, E. D., Awulachew, S. B., Adgo, E., and Ahmed, A. A.: Predicting discharge and sediment for the Abay (Blue Nile) with a simple model, Hydrological processes, 23, 3728–3737, 2009.

Stehr, A., Debels, P., Romero, F., and Alcayaga, H.: 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.

Swenson, S. and Wahr, J.: Post-processing removal of correlated errors in GRACE data, Geophysical Research Letters, 33, 2006.

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., and Wilson, C. R.: Analysis of terrestrial water storage changes from GRACE and GLDAS, Water Resources Research, 44, 2008.

Tarpanelli, A., Brocca, L., Barbetta, S., Faruolo, M., Lacava, T., and Moramarco, T.: Coupling MODIS and radar altimetry data for discharge estimation in poorly gauged river basins, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8, 141–148, 2015.

Taye, M. T. and Willems, P.: Influence of climate variability on representative QDF predictions of the upper Blue Nile basin, Journal of hydrology, 411, 355–365, 2011.

Teferi, E., Uhlenbrook, S., Bewket, W., Wenninger, J., and Simane, B.: The use of remote sensing to quantify wetland loss in the Choke Mountain range, Upper Blue Nile basin, Ethiopia, Hydrology and Earth System Sciences, 14,(12), 2010.

Tekleab, S., Uhlenbrook, S., Mohamed, Y., Savenije, H., Temesgen, M., and Wenninger, J.: Water balance modeling of Upper Blue Nile catchments using a top-down approach, Hydrology and Earth System Sciences, 15,(7), 2011.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., and De Roo, A.: Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin, Journal of Hydrology, 499, 324–338, 2013.

Uhlenbrook, S., Mohamed, Y., and Gragne, A.: Analyzing catchment behavior through catchment modeling in the Gilgel Abay, upper Blue Nile River basin, Ethiopia, Hydrology and Earth System Sciences, 14, 2153–2165, 2010.

Van Dijk, A. I., Brakenridge, G. R., Kettner, A. J., Beck, H. E., De Groeve, T., and Schellekens, J.: River gauging at global scale using optical and passive microwave remote sensing, Water Resources Research, 52, 6404–6418, 2016.

Viste, E., Korecha, D., and Sorteberg, A.: Recent drought and precipitation tendencies in Ethiopia, Theoretical and Applied Climatology, 112, 535–551, 2013.

Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., and Higdon, D.: Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, International Journal of Nonlinear Sciences and Numerical Simulation, 10, 273–290, 2009.

Wale, A., Rientjes, T., Gieske, A., and Getachew, H.: Ungauged catchment contributions to Lake Tana's water balance, Hydrological processes, 23, 3682–3693, 2009.

Wang, H., Guan, H., Gutiérrez-Jurado, H. A., and Simmons, C. T.: Examination of water budget using satellite products over Australia, Journal of Hydrology, 511, 546–554, 2014.

Wosenie, M. D., Verhoest, N., Pauwels, V., Negatu, T. A., Poesen, J., Adgo, E., Deckers, J., and Nyssen, J.: Analyzing runoff processes through conceptual hydrological modeling in the Upper Blue Nile Basin, Ethiopia, Hydrology and Earth System Sciences, 18, 5149–5167, 2014.

Zambrano-Bigiarini, M.: hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series, R package version 0.3-7, 2013.