# Rebuttal Letter related to manuscript hess-2016-290:
# “Water budget modeling of the Upper Blue Nile basin using the JGrass-NewAge model system and satellite data”

By Wuletawu Abera; Giuseppe Formetta; Luca Brocca and Riccardo Rigon

Dear Editor Professor Dominic Mazvimavi,

We would like to thank you for your comments and suggestions, which gave us the opportunity to further 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.

**Editor’s comment:**

**Comments to the Author:**
**The authors have not adequately addressed the comments of the reviewers. I agree with the comments provided in the second review that some important issues have not been addressed. The manuscript still contains a lot of unclear statements or phrases. Some specific comments are given below.**

Dear Editor,

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

---

**1. The authors should avoid paragraphs made up of one sentence.**

This was done. There are no more one-sentence paragraphs in the new manuscript. They were joined as necessary.

**2. The authors need to check the correct usage of the colon and semi-colon.**

Usage of the column and semicolon was revised.

**3. The authors still maintain the phrase that their model estimates “all components” of the water budget, when they are only dealing with precipitation, evapotranspiration, runoff, and subsurface storage of water. Other components of the water budget such as interception, groundwater recharge, groundwater**

**discharge, etc. are not covered in the model. One of the reviewers pointed out this, but the revised manuscript still uses “all components”.**

We are sorry for having neglected this. We modified the manuscript by specifying which components of the budget we were talking about. However, we believe that what the water budget is composed depends on the control volume used to estimate it. At the scale we are working the fluxes are those we estimate with the exception of the groundwater flow at the outlet, which is assumed negligible. In our case, interception is an internal flux and does not appears in the budget, as well as groundwater recharge.

**5. What does “To this scope” in Line 7 of the Abstract mean?**

The scope we are talking about in the old manuscript is to develop a new methodology about improving the estimation of the water budget components that appears in our equation (1). We changed the wording as: “To obtain the water budget closure”

**6. Line 12 of the Abstract, what does “long term mean budget” mean?**

The words “long term” are unessential, and we cut them out.

**7. Line 13 of the Abstract, what does “Evapotranspiration covers 56% of the yearly budget” mean? Same applies to the % given in this line**

Possibly we do not properly interpret this observation. The symbol “%” means “per cent or percentage” (https://en.wikipedia.org/wiki/Percentage). This was specified at the first appearance and the symbol was replaced everywhere with this phrase. In place of “budget” we wrote “water budget” to mean percentage of what. In case if the word “cover” is the problem, we replace with “account for”. So now the sentence is “Evapotranspiration accounts for 56% (per cent) of the annual water budget, runoff is 33%, storage varies from minus 10% to plus 17% of the water budget.”

**8. Page 1, line 21, what does “2000 mm per year” refer to? Presumably precipitation. It is also possible to receive inflows equivalent to 2000 mm/yr**

Yes, the reviewer is correct, it refers to precipitation. It was added to the text: “of precipitation”.

**9. Page 1 line 22, the claim that the basin is one of the most complex in the world has no basis.**

The six points listed in the old manuscript are the base for saying UBN basin is probably one of the most complex. But we removed the phrase and revised the paragraph as suggested. (See specific comment 11)

**10. Page 1, last line, 85% of what?**

It refers to discharges, and now this is specified in the text.

**11 Page 2, lines 1 – 8 are not well written. What do you mean by "diplomatic discussions" when referring to management of transboundary water resources?**

We changed the whole paragraph as follows:

"In Ethiopia, UBN is inhabited by 20 million people whose main livelihood is subsistence agriculture (Population Census Commission 2008). The Ethiopian government, therefore, 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. However, as the principal contributor (i.e 51% of discharge) to the main Nile basin, UBN also supports hundreds of millions of people living downstream, and it is referred to as the "Water Tower" of northeast Africa. Therefore UBN is a part of trans-boundary river, and its development and management require obtaining agreements between many national governments and also non-governmental organizations, each involving different policies, legal regimes, and contrasting interests."

**12. Page 2, line 9, what are you referring to by "all these facts".**

We mean the facts and challenges enumerated in the previous paragraph (in the old version). However, in the new manuscript, the entire paragraph was rewritten, and this particular sentence is revised as:

"Tackling all these complexities and developing better water resource development strategies is only possible by gathering quantitative information (Hall et al., 2014)."

**13. Page 3, line 9, what are you referring to by "aforementioned problems"?**

We specified as: "aforementioned management problems by resolving"

**14. Page 3, line 10, "resolve the water budget" is not a very clear phrase. If you intend to estimate the components of the water budget, then clearly state this.**

This was accomplished, as can be seen, in the new version of the manuscript. It is re-written: "It obtains, at relatively small spatial scales and at daily time step, groundwater storage, evapotranspiration, discharges in such a way to satisfy the water budget equation."

**15. Page 3, line 11, which previous studies do you wish to improve?**

Clearly the studies are those cited in the paper (page 2, line 9-23), where we collect all what is available, to our knowledge, for the study area. And the limitation of those papers are mentioned on page 2, line 12-23. However, mentioning "previous studies" is unessential here and was eliminated. These improvements were already highlighted in Conclusions.

**16. Page 3, line 11-12, be explicit about the components of the water budget that you are attempting to estimate.**

In the revised manuscript we are more explicit (see the specific comment 14).

**17. Page 3, line 24, you state that the basin covers "17% of the total area of the country", which country are you referring to since you are dealing with a transboundary river?**

Country is clearly Ethiopia, since the Blue Nile area studied is entirely in Ethiopia, even if the Nile is a transboundary basin. We correct the text accordingly.

**18. Page 3, line 26, you state that "the UBN basin has the lion's share of the total Nile flow". Do you mean that this basin receives or gets the largest amount, because this is what this phrase means?**

We change the phrase as: "Since the UBN basin gives the largest contribution to the total Nile flow, it is the economic mainstay of downstream countries (i.e. Sudan and Egypt)."

**19. Page 3, line 29, what is a topographic distribution?**

We mean the topography, so we changed the phrase into:
"the maps of elevation"

**20. Page 4, is the information given in line 1 - 8 relevant to this paper?**

The section was describing what controls regional precipitation. In the revised manuscript, we removed it accordingly.

**21. Page 4, Equation (1), while authors have the liberty to use any letters or symbols to represent variables, it helps readers if authors use commonly used letters and symbols. It is common practice to represent precipitation by P. the use of J is rather unusual, although there is nothing wrong with this. This just gives readers a hard time remembering what J represents in your paper.**

We understand this. However, as stated before, we adopted J for precipitation to be consistent with other papers of our research group, and some of the papers cited in the manuscript (for instance, Rigon et al. 2016, Abera et al. 2017).

**22. Page 5 line 1, it would have been much easier to refer to Qki(t) as inflows from upstream HRUs.**

We have changed the text from "Qki(t) is the discharge from the contributing streams." to "Qki( t) are inflows from upstream HRUs."

**23. Page 5, reviewers highlighted that it is not clear whether the "JGrass-NewAge system" is a new model the authors developed or are utilising. The revised manuscript does not enable readers to understand whether this is a modelling system that the authors have developed or not.**

List of papers are cited showing the development of JGrass-NewAge system. Some of the authors developed the model system, but in this manuscript an application of the model system is described. We believe that in this paper we did not mention anything that indicates the development of the model system. The focus of the paper is better described also in the general answer to Reviewer #1.

**24. Page 6, line 5, what were the criteria used to define a Hydrological Response Unit. The definition of an HRU will differ depending on the type of model used, and the problem to be solved. As this is the fundamental modelling unit, the manuscript should have provided this information. It seems that the authors subdivided the basin into sub-basins using a DEM. These sub-basins cannot be said to be HRUs. An HRU has a very specific meaning and not necessarily the same as a sub-basin.**

While the term "subbasin" and "HRU" are defined in various models in different ways, here we mean that the subbasin partition used for modeling. In our paper it is the computational unit assumed to be homogeneous in estimating the terms of the water budget we deal with.
The level of details of our HRUs is based on the problem under investigation and the data availability. For instance, if forcing input is available at few meter resolutions or enough in-situ observation density is available to generate spatially interpolated input data, then the HRU can be as small as few kilometers.

The general procedure to obtain sub basins (HRUs) is more or less the same in all GIS environments. The partitioning of the basin into subbasin particular to specific procedures for JGrass-NewAge is described in detail in other studies (Formetta et al., 2014 and Abera et al., 2014). So we thought that adding description of each step of subbasin extraction is not necessary for this paper.

**25. Page 6, line 8, what was the rationale for dividing the basin into 402 sub-basins? Reviewers made the same comment which has not been addressed. Is a sub-basin**

**similar to an HRU? In some models, a sub-basin is made up of possibly more than one HRU.**

The rationale for dividing the basin into 402 subbasin is to capture the spatial heterogeneity of precipitation data provided by the satellite product used. The following sentence was already available (page 6: line 9-10) to describe this rationale:

"This spatial partitioning may not be the finest scale possible, however, considering the size of the basin and model input data resolutions, and the resolution of satellite products, it can be considered an acceptable scale to capture the spatial variability of the water budget."

However, to make our thinking more clear, we wrote in the revised manuscript:

"This spatial partitioning may not be the finest scale possible, however, considering the size of the basin, it is consistent with input data resolution, including satellite products, meaning that a finer subdivision would imply uniform inputs for adjacent HRUs, and a coarser one would average out inputs variability."

Yes, in SWAT HRUs are constituted of a subbasin. In NewAge, HRU and subbasin are the same and are interchangeably used in this manuscript. For clarity, the following sentence has been added in section 3.1:

"In this paper, the term HRU actually identifies subbasin."

**26. Page 3 Section 3.2 Precipitation. The authors should have provided information about how accuracy of the precipitation product they used. Provide the error level established in similar environment by previous studies.**

The precipitation product used (SM2R-CCI) is assessed in various regions. SM2R-CCI for the particular study area is already the subject of our previous paper (Abera et al 2016).

In the original MS this information was given as:
"Recently Abera et al. (2016) compared five of them with high spatial and temporal resolutions 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."

Based on the suggestion given to add more information on quality of SMR-CCI precipitation product, more citations are added as follows:

"Currently there are several satellite rainfall estimates (SREs) available for free, and Abera et al. (2016) compared five of them with high spatial and temporal resolutions 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. Regards to the quality of

SM2RAIN-based products, recent studies positively assessed their accuracy on a regional (Brocca et al. 2016; Ciabatta et al. 2017) and a global (Koster et al. 2016) scale."

**27. Page 6 and 8, does ET in Equation (2) refer to reference evapotranspiration? In Equation (3) ET(t) seems to represent actual evapotranspiration. You need to clearly state this as you are using almost the same letters to represent different variables.**

Thank you for this, we have changed *ET* in equation 2 into *PET* this to represent the Priestly and Taylor potential evapotranspiration, and it is given in the description of the equation terms.

**28. Page 8, line 6, S(t) is not groundwater but subsurface storage of water which includes soil water. Smax should be maximum subsurface storage capacity, not just maximum storage capacity.**

We changed the text accordingly to the Editor suggestions. However, in our case, S(t), by construction, includes also the groundwater storage.

**29. Page 8, line 11, how was that automatic calibration done? What was the objective function used?**

Particle swarm (PS) optimization is used to optimize both the rainfall-runoff and ET alpha parameters. It was used to fit the discharge observations and optimize ET to close the water budget after a certain time, named Budyko time (and indicated by $T_B$). Essentially, two objective functions are used, KGE for optimizing discharge, S(t) = 0, for closing the water budget after $T_B$. The description for discharge calibration was given at page 9, lines 14-19. In this section, the calibration procedures for two parameters of ET, $\alpha_{PT}$ and Smax , were described instead. It was:

"Once $T_B$ is fixed, particle swarm (PS) automatic calibration can be set to produce the set of parameters, including $\alpha_{PT}$ and Smax, for which, besides discharge is well reproduced, is also S(TB) = S(0) ."

Now we have revised the sentence
"Once $T_B$ is fixed, automatic calibration set to produce the set of parameters, including $\alpha_{PT}$ and Smax, for which, besides discharge is well reproduced, is also S(TB) = S(0)."

**30. Page 9, 3.4 Discharge. The authors should have given a basic description of the ADIGE model used. It does not help readers to refer to a paper that has not been published. The reviewers highlighted this problem, and this has not been addressed.**

ADIGE model is based on the well known Hymod model, and we thought that detailed description of the model was not useful in this paper. The paper Abera et al. 2017 is now published on-line in Advances in Water Resources (and scheduled for the June printed version of the Journal). However, based on the reviewer suggestion, we have added some level of details for Hymod description in the Appendix A, and cited the Appendix in the main text as follows:

At the main text, it was given as follows:

"Detailed descriptions of HYMOD implementations in the NewAge model system are given at Formetta et al. (2011) and Abera et al. (submitted)."

And now it is revised as follows:
"Detailed descriptions of HYMOD implementations in the NewAge model system are given in Formetta et al. (2011), Abera et al. (2017) and summarized in Appendix A"

Then, in appendix A, the following details about Hymod has been given:

"The NewAge system executes one Hymod model for each HRU, and routes water downslope. A detailed description of Hymod model is provided in many previous studies (Moore, 1985; Van Delft et al., 2009; Boyle et al., 2001; Formetta et al., 2011). In Hymod, each HRU, is supposed to be a composition of storages of capability *C[L]* according to distribution (Moore, 1985):

$$F(C < c) = 1 - (1 - \frac{c}{C_{max}})^{B_{exp}} \qquad \text{(A1)}$$

Where *F(C)* represents the cumulative probability of a certain water storage capacity (*C*); $C_{max}$ is the largest water storage capacity within each HRU and $B_{exp}$ is the degree of variability in the storage capacity. As shown in the schematic diagram (figure 11), the precipitation exceeding $C_{max}$ is send directly to the volume generating surface runoff. If we call the precipitation volume in a time interval $\Delta$t , J(t) := P(t) $\Delta$t , then this "direct" runoff can be estimated according to:

$$R_H(t) = \max(0, J(t) + C(t) - C_{max}) \qquad \text{(A2)}$$

Where *C(t)* defines the fraction of storages already filled at time t . The latter equation is true for any precipitation and storage level, even when the maximum storage $C_{max}$ is not exceeded. When precipitation does not exceeds $C_{max}$ runoff volume can be produced by filling some of the smaller storages. To which extent this happens, can be derived by the knowledge of the storage distribution, eq. (A1), the initial storage *C(t)* and the precipitation *J(t)*. This residual runoff is, in fact, given by:

$$R(t) = \int_{C(t)}^{\min(C(t)+J(t), c_{max})} F(c)\, dc \qquad \text{(A3)}$$

An analytic expression for the integral in eq. (A3) is available, which makes the computation easier. Water in storage is made available to evapotranspiration. Water going into the runoff volume, i.e. *R(t)* and *RH(t)*, is further subdivided into a surface runoff volume and subsurface storm runoff. Surface runoff, in turn, is composed by the whole of $R_H(t)$ and part of *R(t)* , and *R(t)* is split according to a partition coefficient *α* such that the part $\alpha R(t)$ goes into surface runoff volume and $(1-\alpha)$ into the subsurface runoff volume. In Hymod, $\alpha$ is a calibration coefficient.

Finally, surface runoff volumes are routed through three linear reservoirs, and, subsurface storm runoff volume is routed through a single linear reservoir. A summary of equations for the surface runoff is therefore:

$$\frac{dS_1(t)}{dt} = \alpha R(t) + R_H(t) - kS_1(t) \qquad Q_1(t) = \frac{S_1(t)}{k} \tag{A4}$$

where $S_1[L^3]$ is the storage in the first of the linear reservoirs, and $k[T]$ is the mean residence time in each of the reservoirs. Then:

$$\frac{dS_i(t)}{dt} = Q_{i-1}(t) - kS_i(t) \quad Q_i(t) = \frac{S_i(t)}{k} \tag{A5}$$

for the other two reservoirs, where $S_i$ [L] with i = 2;3 is the storage in the two remaining surface reservoirs. Subsurface storm runoff is then modeled by:

$$\frac{dS_{sub}}{dt} = (1-\alpha)R(t) - k_{sub}S_{sub}(t) \tag{A6}$$

where $S_{sub}$ $[L^3]$ is the storage in the subsurface storm-flow system and $k_{sub}[T]$ is its mean residence time. A budget equation can be written for the groundwater system as:

$$\frac{dS_g(t)}{dt} = (J(t) - R(t) - R_H(t)) - ET(t) - Q_g(t) \tag{A7}$$

where $S_g(t)[L^3]$ is the groundwater storage, and $Q_g(t)$ the groundwater flow which becomes surface flow at the closure of the HRU.

Summarizing, Hymod subdivides each HRU into three reservoirs: a groundwater reservoir, from where evapotranspiration and groundwater flow is allowed, a subsurface storm-water reservoir, and a surface runoff reservoirs set. Partition of precipitation into the three reservoirs is obtained by a calibration coefficient ($\alpha$) and the use of a probability distribution function of storages' capacity, $F(c)$.

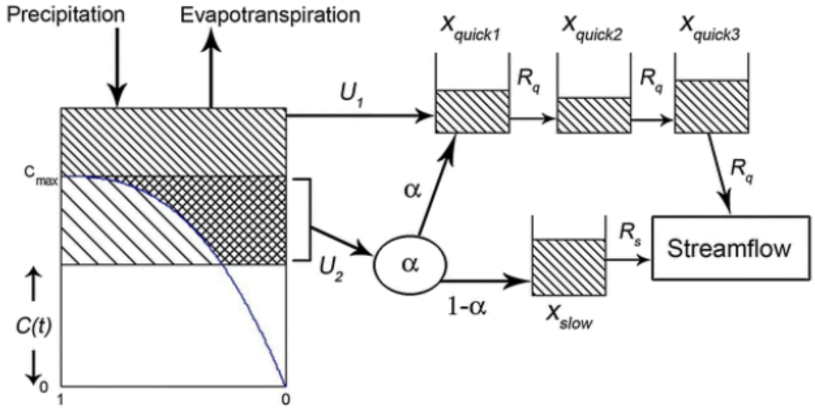

**Figure 11.** Schematic diagram of hymod model (adapted from Van Delft et al. (2009))

”

**31. Page 9, line 18, explain the exact objective function used. At this stage, readers do not have knowledge about which parameters were being calibrated.**

In the revised MS, we have added brief description of Hymod model and its parameters in the appendix section, and this is cited on page 9 line 13.

It was
“The ADIGE rainfall-runoff has five calibration parameters, and the calibration is performed using the particle swarm (PS) optimization.”

It is revised as follows:

“The ADIGE rainfall-runoff has five calibration parameters ($C_{max}$, $B_{exp}$, $\boldsymbol{\alpha}_{Hymod}$, $R_s$, $R_q$, see the details in Appendix A), and the calibration is performed using the particle swarm (PS) optimization method.”

**31. Page 10, line 6,how many rainfall stations were used to correct the precipitation estimates? Was this done in this study?**

In the basin and around, there are about 33 stations, and the procedure of error correction for various satellite precipitations was presented in our previous paper (Abera et al. 2016). Here, the same procedure is applied to SM2R-CCI that is used as model input.

**32. Page 10, line 12 -13, be explicit regarding the data used for validation. Reviewers highlighted this problem. How many river flow stations were used for validation? How were they selected?**

For discharge validation, we used all the available measurement stations, except the one at the basin outlet (Ethiopia-Sudan border), which is used for calibration. They are about 15 stations. They are portrayed at page 4 (figure 1a). Validation results at page 15 (table 4). We believe this could be enough but, we can highlight the number of stations used for validation in this particular sentence as follows:

It was:
"In addition, the simulation of NewAge at the internal links is validated where in situ data are available."

Now its is revised as follows:

"In addition, the simulation of NewAge at the internal links is validated where in situ data are available (15 discharge measurement stations)."

**33. Page 12, Figure 3b. The use of % share to express amount of rainfall received in a season is rather confusing. Is the seasonal contribution important? This is also done on Page 11 when describing precipitation**

Yes, seasonal variability is really important, probably more important than spatial variability. The contribution of each season to the total amount of water received is a topic on which other studies are less explicit about.

**34. Page 13, given that the estimation of actual evapotranspiration involves some uncertainty, which ET rates are acceptable, GLEAM or your modelled values, and why?**

We believe that given (1) improved procedure to chose the best rainfall satellite products (i.e. based on Abera et al. 2016), (2) the bias correction procedures applied, (3) the good model performances obtained at the 15 internal stations where discharge is measured, (4) the procedure of water budget closure to conserve mass, and (5) the independent evaluation using GRACE storage change, our modeled ET is expected to be more accurate than GLEAM estimates.

**35. Page 14, Table 3, since the structure of the model has never been presented, a reader does not have an idea about the meaning of the parameter values given in this table.**

As mentioned above, we have added brief note about Hymod model and its parameters in Appendix A. The following phrase has been added to Figure's caption: "Parameters' physical meaning is explained in Appendix A"

**36. Page 15, the manuscript describes some forecasting done. It is not clear how this was done. Reviewers also raised this point.**

The word “forecasting” is used to mean that given precipitation estimates and the model calibrated at some location (in our case the outlet), we give predictions of the water discharge at internal basin points. However, this comment tells us that it could provide misunderstanding to the reader; hence, we changed “forecasting” with “simulation”.

---

**Reviewer's comment:**

**The authors have addressed some comments/suggestions and made changes to the paper, which has improved its quality considerably. However, I still have some major concerns:**

**1.a Based on the following response and the revised manuscript, I cannot agree with the authors that their contribution was linked all the components in one model.**

Obviously, if the contribution of this paper had been just "linking components" our contribution would have been poor. As we stated better in the revised manuscript, the contribution of this paper is to estimate the water budget flows and storages in a semi-distributed way in a poorly gauged area with the help of sound modeling and remote sensing products (RS). Sound modeling means, in our case, using JGrass-NewAge modeling components, each one tested separately in several other studies, and search for the closure of the water budget. The latter information, in our opinion, adds credibility and consistency to ours results.

Specifically to UBN,

- Such tasks were never accomplished in literature at the spatial and temporal resolution we use, if not for limited parts of the basins.

In order to obtain this, we used all the resources types one can have:

- Various remote sensing precipitation products (eventually using a single one after a comparative study, i.e. Abera et al., 2016), supported by 33 in situ meteorological stations.
- Discharge observation at 16 locations,
- Evapotranspiration from two satellite products (MODIS and GLEAM),
- And storage change at the basin scale from GRACE RS.

We claim that no better can be done for UBN in this moment, and, vice-versa, our study can be the seed to promote more measurements to support more refined studies.

- Besides we claim that the procedure we followed can be easily transported in any other poorly gauged basin.

We added to the Conclusions the following statements:

"The study covered 16 years from 1994-2009 at a finer spatial and temporal resolution than in previous studies. In order to obtain this result, we used, various remote sensing products, rainfall from SM2R-CCI, cloud cover from SAF EUMETSAT CFC, evapotranspiration from GLEAM and MODIS (used for comparison), and storage change from GRACE. We also used all the ground data currently available, i.e. sixteen discharge time series, and thirty-five ground based meteorological stations."

And also:

"We claim that the procedure we followed can be easily transported in any other poorly gauged basin with benefits for the hydrological knowledge of any region on Earth."

**1.b Because there's no field observation to validate the modeled water components (except runoff) in this manuscript, evapotranspiration (ET) was estimated using the Priestley and Taylor (PT) Formula. But GLEAM, which was also estimated based on the PT method, was adopted to validate the estimation. If estimates of each component have large uncertainty, who believe the final output?**

Saying that GLEAM estimates evaporation with Priestly Taylor (PT) equation is true, but, at the same time, an approximation, because what characterizes any of the cited products is not only the choice of PT equation but the way its terms and parameters are evaluated.

GLEAM uses PT estimates for potential ET (PET), but the alpha coefficient is set to a particular value (0.98) got from literature, and above all, potential ET is additively increased, by intercepted rainfall estimated according to a version of the Gash model, and multiplicatively decreased by a "stress coefficient" made dependent on five soil cover types (bare soil, snow, tall vegetation, two levels of low vegetation). The stress coefficients are evaluated through various remote sensing products (RS), which are described in the paper by Martens et al., 2017, and applied to a "three reservoirs schematization" on any pixel of 0.25 degree in latitude and longitude (~28 km of side or ~800 square kilometers of area). Eventually also the radiation term is estimated by another different RS.

On the contrary, in our case, the alpha coefficient is estimated by closing the water budget, our stress coefficient does not depend on vegetation type (which could be seen as a limitation) but only on the overall water content in the HRU (of average size of ~420 square kilometers, approximately twice the resolution of GLEAM data). Radiation in our case is estimated locally (and averaged over a representative number of points for each HRU) according to various topographic features present in the HRU (slope, aspect, sky view factor, shadows).

In turn, while we have no means to validate PET or AET on UBN sites, GLEAM was validated in various (64) Fluxnet sites around the world with consistent results (e.g. Martens et al., 2017). Many other differences between GLEAM and JGrass-NewAge can be easily deduced from reading our and Martens paper.

Therefore we can conclude that:

- As an educated best guess, GLEAM supposedly behaves properly in Ethiopia with statistics similar to those it reproduces for other parts of the world, and therefore, it can be used as reference
- GLEAM and our model can be thought as quite independent estimates of the same quantity, even if using the same PT formula, and, therefore, the comparison between our results and GLEAM result does not produce agreements just because they use the same algorithm, as the comment of the Reviewer could imply.

Certainly the word "validation" has to be probably substituted by "comparison" . In the revised manuscript we added part of this information.

We added the following statements in the new manuscript:

“GLEAM, as well NewAge, uses the PT scheme for estimating ET. However, all inputs of the formula in GLEAM and NewAge are according to different strategies and RS tools. GLEAM sets $\boldsymbol{\alpha}_{PT} = 0.98$ while in NewAge it has been calibrated. In GLEAM PET is additively increased, by intercepted rainfall estimated according to a version of the Gash model (Gash, 1979), and multiplicatively decreased by a stress coefficient depending on five soil cover types (bare soil, snow, tall vegetation, two levels of low vegetation) and has a different expression for anyone of the storages. Moreover, according to the case, the stress coefficients are evaluated through various RS, according to procedures which are described in the paper by Martens et al, 2016.”

“The most recent version of GLEAM was validated globally over sixty-four Fluxnet sites (Martens et al., 2016) with consistent results, letting us guess that it behaves properly also in Ethiopia. The differences between NewAGE estimation and GLEAM's one allow to assume that the our results and their results can be seen as largely independent.”

**2. As mentioned in my previous comments, the discussion section should be enhanced. However, the authors revised limited content in section 4 (Results and Discussion) in this revision (track change in ‘hess-2016-290-author_response-version1.pdf’). For example, it seems that the authors kept discussing some water components in other submitted manuscript. Since the authors think combine modeling all the water components is important, why not focus on this study and discuss the possible limits of such a solution? I know the study area is data-scarce and validation using field observation is impossible. However, reasonable discussion on the comparison and the possible errors may persuade the readers (e.g., see specific comments 5, 9, 11, 14, 15, and 18 for detail). Otherwise, the readers may question about the modeled water components and finally your method.**

We thank the reviewer for the suggestions given to enhance the discussion section, and we are sorry if we did not adequately treat all the comments given in the first review. In this version of the MS, we used the specific comments mentioned to enhance our discussion. Detailed comments are reported along with points 5,9,11,14,15 and 18.

**3. Both reviewers mentioned figure quality. However, the authors may have paid little attention. For example, although units were added to axes in Fig. 1, units for axes were still missing in the rest of figures (e.g., Figs. 3, 4, 7, 8).**

The label latitude and longitude is enough to tell that the maps are given in geographical coordinate system, representing the axis in degree. But, given the reviewer suggestion, we have added the unit (degree) in all the figures.

**Following specific comments may help the authors understanding my concerns and improving the manuscript.**
**The numbers in front of the comments indicate page and line number.**

**1.1-6. 'to obtain the estimates of all the components of the hydrological cycle (precipitation, evapotranspiration, discharge, and storage)'. It would be better to revise the claim, because precipitation, evapotranspiration, discharge, and storage are main components. For example, interception and infiltration are also components of the hydrological cycle, which you did not address.**

We are sorry for having neglected this. We modified the manuscript by specifying which components of the budget we were talking about. However, we believe that what the terms present in the water budget depend on the control volume used to estimate it. At the scale we are working the fluxes are those the Reviewer list among parentheses with the exception of the groundwater flow at the outlet, which is assumed negligible. In our case, interception is an internal flux and does not appears in the budget, as well as groundwater recharge.

**2. 4-3. 'Specifically studies' contains grammatical errors. I encourage the authors to check the entire manuscript carefully to avoid such mistakes.**

This paragraph is deleted following editor's specific comment 20.

And we have checked the whole manuscript for any grammar errors.

**3. 4-8. Please correct the unit for temperature.**

Thank you, we corrected the temperature unit. We have changed $18.5^{o}$ to $18.5^{o}$C.

**4. 5-3. Table 1. I cannot found any description on 'JAMI' or 'three temperature' in the manuscript. If you decide putting such information in the manuscript, please be sure the readers can understand it or find relate context.**

Thank you, it is true that there is no literature about JAMI, hence we removed it from the manuscript. The "three temperature" also revised as "Modelling snow melting using three types of temperature and radiation based algorithms."

**5. 8-5. Eq. (3). In table 1, the author said PT method was used as one method to**

**estimate ET. However, 'S(t) and Smax' were added to Eq. (3). Is Eq. (3) valid for this study or is it used for all JGrass-NewAGE application? Furthermore, what's the advantage of using water storage information when estimating ET, especially in data-scarce regions?**
**In addition, GLEAM estimates are based on the PT method. Can it be used to 'validate' Eq. (3)? Which version of GLEAM did you use?**

The addition to a "stress term" to Priestly-Taylor equation is, in fact quite a common practice. This has been popularized, for instance in works by Rodriguez-Iturbe and coworkers, since 1999, but has its roots in the work of Feddes and coworkers. GLEAM itself uses a stress term, even if its estimation differs from our.

We added the following phrase to the revised manuscript:
"The ratio *S(t)/Smax* a stress coefficient which became very popular since the work of Feddes and coworkers (Feddes et al., 2001)."

In our case, we already used previously a similar scheme in Abera et al., 2017. PT formula in fact is used in literature for estimating mostly potential evapotranspiration, while when dealing, as in our case with the actual evapotranspiration, the action of resistances has to be accounted for.

In our case, modeling surface water with a customized version of HYMOD [Moore et al, 1985] we need an estimation of the actual evapotranspiration, AET, to be subtracted to the water available in storages to produce runoff.

Regarding the GLEAM ET in the manuscript we used GLEAM_v3a_BETA. As said with more detail at the beginning of these answers, we believe that GLEAM estimates are quite independent from ours. This and other information was added in the main text.

**6. 9-9. Are 'the ADIGE model', 'the well-known HYMOD model', and 'The NewAge Hymod' the same? If yes, please consider unify the description.**

ADIGE model uses HYMOD model for each HRU. Therefore, Adige model is an assembly of HYMODs, each one for HRU. Detail notes about the Adige model is available in the papers cited. So, we revised the text to clearly show this idea.

**7. 9-12. I'm confused by the description that 'The main inputs for the ADIGE model are J(t) and ET(t)', because 'Q is modelled as functions of basin water storage'.**
**How did you get the water storage? What are the five calibration parameters?**

We have added Appendix A to describe better the ADIGE model and the parameters. Water storage here is assumed to be the sum of the water contained in all the HYMOD storages, for all the HRUs.

**8. 10-17. ET validation is questionable. See specific comment 5.**

Thank you. We already expressed our opinion in answering the general comment. However, we have removed the term “validation”, and replaced with “comparison”. Please see reply for major comment 1, and for specific comment 5.

**9. 10-21. ds/dt validation. Similar to ET validation, lacking of discussion on GRACE product and the modeled ds/dt. Did you use the GRACE product directly or perform any correction? As reported by studies (e.g., Long et al., 2015, Water Resour. Res., 51, 2574–2594), GRACE data are noisy in smaller basins less than the GRACE footprint (~200,000 km2), as well as in areas with intensive irrigation. Considering the UBN is approximately 176,000 km2 and the highlands have high water demands for irrigation, the product may include typical errors. The author should discuss such uncertainty and the possible impact on the modeled ds/dt.**

The description about GRACE available at page 9 (line 24-33) and page 10 (line 1-4) is enough from our perspective. The area of UBN basin is almost the same size of GRACE footprint (176,000 km2), and, in fact, we used the comparison at the whole basin scale not at subbasin scale, as made more explicit in various parts of the new manuscript. The UBN basin has high water demands for irrigation, but actual irrigation is very minimal, and therefore not affecting our estimates.

**10. 10-26. The headings in section 4 are the same as those in section 3, and the authors claimed that in this way there is a clear relation between the topics of the two sections. I cannot agree with them. For example, in rainfall section, the spatial distribution was described and then compared with some published results. But for ET section, both spatial and temporal distributions were presented, as well as ‘validation’. I don’t think the headings in results section reflecting any useful information.**

We are sorry if we were not clear when we say that the one-to-one correspondence of the title in the method and in the result sections. But, the idea is simple. In the methodological section, we presented how each component is estimated for each HRU level, and in the result section we analyzed and discussed the result of the component use. Of course, the discussion may not follow exactly the same pattern for four water budget components. For instance, since we have some published work, for precipitation, in addition to the spatial and temporal analysis, comparison of our estimates with some published work is to strengthen our discussion. But, when it comes to ET, we do not have studies in the area, and we instead focus on comparison of modeled ET with GLEAM. We are sorry if we still miss the reviewer’s idea!

**11. 11-1. Table 2. The unit for SM2R-CCI’s spatial resolution is missing. Could you provide time periods for these data used? Little information about the used data was presented.**
**In section 4.1, can you discuss what’s the difference between corrected and uncorrected SM2R-CCI products? That is how systematic error (bias) of SM2RCCI affects rainfall amount, as some ungauged basin has no in-situ observation to perform the correction.**

Thank you, we have added unit ‘degree’ for SM2R-CCI’s spatial resolution. The time period used for the products is given in the text, for instance, SM2R-CCI (in the old manuscript page 6, line 21), for CFC (old manuscript page 8, line 24), and for GLEAM (old manuscript page 10, line 19). However, based on the suggestion, and for clarity, in the revised manuscript, we have added a column in Table 2 for providing the time period used.

The impacts of bias correction on rainfall estimation and on discharge estimation are not the subject of this paper, and can be reviewed from other studies (Habib et al 2014; Gumindoga et al. 2016; Bitew et al 2012; Najmaddin et al. 2017; Valdés-Pineda et al, 2016; among many). Regards to the impacts on discharge, for instance, in Koga basin and Gilgel Abbay (subbasins of the UBN basin), Bitew et al. 2014 and Habib et al. 2014 showed that bias correction procedure improved model discharge prediction performances for all satellites used (CMORPH, PERSIANN, TRMM- 3B42RT, TRMM-3B42).

**12. 14-6. The section is to ‘validate’ NewAge ET. I’m not sure why the authors talk about GLEAM estimation.**

Given that NewAge estimates ET consistently with the water budget closure, a sentence on GLEAM from this perspective does not harm, in fact, it could be a useful suggestion for improvement of GLEAM in the region.

**13. 14-10. A better correlation between NewAge ET and GLEAM is because they both estimated using the PT method.**

Please see reply for general comment 1, specific comment 5

**14. 14-25. Table 3. It’s difficult for readers to know what these parameters represent. Furthermore, aPT has a value of 2.9. It’s relatively high compared with the commonly used value (1.26) in the PT method, or the value (1.5-1.8) recommended for estimating ET in more arid regions (ASCE. 1990. Evapotranspiration and irrigation water requirements. ASCE. Manuals and reports on engineering practice. No. 70. New York, NY, USA). In this case, ET may be overestimated. It can be seen from Fig. 4 and Fig. S3 that the NewAge ET higher than GLEAM and MODIS ET, especially the peak value. If ET is overestimated, runoff should be underestimated when precipitation unchanged. Fig. 5 did show that in most cases, the modeled runoff is smaller than the observation, and obvious difference occurred also at peak values. The authors should discuss such physical processes that may cause model uncertainty. Only insightful analyses and discussion on the mechanism behind can highlight the scientific merit of the manuscript.**

Regarding the parameters description, we have added them in Appendix A.

It is true that PT alpha is higher than the literature. However, that literature does not include a stress factor as we do. Besides, as reported by Cristea et al. (2012) literature values of PT alpha varying from 0.6 to 2.4, making literature quite useless in applications in ungauged basins. In fact, the Reviewer should also notice that most of literature regarding PT estimates does not search for water budget closure, as we do, and their estimates, seen the large errors implied in ET measures, can contain very large bias or not close the water budget at all (Examples of studies really closing the water budget are pretty limited, and, summarized, for instance, at this link: https://www.authorea.com/users/24891/articles/142520-a-list-of-papers-that-perform-the-water-budget-estimation/_show_article ).

The reviewer is right when s/he argues about the reciprocal (inverse) dependence of ET and Q, but the interpretation can be simplistic, since the role of the variation in storage is not trivial: the three not the two quantities have to match, and Figure 10 shows exactly this. Besides our discharges are estimated by calibration (at least at selected gauge station), taken ET into account. Therefore the argument of the reviewer can be, in case, reversed, in the sense that discharge underestimation causes ET overestimation and not vice versa. Actually in our procedure we cannot really disentangle ET and Q estimates. Besides, as it can be read in the manuscript underestimation of discharges (as total volume) can vary spatially, thus making weaker his/her arguments.

Finally, in our experience deriving from many simulations, in basins of various size, we observed that our models, and models similar to our, tend systematically to underestimate most of the peak flows. At present, our investigations cannot discriminate if the problem is actually in models or in measures. The latter in fact are deduced as extension of looping rate curves behavior from normal discharges to high discharges, and it is not actually known how reliable this approach is.

Definitely all the questions raised by the reviewer comment deserve more attention in future work and papers.

The following comments has been added in the revised manuscript (in the discharge section 4.3):

“Additional source of error can also be caused by model inconsistency due to averaging out input data over large areas or from some inadequacy in stage-discharge curves used to obtain discharges from water levels. The slight underestimation of runoff could result from the overestimation of evapotranspiration. However, in this case, GLEAM (or MODIS) would cause larger discrepancies.”

The following statement has been added in the revised manuscript at the end of section 4.5:

“The same Figure also shows the complex interplay between discharges, (variation of) storages and evapotranspiration. A first look at Figure 4 and 5 could bring to the conclusion that overestimation of ET brings in underestimation of Q. However, Figure 10 shows that the role of ds/dt is not negligible at all.”

**15. 15-1. Table 4. There may be something interesting, i.e., KGE varied with basin area and may have a poor correlation with area. It's often taken for granted that a hydrological model will perform much better in relative smaller basins than in larger ones. Can you discuss why some times the JGrass-NewAGE System performed good or bad in sub basins with similar area?**

We do not share the Reviewer opinion that hydrological models perform better in smaller basins. Our experience, in fact, is the opposite (regarding the only discharge, indeed). However, there are many factors that could affect the model performance. These basins are distributed in various climatic zones, with different topography, and different vegetation distribution. All these factors can affect the model performance, not just the basin size. For instance, taking lake Tana station the model performance is relatively weak most likely due to the water regulation by the lake. This is commented on page 17 line 7 (old manuscript). As suggested, we have added the following general sentence to indicate this:

"Model performance varies with basins and a consistent behavior with respect to basin size, climate, vegetation density and topographic complexity is not found. Indeed, there are many factors that affect the model performance, including uncertainties in input observations."

**16. 16-1. Fig. 5. Is there any observation used in Fig. d?**

There is no observation for this site. This is plotted to show the model forecast at any link whether we have observation or not.

**17. 18-13. What does S mean? Please consider defining the abbreviation.**

Sorry for this error, we have changed S into ds/dt, as it should be.

**18. 21-1. Fig. 9. It would be better to change the water components to percentage.**

The percentage has a limitation. It does not show the actual volume, the annual variability between the components and the total water budget. If we change the bar into percentage, all years will have the same bar length. In the present form, we can see the total volume of each component and the total water budget. We already have annual mean percentage for each component reported in the text.

**19. 22-1. Fig. 10. I'm curious why ET was so low in the hot season. Supposing most of the rainfall infiltrated into the deep soil (high ds/dt values) in the hot and wet seasons, can they evaporate easily in the dry season? Again, more discussion may be required to persuade the readers about the modeling results.**

In the first months of hot season i.e. September, October, November, and December, ET is very high for the reasons mentioned by the reviewer, i.e., due to accumulation of storage in rainy season (June, July, and August) and lack of cloud causes high evapotranspiration. But, during the last months of hot season (March, April and May), lack of storage limits ET despite atmospheric water demand is high due to high net radiation.

It was:

"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 ds/dt (figure 10)."

We have added more discussion on this and revised as follows:

"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 ds/dt. During summer months, J, Q, and ds/dt shows high magnitude. ET is not high in June, July and August, but in October and December. The S(t) accumulated in the summer season feeds high ET in autumn, and causes very high drops in ds/dt (figure 10). The seasonal trend between J and ET is slightly out-of-phase, i.e., the highest energy to evaporate water (March, April, May) occurs during low precipitation months. Due to this slight out-of-phase, ET is minimal and Q and ds/dt is enhanced during wet months (figure 10), thus revealing that ET is water limited more than energy limited."

## References

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, 104, 37–54, 2017.

Allam, M. M., Jain Figueroa, A., McLaughlin, D. B., & Eltahir, E. A. B. (2016). Estimation of evaporation over the upper Blue Nile basin by combining observations from satellites and river flow gauges. Water Resources Research, 52(2), 644–659. http://doi.org/10.1002/2015WR017251

Bitew, M. M., Gebremichael, M., Ghebremichael, L. T., & Bayissa, Y. A. (2012). Evaluation of high-resolution satellite rainfall products through streamflow simulation in a hydrological modeling of a small mountainous watershed in Ethiopia. Journal of Hydrometeorology, 13(1), 338-350.

Brocca, L., Pellarin, T., Crow, W.T., Ciabatta, L., Massari, C., Ryu, D., Su, C.-H., Rudiger, C., Kerr, Y. (2016). Rainfall estimation by inverting SMOS soil moisture estimates: a comparison of different methods over Australia. Journal of Geophysical Research, 121(20), 12062-12079, doi:10.1002/2016JD025382.

Ciabatta, L., Marra, A.C., Panegrossi, G., Casella, D., Sanò, P., Dietrich, S., Massari, C., Brocca, L. (2017). Daily precipitation estimation through different microwave sensors: verification study over Italy. Journal of Hydrology, 545, 436-450, doi: 10.1016/j.jhydrol.2016.12.057.

Cristea, N. C., Kampf, S. K., Burges, S. J., 2012. Revised coefficients for PT and Makkin-Hansen equations for estimating daily reference evapotranspiration. Journal of Hydrologic Engineering 18 (10), 1289– 1300

Feddes, R. A., et al. 2001, Modeling root-water uptake in hydrological and climate models, Bull. Am. Meteorol. Soc., 82(12), 2797–2809.)

Gash, J. H. C.: An analytical model of rainfall interception by forests, Quarterly Journal of the Royal Meteorological Society, 105, 43–55, doi:10.1002/qj.49710544304, 1979

Gumindoga, W., Makurira, H., Phiri, M., & Nhapi, I. (2016). Estimating runoff from ungauged catchments for reservoir water balance in the Lower Middle Zambezi Basin. Water SA, 42(4), 641-649.

Habib, E., Haile, A. T., Sazib, N., Zhang, Y., & Rientjes, T. (2014). Effect of bias correction of satellite-rainfall estimates on runoff simulations at the source of the Upper Blue Nile. Remote Sensing, 6(7), 6688-6708.

Martens, B., Miralles, D. G., Lievens, H., Schalie, der, R., De Jeu, R. A. M., Fernández-Prieto, et al. (2016). GLEAM v3: satellite-based land evaporation and root-zone soil moisture. Geoscientific Model Development, 1–36. http://doi.org/10.5194/gmd-2016-162

Najmaddin, P. M., Whelan, M. J., & Balzter, H. (2017). Application of Satellite-Based Precipitation Estimates to Rainfall-Runoff Modelling in a Data-Scarce Semi-Arid Catchment. Climate, 5(2), 32.

Koster, R.D., Brocca, L., Crow, W.T., Burgin, M.S., De Lannoy, G.J.M. (2016). Precipitation Estimation Using L-Band and C-Band Soil Moisture Retrievals. Water Resources Research, 52(9), 7213-7225, doi:10.1002/2016WR019024.

Valdés-Pineda, R., Demaría, E. M. C., Valdés, J. B., Wi, S., and Serrat-Capdevilla, A.: Bias correction of daily satellite-based rainfall estimates for hydrologic forecasting in the Upper Zambezi, Africa, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-473, 2016.

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

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

Wuletawu Abera[1, 2], Giuseppe Formetta[3], Luca Brocca[4], and Riccardo Rigon[1]

[1]Department of Civil, Environmental and Mechanical Engineering, University of Trento, Italy
[2]Department of Geography and Environmental Studies, Mekelle University, Ethiopia
[3]Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, UK
[4]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 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 obtain the water budget closure, 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 NewAge estimated ET; and GRACE gravimetry data for comparison of the total water storage amounts available in the whole basin. 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 mean budget analysis shows that precipitation of the basin is 1360 ±230 mm per year. Evapotranspiration ~~covers~~ accounts for 56% (per cent) of the annual water budget, runoff is 33%, storage varies from minus 10% to plus 17% of the water 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 of precipitation per year (Johnston and McCartney, 2010). Particularly, the Upper Blue Nile (hereafter UBN) basin is the main sources of water in the region. ~~and, it is most hydro-climatologically and socio-politically complex basins in the world~~.

In Ethiopia, UBN is inhabited by 20 million people whose main livelihood is subsistence agriculture (Population Census Commission 2008). The Ethiopian government, therefore, 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. However, as the principal contributor (i.e 51% of discharge) to the main Nile basin, UBN also supports hundreds of millions of people living downstream, and it is referred to as the "Water Tower" of northeast Africa. Therefore UBN is a part of trans-boundary river, and its development and management require obtaining agreements between many national governments and also non-governmental organizations, each involving different policies, legal regimes, and contrasting interests. Tackling all these complexities and developing better water resource development 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 and for assessing the sustainability of farming systems in the region. In fact, because of the lack of hydrometeorological data and a proper modelling framework, the recent modelling efforts conducted within the basin have evident limitations in addressing these problems. 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) and evapotranspiration (Allam et al., 2016), 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). Spatially distributed information on all the components of the water budget does not exist and 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 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).

This study is an effort contributing to answering the quantitative issues related to the aforementioned management problems by estimating the components of the water budget of the UBN basin using a new hydrological modelling framework (see section 3.1) and remote sensing data. It obtains, at relatively small spatial scales and at daily time step, groundwater storage, evapotranspiration, discharges in such a way to satisfy the water budget equation. It is also a methodological paper, in that it delineates various methodologies to overcome the data scarcity. The paper is organized as follows: firstly, descriptions of the study area is 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~~ length 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 Ethiopia. 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~~ gives the largest contribution to 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 on their own.

The maps of elevation 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

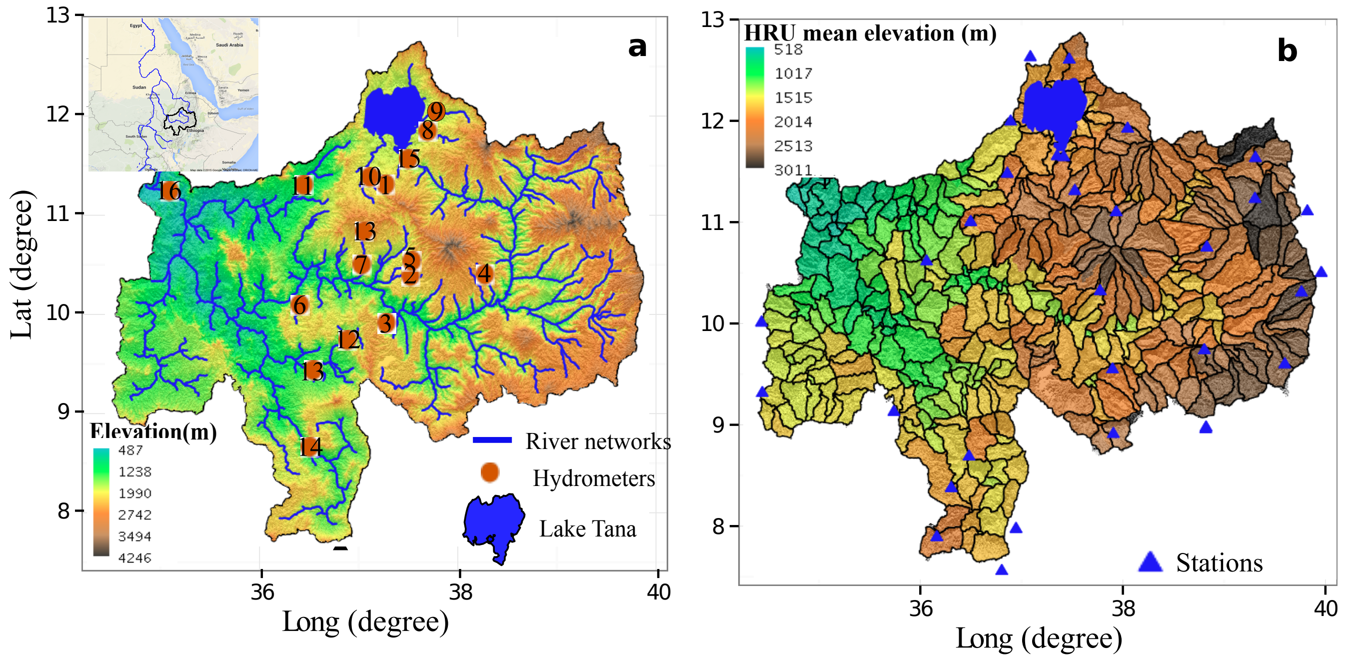

**Figure 1.** The Upper Blue Nile basin digital elevation map, along with the gauge stations present in the basin. Numbers inside the circles designate the river gauging stations whose names are provided in table (a); and subbasin partitions and meteorological stations used for simulation (b). ~~Numbers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

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°C, 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:

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

where $J(t)$ is rainfall, and $ET(t)$ is actual evapotranspiration, $Q(t)$ is discharge, $Q_{ki}(t)$ are inflows from upstream HRUs. 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.

**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 and 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), and 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) described in Appendix A |
| Snow melting | Snow melt | Modelling snow melting using three types of temperature and radiation based 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. |

### 3.1 JGrass-NewAge system set-up

UBN water budget is estimated using the JGrass-NewAge hydrological system, which is, in turn, based on the Object Modelling System framework (David et al., 2013). 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. A similar study using JGrass-NewAge system, but ~~using~~ utilizing mostly in-situ observations, has been conducted in Posina river basin, northeast Italy (Abera et al., 2017).~~, and the model performance is assessed positively~~ A brief descriptions of 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 Geographic Information System (GIS) representation. The basin topographic representation in GIS, as detailed in Abera et al. (2014); Formetta et al. (2011), is based on the Pfafstetter enumeration. 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 (Formetta et al., 2014a). A routing scheme is applied to move the discharges from HRUs to the basin outlet through the channel network is included in the Adige component.

In this study, the UBN basin is divided into 402 subbasins (HRUs of mean area of 430 ± 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,~~ it is consistent with ~~model~~ input data resolution, ~~and the resolution of~~ including satellite products, meaning that a finer subdivision would imply uniform inputs for adjacent HRUs, and a coarser one would average out inputs variability. ~~it can be considered an acceptable scale to capture the spatial variability of the water budget.~~ In this paper, the term HRU ~~and subbasin are used alternatively for the same basin partitioning concept.~~ actually identifies subbasins.

### 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.~~ and ~~Recently,~~ Abera et al. (2016) compared five of them with high spatial and temporal resolutions 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. Regards to the quality of SM2RAIN-based products, recent studies positively assessed their accuracy on a regional (Brocca et al., 2016; Ciabatta et al., 2017) and a global (Koster et al., 2016) scale. 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 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 specialised for UBN byAbera et al. (2016) by using in-situ observations. 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 each of the 402 subbasins.

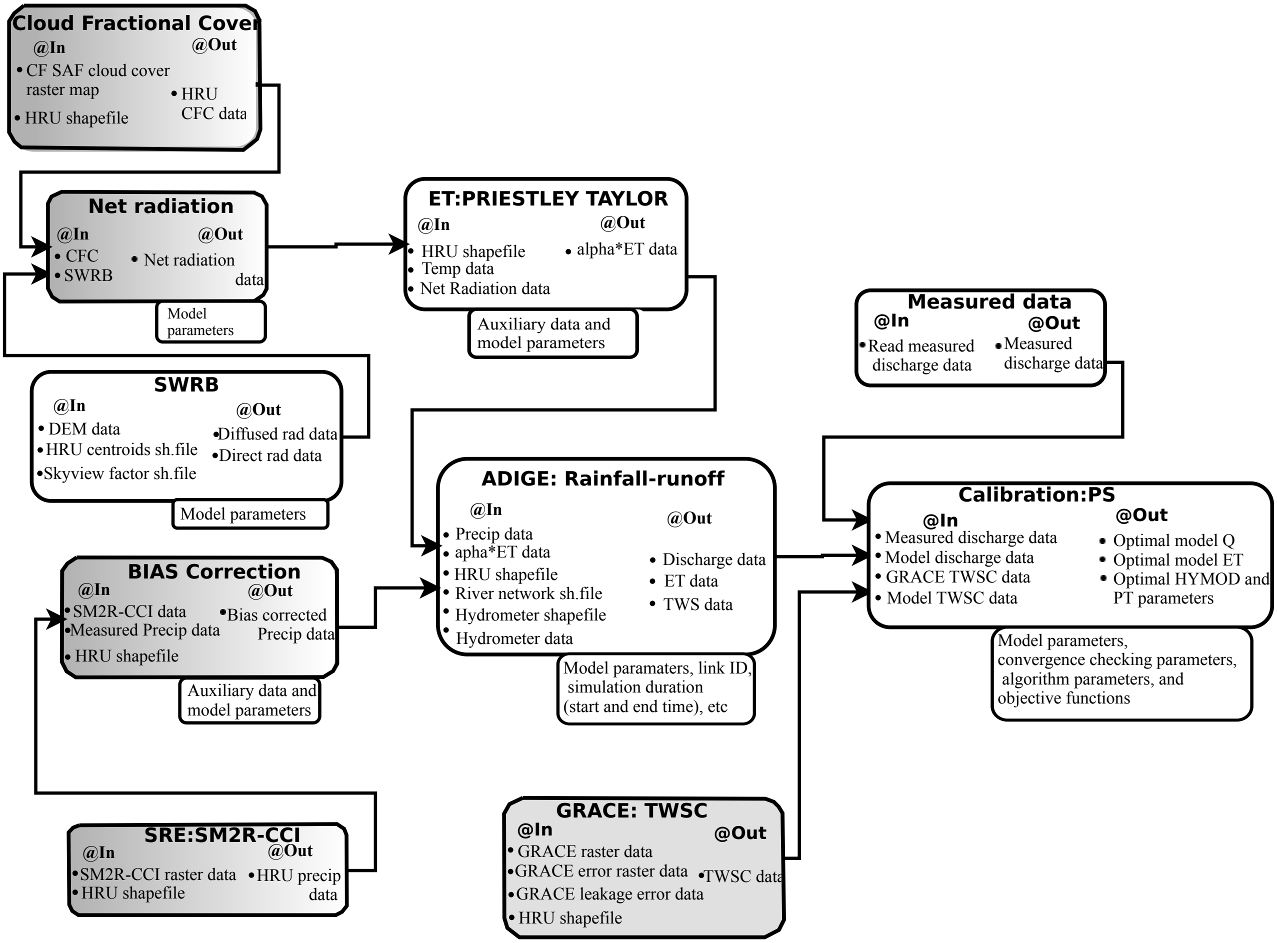

**Figure 2.** Workflow with a list of NewAge components (in white), and remote sensing data processing parts (shaded in grey, not yet included in JGrass-NewAGE and currently performed with R tools) used to derive the water budget of the UBN. It does not include the components used for the validation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

### 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. 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. PT is mainly based on net radiation estimation, $Rn$, ~~simplifying~~ grouping all the unknowns into the $\alpha_{PT}$ coefficient, as shown in Eq. 2.

$$PET = \alpha_{PT} \frac{\Delta}{\Delta + \gamma} Rn \tag{2}$$

~~Where~~ $PET$ is Priestley and Taylor potential evapotranspiration, $\Delta$ is the slope of the Clausius-Clapeyron relation~~s~~ and $\gamma$ is the psychometric constant (Brutsaert, 2005). In this study, however, we need an estimate of the actual evapotranspiration ($ET$), which 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., 2017):

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

where $S(t)$ is the groundwater storage, and $S_{max}$ the maximum storage capacity for each HRU. The important unknown coefficient $\alpha_{PT}$ (Pejam et al., 2006; Assouline et al., 2016) and the $S_{max}$ are calibrated within the rainfall-runoff model component, as explained below. The ratio $S(t)/S_{max}$ represents a stress coefficient which became very popular since the work of Feddes and coworkers (Feddes et al., 2001).

In our 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~~ its initial value (Abera et al., 2017). Once $T_B$ is fixed, the tools for automatic calibration, provided by the Object Modelling System, ~~set to~~ produce the set of parameters in tab 4, including $\alpha_{PT}$ and $S_{max}$, for which ~~besides~~ discharge is well reproduced and 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 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 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 compare it with independently obtained $ET$ estimates or data. In situ $ET$ observations are not ~~available~~ present for this basin, as is the case for most regions. Estimates of $ET$ based on RS have been made available 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, version_v3_BETA) (Miralles et al., 2011a; Martens et al., 2016), a global, satellite-based, $ET$ data set is used. ~~The performance of GLEAM is assessed positively in different studies.~~ GLEAM, as well NewAge, uses the PT scheme for estimating ET. However, all inputs of the formula, in GLEAM and NewAge, are evaluated according to different strategies and RS tools. GLEAM sets $\alpha_{PT} = 0.98$ while in NewAge it has been calibrated. In GLEAM PET is additively increased, by intercepted rainfall estimated according to a version of the Gash model (Gash, 1979), and multiplicatively decreased by a stress coefficient depending on five soil cover types (bare soil, snow, tall vegetation, two levels of low vegetation) and has a different expression for anyone of the storages. Moreover, according to the case, the stress coefficient is evaluated through various RS, according to procedures which are described in the paper by Martens et al. (2016). Differently from the NewAge approach, GLEAM also considers dynamic vegetation information to estimate the stress factor (Miralles et al., 2011a).

GLEAM is available at $0.25^o$ spatial resolution ( 28 km of side or 800 square kilometers of area) and daily temporal resolution, and assessed positively in different studies (McCabe et al., 2016; Miralles et al., 2011b). The most recent version of GLEAM was validated globally over sixty-four Fluxnet sites (Martens et al., 2016) with consistent results, letting us guess that it behaves properly also in Ethiopia. The differences between NewAGE estimation and GLEAM's one allow to assume that the our results and their results can be seen as largely independent . For comparison with NewAge ET, we ~~estimated area weighted~~ averaged 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.~~ 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 a routing component and artificial inflows-outflows management.

~~component.~~ Detailed descriptions of HYMOD implementations in the NewAge model system are given in Formetta et al. (2011); Abera et al. (2017) and summarized in Appendix A. The main inputs for the ADIGE model are $J(t)$ and $PET(t)$, 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 of the 402 HRU routed to the outlet. The ADIGE rainfall-runoff has five calibration parameters ($C_{max}$, $B_{exp}$, $\alpha_{Hymod}$, $R_s$, $R_q$, see the details in Appendix A), and the calibration is performed using the particle swarm (PS) optimization method. PS is a population-based stochastic optimization technique (Kennedy and Eberhart, 1995). ~~inspired by the social behaviour of flocking birds or fish schools 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, the two other goodness-of-fit (GOF) methods (percentage bias (PBIAS) and correlation coefficient) used in this study are described in Appendix B.

### 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 of 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 GRACE 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~~ affects the goodness of fit of all the 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 3.2 , $\alpha_{PT}$ is calibrated by by imposing that $S(0) = S(T_B)$ 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 B), as follows:

- Discharge validation: Discharge simulation is validated ~~for separate time-series data~~ at the outlet ~~at~~ close to the Ethiopian-Sudan border, where the model is calibrated. In addition, the simulation of NewAge at the internal links is validated in 15 discharge measurement stations, where in situ data are available. The evaluations of discharges at the internal links provide an assessment of model estimation capacity at ungauged locations.

- ET ~~validation~~comparison: 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. 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 and at the whole basin scale.

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

| Satellite products | Spatial resolution | Temporal resolution | Data used | Reference | used as |
|---|---|---|---|---|---|
| SM2R-CCI | 0.25 degree | daily | 1994-2009 | Brocca et al. (2014, 2013); Abera et al. (2016) | input for Precipitation |
| GLEAM | 0.25 degree | daily | 1994-2009 | Miralles et al. (2011a); McCabe et al. (2016) | verification for evapotranspiration |
| MODIS ET (MOD16) | 1-km | 8-days | 1994-2009 | Mu et al. (2007, 2011) | verification for evapotranspiration |
| GRACE TWS | 1 degree | 30-days | 2003-2009 | Landerer and Swenson (2012) | Verification for storage change |
| CM-SAF | 0.25 degree | daily | 1994-2009 | 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 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). This spatial pattern is consistent with the results of Mellander et al. (2013) and Abtew et al. (2009). 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.For this location 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, is from point gauge data, while this study is based on areal data. ~~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. In this section we provide mainly discussion about the comparison of NewAGE ET and RS estimates, but further comments on ET can be found in section 4.5 which show that ET is mostly water limited than energy limited. 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. 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 4b. 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 4b). 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). 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

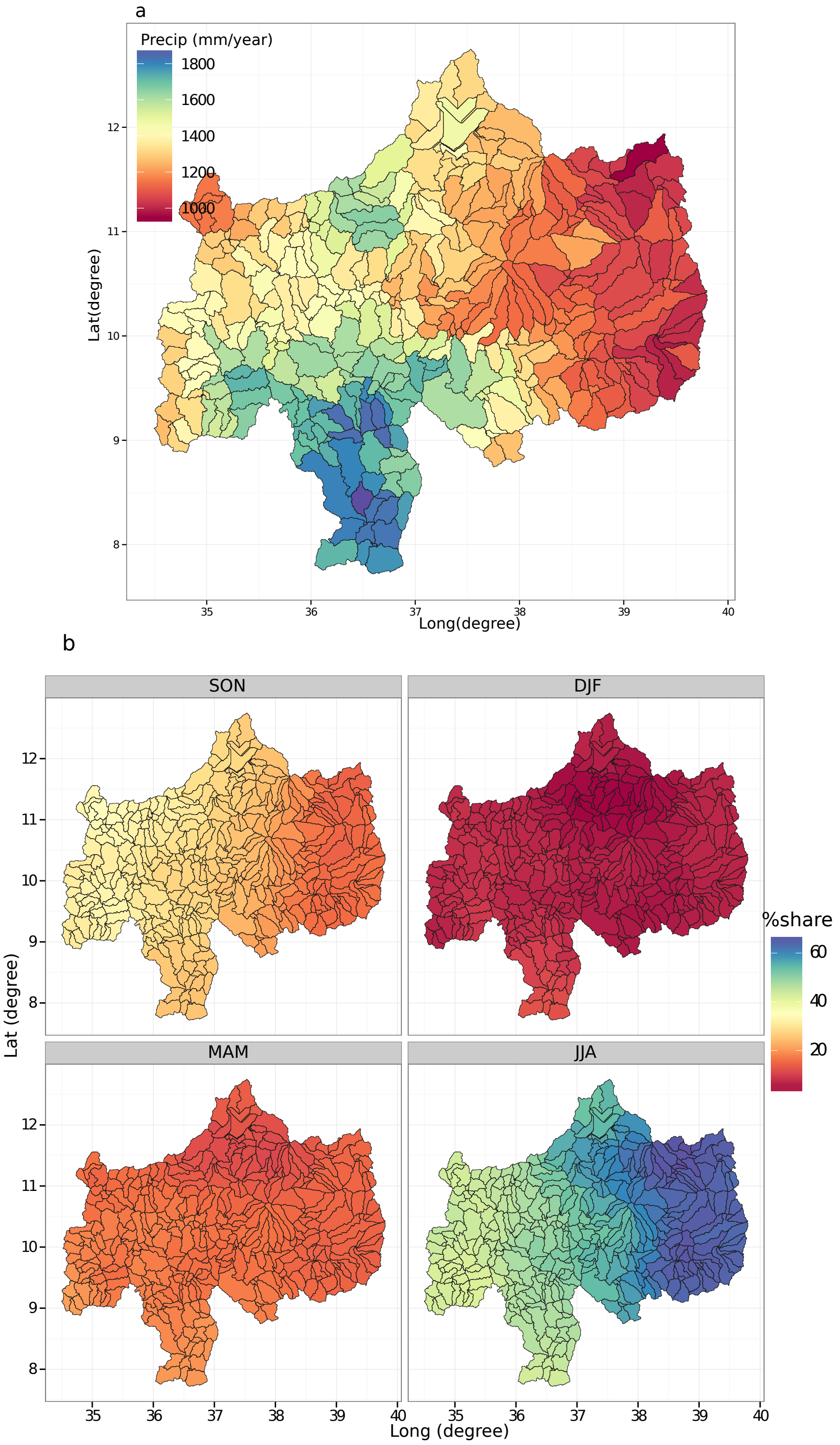

**Figure 3.** The spatial distribution of mean annual rainfall (a), and quarterly percentages share of the total rainfall (b) estimated from long term data (1994-2009): SON (September, October, and November), DJF (December, January, and February), MAM (March, April, and May), JJA (June, July and August).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

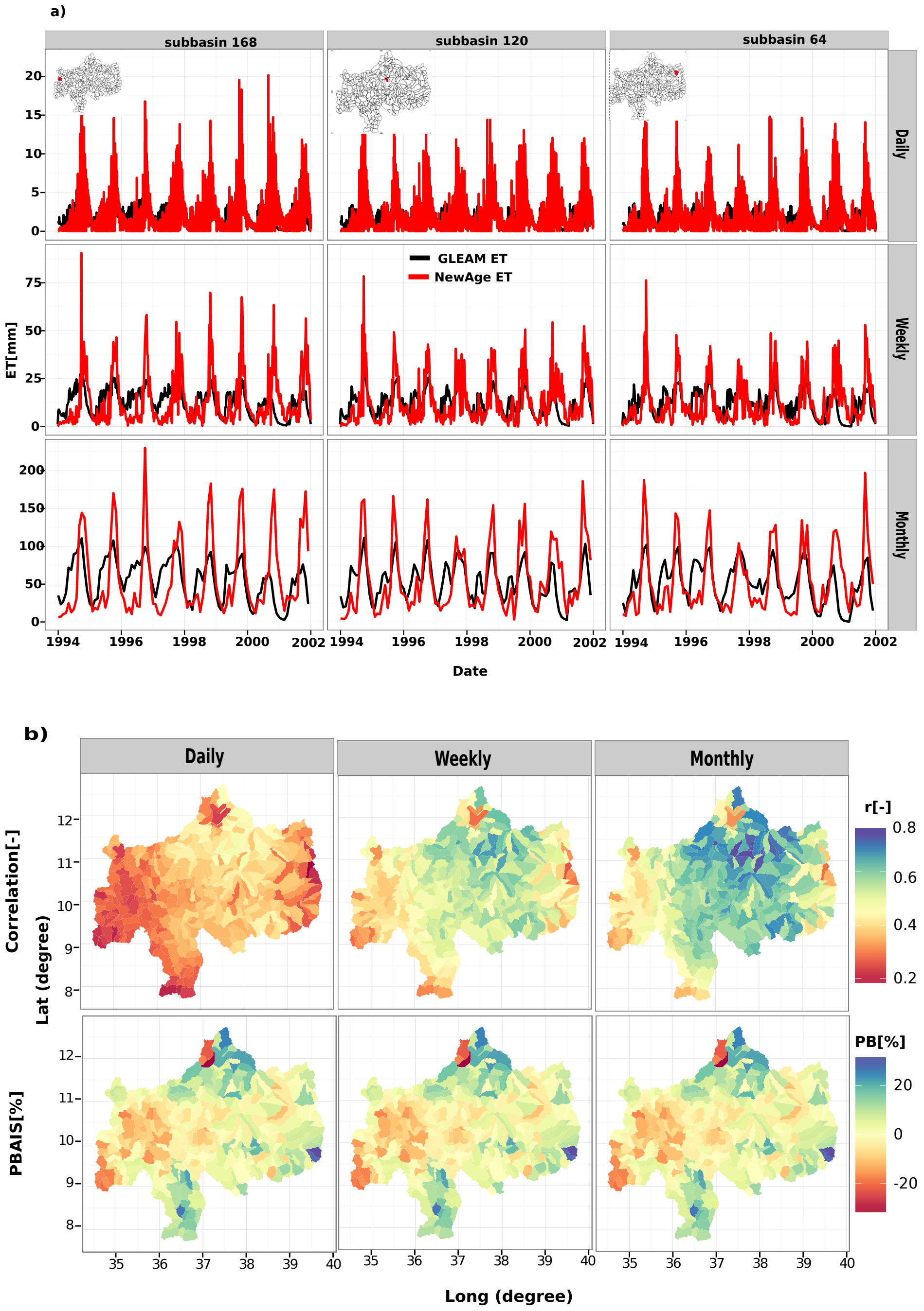

**Figure 4.** a:Time series ET estimation with NewAge and GLEAM for three subbasins: subbasin ID168 , subbasin ID120, and subbasin ID64 at daily, weekly and monthly time steps. The locations of the subbasins are indicated on the maps at the top of each column of plots. b: spatial distribution of correlation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

Product (MOD16), as shown in the supplementary material, 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 KGE=0.92, PBIAS = 2.4, and r = 0.93. Model performances are also evaluated within the basin at the internal catchments outlets (table 4) where stage measurements are available. Figure 5 shows simulated hydrographs along with the observed discharges for some locations. The results show that the performances of the NewAge simulation are a little better than 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 or from some inadequacy in stage-discharge curves used to obtain discharges from water levels. The slight underestimation of runoff could result from the overestimation of evapotranspiration. However, in this case, GLEAM (or MODIS) would cause larger discrepancies.

**Table 3.** Optimized parameters obtained from daily ADIGE simulation during the calibration period (1994-1999). Parameters' physical meaning is explained in Appendix A. 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 simulation, 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, 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

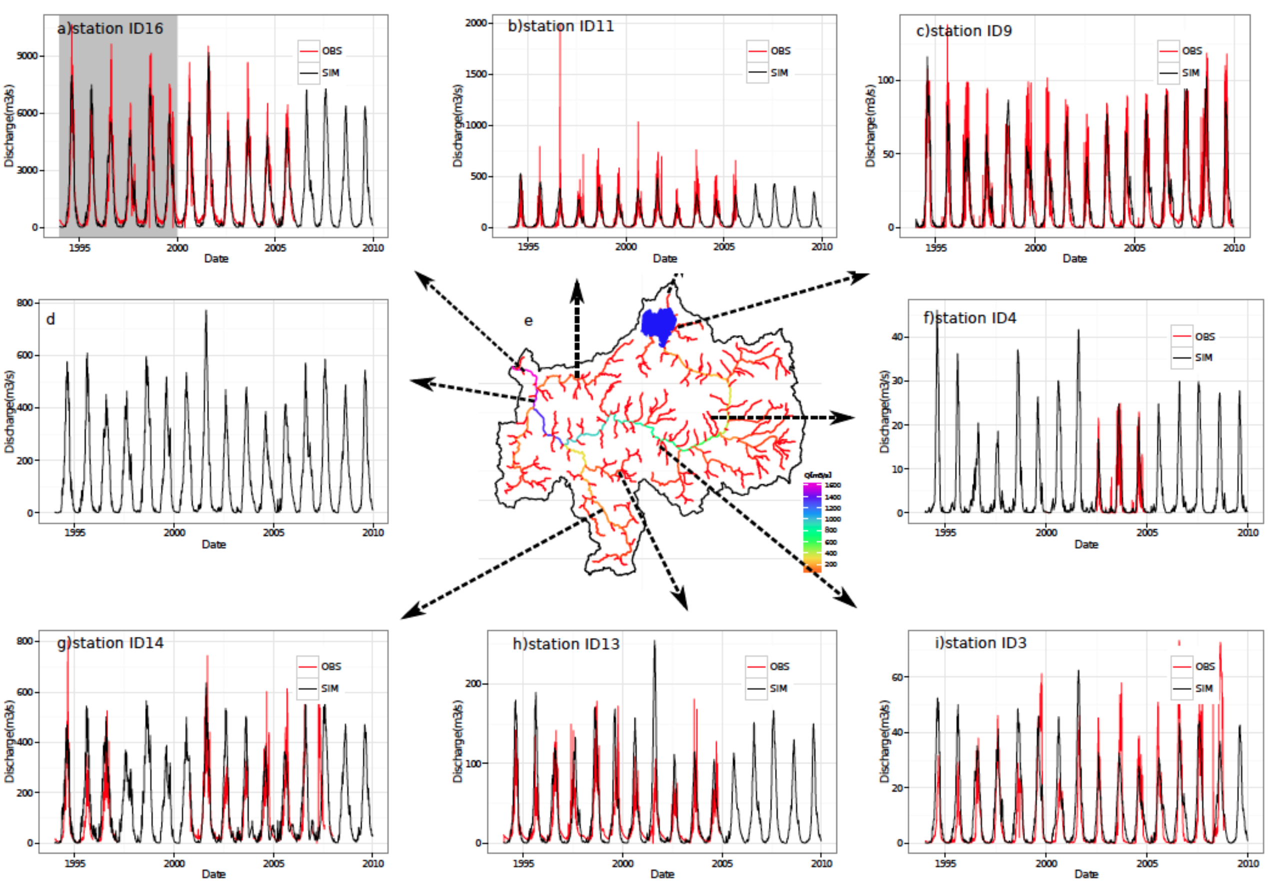

**Figure 5.** NewAge model simulation validation at internal subbasins. The model calibrated (shown by gray shaded period) and validated at El Diem (a) is used to estimate at each channel link and, where discharge measurements are available, they are verified: main Beles bridge (b), Ribb river enclosed 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 simulation 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 |

performance is 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 simulation 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 = 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. An exception is ~~except for~~ Lake Tana, where the discharge is regulated (figure 5 and table 4). Model performance varies with basins and a consistent behavior with respect to basin size, climate, vegetation density and topographic complexity is not found. Indeed, there are many factors that affect the model performance, including uncertainties in input observations. 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

NewAge simulated $ds/dt$ for 16 years for each subbasin is calculated as a residual of the flux terms. The simulated $ds/dt$ is represented and compared with the GRACE-based TWSC in Figure 6. 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 and for the whole basin, is in accordance with the GRACE TWSC both in temporal pattern and amplitude ~~. Over the whole basin a~~ The good correlation coefficient of 0.84 and the general good performances of the $ds/dt$ component 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.However, the spatial distribution of NewAge and GRACE $ds/dt$ estimates can be found in the supplementary material.

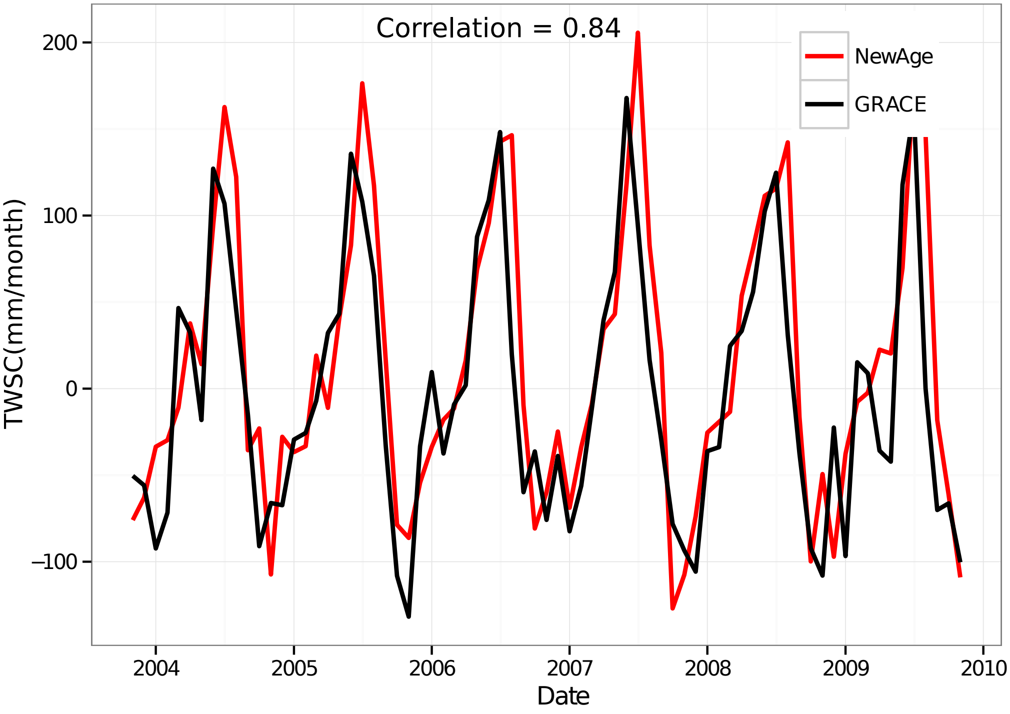

**Figure 6.** Comparison between basin-scale (whole UBN, 176,315 square kilometers) NewAge $ds/dt$ and GRACE TWSC from 2004-2009 at monthly time steps.

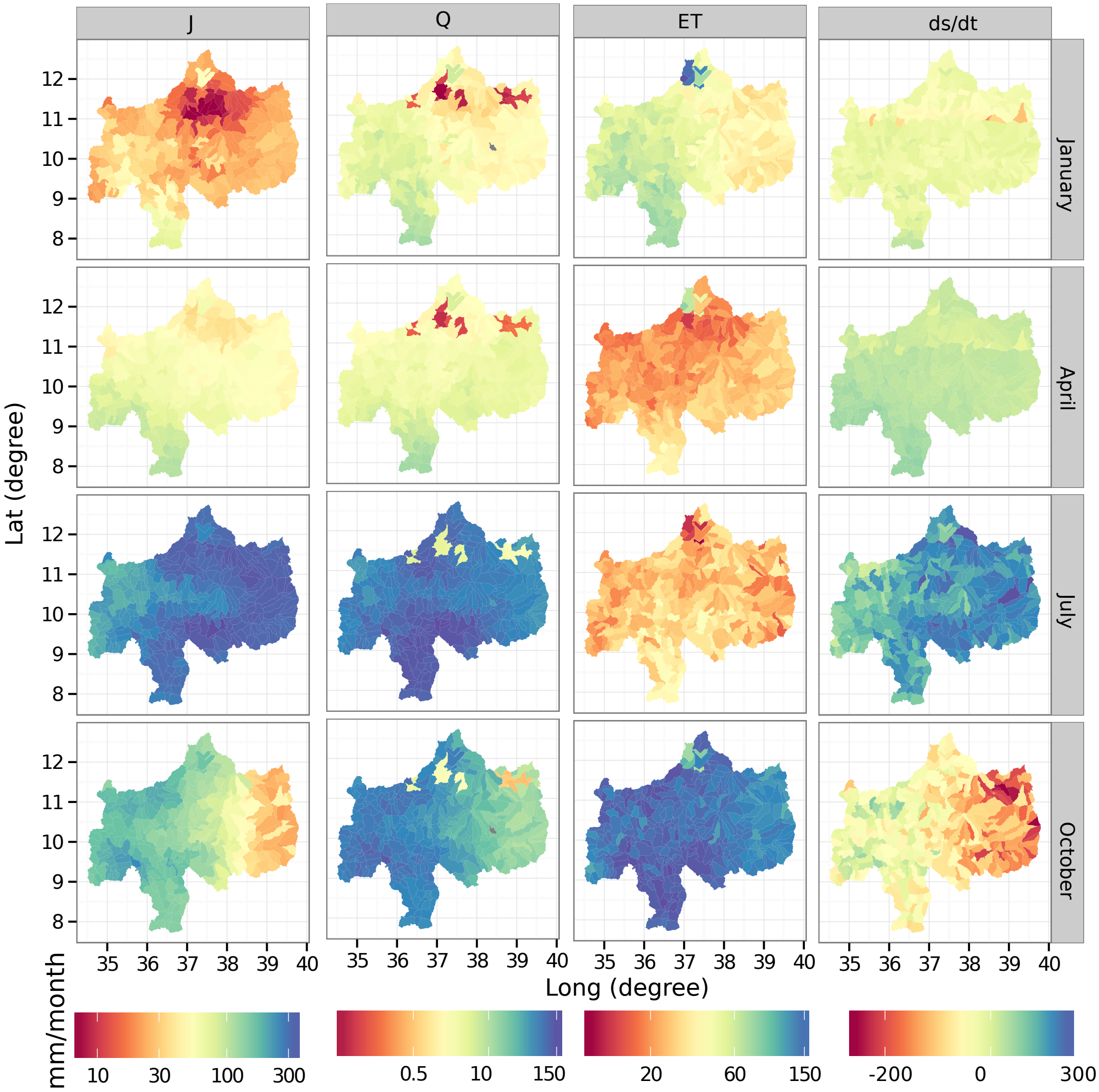

**Figure 7.** Spatial distribution of long term mean monthly water budget (January,April, July and October) in the UBN basin. For the sake of visibility, the legend is plotted separately and on logarithmic scale, except for the storage component.

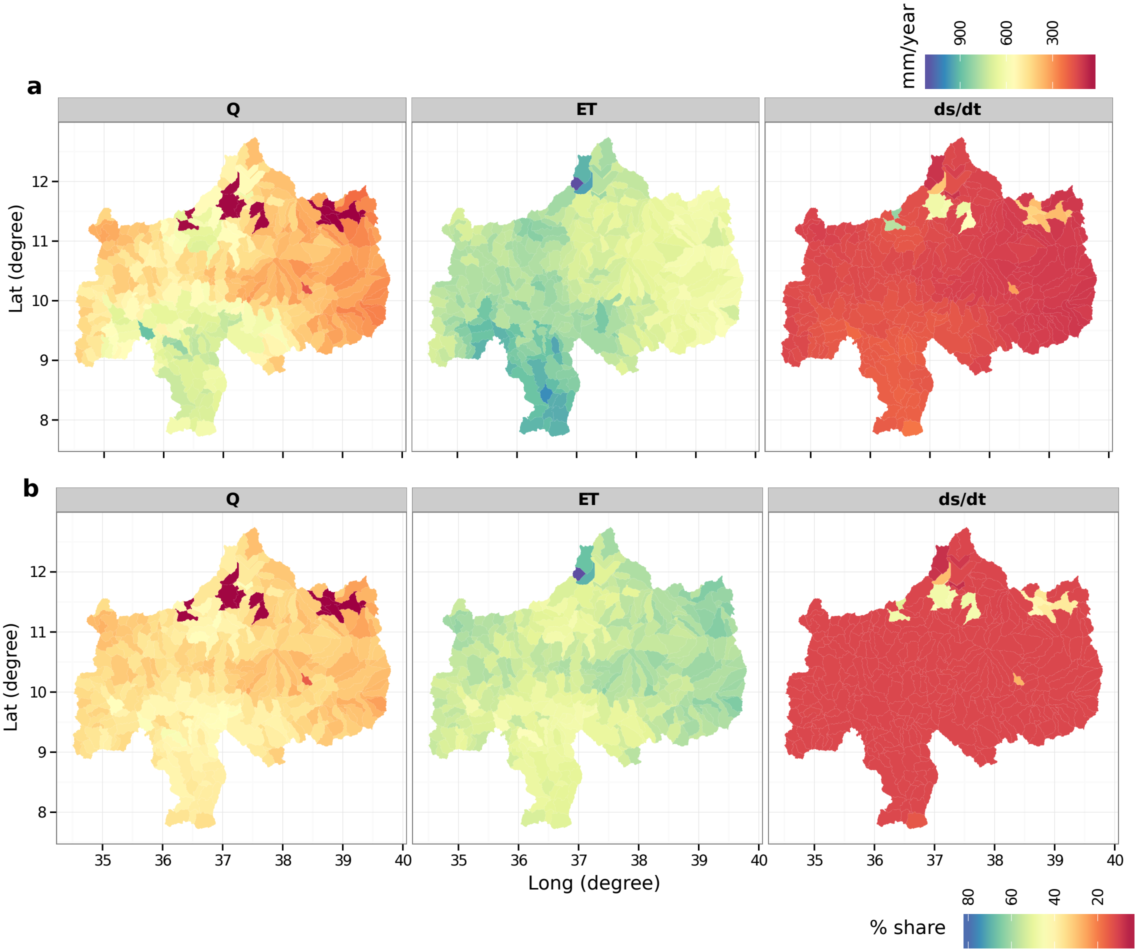

**Figure 8.** The spatial distributions of long term mean annual water budget closure: precipitation in mm (figure 3), the output terms (Q, ET, ds/dt) in mm (a), and the percentage share of the output term (Q, ET, ds/dt) of the total precipitation (b).

### 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.

The variability between the subbasins is also appreciable. Generally, all water budget 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.

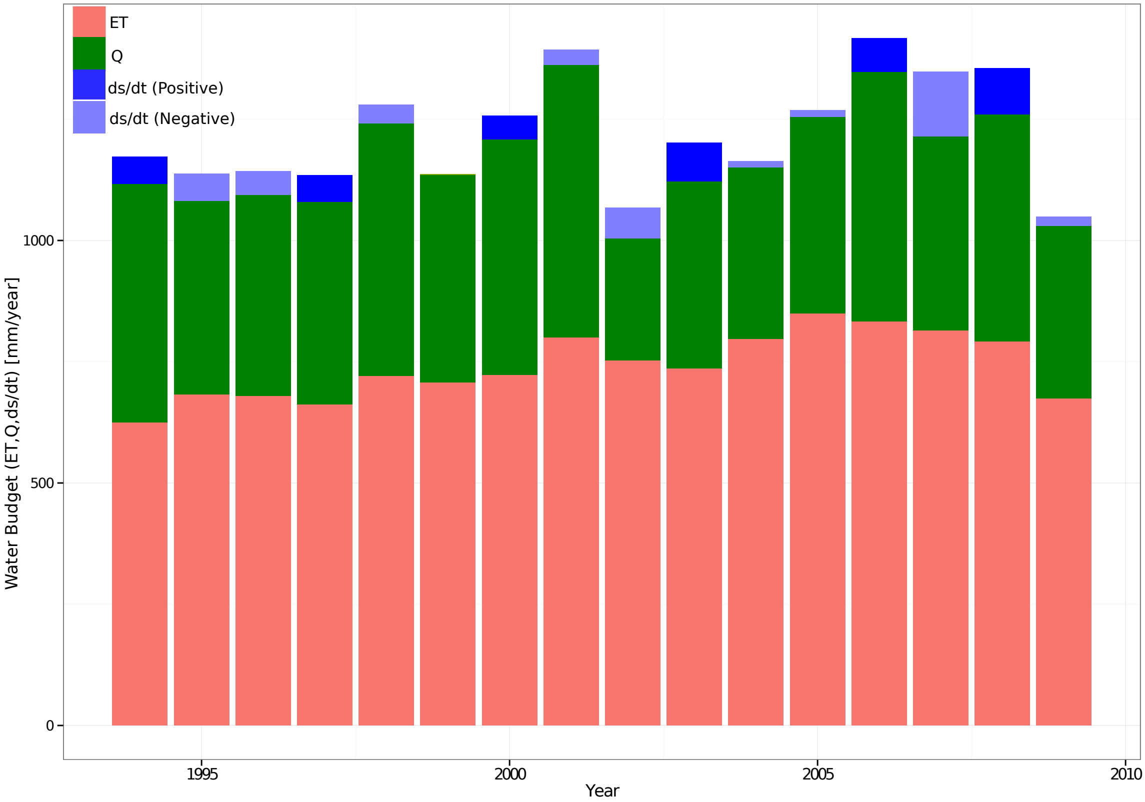

**Figure 9.** Water budget components of the basin and its annual variabilities from 1994 to 2009. The relative share of each of the three components (Q, ET and ds/dt) of the total available water J is represented by the length of the bars (N.B. the total length of the bar minus the negative storage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

The spatial variability of the long term mean annual water budget closure is shown in figure 8. The spatial variability ~~of~~ for 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 (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.

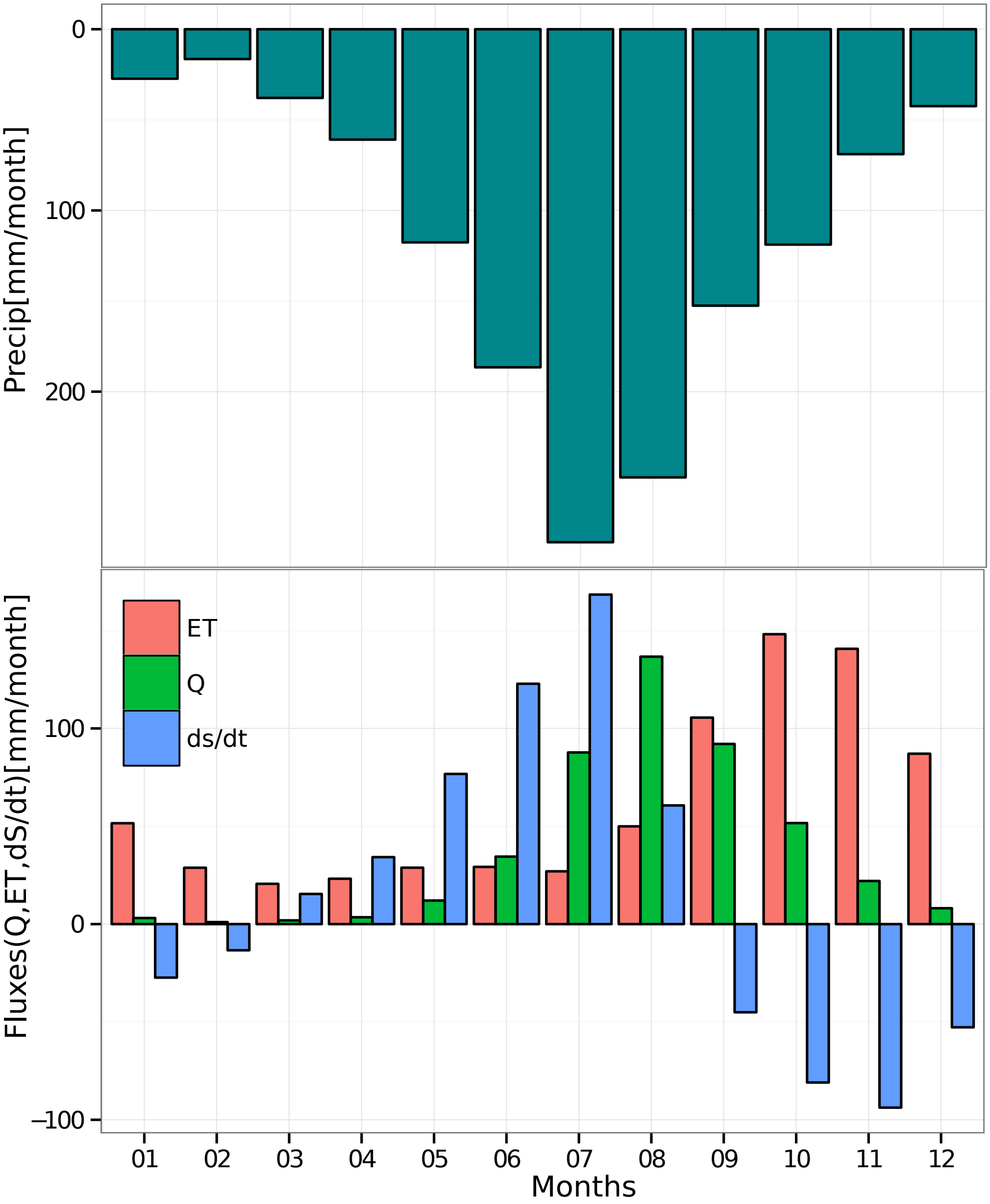

**Figure 10.** Monthly mean Water budget components at basin scale and long term, based on estimates from 1994 to 2009. The relative shares of the three components (Q, ET and dS/dt) of the total available water J are shown .

The long-term basin-average water budget components shows: 1360 ± 230 mm of J, followed by 740 ± 87 mm of ET, 454 ± 160 mm of Q and -4 ± 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 read in Viste et al. (2013).

Figure 10 provides long term monthly mean estimates of water budget fluxes and storage. The average basin scale budget is highly variable. The highest variability is mainly in J and $ds/dt$. During summer months, J, Q, and $ds/dt$ shows high magnitude. ET is not high in June, July and August, but in October and December. The $S(t)$ accumulated in the summer season feeds high ET in autumn, and causes very high drops in $ds/dt$ (figure 10). The seasonal trend between J and ET is slightly out-of-phase, i.e., the highest energy to evaporate water occurs during low precipitation months (March, April, May). Due to this slight out-of-phase, ET is minimal and Q and $ds/dt$ is enhanced during wet months (figure 10) thus revealing that ET is water limited more than energy limited. The same Figure also shows the complex interplay between discharges, (variation of) storages and evapotranspiration. A first look at Figure 4 and 5 could bring to the conclusion that overestimation of ET brings in underestimation of Q. However, Figure 10 shows that the role of $ds/dt$ is not negligible at all.

## 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 system and remote sensing data. The study covered 16 years from 1994-2009 at a finer spatial and temporal resolution than in previous studies. In order to achieve this result, we used~~,~~ various remote sensing products, rainfall from SM2R-CCI, cloud cover from SAF EUMETSAT CFC, evapotranspiration from GLEAM and MODIS (used for comparison), and storage change from GRACE (also used for comparison). We also used all the ground data currently available, i.e. sixteen discharge time series, and thirty-five ground based meteorological stations. The results can be summarized as follows:

- The basin scale annual precipitation over the basin is 1360 ± 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).

- 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 ± 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 than our estimates. 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 supplementary 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 water budget according both simulated and measured discharges.

- 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 ± 160 mm. The verification results at the internal sites where measurements are available reveal that the model can be used for forecasting at ungauged locations with some success.
- The performances obtained are promising (figures 5 and 6, and table 4) and often greatly improve previous results.

  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.

We claim that the procedure we followed can be easily transported in any other poorly gauged basin with benefits for the hydrological knowledge of any region on Earth.

## 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: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: Hymod model in NewAge-JGrass system

The NewAge system executes one Hymod model at each HRU, and routes water downslope. Detailed description of Hymod model is provided in many researches (Moore, 1985; Van Delft et al., 2009; Boyle et al., 2001; Formetta et al., 2011). In Hymod, each HRU, is supposed to be a composition of storages of capability $C$ [L] according to distribution (Moore, 1985):

$$F(C < c) = 1 - (1 - \frac{c}{C_{max}})^{B_{exp}} \tag{A1}$$

where $F(C)$ represents the cumulative probability of a certain water storage capacity ($C$); $C_{max}$ is the largest water storage capacity within each hillslope and $B_{exp}$ is the degree of variability in the storage capacity. As shown in the schematic diagram (figure 11), the precipitation exceeding $C_{max}$ is send directly to the volume available for surface runoff. If we call the precipitation volume in a time interval $\Delta t$, $J(t) := P(t)\Delta t$, then this “direct” runoff can be estimated according to:

$$R_H(t) = \max(0, J(t) + C(t) - C_{max}) \tag{A2}$$

where $C(t)$ defines the fraction of storages already filled at time $t$. The latter equation is true for any precipitation and storage level, even when the maximum storage $C_{max}$ is not exceeded. When precipitation does not exceeds $C_{max}$ runoff volume can be produced by filling some of the smaller storages. To which extent this happens, can be derived by the knowledge of the storage distribution, eq. (A1), the initial storage $C(t)$ and the precipitation $J(t)$. This residual runoff is, in fact, given by:

$$R(t) = \int\limits_{C(t)}^{\min(C(t)+J(t), c_{max})} F(c)\, dc \tag{A3}$$

An analytic expression for the integral in eq. (A3) is available, which makes the computation easier. Water in storage is made available to evapotranspiration. Water going into runoff the runoff volume, i.e. $R(t)$ and $R_H(t)$, is further subdivided into a surface runoff volume and subsurface storm runoff. Surface runoff, in turn, is composed by the whole of $R_H(t)$ and part of $R(t)$, and $R(t)$ is split according to a partition coefficient $\alpha$ such that the part $\alpha R(t)$ goes into surface runoff volume and $(1-\alpha)$ into the subsurface storm runoff volume. In Hymod, $\alpha$ is a calibration coefficient.

Finally, surface runoff volumes are routed through three linear reservoirs, and, subsurface storm runoff volume is routed through a single linear reservoir. A summary of equations for the surface runoff is therefore:

$$\frac{dS_1(t)}{dt} = \alpha R(t) + R_H(t) - kS_1(t) \qquad Q_1(t) = \frac{S_1(t)}{k} \tag{A4}$$

where $S_1$ [L$^3$] is the storage in the first of the linear reservoirs, and $k$ [T] is the mean residence time in each of the reservoirs. Then:

$$\frac{dS_i(t)}{dt} = Q_{i-1}(t) - kS_i(t) \quad Q_i(t) = \frac{S_i(t)}{k} \tag{A5}$$

for the other two reservoirs, where $S_i$ [L] with $i = 2,3$ is the storage in the two remaining surface reservoirs. Subsurface storm runoff is then modeled by:

$$\frac{dS_{sub}}{dt} = (1-\alpha)R(t) - k_{sub}S_{sub}(t) \tag{A6}$$

where $S_{sub}$ [L$^3$] is the storage in the subsurface storm-flow system and $k_{sub}$ [T] is its mean residence time. A budget equation can be written for the groundwater system as:

$$\frac{dS_g(t)}{dt} = (J(t) - R(t) - R_H(t)) - ET(t) - Q_g(t) \tag{A7}$$

where $S_g(t)$ [L$^3$] is the groundwater storage, and $Q_g(t)$ the groundwater flow which becomes surface flow at the closure of the HRU.

Summarizing, Hymod subdivides each HRU into three reservoirs: a groundwater reservoir, from where evapotranspiration and groundwater flow is allowed, a subsurface storm-water reservoir, and a surface runoff reservoirs set. Partition of precipitation into the three reservoirs is obtained by a calibration coefficient, $\alpha$, and the use of a probability distribution function of storages' capacity, $F(c)$.

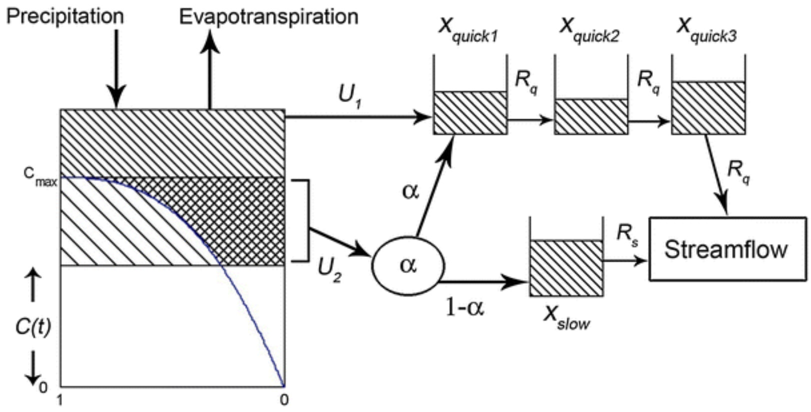

**Figure 11.** Schematic diagram of hymod model (adapted from Van Delft et al. (2009))

## Appendix B: 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).

$$PBIAS = \frac{\sum_{i=1}^{n}(P_i - O_i)}{\sum_{i=1}^{n} O_i} 100 \tag{B1}$$

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

$$KGE = 1 - ED \tag{B2}$$

$$ED = \sqrt{(r-1)^2 + (vr-1)^2 + (\beta - 1)^2} \tag{B3}$$

   where ED is the Euclidian distance from the ideal point, $\beta$ 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 ($\sigma_o$) and modelled ($\sigma_s$) 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, 104, 37–54, 2017.

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.

Allam, M. M., Jain Figueroa, A., McLaughlin, D. B., and Eltahir, E. A.: Estimation of evaporation over the upper Blue Nile basin by combining observations from satellites and river flow gauges, Water Resources Research, 2016.

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.

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z., and Smith, M.: Toward improved streamflow forecasts: Value of semidistributed modeling, Water Resources Research, 37, 2749–2759, 2001.

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.

Brocca, L., Pellarin, T., Crow, W. T., Ciabatta, L., Massari, C., Ryu, D., Su, C.-H., Rüdiger, C., and Kerr, Y.: Rainfall estimation by inverting SMOS soil moisture estimates: A comparison of different methods over Australia, Journal of Geophysical Research: Atmospheres, 121, 2016.

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

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

Ciabatta, L., Marra, A. C., Panegrossi, G., Casella, D., Sanò, P., Dietrich, S., Massari, C., and Brocca, L.: Daily precipitation estimation through different microwave sensors: Verification study over Italy, Journal of Hydrology, 2017.

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.

David, O., Ascough Ii, J., Lloyd, W., Green, T., Rojas, K., Leavesley, G., and Ahuja, L.: A software engineering perspective on environmental modeling framework design: The Object Modeling System, Environmental Modelling & Software, 39, 201–213, 2013.

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.

Feddes, R. A., Hoff, H., Bruen, M., Dawson, T., de Rosnay, P., Dirmeyer, P., Jackson, R. B., Kabat, P., Kleidon, A., Lilly, A., et al.: Modeling root water uptake in hydrological and climate models, Bulletin of the American meteorological society, 82, 2797–2809, 2001.

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.

Gash, J.: An analytical model of rainfall interception by forests, Quarterly Journal of the Royal Meteorological Society, 105, 43–55, 1979.

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.

Koster, R. D., Brocca, L., Crow, W. T., Burgin, M. S., and De Lannoy, G. J.: Precipitation estimation using L-band and C-band soil moisture retrievals, Water Resources Research, 52, 7213–7225, 2016.

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.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev. Discuss, pp. 1–36, 2016.

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.

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 Delft, G., El Serafy, G., and Heemink, A.: The ensemble particle filter (EnPF) in rainfall-runoff models, Stochastic Environmental Research and Risk Assessment, 23, 1203–1211, 2009.

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.