Dear Prof. van den Hurk,

Likewise the companion paper we extended our revisions of this manuscript to better reflect on the comments given by the editor and the respected reviewers.

Below please find our updated detailed response to the reviewer's comments. To elaborate on the changes made to the manuscript we are submitting a marked-up pdf as a supplement file.

We also made changes to fig.8 caption and axis to help making it more self-explanatory.

We hope you find the revisions satisfactory.

Kind regards,
Poolad Karimi
On behalf of all co-authors

Response to anonymous reviewer #1

We thank the anonymous referee for his/her detailed review and critical but constructive comments. In the following, we would like to respond to his/her main points and detailed comments:

This is an interesting paper shows the application of the WA+ to the Awash basin in Ethiopia. The paper attempted to assess the uncertainty involved to support decision making for water resources management at a river basin scale. My main comments on the paper, it still lacks giving accurate assessment of the uncertainty for the decision making. While emphasizing advantages of satellite measurements, the paper misses to discuss accuracy of ground measurements, at least for runoff as the most viable method for checking a closure of water balance in a river basin.

The main goal of this paper is to lay the foundation and introduce a standard method for incorporating and assessing impact of the error in remote sensing measurements on water accounting and information provided to users. The Awash basin is used to showcase and test the proposed method. Discussions on accuracy of ground measurements, although very interesting, is not within the scope of this paper. We will revise the paper to be more explicit on this matter.

Detailed comments:

-P 1126, 18, 19: low reliability of utilized flow, and basin closure fraction? This conclusion of the abstract needs to be clarified further. In fact basin closure is a measure to the water balance closure, if it has high uncertainty it indicates large error of the runoff (blue water). The error seems small for other fractions and water balance components, e.g., rainfall and ET, because typically these are much higher than runoff in river basins such as Awash.

WA+ is by default meant to avoid withdrawal and stream flow information. It is directly computing the incremental ET from differences in rainfall and ET. River flow is essential for estimating the non-consumed outflow from the basin, and this is in Awash virtually zero. Hence, there is - in this particular case - no need for runoff measurements in the field. Instead, there appears to be a strong need for estimating groundwater outflow, and this is explicitly mentioned on a few different locations.

- P1127, l 7: "…….and the data quality from field observatories is questionable", this general statement may not be correct, e.g., runoff measurements can be accurate, and vital term to check the accuracy of the water balance.

Runoff is a vital term indeed, and great to verify the consistence with upstream integrated rainfall and ET estimates. Runoff becomes especially valuable if being measured at multiple locations and with a high accuracy. While we agree with the reviewer that flow can be measured accurately, practical experiences show that a perfect accuracy is often not met.

- P1127, l 19: "……………the average error in land use, precipitation and ET is was 14.5, 18.5, and 5.4% respectively? For which time step. Discuss high accuracy of P, and ET.

Response: These numbers are taken from, and discussed in, the companion paper and are based on an extensive literature review. To better inform the readers we will include few supporting lines and provide more information on them.

- P1127, l 19: "……………Such errors are not worse than classical ground based observations? Too general statement, and may not be accurate.

Response: The sentence will be revised to reflect on this point.

- P 1128, l4: "….. river and canal discharge is often based on water levels or the sound of water flow, rather than direct measurements? This statement is not accurate a discharge derived from discharge rating curve can be considered as measured flow.

But it remains an indirect measurement, because not the entire volume of the river discharge is measured. The same comment applies to remote sensing where key variables are measured that are strongly linked to the process to be quantified. The bottom line is that many measurements in hydrology are indirect if you think closer about it, in situ and remote sensing measurements are both indirect. We agree that the additional uncertainty of remote sensing originates from the spatial variability that is very difficult to verify. This is mentioned in the updated version of the manuscript

- P 1129, first and 2nd paragraph: too detailed description on groundwater, while also missing other important features, e.g., land use land cover.

Response: The paragraphs was shortened and land use of the basin was discussed. Groundwater appears to be the big unknown factor in Awash basin, that requires certainly more attention in future studies on the hydrology and irrigation systems of Ethiopia. This is mentioned explicitly on a few different locations in the manuscript.

- P 1130, l 8: do you mean GDP of Ethiopia? Please mention in the text.

Response: Yes, it should read "GDP of Ethiopia". Will be revised.

- P 1134, l 23: define "agricultural production" and give examples.

Response: Agricultural production refers to rainfed and irrigated crops. The term will be defined in the paper.

- P1136, l2, on what bases is the "constant correction factor", will it not affect the error analysis then?

Response: The sentence will be revised to better communicate the point. The total basin area is 116 449 km2 and in simulations the total area was kept constant for keeping conservations of the volumetric water balances. What changed in 1000 simulations was the distribution the total area among to different land use classes.

- P1136, l9: does standard deviation for rainfall or ET, derived from 3 values, give any meaningful information?

Response: Good point, we agree that standard deviation derived from 3 numbers is statistically incorrect for making any conclusion. The paragraph will be revised.

- P1137, l20: "The only possible outlet is underground basin discharge". This assumes no error in P, or ET, so that the closure of the water balance minus runoff must be groundwater recharge. This is not an accurate conclusion, and need to be supported by first confirming accuracy of both P and ET.

Response: The rainfall product is based on FEWS, and this product from USGS was selected because it is one of the few operational rainfall products in Africa that is calibrated real time against local rain gauges. We will not be able to generate a better rainfall product, unless special field campaigns will be organized with manifold rain gauges in the mountains and flood plains. The ET values cannot be validated because actual ET is not compatible with the reference ET0 that could be -at best - computed from the weather stations in the area that are not installed over well clipped grass being not short on water supply. A better method is to compare the difference between rainfall and ET with streamflow, and this comparison demonstrated that the ET (at least as a bulk value) is realistic.

Although no part of this paper, in a subsequent study more remotely-sensed ET models are compared for the Nile basin to further understand the limitations of using a single model ET data set.

- P1138, l4: the assumption of 50% change needs to be supported.

Response: the assumption will be discussed more in the paper and we admit that this remains to be a subjective assessment that can only be solved after having designed a multi-layer groundwater flow model for Awash Basin. The flow must be rather quasi constant due to the thick vadoze zone and should be somewhere between a 10 to 100 % change of flow. Local faults can affect the inter-annual groundwater flow, and we simply do not know this, and for this reason we have been pragmatic by assigning an average value of 50%. We recognize this uncertainty by adding substantial errors in the MC simulations. As such, it affects the results which state that no more water should be allocated for additional utilization due to uncertainty in non-utilized basin outflow.

Response to anonymous reviewer #2:

We thank the anonymous referee for his/her thorough review and his/her helpful comments. In the following, we would like to respond to his/her major points and detailed comments:

The paper is basically a study of how uncertainties in remote sensing data translate to uncertainties in a number of hydrological variables and indicators. This is interesting in itself; it is not clear what the outcomes say about the proposed water accounting sheets. The usefulness of many of the proposed indicators should follow from the relevance they have for decision making. The fact that there are uncertainties is problematic and should be dealt with, but the absence (presence) of uncertainties does not add to (subtract from) the usefulness of the proposed accounting framework.

I have three major comments:

1. What is the value of annual data on all the indicators presented and estimates of the inaccuracy in the annual data given that water scarcity manifests itself within the year, in specific months?

   Response: The WA+ can be used to generate water accounts with higher temporal resolution that annual (monthly for example). This paper aims to investigate and introduce a standard methodology to assess the impact of RS errors on the reliability of outputs of WA+. The same methodology can be used to capture reliability of monthly values, should the accounting be done at monthly scale.

2. As for me there are a lot of questions on definitions of terms and indicators used in Tables 2-3 and Figs. 5-7. I will detail those questions below.

   Response: The terminology used in this paper is based on the standard WA+ terminology that has been reviewed, approved, and used by peers. While we appreciate the referee's interest and his remarks, for the sake of consistency this paper cannot deviate from the already defined terminology.

3. I am puzzled by the conclusion that it is recommended not to use the estimated "utilized flow" and "basin closure fraction" in policy decisions because of the low reliability of the estimates. These indicators are among the most important indicators, more relevant than the underlying variables such as P, ET etc. I can well imagine that uncertainties accrue in top-level indicators, but is the presence of uncertainties a reason not to base decisions on these indicators? In the end they may be more inaccurate, but they are most relevant.

   Response: The sentence needs to be rewritten. It should read "in case of Awash basin in the study period the "utilized flow" and "basin closure fraction" (the degree to which available water in a basin is utilized) have a high margin of error and thus a low reliability." The paragraph will be revised to better communicate the meaning.

We do agree with the reviewer that unfortunately on of the most relevant indicators is also the most uncertain one. This is however true for this case study, and can be different in other basis where the outflow from the basin or sub-basin is accurately measured. A statement is included in the manuscript to make readers more aware that this finding is typical for Awash basin only.

Detailed comments:

p.1127 line 1-4: Why speak about WA and WA Plus? Simply based on the chosen term it is suggested that WA+ is better than WA, while it seems to me that in many cases WA Plus is complementary to WA, rather than an alternative. The two accounting systems are based on different sources (field measurements versus remote sensing images); I guess that these sources enrich each other in terms of the information they provide.

Response: p.1127 line 1-4 are to provide an understanding of WA+ origins to reader. It is not to say WA+ is better than WA. The paragraph will be revised to communicate this more effectively.

Section 2.2: Nothing is said on the required temporal and spatial resolution level of the data on the four sheets. Later on it appears to me that all data are presented on annual basis. This is a poor basis for decision making, since water use, availability and scarcity strongly vary within the year.

Response: We agree with reviewer that capturing temporal variations and reflecting on them is of great value for decision making. There are two points here; 1-WA+ can be used for such purpose, if input data is, for instance, have monthly temporal resolution the outputs of 4 sheets will be for every month. 2- Annual figures also have their own value and are of great importance for policy making and understanding the river basin water management.

We will reflect on this issue in the paper.

p.1140 line 14 & Fig. 8: I don't understand the two lower graphs in the figure. What is on the x-axes of these two figures? I understand that there is a normal distribution of P and ET (the two upper graphs), but I don't see what is meant by the "distribution of the area of irrigated crops", let alone that I understand that this is a bi-modal distribution.

Response: In simulations the total basin area is 116 449 km2 was kept constant. What changed in 1000 simulations was the distribution the total area among to different land use classes. Two graphs show simulated variability of the area of two example classes "irrigated area" and "Closed to open shrublands".

Table 1: it would be useful to specify the period considered when providing numbers for "long-term" P and ET.

Response: Will be revised to include the period of these estimates.

Comments on terms in Table 2:

- I find the term "exploitable water" a dubious term. From Fig.5 I see that it is simply defined as "runoff", so why not call it runoff then? It's a dubious term, because in practice one cannot exploit all runoff, there will always be flows that are difficult or impossible to capture, e.g. flood flows or deep underground runoff flows. In the proposed terminology, they make the difference between "exploitable" and "available" water, but I don't think this terminology is clear. To me it seems that "available" encompasses more than "exploitable", but it appears to be just the opposite.

Response: While we appreciate the reviewer suggestions, the terminology used in this paper follows standard WA+ terminology that is introduced by Karimi et al. (2013) in a paper published in HESS. The terminology is also endorsed by FAO and IWMI and as such used in the data repository www.wateracounting.org.

- What is the temporal unit for dS? Is that daily, monthly, annual?

Response: dS is annual. The unit will be mentioned.

- The definition of "available water" is rather unclear. Fig.5 gives some more info, but what are "reserved outflows" and "non-utilizable flow"? Without clear definitions on how to calculate those variables, the terms remain empty. Reserved by law, by policy..? Environmental flows are generally not well protected, so in practice this means that they are not "reserved" in any formal sense and thus "available". It is necessary to clearly define non-utilizable outflow.

Response: Definitions of these terms have been discussed in details in the earlier published papers on WA+. However to help readers we will introduce them briefly in the paper.

- The reserved outflow fraction (=reserved outflow/outflow) doesn't indicate the degree of meeting the flows set aside for inter-basin transfer etc., it measures the fraction of outflow reserved for inter-basin transfer etc.

Response: The reserved outflow can be bigger than actual outflow. This happens when committed outflow (including environmental flows, downstream water rights, navigational flows, etc) is bigger than actual outflow. The reserved outflow fraction is meant to reflect on whether commitments are met or the outflow is actually less than committed outflow. For instance the fraction of >1 indicates reserved flows are not met.

Comments on Table 3:

- The term "managed ET" is unclear. According to the explanation a better term would be "manageable ET", but then still it's unclear what is understood under ET that can be managed (manipulated). ET can decrease and increase as a result of a great variety of human factors, some of which are likely to occur, others less likely.

Response: As explained above the terminology used in this paper follows the standard WA+ terminology.

- The irrigated ET fraction equals (ET in irrigated agriculture) / (ET in total agriculture). Is this what is meant indeed? Or is it more useful to look at the ratio: (ET of blue water in agriculture) / (total ET in agriculture)?

Response: Yes, The irrigated ET fraction equals (ET in irrigated agriculture) / (ET in total agriculture). The suggested fraction by the reviewer (ET of blue water in agriculture) / (total ET in agriculture) indeed could be very informative too.

Comments on terms in Fig.5:

- I assume that "managed water use" refers to net water withdrawal (consumptive water use, blue water footprint). Would be useful to refer to the other, more commonly used terms.

Response: In addition to blue water, managed water use includes ET from rainfall over land use classes that belong to this group (defined and discussed in Karimi et al (2013)) such as irrigated crops. It has broader definition than the blue water foot print. Consumptive use of water is ET in general and does not relate to the processes through which water has been evaporated.

Comments on terms in Fig.6:
- What is beneficial E?

Response: Beneficial E is evaporation from natural water surfaces is often beneficial, for example in cases where water bodies serve the purpose of fishing, aquatic birds, buffering floods, water sports, leisure, etc.

- What is non-beneficial T?

Response: While T is generally considered as beneficial, it can be considered non-beneficial in some cases such as weed infestations in cropland or in degraded landscapes, or when there are non-desirable plants.

- P1138, l14 such assumptions “If we assume that one-third of rainfall surplus…” cannot lead to the conclusion given earlier that satellite measurements can be more accurate than ground measurements, see related comments given above.

Response: The aim of this paper by no means is to argue that satellite measurements can be more accurate than ground measurements. The paper acknowledges the error in satellite measurements and is designed to reflect on the impact of these error on the reliability of outputs of water accounting.

- P1140, l10: would be good to briefly describe the method of error analysis here for reference.

Response: Will be included.

Below text was added to the introduction section to briefly describe the method of error analysis:”

Karimi and Bastiaanssen (2014) showed that the average error in land use mapping, and annual or seasonal precipitation and evapotranspiration estimates on the basis of multi-spectral remote sensing data was 14.5, 18.5, and 5.4 % respectively. These figures are based on a comprehensive literature review in which for each variable several numbers of peer-reviewed publications post 2000 were consulted for reported difference of satellite-based estimates from conventional ground measurement. Results of the study show that errors in satellite based estimates are within an acceptable range and comparable to errors reported in conventional ground-based observations”