Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-233-AC4, 2019

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Can we trust remote sensing ET products over Africa?" *by* Imeshi Weerasinghe et al.

Imeshi Weerasinghe et al.

imeshi.nadishka.weerasinghe@vub.be

Received and published: 8 October 2019

Overview

We want to thank the reviewers for their dedication in reviewing the manuscript. We also are thankful for their thoughtful and constructive suggestions and comments. We have addressed all the comments raised by the reviewers and the manuscript has improved from the proposed changes.

Reviewer 2

Summary

The article "Can we trust remote sensing ET products over Africa?" by Imeshi Weeras-

inghe et al. presents an evaluation analysis of the eight satellite - based evapotranspiration (ET) products over selected African river basins against the ET estimates derived from the water balance equation. The main conclusion of the study ranks the selected ET products in accordance with the results of the comparison analyses. The topicality and scientific relevance of the research question addressed in this study is high considering the sparseness of the in situ ET data in the region as well as the urgency of having a high quality ET estimates for the climate related problems in Africa. However, at this point I cannot recommend publication of this article as it (i) – contains a number of significant methodological inaccuracies and (ii) – gives poor explanation and presentation of the performed analyses and graphics. Also, stylistically and structurally the manuscript needs a substantial improvement. I highly recommend a major revision of the manuscript followed up by an internal review prior to the re-submission. Details follow.

Authors Response

We thank Reviewer 2 for his/her time spent on the review and constructive suggestions and feedback that he/she has given to improve the manuscript. Below, we address the issues that were raised for the improvement of the manuscript.

Reviewer 2, General Comment

Generally, the presentation style of the paper makes it often hard to understand the correctness and hence the added value of the illustrated results. The lack of accompanying relevant information in a well-written form along with the multitude of presented data combinations in a variety of forms and at different scales in many cases confuses an understanding of (i) which data sets were used for this concrete calculation, (ii) in which form the data went into the following graphic, (iii) what the estimates were compared to, (iv) how many and which basins were used this time, (v) when data mean was used and over which scales the averaging was done? In my opinion, the paper did not succeed in wrapping up the results in a clear manner. Usage of multiple data

levels, i.e. 3 reference rain datasets, 8 ET products, 27/20 basins with/without trends, different temporal resolutions (from one value to time-series), two spatial levels (from basin integral to pixel-basis) comprises a fairly large number of levels of information which the authors should unwrap and present in a very simple, consequent and logical manner. In the present version of the paper this have not been archived. The presentation style needs a thorough improvement, including restructuring of the manuscript, improvement of English grammar and scientific wiring style itself. One of the major remarks is that the whole manuscript text is written in a very intermittent and superficial manner. The explanations throughout the whole paper are significantly lacking concreteness. Also, confusion and replacement of some terms used throughout the study (e.g. trends as trends or trends as tendency to show certain value, among others) together with the multitude of abbreviations used in the text makes it very hard to follow the presentation (details given below). Finally, some data descriptions and methodological assumptions (in particular, downscaling to the smallest grid and usage of different time periods) raises a number of questions related to their correctness and validity. The details are given further below.

Authors Response

The paper has gone through a substantial restructuring in terms of the presented figures and different sections and sub-sections. One of the main reasons behind this paper was to use different products with varying temporal and spatial scales for evaluation due to the fact that these products are a sample of the products available for use. In terms of which data sets were used for the concrete calculation, this was mentioned in the introduction and methodology sections. The 9 products being evaluated have all been used in the calculation. They have been resampled to the same spatial and temporal resolution. The average of the three precipitation products have been used and the average of the three PET products have been used. The discharge data has been obtained from GRDC and HYDR and have been converted to yearly averages. For each graphic and section, the methodology has been described. Each section de-

scribed clearly what the ET estimates are being compared to both in the methodology section, in the results section and in the figure captions. According to how many basins are being used in the study. This is clearly identified in the results section of the paper including with graphics. The data mean for basins and land cover elements was used and stated. Long-term averaging was found after finding yearly averages for the different basins and land cover elements. The manuscript is written by an English native speaker and has been checked by a native English professor for structure and grammar. More concrete explanations and statements have been corrected to be clearer. The confusion between certain terms in their use in this paper such as trends and tendencies has been corrected. All abbreviations have been opened and no abbreviation has been used without explanation. As to the questions raised due to downscaling of products, these are addressed further into the specific questions raised. In short, this was done as to not lose any information from the high spatial resolution products at the same time not losing information from the coarser scale products. It would have been simpler to resample all products to the coarsest resolution in terms of computational efficiency and storage space. However, this was not done in order to ensure all features of ET products such as, spatial resolution, were evaluated according to the individual products as would not be the case if resampled from high to coarse. The paper addresses the use of different time periods as to the reasons why they are need and the possible reasons why they can be used.

Authors Changes in Manuscript

The entire manuscript should be shown here as most parts have been adapted and changed. Therefore the comments to the changes in the manuscript to the specific comments have been shown.

Reviewer 2, Specific Comment 1 – Scientific Relevance

From the abstract, introduction and methodology sections it remains unclear how new is the water balance (WB) method, how accurate is the method and which other studies

already used it for similar tasks. The abstract even makes a false impression that the authors developed a method and not used the existing one (p1, L6). The introduction in turn makes an impression that the focus of the study is the methodology (p3, L9) and not the evaluation of the existing products. In this view, the authors must provide an overarching literature review of the studies that already used the WB ET estimates for satellite products evaluation, and also studies which evaluated the same satellite products over Africa using the same or other techniques. One of such study examples is the Miralles et al. (2016), which is also referenced in the present manuscript. Note, that Miralles' study also involved the African river basins. This has to be explicitly mentioned in the introduction. The authors should also then place their results into the findings of others. This is not done at the present state of the paper.

Authors Response

The intention was not to give the impression that the catchment water balance methodology was developed within the scope of this study. The paper has been amended so that this false impression is not given. The introduction has been changed to show focus on the evaluation of the products and not the methodology. A literature review of similar comparisons using some products under evaluation in this study among others has been conducted. (P1,L7), (P4,L4), (P4,L10), (P3,L5), (P19,L19)

Authors Changes in Manuscript

Abstract "Thus we conduct a methodological evaluation of nine existing RS derived and other ET products in order to evaluate their reliability at the basin scale."

Introduction "Therefore, this study focuses on evaluating nine existing, mostly open access, ET products using a water balance approach over Africa." "The evaluation of the products will be conducted using a) a comparison of their performance against calculated ETWB, b) a robustness check of their performance against the Budyko curve which provides a reference condition for the water balance assuming it correctly partitions P into Q and c) a spatial variability assessment using specific land cover elements

(forests, water bodies and irrigated areas)." "Satellite observations often give useful information on the spatial variability, however many products tend to suffer from a large bias. With this range of approaches to estimate ET, large differences are observed among the products and therefore, evaluation is required. Keeping in mind limited availability of in situ measurements for evaluation, an alternate approach is to consider the water balance closure at the river basin scale. Only few studies exist comparing different satellite based and gridded ET products at the global and continental scales using this approach among others. In their study, (Miralles et al. 2016) evaluated four commonly used and tested algorithms (the Surface Energy Balance System (SEBS): (Su 2002), the Moderate Resolution Imaging Spectroradiometer (MOD16): (Qiaozhen Mu et al. 2007; Q Mu et al. 2011), the Global Land Evaporation Amsterdam Model (GLEAM): (Martens et al. 2017) and the Priestly-Taylor Jet Propulsion Laboratory model (PT-JPL): (Fisher, Tu, and Baldocchi 2008) to derive ET using a range of methods including water balance closure across a broad range of catchments worldwide. They found that GLEAM and PT-JPL appear more realistic when compared with 837 globally distributed catchments, however find that all products show large dissimilarities in conditions of water stress and drought conditions (Miralles et al. 2016). Another global evaluation of three process-based models (SEBS, Penman-Montieth algorithm (PM-Mu): (Qiaozhen Mu et al. 2007; Penman 1948; Montieth 1965) and Priestly-Taylor based approach (PT-Fi): (Priestley and Taylor 1972; Fisher, Tu, and Baldocchi 2008) in their estimation of ET was conducted by (Vinukollu, Meynadier, et al. 2011) using the water balance approach at twenty six major basins worldwide along with other methods. A Root Mean Square Difference (RMSD) of 118 to 194 mm/year and bias of -132 to 53 mm/year were found between the estimated annual ET and water balance approximations. The LandFlux initiative, supported by GEWEX (http://www.gewex.org/) is a framework aiming to evaluate and compare several global ET data sets (Mueller et al. 2011; Jiménez et al. 2011). With these aims, global merged bench-marking ET products were derived (Mueller et al. 2013) using 40 datasets over a seven year period (1989-1995) and 14 datasets over a seventeen year period (1989-2005) to be used for

evaluation. At the continental scale a study by (Trambauer et al. 2014) compared ET estimates derived using a continental hydrological model (PCR-GLOBWB: (Van Beek and Bierkens 2009) with other independently computed ET products (the European Center for Medium-range Weather Forecasts (ECMWF) Re-Analysis (ERA)-Interim: (Dee et al. 2011), ERA-Land: (Balsamo et al. 2015), MOD16, GLEAM and three other versions of the PCR-GLOBWB model) using visual inspection and statistical methods. By sub-diving the continent into climatic regions, they found that the annual anomalies of ET for each of the products with respect to the multi-product mean was highest in ERA-Interim. GLEAM was in most cases lower than the multi-product mean while PCR-GLOBWB was close to the multi-product mean in nearly all cases."

Discussion "In terms of consistency in results with previous studies conducted on some of the products under evaluation we see similar tendencies. According to (Miralles et al. 2016) GLEAM, MOD16 and other products in their study show divergences in conditions of water stress and drought. Considering large parts of Africa are potentially under water stress due to the semi-arid and arid climate (IPCC 2019; World Bank 2018), this can explain the low ranking of GLEAM and MOD16 in this study. The RMSE and biases found in our study for Africa are comparable with those found by (Vinukollu, Wood, et al. 2011) at the global scale, however comparing different products to that of this study. The range is higher in this study for Africa than the range found at the global scale. In their study, (Trambauer et al. 2014) found GLEAM to underestimate ET in terms of their multi-product mean. This is again consistent with our finding where biases in GLEAM showed large underestimations across the basins in Africa with respected to the calculated ETWB. We used the LandFlux-EVAL benchmark product as an ensemble product without calculating the multi-product mean of the products being used in this study, as it was developed using a large range of ET products. LandFlux-EVAL, with the coarsest spatial resolution, ranked fourth in the final ranking only outranked by the products with the three highest spatial resolutions in this study, CMRSET, SSEBop and WaPOR. Therefore, LandFlux-EVAL performs well overall regardless of it's coarse resolution and is interesting due to being an ensemble

product. Therefore, continuation or commencement of a similar initiative to develop a benchmark product using a range of ET data sets including high resolution products ranked within this study may improve the ensemble product for future use."

Reviewer 2, Specific Comment 2 – Scientific Relevance

The authors should be more careful in formulating their scientific conclusions. The following sentence in the abstract: "However our recommendation would be the three highest ranked products being CMRSET, SSEBop and WaPOR." sounds rather subjective and needs further motivation (especially considering the huge differences in spatial-temporal scales between the products, as well as the manipulations on interpolation, then vice versa - integration - done within the study). Why first three? The same remark applies for the conclusions.

Authors Response

We have changed subjective recommendations within the study to be more conclusive based on the findings. (P1,L18), (P20,L23)

Authors Changes in Manuscript

Abstract "Based on the evaluation criteria in this study the three highest ranked products, CMRSET, SSEBop and WaPOR would suit many of the needs of readers due to low biases and good spatial variability."

Conclusion "Therefore, if we answer our question of whether to trust remote sensing estimates of ET across Africa, the answer is not black and white. Yes, in general we can trust the products under evaluation in this study. CMRSET, WaPOR and SSEBop show low biases in estimations and a good spatial distribution of ET patterns. Each of these products have relatively high resolutions and both CMRSET and SSEBop are global products. Depending on the study under question, whether an early and long time period is needed, whether a higher or lower resolution is required, whether looking at the global or regional scale or whether looking only at land evapotranspiration, a

different product may be more suited than another. However, a large consideration to be kept in mind for Africa, is that the three highest ranked products, CMRSET, SSEBop and WaPOR have low biases and perform well in spatial variability and will suit most needs within a given study."

Reviewer 2, Specific Comment 3 – Scientific Relevance

The study does not mention anything at all about the quality of the reference precipitation data sets, nor about the quality of the final WB- based ET product. The study should also provide or mention some quantitative assessment to the magnitude of the differences which arise only due to application of different rainfall products. This will (i) - substantiate introduction of three different rainfall products in the paper and (ii) – justify better the obtained differences in ET between the products.

Authors Response

We have now used the average of the three precipitation products due to only slight differences found in the calculated mean precipitation between the three products also considering that we use the average of the three products for ranking purposes. The comparison between the three precipitation products have been included in the paper. We have also tried to evaluate the reasons behind the differences between the ET products within the conclusion. This was already mentioned somewhat in the discussion (P5,L23), (P9,L25), (Figure 5), (P20,L15).

Authors Changes in Manuscript

"Since direct observations of precipitation from gauges were not used, precipitation was taken as the average of the three data products EWEMBI, CHIRPS and MSWEP."

"Precipitation and potential evapotranspiration were taken as the average of three products. Here we compare the results of the different P and PET products for the basins being analysed. We see that the three precipitation products show little differences in their estimations of long-term average P across the basins. No large outliers can be

seen (Figure 5)."

Conclusion "A big difference between the top three ranked products and the others is the high spatial resolution as well as the estimation of ET as a whole rather than only land ET in most other cases. However, no pattern can be found between the product ranking and the forcing or ET calculation methods."

Reviewer 2, Specific Comment 1 – Data Products

The data section needs a major elaboration. Many paragraphs appear rather like a snippets of information with lack of logical sequence, and hence they often fail to deliver main message of the paragraph or peculiarity of the concrete product. My suggestion would be: i) to either extend the data product descriptions to make them more complete and understandable or vice versa, provide only a reference links to the web sources and main papers of the products, and use instead the data section to discuss / group the products by their similarities and differences, advantages and disadvantages which can further help interpreting the paper results. ii) to omit the repetition of the time period and resolution information since they are already given in the table; iii) to place all products into the tables for consistency and clarity, i.e. also precipitation products, discharge data and reference data should be summarised in the same or separate table.

Authors Response

We have decided to take the advice of the reviewer to provide only reference links to the sources and papers of the products with a brief description on the similarities, differences, advantages and disadvantages. All other products have also been summarised in tables consistent with the ET products. (P4,L17), (Table 1), (Table 2), (Table 3), (Table 4)

Authors Changes in Manuscript

"The derived ET products being evaluated in this study include CMRSET, ETMoni-

tor, GLEAM, LandFlux-EVAL, MOD16, MTE, SSEBop, WaPOR and WECANN. Overall there are large differences between the products which results in certain advantages and disadvantages between products. All products have a global spatial coverage (advantage) except for WaPOR (disadvantage). All products are openly accessible (advantage) except for ETMonitor (disadvantage). GLEAM and ETMonitor have a daily, CMRSET has an 8-daily and WaPOR has dekadal temporal resolution (advantage) over other products which have monthly or yearly resolutions(disadvantage). Most products are still ongoing (advantage) except for ETMonitor, LandFlux-EVAL and MTE (disadvantage). GLEAM, MTE and LandFlux-EVAL have data available prior to 1990 (advantage) with all other product data available after 1999 (disadvantage). These different ET products give a good sample of the available data sets to choose from with their many advantages and disadvantages. "

Reviewer 2, Specific Comment 2 – Data Products

Many product descriptions and references miss version numbers. Those must be included, since depending on the product version there might be some already known issues related to a parameter derivations.

Authors Response

We have now included version numbers of the products in Table 1. (Table 1)

Authors Changes in Manuscript

Refer to Table 1 under Reviewer 2, Specific Comment 1 – Data Products

Reviewer 2, Specific Comment 3 – Data Products

Check carefully the correct citations, once you add the product version numbers. I am more familiar with the GLEAM product, and I know that for GLEAM v3 (if you used that version) the correct references are Miralles 2011 (HESS) and Martens 2016 (GMD).

Authors Response

We have carefully checked the references and have included and amended these in Table 1. (Table 1)

Authors Changes in Manuscript

Refer to Table 1 under Reviewer 2, Specific Comment 1 – Data Products

Reviewer 2, Specific Comment 4 – Data Products

It is also a rule of a good scientific practice to provide/cite the data source: a webpage, ftp or a personal communication. No data sources are mentioned in the current manuscript version. For GLEAM, for example it should be the web portal: www.gleam.eu; For MSWEP: http://www.gloh2o.org/ (?), etc.

Authors Response

We have now also added data sources to the manuscript in Table 1. (Table 1)

Authors Changes in Manuscript

Refer to Table 1 under Reviewer 2, Specific Comment 1 – Data Products

Reviewer 2, Specific Comment 5 – Data Products

One of my major remarks here concerns inaccurate or sometimes even false information in data set descriptions. That is unacceptable. Please, check carefully all product descriptions you are giving! On the example of GLEAM: - GLEAM is not a physically-based model, Prietsley-Taylor, the interception loss model, the stress module, and the water-balance model in GLEAM which form the core of GLEAM are all empirical! One can call it a process-based model, as it empirically describes the process needed to estimate E from satellites; - Table 1: GLEAM does not use CMORPH at all! That is simply wrong information. - Alemohammad et al., 2017 is the reference to a paper where they describe another method of deriving E. It is not clear why this is included in this GLEAM section?.

Authors Response

We apologise for the incorrect description of some of the models and have made sure to change or take out the descriptions. The incorrect information has been deleted with the reconstruction of the data section. This should be reflected in Table 1. (Table 1)

Authors Changes in Manuscript

Refer to Table 1 under Reviewer 2, Specific Comment 1 – Data Products

Reviewer 2, Specific Comment 1 – Methodology and results

As it was already mentioned earlier, the presentation of the calculation steps is done in a rather superficial manner. Lots of information is not given or remains unclear. E.g.: - which concrete quality control steps were involved in the selection of the basins and, which additional analyses were done and by whom? (e.g. p8, L7-10) - how were the basin boundaries defined? - what is the time-period of available discharge data? - how the integration over the basins is exactly done? were the simple mean or the areaweighted mean of ET or P fields used when averaging over the basin area? - which manipulations were done with the precipitation data prior averaging it over the basins? Were the data also re-scaled to the 0.0022 deg resolution and then averaged over the basins? Never mentioned. - In their paper Miralles et al., 2016 applied additional quality control check on the difference between the GRDC-reported area and the area calculated from basin boundaries. Would not it be also relevant for the present study? - How was the MPM calculated? The products have big differences in resolution. The averaging of the products to get the MPM without applying corresponding weights can be a source of errors.

Authors Response

We mention in the paper that we select the basins considering the availability of the discharge data at the outlet of a particular basin. This was how the initial 27 basins were selected. This was mentioned in the previous version and has not changed in

this current version of the manuscript. A complete restructure of the methodology section has been done in order to incorporate the suggestions from this particular comment. Specifically, we now mention how the basin boundaries are defined (from the 'Major River Basins of the World' (MRBW) shapefile (World Bank 2017)). The available time period of the discharge data has been added and can be found in Table A1 in Appendix A. Integration over the basin has been done according to the MRBW shapefile boundaries for each basin with the mean of each basin within the shapefile recorded. This description has been added to the manuscript. The basin area weighted mean for averaging over a basin was only used in the statistics for RMSE and bias, otherwise the simple mean was used as stated in the manuscript previously and in this new version. No manipulations were done to the precipitation data prior to averaging over the basins except for obtaining yearly averages. This is now mentioned. Yes the data was rescaled to 0.0022 deg resolution and then averaged over the basins. This is now mentioned. We have also conducted an analysis on the difference between the MRBW shapefile areas taken for the basins and the area reported by the GRDC and mentioned only the potential problematic cases. We have now not used the MPM but an existing benchmark ensemble product, LandFlux-EVAL. (P5,L20), (Table A1), (P5,L24), (P5,L1), (P4,L26),

Authors Changes in Manuscript

"Catchment or basin areas were taken from the 'Major River Basins of the World' (MRBW) shapefile (World Bank 2017). Discharge was converted from cubic meters per second to millimeters per year using the above mentioned catchment areas for all years of data availability for each basin."

"Basin average precipitation was calculated for the years 1979-2016 according to the MRBW shapefile boundaries recording the basin mean."

"Precipitation products were averaged at yearly temporal resolution for the purposes of this study."

“All products have been projected and gridded on a 0.0022 deg resolution geographic grid and averaged at yearly temporal resolution for the purposes of this study.”

“The MRBW shapefile area did not differ greatly with the drainage area reported by the GRDC except in two cases. Here we found the ETWB calculated using the two areas only differed by 2.5 percent and 3.3 percent and thus kept these basins in the analyses.”

Reviewer 2, Specific Comment 2 – Methodology and results

(i) - It remains not very clear from the text over which values the correlation analysis is performed? Over time-series of annual means? Over multi-year averages of different basins? Should be put more clear. (ii) - Units of the correlations are not common, and confuse the interpretation. Correlation should rank from -1 to 1. Besides, it is never clear from all the graphs with percentages, by what value the normalisation was done. (iii) – Correlations should always provide significance measure, or the latter should be mentioned in the text.

Authors Response

(i) - The correlation was performed over multi-year averages (long-term averages) across all basins under evaluation (20 basins). This has been mentioned in the manuscript. (P6,L16) (ii) – Correlations have been changed to -1 to 1 without units. The normalisation for each basin was conducted based on the calculated ETWB for each basin. These have been adapted in the manuscript. (Table 5), (P11,L15) (iii) – This has been adapted in the manuscript to reflect the significant measure. (P11,L11)

Authors Changes in Manuscript

(i) - “Correlations were calculated based on long-term averages across all basins.”

(ii) - Table 5. “Percentage biases were normalised based on the calculated ETWB for each basin.”

(iii) – “There is a significant positive correlation for all products ranging from 0.89-0.97 with GLEAM and LandFlux-EVAL showing the strongest relationships with ETWB”

Reviewer 2, Specific Comment 3 – Methodology and results

The choice of the highest resolution is one of the two major remarks that I have to the methodological part: (i) – Generally, it is not common to interpolate products to the highest resolution, especially when the difference between the highest and the coarsest resolution is that high. It would be more correct to upscale the higher resolved data to the coarser estimates to minimise the bias. (ii) – Besides, the fact that the comparison of the products is mostly done at the basin level, the downscaling does not seem to make sense at all. First you interpolate the coarse data to the very high resolution, and then, you integrate it back again over the river basin. This clearly can be a source of additional biases and errors, which also raises my doubts about the validity of the ranking results. (iii) – All the above inter alia also raises a question of what is the minimal area of the smallest basin you have, and whether it is resolved by the products with the coarsest resolution at all?.

Authors Response

(i) - We did not upscale the higher resolution products to coarser resolution products as our aim was not to minimise bias between products but to evaluate each product independently of the other according to their features (e.g. spatial resolution) regardless of whether they were advantageous or disadvantageous. I believe if we were trying to find the effects of ET calculation methods or forcing data and looking at the different products for comparison this would make sense that we would need to minimise other biases between the different products. However, as the goal of our manuscript is to evaluate different offerings of ET products regardless of temporal and spatial resolutions or coverage, forcing data, ET calculation method, etc, it does not make sense for us to minimise bias between products. We want to calculate the bias of each product with respect to a reference (ETWB). In this respect we do not lose any information by

downscaling from coarse to high resolution and in fact we found that for basin boundaries to be as close to the shapefiles as possible, the data set of GLEAM, for example, showed very little difference (approximately 0-5mm/year) in the estimation of mean ET for different basins when using the coarse resolution or the resampled high resolution dataset. Whereas a slightly larger difference range was found for the CMRSET product (0-50mm/year) when comparing the high resolution product to the resampled coarse resolution for a sample of basins. In this regards we did not believe that there was a disadvantage by resampling to the highest resolution of the products. Also, since we were trying compare individual products on their own merits of each feature, this enable a more accurate comparison without minimising bias between the products. (ii) – As with point (i) above, we did not find that there were any additional biases and errors from this method of evaluation and found very small differences when looking at the two different methods of up or downscaling resolutions. In fact, the biases were greater when resolving from high resolution to coarse resolution and the biases slightly smaller when resolving from low resolution to high resolution when looking at certain basin means according to the reference used (ETWB). Although since very small differences were seen, they were almost negligible. Therefore the validity of the ranking results still hold. (ii) – The smallest basin is >30000km2 which for certain products such as LandFlux-EVAL and WECANN, this would not be fully resolved. However the intent was to also include basins that were smaller in our analysis to: 1. Have a good spatial coverage across Africa and 2. Have a range of basin sizes to evaluate the products on. Even though we do not see any spatial variability in products such as LandFlux-EVAL and WECANN in the smaller basins. Their prediction of long-term average ET for smaller basins, especially LandFlux-EVAL showed lower biases than higher resolution products such as ETMonitor.

Authors Changes in Manuscript

For the above reasons, nothing with regard to this point was changed within the manuscript.

Reviewer 2, Specific Comment 4 – Methodology and results

The second and the most major remark of mine is related to the application of the analyses at different time-periods. (i) - First of all it never comes clear what is the time period of available discharge data for every basin; (ii) – From the Tabel 2 it appears that for most of the basins discharge data does not extend the whole period of available precipitation data at all, and the spread of data periods is huge among the basins. In this view I do not understand at all how the analyses tests were done? (ii) – The test for the effect of temporal variability on annual means mentioned in the discussion section was done only for the four basins, while 20 basins are analyzed throughout the study. Moreover, Congo - one of the four tested basins - has data only till 2010, while remote sensing ET products span up to 2017. In this view, I would not be able to call it a fair validity test! (iii) – Clearly, the exclusion of periods with trends does not account for the temporal variability of data which can still result in the pretty different annual means. So, the effect of temporal variability on annual means must be done for all the basins, which are used for the evaluation of the satellite ET products in order to draw a fair conclusions. (iv) – Calculating trends only for the WB ET reference data set is not a complete analysis. If a satellite data product has a trend, this also has to be mentioned, and maybe even that product should not participate in the validation (?) To conclude, if the tests will show that the variability indeed matters, then none of the performed analyses is valid since they will all be affected by the differences due to variability.

Authors Response

(i) – The period of available discharge data for each of the 27 basins where ETWB was initially calculated for has been added and can be found in Table A1 within the manuscript. (ii) – Here we used the long-term average precipitation data minus the long-term average discharge data to calculate the ET. This is mentioned in the manuscript. We also calculated the ET based on the average precipitation and discharge based on the overlapping periods and found a maximum of 5% difference in ET from both methods. In most cases 0% difference was found. (P5,L26) (ii) – the test for

temporal variability was conducted on the selected four basins as they were the only basins with long enough time periods to conduct this test. We used these as samples to show that in all cases tested the difference in calculated ET was minimal. Congo was also used, although it only had data until 2010 so that we could test more than just three basins. The corresponding periods were used in the remote sensing products so we made sure the periods tested overlapped. In this regard we believe this was a fair validity test. (iii) – We agree that if possible the test for temporal variability should have been conducted on all basins, but we are unable to do this based on the available time periods of the data. Therefore we took the four basins as a sample and surmised that due to finding minimal differences in calculated ET, we could compare long-term averages from different time periods. (iv) – Considering the point is evaluating the prediction of ET estimates by different ET products, whether they do or do not have trends is not a basis for inclusion or not in this analysis. It would be interesting to find out whether the different products do show trends for particular basins but this was not in the scope of this study. Our goal was to evaluate based on a reference which we needed to calculate in an accurate manner, which would mean no trends when looking at long-term averages of different time periods. Therefore our 20 basins under analysis without trends in their calculated long-term ETWB was used as a reference to compare with ET product estimations. If an ET product had a trend in their estimation of ETWB for one of the 20 basins under analysis, then this would most likely have a higher bias when compared with the reference ETWB. It was not relevant for us to see if ET products had trends or not, only to evaluate their ET estimates.

Authors Changes in Manuscript

"Long-term ETWB was calculated by using the long-term average discharge and precipitation data for each catchment."

Nothing else for this point was changed in the manuscript.

Reviewer 2, Specific Comment 5 – Methodology and results

My advice would be to not use percentages for all the figure results. This only confuses the interpretation. Use -1 to 1 scale for correlation and differences.

Authors Response

We did not use percentages for the overall statistics in Table 5. However, we still used percentages when looking at bias and basin area weighted bias as subjectively we found these results easier to interpret and was given this advice from different advisers of the manuscript. (Table 5) Authors Changes in Manuscript

Refer to Table 5 under Reviewer 2, Specific Comment 2 – Methodology and results

Reviewer 2, Specific Comment 6 – Methodology and results

The raking of products based on visual inspection is rather speculative for me. For ex on Fig 9 it was impossible to follow the text conclusions: I did not see where irrigation C7 HESSD Interactive comment Printer-friendly version Discussion paper area is, to which reference product other products are compared, why MPM and some other products have no data and why GLEAM is concluded to perform worst? The same for Fig 10..

Authors Response

We re-did this section to include forest as an additional land cover type for inspection as well as using highlighted areas across the entire African continent rather than zooming into selected areas shown in figure 5. We agree that this is a rather subjective method, however, we also believe it is relatively visible to see the difference between the products especially when using different scales to be able to visually interpret the results. We believe these maps are an added advantage for reader interested in spatial characteristics of ET. However, due to its potential subjectivity in ranking, we do two final rankings, with and without visual inspection to minimise this subjectivity. (P13,L4), (Figure 10)

Authors Changes in Manuscript

“Figure 10 shows ET across Africa for all ET products with the specific land cover elements (forest, irrigated areas and water bodies) highlighted. Two different scales are used in order to be able to visually compare the products according to spatial variability rather than magnitude of ET. For products where large biases were found, a scale of 0-1200 mm/year was used and for the remaining products a scale of 0-1800 mm/year was used. Visually, all products capture the forested area. Irrigated areas are also captured well by most products. GLEAM and LandFlux-EVAL do not capture the majority of selected irrigated areas. CMRSET, ETMonitor, SSEBop and WaPOR capture most of the selected irrigated areas while the remaining products capture a few. GLEAM, LandFlux-EVAL, MOD16, MTE and WECANN only estimate land ET and thus do not have ET across water bodies. The remaining products capture the water bodies well, with CMRSET and ETMonitor showing larger differences in their estimations of ET across water bodies than the surrounding areas over SSEBop and WaPOR.”

Reviewer 2, Specific Comment 7 – Methodology and results

The analyses of comparing products over one irrigation area, and one lake, where some products have no data, and others do not even resolve the region does not make much sense to me, nor it is complete enough to make a serious conclusion on which product is better or worst.

Authors Response

Please refer to answer from section Reviewer 2, Specific Comment 6 – Methodology and results.

Reviewer 2, Specific Comment 8 – Methodology and results

It maybe due to the presentation style, but I could not follow the result section presenting the crop coefficients very well. It has to be structured and presented in a more clear manner: objective and reasoning for location, crop types, etc, data used, hypothesis to prove, which products in which form are tested, and what do results show.

Authors Response

This section has been re-written to try and present this concept in a more structured and clear manner in the manuscript. (P8,L11)

Authors Changes in Manuscript

“For irrigated areas, the crop coefficient (kc) was used. The crop coefficient is a property of a plant that aids in determining ET and can be calculated using equation 3.

kc = ET/PET

Where kc is the coefficient for crops growing under conditions of optimum fertility and soil moisture and achieving full production potential (Allen et al. 1998). In reality optimal conditions are rarely met, however this measure was used to evaluate how well the ET products determined ET across irrigated areas. Average crop coefficients for maize, wheat and sugarcane estimated by FAO were used as a reference. The long-term annual average mean ET estimates across irrigated areas were divided by the long-term annual average mean PET estimates across irrigated areas to find the average crop coefficient (kc) across irrigated areas. These estimates were found by looking at the mean of the area according to the AEIai shapefile. The bias between the reference kc from FAO and estimated kc using individual ET product estimates and PET derived using the mean of the three PET products was recorded.”

Reviewer 2, Specific Comment – Stylistic and structural

I highly recommend to look into the papers of Zeng et al., 2012 (ERL) and Miralles et al., 2016 (HESS) as an example of a good presentation style, and especially of the methodological part, as well as their choice of graphs.

Authors Response

The mentioned papers and others as well as internal discussion was conducted and the paper revised along with the figures and graphs. There is a substantial change in the

entire manuscript as well as most figures and graphs to be more clear and structured.

Authors Changes in Manuscript

Changes are found throughout the entire manuscript.

Reviewer 2, Specific Comment 1 – Stylistic and structural

Usage of term throughout the paper changes and confuses the reader. For ex, the term trend is first used as trend itself, but also to indicate tendencies if I understood it correctly (p13, L5 or p16,L9).

Authors Response

The term trend has been looked into and changed according to trend and tendency. As have other terms which we found to be confusing or arbitrary. (P12,L2), (P19,L19)

Authors Changes in Manuscript

“The calculated ET for most of the ET products and also for the majority of basins falls under the curve showing a tendency for products to underestimate basin ET.” “In terms of consistency in results with previous studies conducted on some of the products under evaluation we see similar tendencies.”

Reviewer 2, Specific Comment 2 – Stylistic and structural

Typos are also present throughout the paper (e.g. p3 L2, p7 L17, p15 L7, p20 L3). A proper internal resew is required.

Authors Response

Any typos we have found have been corrected in the manuscript. A spell check and internal review has been conducted again with the new version of this manuscript.

Reviewer 2, Specific Comment 3 – Stylistic and structural

Discussion section rather reads like methods and should be incorporated to methods.

Instead, the discussion section should place the paper findings into the existing knowledge as was already mentioned earlier.

Authors Response

We agree that the discussion should place the paper findings within existing knowledge and have added this to the manuscript. However, we also feel that the discussion section discusses certain assumptions and findings that requires further explanations. We found this style quite useful and interesting to read and did not incorporate these findings into the methods section. (P19,L19)

Authors Changes in Manuscript

“In terms of consistency in results with previous studies conducted on some of the products under evaluation we see similar tendencies. According to (Miralles et al. 2016) GLEAM, MOD16 and other products in their study show divergences in conditions of water stress and drought. Considering large parts of Africa are potentially under water stress due to the semi-arid and arid climate (IPCC 2019; World Bank 2018), this can explain the low ranking of GLEAM and MOD16 in this study. The RMSE and biases found in our study for Africa are comparable with those found by (Vinukollu, Wood, et al. 2011) at the global scale, however comparing different products to that of this study. The range is higher in this study for Africa than the range found at the global scale. In their study, (Trambauer et al. 2014) found GLEAM to underestimate ET in terms of their multi-product mean. This is again consistent with our finding where biases in GLEAM showed large underestimations across the basins in Africa with respected to the calculated ETWB. We used the LandFlux-EVAL benchmark product as an ensemble product without calculating the multi-product mean of the products being used in this study, as it was developed using a large range of ET products. LandFlux-EVAL, with the coarsest spatial resolution, ranked fourth in the final ranking only outranked by the products with the three highest spatial resolutions in this study, CMRSET, SSEBop and WaPOR. Therefore, LandFlux-EVAL performs well overall regardless of it’s coarse

resolution and is interesting due to being an ensemble product. Therefore, continuation or commencement of a similar initiative to develop a benchmark product using a range of ET data sets including high resolution products ranked within this study may improve the ensemble product for future use."

Reviewer 2, Specific Comment 4 – Stylistic and structural

Nothing is mentioned about the Budyko result in Discussion or Conclusions.

Authors Response

We have now included conclusions regarding the Budyko analyses in the conclusion section. (P20,L7)

Authors Changes in Manuscript

"According to the comparison of the ETWB with ETBudyko, we see that ETWB follows the Budyko curve and has an overall low bias across the basins. This indicates the calculated ETWB is a sound reference condition to use for analyses."

Reviewer 2, Specific Comment 5 – Stylistic and structural

Nothing is mentioned about the differences between using three precipitation products.

Authors Response

We now use the average of the three precipitation products and not evaluations based on the individual products and thus do not mention this.

Reviewer 2, Specific Comment 6 – Stylistic and structural

Many abbreviations are not opened. Add abbreviation table to the paper. Use less abbreviations, i.e. if possible leave it open. Very hard to follow.

Authors Response

We have now opened all abbreviations but have not added an abbreviation table to the

paper as all abbreviations are opened.

Reviewer 2, Specific Comment 7 – Stylistic and structural

Explanation of the results in Figures are often not complete. Not clear which products were used, which reference, etc. Figures are too small.

Authors Response

Figures have mostly all been changed being high resolution and easier to read. Explanation of results of these figures tries to be more complete.

Authors Changes in Manuscript

Changes throughout the entirety of the manuscript

Reviewer 2, Specific Comment 8 – Stylistic and structural

Lots of sentences are too wake, "Based on the elements being analysed. . . p21, L30" What is meant? Be more concrete.

Authors Response

We have tried to be stronger in our sentences. Instead of using 'based on the elements being analysed' we have used 'based on the selected land cover elements being analysed' for example.

Authors Changes in Manuscript

Changes throughout the entirety of the manuscript

Reviewer 2, Specific Comment 9 – Stylistic and structural

Titles of the sections should be reconsidered.

Authors Response

Titles of the sections have been reconsidered and changed. Previous headers and

figure captions:

1. Introduction

2. Data

2.1. Remotely Sensed ET products

2.1.1. GLEAM

2.1.2. WaPOR

2.1.3. MOD16

2.1.4. SSEBop

2.1.5. WECANN

2.1.6. FLUXNET-MTE

2.1.7. ETMonitor

2.1.8. CMRSET

2.1.9. Multi-Product Mean

2.2. Precipitation data

2.2.1. EWEMBI

2.2.2. CHIRPS

2.2.3. MSWEP

2.3. Discharge data

2.4. Reference potential evapotranspiration data

3. Methodology

3.1. Preprocessing and data analsysis

3.2. Comparison using WB inferred ET estimates

3.3. Performance with characteristics land cover elements

3.4. Evaluation using the Budyko curve

4. Results

4.1. Catchment water balance

4.1.1. Preprocessing and data analsysis

4.1.2. Comparison using WB inferred ET estimates

4.1.3. Performance with characteristics land cover elements

4.1.4. Evaluation using the Budyko curve

5. Discussion

6. Conclusions

Authors Changes in Manuscript

“. . .

1. Introduction

2. Data and Methods

2.1. Data

2.1.1. Evapotranspiration products

2.1.2. Precipitation products

2.1.3. Discharge data

2.1.4. Reference potential evapotranspiration products

2.2. Methods

2.2.1. Catchment water balance evapotranspiration (ETWB)

2.2.2. Evaluation using the Budyko curve

2.2.3. Spatial variability assessment

2.2.4. Assessment of similarity

3. Results

3.1. Catchment water balance

3.1.1. Comparison of precipitaton and potential evapotranspiration products

3.1.2. Basins used in analyses

3.1.3. Catchment water balance comparison

3.2. Evaluation using the Budyko curve

3.3. Spatial variability assessment

3.4. Product similarity assessment

3.5. Ranking of products

4. Discussion

5. Conclusion

. . . "

References:

Allen, Richard G, Luis S Pereira, Dirk Raes, Martin Smith, and W Ab. 1998. "Guidelines for Computing Crip Water Requeriments-FAO Irrigation and Drainage Paper 56." Crop Evapotranspiration. Rome. doi:10.1016/j.eja.2010.12.001.

Balsamo, G, C Albergel, A Beljaars, S Boussetta, E Brun, H Cloke, D Dee, et al. 2015.

"ERA-Interim/Land: A Global Land Surface Reanalysis Data Set." Hydrol. Earth Syst. Sci 19: 389–407. doi:10.5194/hess-19-389-2015.

Beek, L.P.H. (Rens) Van, and Marc F P Bierkens. 2009. "The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification Report." Utrecht, Netherlands. http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, et al. 2011. "The ERA-Interim Reanalysis: Configuration and Performance Ofthe Data Assimilation System." Royal Meteorology Society 137: 553–97.

Fisher, Joshua B., Kevin P. Tu, and Dennis D. Baldocchi. 2008. "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 (3): 901–19. doi:10.1016/j.rse.2007.06.025.

IPCC, Intergovernmental Panel on Climate Change. 2019. "IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security." https://www.ipcc.ch/site/assets/uploads/2019/08/Fullreport-1.pdf.

Jiménez, C., C. Prigent, B. Mueller, S. I. Seneviratne, M. F. McCabe, E. F. Wood, W. B. Rossow, et al. 2011. "Global Intercomparison of 12 Land Surface Heat Flux Estimates." Journal of Geophysical Research Atmospheres 116 (2): 1–27. doi:10.1029/2010JD014545.

Martens, Brecht, Diego G Miralles, Hans Lievens, Robin Van Der Schalie, Richard A M De Jeu, Diego Fernández-Prieto, Hylke E Beck, Wouter A Dorigo, and Niko E C Verhoest. 2017. "GLEAM v3: Satellite-Based Land Evaporation and Root-Zone Soil Moisture." Geosci. Model Dev 10: 1903–1925. doi:10.5194/gmd-10-1903-2017.

Miralles, D G, C Jiménez, M Jung, D Michel, A Ershadi, M F Mccabe, M Hirschi, et al. 2016. "The WACMOS-ET Project-Part 2: Evaluation of Global Terrestrial Evaporation Data Sets." Hydrol. Earth Syst. Sci 20: 823–42. doi:10.5194/hess-20-823-2016.

Montieth, L. J. 1965. “State and Movement of Water in Living Organisms.” In 19th Symposium of Evaporation and the Environment. Swansea, London: Cambridge University Press.

Mu, Q, F A Heinsch, M Zhao, and S W Running. 2011. “Improvements to a MODIS Global Terrestrial Evapotranspiration Algorith.” Remote Sensing of Environment 115: 1781–1800. doi:10.1016/j.rse.2007.04.015.

Mu, Qiaozhen, Faith Ann Heinsch, Maosheng Zhao, and Steven W Running. 2007. “Development of a Global Evapotranspiration Algorithm Based on MODIS and Global Meteorology Data.” Remote Sensing of Environment 111: 519–36. doi:10.1016/j.rse.2006.07.007.

Mueller, B., S. I. Seneviratne, C. Jimenez, T. Corti, M. Hirschi, G. Balsamo, P. Ciais, et al. 2011. “Evaluation of Global Observations-Based Evapotranspiration Datasets and IPCC AR4 Simulations.” Geophysical Research Letters 38 (6): 1–7. doi:10.1029/2010GL046230.

Mueller, B, M Hirschi, C Jimenez, P Ciais, P A Dirmeyer, A J Dolman, J B Fisher, et al. 2013. “Benchmark Products for Land Evapotranspiration: LandFlux-EVAL Multi-Data Set Synthesis.” Hydrol. Earth Syst. Sci 17: 3707–20. doi:10.5194/hess-17-3707-2013.

Penman, H L. 1948. “Natural Evaporation from Open Water, Bare Soil and Grass.” Source: Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences. Vol. 193.

Priestley, B., and R. Taylor. 1972. “On the Assessment of Surface Heat Flux and Exporation Using Large-Scale Parameters.” Monthly Weather Review 100 (2): 81–92.

Su, Z. 2002. “The Surface Energy Balance System (SEBS) for Estimation of Turbulent Heat Fluxes.” Hydrology and Earth System Sciences. Vol. 6. https://www.hydrol-earth-syst-sci.net/6/85/2002/hess-6-85-2002.pdf.

Trambauer, P, E Dutra, S Maskey, M Werner, F Pappenberger, L P H Van Beek, and

S Uhlenbrook. 2014. “Comparison of Different Evaporation Estimates over the African Continent.” Hydrol. Earth Syst. Sci 18: 193–212. doi:10.5194/hess-18-193-2014.

Vinukollu, Raghuveer K., Remi Meynadier, Justin Sheffield, and Eric F. Wood. 2011. “Multi-Model, Multi-Sensor Estimates of Global Evapotranspiration: Climatology, Uncertainties and Trends.” Hydrological Processes 25 (26): 3993–4010. doi:10.1002/hyp.8393.

Vinukollu, Raghuveer K, Eric F Wood, Craig R Ferguson, and Joshua B Fisher. 2011. “Global Estimates of Evapotranspiration for Climate Studies Using Multi-Sensor Remote Sensing Data: Evaluation of Three Process-Based Approaches.” doi:10.1016/j.rse.2010.11.006.

World Bank. 2017. “Major River Basins Of The World.” World Bank Data Catalogue. https://datacatalog.worldbank.org/dataset/major-river-basins-world.

World Bank, 2018. Beyond Scarcity Water Security in the Middle East and North Africa. Edited by World Bank. MENA Development Report: Washington DC: MENA Development Report: doi:10.1192/bjp.111.479.1009-a.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-233, 2019.

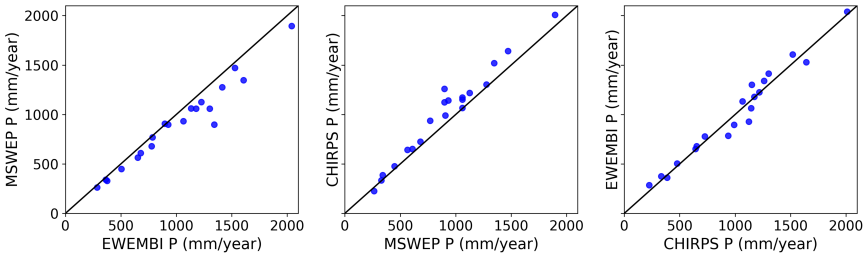

Fig. 1. Figure 5: Comparison of EWEMBI, MSWEP, CHIRPS precipitation products on their prediction of mean P across the basins

Table 1. Characteristics of remotely sensed ET products

Product	Temporal Coverage	Spatial Coverage	Temporal Resolution	Spatial Resolution	Estimation Approach	Input Data Source	Reference
CMRSET (v20140423)	2000-2013	Global	8-daily	$0.0022^\circ \times 0.0022^\circ$	P-T Equation, relationship between EVI and GVMI	MODIS	(Guerschman et al., 2009)
	Access: http://remote-sensing.nci.org.au/u39/public/html/wirada/index.shtml						
ETMonitor	2008-2013	Global	daily	$0.005^\circ \times 0.005^\circ$	P-M, Gash model, Shuttleworth-Wallace	MODIS	Zheng et al. (2016)
	Access: email first author in reference						
GLEAM (v3.2a)	1980-2016	Global	Daily	$0.25^\circ \times 0.25^\circ$	P-T Equation, soil stress factor	AMSR-E, LPRM, TRMM	Martens et al. (2017); Miralles et al. (2011)
	Access: www.gleam.eu						
LandFlux-EVAL	1989-2005	Global	Monthly	$1^\circ \times 1^\circ$	Ensemble Approach	See reference	Mueller et al. (2013b)
	Access: https://iac.ethz.ch/group/land-climate-dynamics/research/landflux-eval.html						
MOD16 (vA3)	2000-2014	Global	Monthly	$0.0083^\circ \times 0.0083^\circ$	P-M Equation, surface conductance model	MODIS	Mu et al. (2011, 2007)
	Access: https://modis.gsfc.nasa.gov/data/dataprod/mod16.php						
MTE (vMay12)	1982-2012	Global	Monthly	$0.5^\circ \times 0.5^\circ$	MTE approach, training using in-situ observations, flux tower data	Eddy Covariance, in-situ	Jung et al. (2011)
	Access: https://climatedataguide.ucar.edu/climate-data/fluxnet-mte-multi-tree-ensemble						
SSEBop (v4)	2003-2017	Global	Monthly	$0.0096^\circ \times 0.0096^\circ$	P-M Equation, ET fractions from T_s estimates	MODIS	Senay et al. (2013)
	Access: https://earlywarning.usgs.gov/fews/search						
WaPOR (v1.1)	2009-2017	Africa	Dekadal	$0.0022^\circ \times 0.0022^\circ$	P-M Equation, calculates E, T and I separately	MODIS, GEOS-5/MERRA	FAO (2018)
	Access: https://wapor.apps.fao.org/home/1						
WECANN (v1.0)	2007-2015	Global	Monthly	$1^\circ \times 1^\circ$	ANN approach, training using observations and model based LE	GOME-2	Alemohammad et al. (2017)
	Access: https://avdc.gsfc.nasa.gov/pub/data/project/WECANN/						

Fig. 2. Table 1: Characteristics of Remotely sensed ET products

Table 2. Characteristics of precipitation products

Product	Temporal Coverage	Spatial Coverage	Temporal Resolution	Spatial Resolution	Input Data Source	Reference
EWEMBI (v1.1)	1979-2016	Global	daily	0.5° × 0.5°	ERA-Interim, WFDEI: (Weedon et al., 2014), E2OBS: (Calton et al., 2016)	(Lange, 2019)
	Access: http://doi.org/10.5880/pik.2019.004					
CHIRPS (v2.0)	1981-2019	quasi-Global	daily	0.05° × 0.05°	in situ precipitation gauges, TRMM: (Huffman et al., 2007), CMORPH: (NCAR, 2017)	Funk et al. (2015)
	Access: https://chc.ucsb.edu/data/chirps					
MSWEP (v2.2)	1979-2017	Global	3-hourly	0.1° × 0.1°	in situ precipitation gauges, CMORPH, TRMM, GSMaP: (JAXA, 2009), IRA-Interim	Bai and Liu (2018)
	Access: http://www.gloh2o.org/					

Fig. 3. Table 2: Characteristics of precipitation products

Table 3. Characteristics of discharge data

Product	Temporal Coverage	Spatial Coverage	Temporal Resolution	Spatial Resolution	Input Data Source	Reference
GRDC (v1.1)	1806-2019	Global	daily	point data	in situ discharge gauges	
	Access: https://www.bafg.de/GRDC/EN/Home/					
HYDR-VUB	1932-2018	Global	daily	point data	in situ discharge gauges	
	Access: on request to http://www.hydr.vub.ac.be/					

Fig. 4. Table 3: Characteristics of discharge data

Table 4. Characteristics of potential evapotranspiration products

Product	Temporal Coverage	Spatial Coverage	Temporal Resolution	Spatial Resolution	Input Data Source	Reference
Hargreaves	1979-2012	Global	daily	0.05° × 0.05°	WFDEI, SRTM DEM	(Sperna Weiland et al., 2015)
	Access: https://wci.earth2observe.eu/					
Penman-Montieth	1979-2012	Global	daily	0.05° × 0.05°	WFDEI, SRTM DEM	(Sperna Weiland et al., 2015)
	Access: https://wci.earth2observe.eu/					
Priestly-Taylor	1979-2012	Global	daily	0.05° × 0.05°	WFDEI, SRTM DEM	(Sperna Weiland et al., 2015)
	Access: https://wci.earth2observe.eu/					

Fig. 5. Table 4: Characteristics of portential evapotranspiration products

Basin	Variable	Data Availability	Trend	hypothesis	p-value	z-value	no. of samples
Awash	ET	1990-2004	no trend	false	0.2496	+1.1514	14
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1990-2004					15
Bandama	ET	1979-1996	no trend	false	0.7619	-0.3030	18
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1970-1996					27
Blue Nile	ET	not enough ET data points to conduct MK test on calculated ET					4
	P	1979-2016	no trend	false	0.6875	-0.4023	38
	Q	1900-1982	decreasing	true	0.0009	-3.3271	83
Buzi	ET	not enough ET data points to conduct MK test on calculated ET					5
	P	1979-2016	no trend	false	0.4210	-0.8046	38
	Q	1957-1983	no trend	false	1.0	0.0	23
Cavally	ET	1979-1996	no trend	false	0.54449	-0.6060	18
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1970-1996					27
Congo	ET	1979-2010	no trend	false	0.0830	-1.7336	31
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1903-2010					108
Cunene	ET	1980-2015	increasing	true	0.0003	3.5823	36
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1980-2015					36
Gambia	ET	not enough ET data points to conduct MK test on calculated ET					5
	P	1979-2016	no trend	false	0.2579	1.1315	38
	Q	1979, 1981-82, 1984,1988	no trend	false	0.8065	0.2449	5
Groot	ET	1979-2015	no trend	false	0.1697	1.3733	37
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1964-2014					51
Kamoe	ET	1979-1996	no trend	false	0.3633	-0.9091	18
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1970-1996					27
Lake Chad	ET	not enough ET data points to conduct MK test on calculated ET					4
	P	1979-2016	increasing	true	0.0194	2.3384	38
	Q	1983-1986	no trend	false	0.3081	-1.0190	4
Maputo	ET	not enough ET data points to conduct MK test on calculated ET					5
	P	1979-2016	no trend	false	0.3393	-0.9555	38
	Q	1953-1983	no trend	false	0.1261	-1.5297	31
Mono	ET	1979-2007	no trend	false	0.5115	-0.6565	29
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1944-2007					64
Niger	ET	1979-2006	no trend	false	0.6214	0.4939	28
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1970-2006					37
Nile	ET	not enough ET data points to conduct MK test on calculated ET					6
	P	1979-2016	no trend	false	0.2909	1.0560	38
	Q	1912-1984	no trend	false	0.0693	1.8164	56
Okavango	ET	1979-2014	increasing	true	0.0127	2.4926	36
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1950-2014					65
Olifant	ET	1979-2014	no trend	false	0.9457	0.0681	36
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1927-2014					88
Orange	ET	1979-2016	no trend	false	0.6691	0.4274	38
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1936-2014					79
Queme	ET	1979-80, 1982-84, 1990-2005, 2007	no trend	false	0.3377	0.9587	22
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1948-2007					60
Rufiji	ET	not enough ET data points to conduct MK test on calculated ET					0
	P	1979-2016	no trend	False	0.6508	-0.4526	38
	Q	1954-1978	no trend	False	0.9741	-0.0324	20
Sassandra	ET	1979-1996	no trend	false	0.8796	0.1515	18
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1970-1996					27
Save	ET	not enough ET data points to conduct MK test on calculated ET					3
	P	1979-2016	no trend	False	0.8801	0.1509	38
	Q	1968-1981	increasing	True	0.0118	2.5183	14
Senegal	ET	1979-1989	no trend	false	0.2129	1.2456	11
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1979-1989					11
Tana	ET	not enough ET data points to conduct MK test on calculated ET					0
	P	1979-2016	decreasing	True	0.0006	-3.4447	38
	Q	1975-1978	no trend	False	0.7341	-0.3397	4
Upper Blue Nile	ET	not enough ET data points to conduct MK test on calculated ET					8
	P	1979-2016	no trend	False	0.6875	-0.4023	38
	Q	1961-1983	no trend	False	0.1339	+1.4988	26
Void	ET	not enough ET data points to conduct MK test on calculated ET					3
	P	1979-2016	no trend	False	0.1251	+1.5338	38
	Q	1979-1981	increasing	True	0.0483	1.9748	7
Zambezi	ET	1979-1990	no trend	false	0.5371	0.6172	12
	P	1979-2016	MK test not conducted, no trend found in ET				38
	Q	1960-1990					31

Fig. 6. Table A1: Mann-Kendall test results for all basins on evapotranspiration, precipitation and discharge

Table 5. Calculated statistics, bias, $bias_{aw}$, RMSE, $RMSE_{aw}$ and r, for the comparison of the long-term annual average ET_{WB} versus ET_{RS}

	CMRSET	ETMonitor	GLEAM	LandFlux-EVAL	MOD16	MTE	SSEBop	WaPOR	WECANN
bias	-19	156	254	115	131	146	12	-3	139
$bias_{aw}$	-18	237	313	148	266	183	30	-46	223
RMSE	113	211	273	152	199	184	163	104	189
$RMSE_{aw}$	187	502	594	304	590	424	123	165	520
r	0.94	0.91	0.97	0.97	0.91	0.95	0.89	0.96	0.95

Fig. 7. Table 5: Calculated statistics bias, bias_aw, RMSE, RMSE_aw and r for the comparison of the lon-term annual average ET_WB versus ET_RS

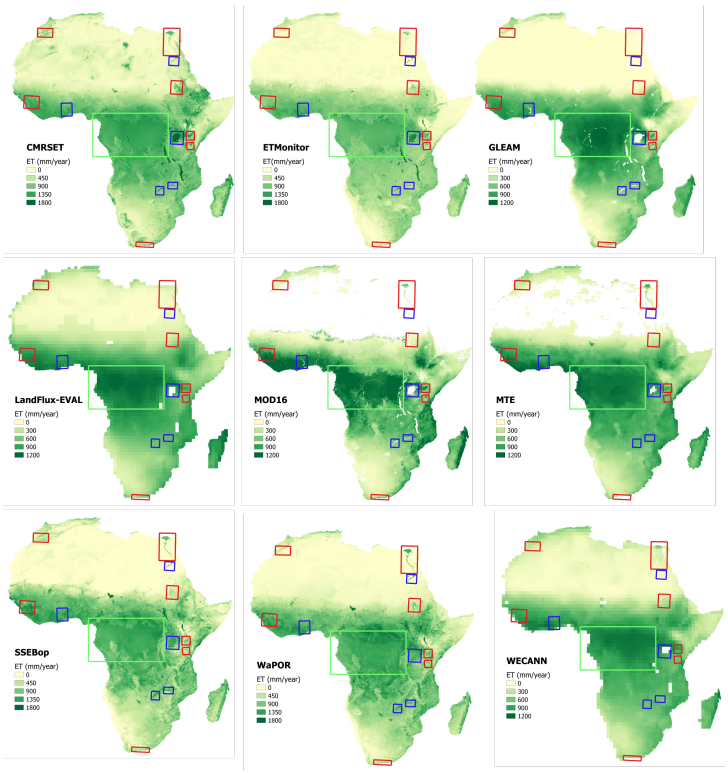

Fig. 8. Figure 10: Spatial assessment across Africa of each ET product based on selected land cover elements, forest, irrigated areas and water bodies