Reply to comments of Anonymous Referee #2

We thank anonymous referee #2 for reviewing our manuscript and for the positive comment. We are especially grateful for the many constructive comments and valuable suggestions, as these comments will lead us to improve the paper. We have to the best of our abilities responded to them and address the referee's comments in the following point by point response. Note the following conventions: RC = referee comments, AC = authors comments (replies) printed in italic.

Major comments

RC (a): Literature Review: it lacks significant contributions in the context of large-scale hydrological model simulation in data scarce area and it mainly focused on previous studies based on SWAT. It could be good to mention and discuss other approaches even if performed in different study areas but with the same problems (data scarce areas) (Kim et al., 2008; Kim and Kaluarachchi, 2009; Gebremicael et al., 2013; Tekleab et al., 2011; Abera et al., 2016 which applied a different hydrological modeling approach (Formetta et al., 2014))

AC (a): Agreed. We will add other related references and discuss other approaches in the context of large-scale hydrological model simulation in data scarce area as suggested.

RC (b): I feel that the authors should acknowledge explicitly that the analysis presented needs to be tested against observed data and that the satellite data are them self-based on modeling assumptions, which may or may not be plausible in some areas. Of course, they provide a huge help and the way in which they are used in the paper nicely show it, but probably assuming them as "measured" can be misleading. At least can be specified once in the text that "measured AET" doesn't mean proper eddy-covariance data

AC (b): We thank the reviewer for this suggestion. In the revised manuscript, we will strengthen the fact that we only make use of two satellite derived AET products. We will emphazise that we do not have any e.g. EC-based local measurments within our catchment. However, the satellite products have been tested elsewhere and we will briefly summarize study results that are relevant to our study (similar climate conditions) in a revised version of the manuscript. .

RC (c): In the paper is claimed the importance of the Curve Number parameter but nothing is said about soil moisture evolution and runoff. I wonder why the authors do not use runoff-measured data as independent validation. This will show the effects of the different ET calibration on the runoff dynamic. The two processes are strongly related and the sensitivity of the CN parameter confirms this. This will be an important added value to the paper. Again, the authors claim: "The average long-term annual of the water balance at the outlet of the study area shows a satisfactory percentage error of closure". Is this referred to modelled data or modelled and measured? The use of measured streamflow data would help to better understand this part as well.

RC (c): It is true that we emphases importance on the curve number parameter because it is found to be sensitive for all the six calibrations. As suggested by the referee, we agree, to mention and discuss other parameters relating to soil moisture and runoff considered in the calibration process.

We agree that AET calibration and runoff dynamics are strongly related. In the manuscript, we do not consider runoff-measured data as independent validation because it is not available for the study area and that is main reason we considered AET derived from satellite products as an alternative option for SWAT hydrologic model calibration. We believe using a freely available AET products (GLEAM & MOD16), that have been heavily tested in the past by calibration and validation experiments by a large number of scientist/research groups is one solution in setting up a hydrological model that will be used *as a decision support tool in such a data scarce region. These points will be emphasis in the revised manuscript.*

RC (d): Because one of the main points in discussion/conclusion is the fact that: "Hargreaves equation had a superior model performance of the Penman Monteith and the Priestly-Taylor" the authors should add their equations in the text. This would help to visualize the variables in input for each method, the variables that have been chosen for calibration and the variables that have been excluded.

AC (d): We will include the Hargreaves, Penman-Monteith and Priestly-Taylor PET equations in the paper for visualizing of each variables input for each method.

Specific comments:

RC1: Page 1 line 20: remove space in the number: River Basin (20 292 km2)

AC1: Thank you for this point. Although we haven't make the change yet, because checking the HESS manuscript preparation guidelines for authors, it is mentioned in the last sentence (h.) under heading "physical dimensions and units", that:

Numerals should also be typeset using upright fonts. The symbol for the decimal marker is the dot. To facilitate reading, numbers may be divided in groups of three using a thin space (e.g. 12 345.6), starting with the ten-thousand digit. Neither dots nor commas are permitted as group separators

RC2: Page 1 line 21: "The novelty of the study is the use of freely available satellite derived AET data for calibration/validation of each of the SWAT delineated subbasins, thereby obtaining a better performing model at the local scale as well as at the whole watershed level": sounds like this is the first time the gleam dataset have been used to validate/calibrate swat, which is a strong sentence. May be in the study area?

AC2: We agree with the referee that it is a strong sentence and we acknowledge that the novelty of the study needs further clarification. To this effect we will rewrite the statement in the revised manuscript to read as follows:

"The novelty of the study is the use of two freely available satellite derived AET dataset to calibrate/validate SWAT setups using three different estimated AET method for each of the 53 SWAT delineated subbasins, to improve hydrological model performance for a data scarce watershed in Nigeria."

Although the 3 different PET methods and the corresponding actual evapotranspiration estimate in SWAT have been tested (Wang et al. (2006), Franco and Bonumá (2017), Samadi (2017), Ha et al. (2018)), but study of calibrating each of the 3 SWAT estimated AET with two different freely available remotely sensed derived AET for each delineated subbasin within SWAT framework in order to know the highest performing model for a particular region/basin is yet unknown.

We believe this is a new contribution both to research community and in the study area.

RC3: Page 1 line 24: "Three different structures of the SWAT model were used in which each model structure was a set-up of SWAT with a different potential evapotranspiration (PET) equation": I would say that three different PET equations are tested: the model setup (in term of all the single components is the same except the pet).

AC3: We agree with the referee. As suggested, we will make the changes throughout the manuscript.

RC4: Page 2: mechanistic, what the authors mean? Please explain.

AC4: We agree with the referee that the terminology needs further clarification in the manuscript. In the context of this study mechanistic model such as of SWAT (Arnold et al., 1998), SHETRAN (Ewen et al., 2000), HSPF (Bicknell et al., 1997) means structural models that aim to describe which driving processes are present in a system and are able to make detailed predictions in both time and space. Hence, mechanistic is meant in a sense that relevant processes are described in some detail using physical and geochemical principles and process descriptions. The added statement will be inserted for more clarification in the revised manuscript.

RC5: Page 3 line 25: results showed a good Nash-Sutcliffe efficiency (NSE) and Coefficient of determination (R2) value for monthly average: quantify what good means for the authors and the values obtained.

AC5: We agree with the referee that the statement needs further quantification and clarification. To this effect we will rewrite the statement in the revised manuscript to read as follows:

"*The model results showed a good Nash-Sutcliffe efficiency (NSE) of 0.72 and Coefficient of determination (R^2) of 0.76 during the calibration periods. For the validation periods, a good model performance result showing R^2 of 0.71 and NSE of 0.78 values for monthly average streamflow were also obtained.*

RC6: Pag 5: The mean annual rainfall for the watershed is 1224 mm year-1 and the mean annual temperature is about 27o C. Mean annual potential evapotranspiration (PET) estimated by Hargreaves method (Hargreaves and Samani, 1985) is 1720 mm year-1 and the mean AET is about 692 mm year-

AC6: Many thanks for highlighting this points for clarification. We will rewrite the statement in the revised manuscript to read as follows:

The mean annual rainfall (1984-2012) obtained from measured data of Ogun watershed is 1224 mm yr^{-1} and the mean annual temperature obtained from measured data is about 27^o C. Mean annual potential evapotranspiration (PET) estimated by Hargreaves method (Hargreaves and Samani, 1985) using measured minimum and maximum temperature is 1720 mm yr^{-1} and the mean AET derived from SWAT output(1989-2012) for this study area is 692 mm yr^{-1}.

RC7: Page 5 typos: is 1224 mm year-1

AC7: Thank you. We will make the corrections throughout the manuscript.

RC8: Page 5: of 4103 ha, and please convert in km2 because all the other areas are in km

AC8: Thank you. We will make the changes as suggested in the revised manuscript

RC9: 30 m spatial resolution digital elevation model (DEM), 17 soil classes, 17 landuse classes, 3 slope categories, meteorological data and landuse with its management (Table 1). Please specify if those data are available, from which web-site, and the accessed date

AC9: Thank you. We will add the additional information to Table 1 in the revised manuscript as suggested.

RC10: Page 7 line 10: "The topHRU program allows the identification of a pareto-optimal threshold which minimizes the spatial error to 0.01 ha for a given number of HRUs and thereby minimizes the trade-off between SWAT computation time and number of HRUs. In this case, topHRU determined the optimum number of HRUs to be 1397 for the Ogun River basin. Thresholds of 0 ha for landuse, 150 ha for soil and 250 ha for slope were used in the SWAT set-up". What are the physical consequences of the thresholds? What happens if you use larger or lower values? How you define them?

AC10: As explained in the manuscript SWAT uses HRUs, whereby a watershed is subdivided into homogenous hydrologic response units having unique soil, slope, and land use properties as the basic unit of all SWAT model calculations. For SWAT, threshold specification of landcover, soil and slope is allowed and the physical consequences of the thresholds is to improve the computational efficiency of simulations while keeping key landscape features and information of a watershed in the hydrologic modelling. In the paper, we selected thresholds of 0 ha for landuse, 150 ha for soil and 250 ha for the slope. This means that HRUs should be created for all the area occupied by the landuse classes. For soil it means that any homogenous soil class occupying less than 150 ha should not be considered when determining HRUs. For slope, it means that any homogenous slope class that occupies less than 250 ha should not be considered. This allows us to define how detail the watershed will be represented by selecting the desired threshold values.

If we select larger values, then we eliminate key landscape features and their processes out of the system which may lead to considerable loss of information about the watershed landscape, resulting in model output that are less representative of the watershed as a whole and if we select lower values then we retain as many landscape features (spatial data) in the model thereby increasing the computational time of SWAT.

Therefore, we defined this threshold by selecting the desired threshold values using topHRU tool and its concept as explained in the paper while minimizing the spatial error to 0.01 ha for a given number of HRUs. Moreover, the criterial for selecting 0 ha for landuse was based on our desired to retain all the landuse area for all the landuse classes without losing out any information for future research needs.

We will try to explain the process of threshold selection more clearly in the revised manuscript.

RC11: Page 7: "delineated into 53 subbasins, with the main outlet in Abeokuta". Can you please give some summary statistics about them: min max average area, elevation, etc. Daily precipitation 5 data (1984-2012) and minimum and maximum temperature data (1984-2012) at four weather stations (Fig. 1) were used as observed input data. Are you only using 4 stations for the whole basin (20292 km2)? Why not considering satellite products for a variable (precipitation), which sometimes could be even more important than etp? The authors should include this in the discussion. The missing values of daily precipitation and minimum and maximum temperatures, along with solar radiation, wind speed and relative humidity were simulated by the ArcSWAT CSFR_World weather generator: it is clear that the ArcSWAT CSFR_World is used for gap filling of precipitation and temperature. The authors should specify: 1) how did you use the dataset for solar radiation, wind speed and relative humidity? 2) At which time resolutions are you specifying that input? 3) For which hydrological processes did you use these “simulated” forcing variables and how this affects your results?

AC11: Many thanks for raising this comment for clarification.

In SWAT, the Ogun River Basin was delineated into 53 subbasin for this study. The summary statistics of the 53 delineated subbasins is as follows; the minimum and maximum elevation are 23 m and 624 m respectively, while the mean elevation is 289.1 m. The minimum and maximum subbasin area are 72.4 km^2 and 853.1 km^2 respectively, while the mean is 382.8 km^2. The minimum and maximum subbasin length are 72633 m and 269744 m while the mean length is 153528 m. We will add this information in the manuscript.

Four weather station data from the Nigerian meteorological agency were used as observed input data because that is the only available ground weather stations in the study area. Initially we considered using satellite precipitation products but comparing the observed with the satellite data downloaded yielded a huge bias. Thus we focused on the limited ground truth stations weather data. The weather stations are distributed evenly within the watershed, and one station is located outside the basin. All of the stations show high consistency. Looking at the elevation, no orographic effect needed to be considered for correcting the precipitation values. The reason for using only 4 weather stations in this study will be included in the revised manuscript.

SWAT requires daily values of solar radiation, relative humidity and wind speed in addition to the daily precipitation, minimum and maximum temperature as weather input in SWAT. One out of many options in SWAT to generate this input variables, is to use “ WGEN_CFSR_World (ArcSWAT CSFR_World weather generator)” which is an MS Access file containing long-term monthly weather statistics covering the entire globe. In this study, the long-term monthly weather statistics developed using The National Centres for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) global dataset were used to simulate daily solar radiation, relative humidity and wind speed using the WGEN CFSR World. It is also used for filling gaps in measured climate data.

The simulated variables were used as input variables into Penman-Monteith and Priestly-Taylor equations for obtaining the different PET estimates from SWAT.

The simulated variables allow options for different evaporation estimates which actually affect the results of the model performance during the calibration and validation period as shown in the manuscript.

RC12: Page7 line22-27: it sounds slightly repetitive: please consider to write the full sentence only of one model structure and to generalize for the other 2.

AC12: We agreed to make the suggested changes in page 7 from line 16-27. We will rewrite the statement in the revised manuscript to read as follows:

The SWAT model was run three times whereby three different PET equations were tested in SWAT: the model setup (in term of all the single components were the same except the PET). The SWAT model in which three different AET estimates were obtained respectively with each PET equation (SWAT_HG, SWAT_P-M, SWAT_P-T) were used to evaluate the model performance by comparing the calibrations, validations, and the reference run (uncalibrated SWAT) with two global AET products (GLEAM_3.0a and MOD16), thus allowing for six calibrations of SWAT (GS1 through MS6) and six reference runs with SWAT default parameters (RGS1 through RMS6). Calibration GS1 through GS3 refers to SWAT simulated AET when SWAT_HG, SWAT_P-M, SWAT_P-T were selected respectively and were calibrated and validated with the AET from GLEAM_v3.0a. Calibration MS4 through MS6 refers to SWAT simulated

AET when SWAT_HG, SWAT_P-M, SWAT_P-T were selected respectively and were calibrated and validated with the AET from MOD16. Reference run (RGS1 through RGS3) refers to SWAT simulated AET obtained with SWAT default parameters when SWAT_HG, SWAT_P-M, SWAT_P-T were selected respectively and were compared with GLEAM_v3.0a. Reference run (RMS4 through RMS6) refers to SWAT simulated AET obtained with SWAT default parameters when SWAT_HG, SWAT_P-M, SWAT_P-T were selected respectively and were compared with MOD16 AET.

The changes will be made in the revised manuscript as requested. The figure 2 as shown in the manuscript will also be updated.

RC13: Page 8 line 10: please explain what are the main differences between the two dataset GLEAM_v3.0a and GLEAM_v3.0b and justify why you selected one of the two.

AC13: Many thanks for raising this comment for more clarification.

The two datasets differ in their forcing variables and their temporal coverage (Martens et al., 2016).

GLEAM_v3.0a is a global dataset that is based on reanalysis net radiation and air temperature, satellite-based vegetation optical length, and a combination of gauge-based, reanalysis satellite-based precipitation. It is a dataset spanning the 35-year period 1980-2014. For this study, we preferred and selected GLEAM_v3.0a dataset spanning 24-year period 1984-2012 because of its long-term availability that allows reasonably selection and splitting of calibration and validation periods that are not substantially different in climatic condition i.e., wet, moderate, and dry years occur in both periods and which covers our SWAT simulation output period (1989-2012).

GLEAM_v3.0b is a global dataset driven by satellite data only and spanning 13-year period 2003-2015. We considered this dataset for the verification of SWAT simulated AET because there are no ground truth AET data in the study area and also, because of its different forcing variable, which categories it as an independent dataset not considered in the calibration and validation period.

The added statement will be used to clarify further the differences between GLEAM_v3.0a and GLEAM_v3.0b in the revised manuscript.

RC14: Page 8 line 25: "was implemented in SWAT-CUP. SWAT-CUP (Abbaspour, 2015)" move the citation when you firstly introduce SWAT-CUP.

AC14: Thank you. We will make the changes as suggested

RC15: Pages 8 line 28 to page 9 line 6: Please specify the parameter set that you started the sensitivity analysis with, at least the processes to which they are related. Moreover, specify the list of the parameters that resulted sensitive and how you define a parameter as "sensitive".

AC15: We started the global sensitivity analysis for each of the six-calibration run with the same 50 parameters as shown in the table below. The table below only represent the result of the GS1 parameter sensitivity analysis but the same 50 parameters were used for GS2 through MS6 sensitivity analysis.

In this study as described in the manuscript, the parameter sensitivity is determined by numerous rounds of Latin Hypercube sampling and we defined a parameter as being sensitive when (considering the absolute values) large values of t-stat and smaller values of p-value were obtained, then, the more sensitive the parameter.

The processes to which the parameters are related hydrological are runoff, evaporation, interception, transpiration. In short, each of the hydrological parameters are represented in the 50 parameters we started the initial global sensitivity with, as shown in table 1.

Table 1: The 50 parameters consider in the initial global sensitivity analysis and their relative significance. The table below only show the global sensitivity analysis result of the GS1

Parameter Name	t-Stat	P-Value
45:V__BMX_TREES{..}.plant.dat	0.68	0.62
23:R__SOL_CBN(..).sol	-0.99	0.50
47:V__TMPMX(..).wgn	1.84	0.32
48:V__PCPMM(..).wgn	-2.27	0.26
35:V__LAI_INIT.mgt	2.28	0.26
31:V__SURLAG.bsn	-3.03	0.20
28:R__ALPHA_BF_D.gw	3.21	0.19
19:A__GWQMN.gw	-3.32	0.19
24:R__SOL_ALB(..).sol	-3.82	0.16
37:V__FLOWFR.mgt	-3.86	0.16
40:R__SOL_ZMX.sol	-4.06	0.15
50:R__SOL_Z(..).sol	4.08	0.15
39:V__TLAPS.sub	4.27	0.15
15:V__ESCO.hru	-4.35	0.14
27:V__RCHRG_DP.gw	4.55	0.14
42:V__ALPHA_BF_D.gw	-4.70	0.13
25:R__USLE_K(..).sol	-5.73	0.11
26:V__SOLARAV(..).wgn	-5.96	0.11
38:V__TDRAIN.mgt	6.02	0.10
34:V__BIO_MIN.mgt	6.03	0.10
46:V__TMPMN(..).wgn	-6.18	0.10
3:V__REVAPMN.gw	-6.19	0.10
43:V__RADINC(..).sub	-6.20	0.10
36:V__BIO_INIT.mgt	6.25	0.10
13:V__EVRCH.bsn	-6.91	0.09
20:R__HRU_SLP.hru	6.94	0.09
21:V__GW_DELAY.gw	6.96	0.09
14:V__GW_REVAP.gw	-7.01	0.09
44:V__HUMINC(..).sub	-7.22	0.09
29:V__SHALLST.gw	-7.34	0.09
16:V__CH_N2.rte	7.59	0.08
22:V__ALPHA_BNK.rte	7.66	0.08
17:V__CH_K2.rte	-7.66	0.08
41:V__CH_N1.sub	7.70	0.08
49:V__DEEPST.gw	7.98	0.08
12:V__EVLAI.bsn	8.03	0.08
11:V__OV_N.hru	8.12	0.08
18:V__SFTMP.bsn	8.49	0.07

33:V__GWHT.gw	8.74	0.07
10:V__FFCB.bsn	-8.78	0.07
7:R__SOL_BD(..).sol	-8.91	0.07
6:V__EVRSV.res	-8.95	0.07
32:V__GSI{..}.plant.dat	-9.78	0.06
2:V__EPCO.hru	10.14	0.06
4:R__SOL_K(..).sol	10.51	0.06
9:V__ALPHA_BF.gw	-10.52	0.06
1:V__ESCO.hru	-10.71	0.06
30:V__CANMX.hru	11.52	0.06
5:R__SOL_AWC(..).sol	12.37	0.05
8:R__CN2.mgt	15.79	0.04

After choosing the 11 most sensitive parameters based on the t-stat and p-values. We tried many combinations of the most sensitive analysis e.g. we started with $P < 0.09$ (37 parameters) the result was not as good at that of the 11 parameters combination. So we decided to take the 11 most sensitive parameters and we run another global sensitivity analysis to further identify the relative significance of each parameter before calibration. Using 11 parameters gave the most reasonable results. This methodology was extended to all the remaining calibration runs.

We will further clarify the number of parameters we started with, the process to which they are related and how we define the parameter sensitive in the revised manuscript.

Table 2: The 11 most sensitive parameters consider in the final calibration and validation of all the six calibration runs. The table below only show the sensitivity analysis of the GS1 calibration results.

Parameter Name	t-Stat	P-Value
9:R__SOL_AWC(..).sol	0.01	0.99
2:V__EPCO.hru	-0.21	0.83
7:V__FFCB.bsn	-0.27	0.79
4:V__GSI{..}.plant.dat	0.40	0.69
6:V__EVRSV.res	-0.54	0.59
10:R__SOL_K(..).sol	-0.89	0.37
5:V__ALPHA_BF.gw	1.43	0.15
11:R__SOL_BD(..).sol	-2.01	0.04
3:V__CANMX.hru	2.19	0.03
1:V__ESCO.hru	-2.86	0.00
8:R__CN2.mgt	-23.93	0.00

RC16: Page 9 line 16: A metric among the six can be considered an objective function if it is optimized in the calibration procedure; it can be considered as goodness of fit metric if it is used to quantify how well or bad the model reproduces the measured data. Are those goodness of fit metrics? Which one of these six metrics has been optimized in the calibration procedure? Have you used all of them also as objective function? This is not fully clear.

AC16: Many thanks for raising this comment for more clarification.

The Nash-Sutcliffe is the selected objective function that was optimized during the calibration procedure .This statement will be added for clarity in the revised manuscript.

RC17 Page 9 line 20 –Page 10 line 20: Consider to: i) just spell in the text the statistics used, their ranges and their optimal values and ii) move in appendix the explanation of each statistics because they are well known.

AC17: Thank you. We will make the changes as suggested.

RC18: Page 10 line 23-26: please specify how the uncertainty is quantified: what are the parameters that are changed/sampled the LHS, what are their ranges?

AC18: Thank you for highlighting this point for further clarification

As described in the manuscript the uncertainty was quantified by: 1.) assessing the percentage of GLEAM_3.0a AET data bracketed by the model output 95% predictive uncertainty band, the index used for the quantification is P-factor, and 2.) assessing the ratio of the average width of the 95ppu and the standard deviation of the GLEAM_3.0a, the index used for the quantification is R-factor.

The parameters that are changed are listed in Table 2 above by implementing numerous rounds of Latin Hypercube sampling. The optimum values are presented in the manuscript but their ranges are presented in Table 3a-b below:

Table 3a: Eleven parameter set and their ranges used for the first 1000 simulations (1st iteration during calibration) in this study for all the six calibration runs.

Parameter name	Minimum range	Maximum range
v__ESCO.hru	0.00	1.00
v__EPCO.hru	0.00	1.00
v__CANMX.hru	0.00	100.00
v__GSI{2,4,5}.plant.dat	0.00	5.00
v__ALPHA_BF.gw	0.00	1.00
v__EVRSV.res________17,50	0.00	1.00
v__FFCB.bsn	0.00	1.00
r__CN2.mgt	-0.25	0.85
r__SOL_AWC().sol	0.23	0.95
r__SOL_K().sol	-0.06	0.95
r__SOL_BD().sol	-0.41	0.95

Table 3b: Eleven parameters set and their ranges used for the second 500 simulations of the GS1 calibration run (2nd iteration during calibration) showing how the parameter range changes from the 1st iteration of GS1 calibration run.

Parameter name	Minimum range	Maximum range
v__ESCO.hru	0.00	0.59
v__EPCO.hru	0.31	0.95
v__CANMX.hru	0.00	59.46
v__GSI{2,4,5}.plant.dat	2.35	5.00
v__ALPHA_BF.gw	0.38	1.00
v__EVRSV.res________17,50	0.43	0.60

v__FFCB.bsn	0.36	1.00
r__CN2.mgt	-0.47	0.40
r__SOL_AWC().sol	0.52	0.90
r__SOL_K().sol	-0.12	0.59
r__SOL_BD().sol	-0.39	0.50

All the necessary additional information will be added in the revised manuscript for further clarification.

RC19: Page 11: please show a figure of the river basin with the subbasin polygons and the pixel of MODIS and GLEAM. This will help the reader to understand how many pixels of MODIS and GLEAM cover your basins. Each sub-basins has its own model AET. How did you choose the MODIS or GLEAM pixel to compare with and compute the NSE, R^2, etc.

AC19: Thanks for the suggestion, we prepared a figure showing the river basin with the subbasin polygon and the pixel of MOD16 and GLEAM AET. We will insert the figures as suggested in an Appendix.

To compare MODIS pixel value to SWAT simulated AET values from each subbasin for computing the NSE, R^2, PBIAS, KGE, an area-weighted averaging scheme was performed in ArcGIS to create aggregated monthly time-series of AET data of each subbasin.

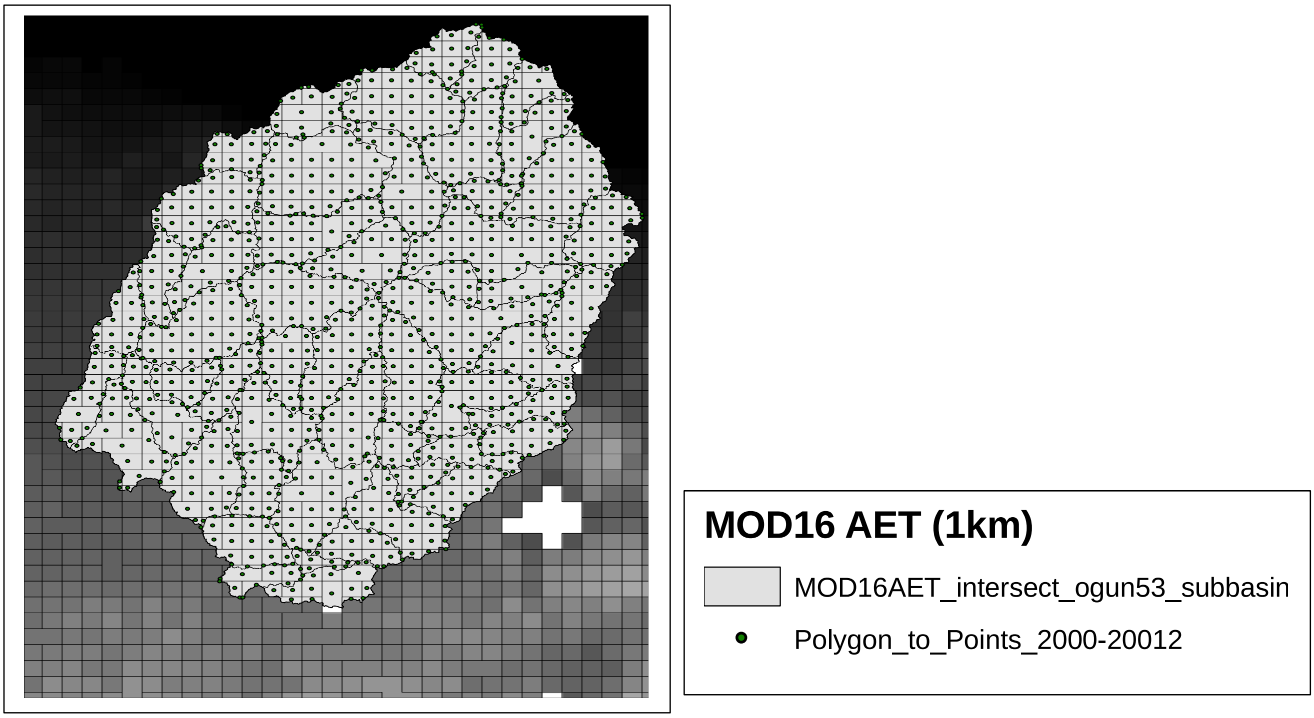

Figure 1: Ogun River Basin with its 53-subbasin polygons intersected by the pixel of MOD16 AET.

The GLEAM_3.0a and GLEAM_3.0b AET is provided in netcdf format, with one file per year and variable. The datasets are available on a 0.25^0 latitude- longitude regular grid and at daily temporal resolution.

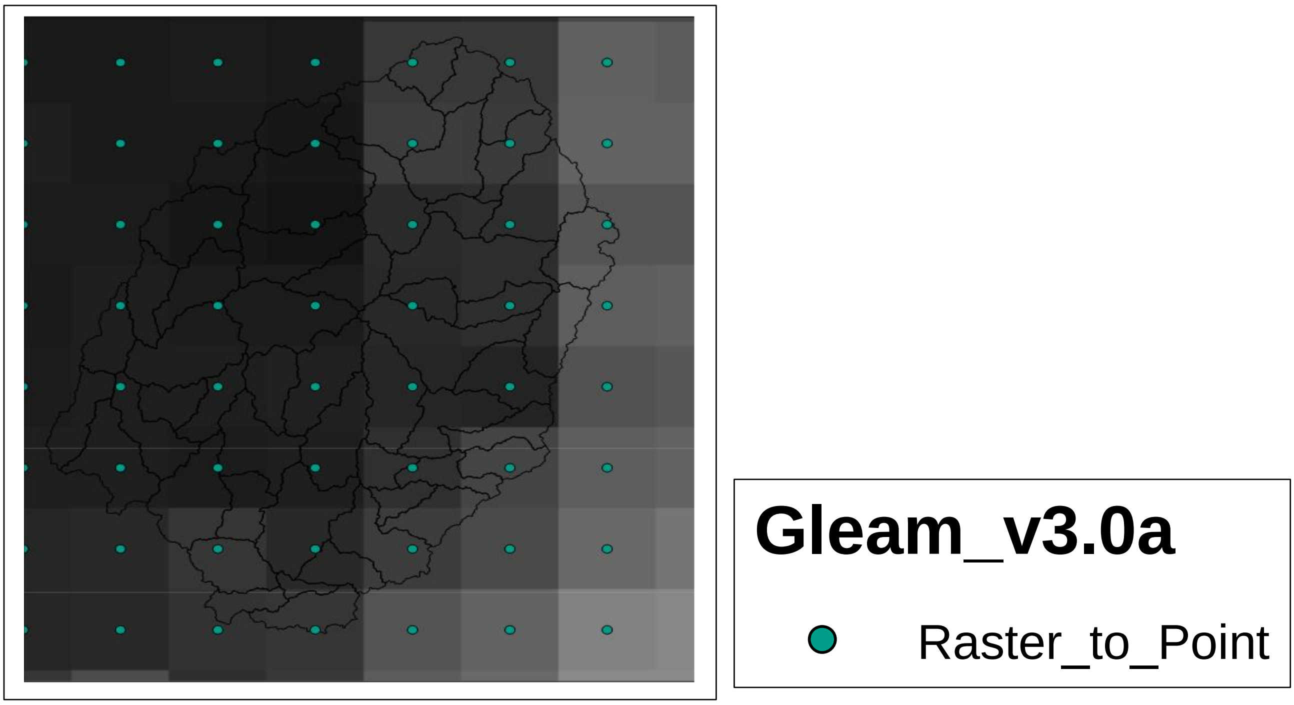

Figure 2: Ogun River Basin with its 53-subbasin polygons intersected by the pixel of GLEAM AET

We used "make NetCDF raster layer" tool in ArcGIS to convert the NetCDF file into a raster layer to view how many pixels of GLEAM cover our subbasins (Figure 1). We realized some points from which the data will be extracted from the pixel are not located in some subbasins (Fig.3). Therefore, we decided to create a point at the center of each subbasin in ArcGIS (Fig. 4). The coordinates of each point at the center of the subbasins were obtained.

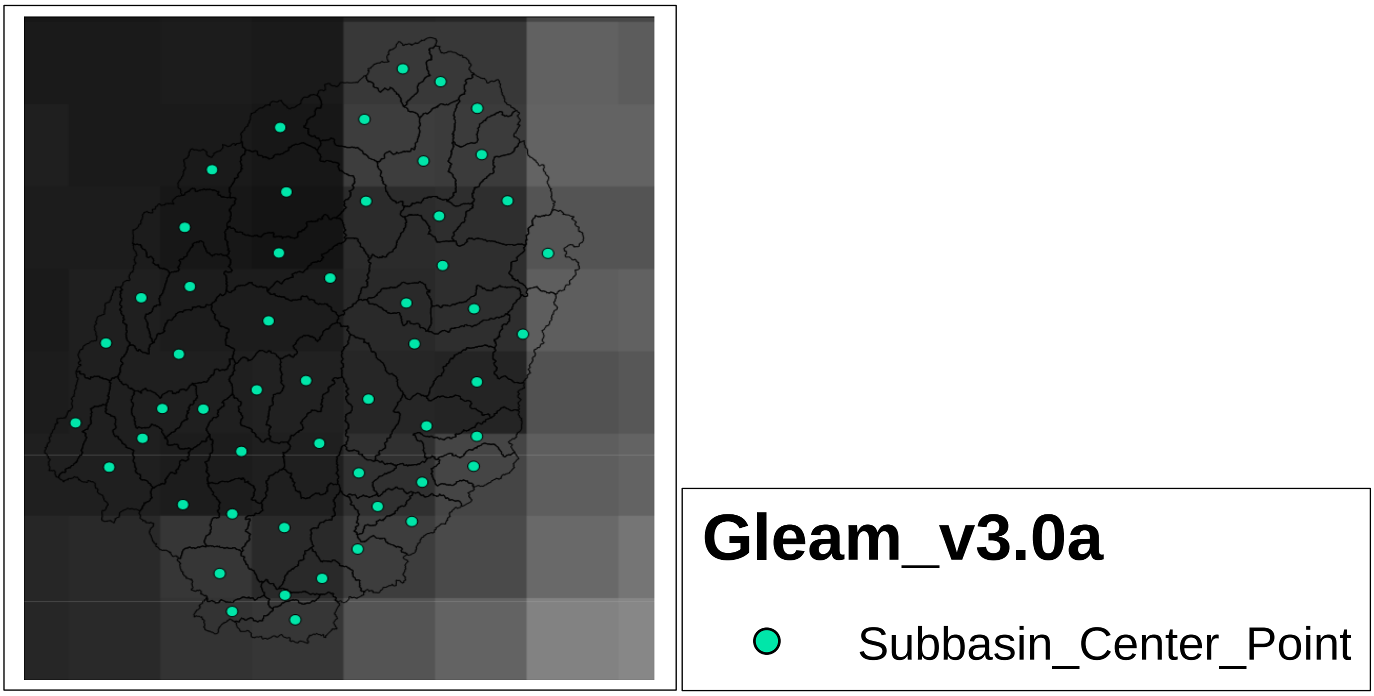

Figure 3: Ogun River Basin with its 53-subbasin polygons intersected by the pixel of GLEAM AET with a point at the center of each subbasin.

To enable us to compare GLEAM pixel value to SWAT simulated AET values for each subbasin for computing the NSE, R2, PBIAS, KGE, we went back to the NetCDF file of the daily time series (1989-2012) and extracted the GLEAM AET value for each subbasin using "Make NetCDF table view. The values extracted were aggregated to monthly values. The same procedure was carried out for GLEAM_3.0b version.

RC20: Page 11: figure 3, 4, 5, and 6 please consider to add a 4th class for the range of KGE and NSE (<0). This indicates where just the mean of the observed data will be more performing than the model itself

AC20: Many thanks for the suggestion that needs further clarification.

We divided the results into 5 classes. Wish we think the performance rating classes represented with figure 3,4,5,6 for each subbasins are well represented and sufficient

-6.0 – 0 = red

0.10- - 0.50 = purple

0.51 – 0.60 = yellow

0.61-0.70 = orange

0.71 -1.0 = green

The KGE and NSE figure that do not contain a class -6.0 – 0 (<0) do not contain values <0 (fig.3, 4 in the manuscript) and figures that contains -6.0 – 0 (<0) has a model performance that is less than 0 (fig. 5 and fig.6 in the manuscript)

We do not think is necessary to include a class in the legend without having its value in the subbasin polygon.

RC21: Page 10 line 15: equation is not correct please revise it.

AC21: Many thanks for the point. We will correct and insert it in the revised manuscript.

RC22: Page 11: figure 3, 4, 5, and 6 please add the North symbol and the scale bar in the maps.

AC22: Many thanks for the suggestion. We will include North symbol and the scale bar in the revised manuscript.

RC23: Page 11 line 23: "The results of global sensitivity analysis revealed that the SCS runoff": please specify from where the reader can see this.

AC23: Thank you. We will include Table 2 at the end of the statement to read as follow:

"The results of global sensitivity analysis revealed that the SCS runoff curve number (CN2.mgt) is the most sensitive parameter to SWAT simulations of AET for all the six calibrations (Table 2)"

RC24: Page 12: all the page can be summarized by just one figure reporting on the x-axis the model configurations and on the y axis the percentage of sub-basin for a given class of the goodness of fit index (NSE, R2, etc).

AC24: Many thanks for the suggestion. The results presented in page 12 are already summarized in figure 3, 4,5,6,7, and 8. We believe this figures depicts and summarized well the SWAT model performance for each subbasin and the percentage covered (using pie chart). We believe having another figure to summarize page 12 is not needed.

We agree to reduce the presentation of results described on page 12 because the figures already shown them.

RC25: Page 13 the paper goes from section 3 to subsection 3.3: 3.1 and 3.2 are missing.

AC25: Thank you for highlighting this point. We will include subsection 3.1 and 3.2 in the revised manuscript

RC26: Page 14: increases in February from 55mm to 76mm as the space between 55 and mm

AC26: Thank you for highlighting this point. We will make the corrections in the revised manuscript

RC27: Page 15 line 5: "Using the guidelines in Moriasi et al. (2007, 2015) and Kouchi et al. (2017) for" probably these guidelines were drawn for runoff? Is it correct to use it for others hydrological processes? Is this been done in the past? If yes, please add a citation otherwise just clarify this aspect.

AC27: *Many thanks for raising this point for further clarification.*

the general hydrologic model performance ratings for recommended statistics (NSE, PBIAS, R^2) performed at a monthly time and recommended by Moriasi et al. (2007, 2015) and Gupta et al are mostly drawn for runoff, sediment and nutrients because when these articles were published, the use of remotely sensed evapotranspiration datasets for hydrologic model calibration/validation did not gain much ground.

In this paper, we also conducted reviewed literatures on model evaluation methods and ratings for model calibration using satellite or non-satellite derived evapotranspiration. Ha et al. (2018) presented a study of calibration of spatially distributed hydrological processes and model parameters in SWAT using remote sensing data and an autocalibration procedure. NSE, R^2, and KGE criteria were used to assess the model performances.

Djman (2016) in their study of evaluation, calibration and validation of six reference ET_0 equation for Senegal River Delta using Penman-Monteith derived ET_0 obtained at saint louis station (1960-2012), uses R^2 and other statistical measures to perform the evaluation. Lopez et al. (2017), calibrated a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. They evaluated the model performance using NSE, PBIAS, KGE and R. Samadi et al. (2016) presented a study on assessing the sensitivity of SWAT physical parameters to potential evapotranspiration estimation methods over a coastal plain watershed in the southeastern United States, NSE, and KGE statistical measure were used for the model performance assessment.

All the reviewed literatures set their performance ratings for recommended statistics (NSE, PBIAS, R, KGE, R^2) based on Moriasi et al. (2007, 2015) guidelines.

In this study, we follow Lopez et al. (2017) and others as reviewed above to base our performance rating criteria for judging the model performance by using NSE, R^2, PBIAS and KGE.

RC28: Page 15 line 20: “From our results, we agree that the AET from MOD16 tends to overestimate AET”. Overestimate against what? This is a strong statement mainly because there is not direct comparison against measured AET data

AC28: Thank you for highlighting this point. We will make the corrections in the revised manuscript

We agree this is a strong statement that needs to be revised. Actually, we meant that AET from MOD 16 values are higher than that of GLEAM AET and SWAT simulated AET and that this finding agrees with other studies carried out in tropical regions. Since we are using the satellite based MOD 16 to calibrate the SWAT AET simulation, the statement needs correction and we agree to write it to read as follows:

“From our results, we agree that AET from SWAT tends to underestimate AET, when calibrated the model with MOD16 AET and this finding agrees with other studies carried out in the tropical regions (Ruhoff et al., 2013)”.

Reference

Arnold, J. G., Srinivasan, R., Muttiah, R. S. and Williams, J. R.: Large area hydrologic modeling and assesment Part I: Model development, JAWRA J. Am. Water Resour. Assoc., 34(1), 73–89, doi:10.1111/j.1752-1688.1998.tb05961.x, 1998.

Bicknell, B. R., Imhoff, J. C., Kittle, J. L. J., Donigian, A. S. J. and Johanson, R. C.: Hydrological Simulation Program--Fortran: User’s manual for version 11, , 755, doi:EPA/68/C-01/037, 1997.

Djaman,K., Tabari, H., Baide,A.B., Diop,L., Futakuchi,K., and Irmak,S.: Analysis, calibration, and validation of evapotranspiration model to predict grass reference evapotranspiration in Senegal River Delta. Journal of Hydrology: Regional Studies 8, 82-94,2016.

Ewen, J., Parkin, G. and O’Conell, P. E.: SHETRAN : Distributed River Basin Flow Modeling System, J. Hydrol. Eng., 5(JULY), 250–258, doi:doi:10.1061/(ASCE)1084-0699(2000)5:3(250), 2000.

Franco, A. L. and Bonumá, N.B.: Multi-variable SWAT model calibration with remotely sensed evapotranspiration and observed flow. *RBRH*, *22*, e35, 2017.

Ha,L.T, Bastiaanssen, W.G.M, van Griensven, A, van Dijk, A.I.J.M, Senay, G.B.: Calibration of Spatially Distributed Hydrological Processes and Model Parameters in SWAT Using Remote Sensing Data and an Auto-Calibration Procedure: A Case Study in a Vietnamese River Basin. *Water*. 10(2):212, doi: 10.3390/w10020212, 2018.

Hargreaves, G. H. and Samani, Z. A.: Reference Crop Evapotranspiration from Temperature, Appl. Eng. Agric., 1(2), 96–99, doi:10.13031/2013.26773, 1985.

Kouchi, D. H., Esmaili, K., Faridhosseini, A., Sanaeinejad, S. H., Khalili, D. and Abbaspour, K. C.: Sensitivity of calibrated parameters and water resource estimates on different objective functions and optimization algorithms, Water (Switzerland), 9(6), 1–16, doi:10.3390/w9060384, 2017.

Lopez Lopez, P., Sutanudjaja, E., Schellekens, J., Sterk, G. and Bierkens, M.: Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products, Hydrol. Earth Syst. Sci. Discuss., (January), 1–39, doi:10.5194/hess-2017-16, 2017.

Martens, B., Miralles, D., Lievens, H., Fernández-Prieto, D., and Verhoest, N.: Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture, Int. J. Appl. Earth Obs., 48, 146–162, 2016.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Binger, R. L., Harmel, R. D. and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Trans. ASABE, 50(3), 885–900, doi:10.13031/2013.23153, 2007.

Moriasi, D. N., Gitau, M. W., Pai, N. and Daggupati, P.: Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria, Trans. ASABE, 58(6), 1763–1785, doi:10.13031/trans.58.10715, 2015.

Ruhoff, A. L., Paz, A. R., Aragao, L. E. O. C., Mu, Q., Malhi, Y., Collischonn, W., Rocha, H. R. and Running, S. W.: Assessment of the MODIS global evapotranspiration algorithm using eddy covariance measurements and hydrological modelling in the Rio Grande basin, Hydrol. Sci. J., 58(8), 1658–1676, doi:10.1080/02626667.2013.837578, 2013

Samadi,S.Z.: Assessing the sensitivity of SWAT physical parameters to potential evapotranspiration estimation methods over a coastal plain watershed in the southeastern United States. IWA, 48 (2) 395-415, doi: 10.2166/nh.2016.034, 2017.

Wang, X., Melesse, A. M. and Yang, W.: Influences of potential evapotranspiration estimation methods on SWAT's hydrologic simulation in a northwestern Minnesota watershed, Trans. ASABE, 49(6), 1755–1771, doi:10.13031/2013.22297, 2006.